Kernel Estimators in Industrial Applications

Piotr Kulczycki

Systems Research Institute, Polish Academy of Sciences
ul. Newelska 6, PL-01-447 Warsaw, Poland
kulczycki@ibspan.waw.pl
Cracow University of Technology, Department of Automatic Control
ul. Warszawska 24, PL-31-155 Cracow, Poland
kulczycki@pk.edu.pl

1 Introduction

The specification, based on experimental data, of functions which characterize an object under investigation, constitutes one of the main tasks in modern science and technological problems. A typical example here is the estimation of density function of random variable distribution from any given sample. The classical procedures rely here on arbitrary assumption of the form of this function, and then in specification of its parameters. These are called parametric methods. A valuable advantage is their theoretical and calculational simplicity, as well as their being commonly known and present in subject literature. Nowadays – along with the dynamic development of computer systems – nonparametric methods, whose main feature constitutes a lack of arbitrary assumptions of the form of a density function, are used more and more often. In a probabilistic approach, kernel estimators are becoming the principal method in this subject. Although their concept is relatively simple and interpretation transparent, the applications are impossible without a high class of computer which, even until recently, significantly hindered theoretical, and especially practical research.

In this chapter, first – in Section 2 – the basics of kernel estimators methodology are presented in a form suitable for researchers without thorough knowledge in the area of advanced statistical methods. So, the fundamental definitions of a kernel estimator are described, as are one-dimensional, and also radial and product kernels for the multidimensional case, suboptimal – in a mean-square sense – methods for calculation of functions and parameters occurring there, as well as procedures of smoothing parameter modification and linear transformation. Thanks to today's availability and the possibilities of contemporary computer systems as well as the automation of metrological and data gathering processes, the universal character of kernel estimators allows for their broad application in various problems of modern science and technology, particularly those of an industrial nature. In Section 3 of this chapter, uses of kernel estimators are described for the following subjects:

- data analysis and exploration – recognition of atypical elements (outliers), clustering, and classification – applied to the detection and diagnosis of devices working in real-time, and planning of strategy for mobile phone operators;
- parametric identification illustrated in automatic control applications and by tasks of sharpening of imprecise information;

B. Prasad (Ed.): Soft Computing Applications in Industry, STUDFUZZ 226, pp. 69–91, 2008.
springerlink.com

- definition of spatial distribution of demand based on fuzzy data for the needs of a problem from the telecommunications field.

In addition, the tasks of data compression and dimensionality reduction, based on artificial neural networks and evolutionary algorithms are also commented upon. These among others efficiently reduce calculation time in the above presented tasks.

The detailed description of the basics of kernel estimators methodology can be found in the monographs (Kulczycki 2005, Silverman 1986, Wand and Jones 1994). Many applicational aspects are also found in the papers (Kulczycki 2000-2002; Kulczycki and Charytanowicz 2005, Kulczycki and Daniel 2006, Kulczycki and Mazgaj 2005; Kulczycki and Waglowski 2005, Kulczycki and Wisniewski 2002). The preliminary version of this text was presented as (Kulczycki 2007).

2 Methodology of Kernel Estimators

Let the *n*-dimensional random variable *X*, with a distribution having the density *f*, be given. Its kernel estimator $\hat{f} : \mathbb{R}^n \to [0,\infty)$ is calculated on the basis of the *m*-elements random sample

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

experimentally obtained from the variable *X*, and is defined in its basic form by the formula

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

where the measurable, symmetrical with respect to zero and having a weak global maximum in this point, function $K : \mathbb{R}^n \to [0,\infty)$ fulfils the condition $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x = 1$ and is called a kernel, whereas the positive coefficient *h* is referred to as a smoothing parameter.

The interpretation of the above definition is illustrated in Fig. 1, taking a one-dimensional random variable (i.e. when $n = 1$) as an example, for a 9-element sample (i.e. $m = 9$). 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 realizations 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 $\int_{\mathbb{R}^n} \hat{f}(x)\,\mathrm{d}x = 1$, required of the density of a probability distribution.

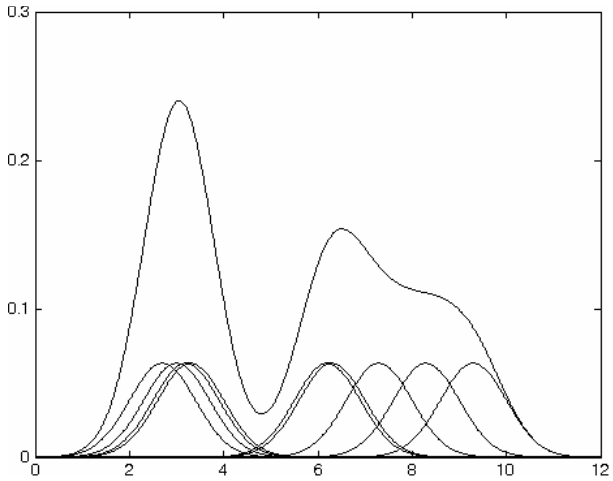

Fig. 1. Kernel density estimator (2) of one-dimensional random variable distribution

It is worth noting that a kernel estimator allows the identification of density for practically every distribution, without any assumptions concerning its membership to a fixed class. Atypical, complex distributions, also multimodal, are regarded here as textbook unimodal. In the multidimensional case, i.e. where $n > 1$, this enables, among others, the discovery of complete dependences between particular coordinates of the random variable under investigation.

Setting the quantities introduced in definition (2), i.e. choice of the form of the kernel K as well as calculation of the value for the smoothing parameter h, is most often carried out according to the criterion of minimum of an integrated mean-square error. These problems will be discussed in Sections 2.1 and 2.2.

2.1 Choosing the Kernel Form

From the statistical point of view, the form of the kernel appears to have no significance, and so it becomes possible for the choice of the function K to be arbitrary, taking into account above all the desired properties of an obtained estimator, for example its class of regularity, assuming (strictly) positive values or other features important in a considered problem, including calculational suitability. This is of particular importance in the case of complicated problems of modern industrial challenges, where the kernel estimator constitutes the most common basis for further comprehensive and complex investigations. Its obtained properties can then not only simplify further procedures, but often actually allow a concrete result, suitable for application, to be reached.

In the one-dimensional case, the function K assumes the classic forms of densities of probability distributions, normal, Cauchy, and triangular among others, as well as – in specific tasks – their linear combinations. In the sense of the integrated mean-square error criterion, the so-called Epanechnikov kernel

$$K(x)=\begin{cases}\dfrac{3}{4}(1-x^2) & \text{for} \quad x\in[-1,1]\\ 0 & \text{for} \quad x\in(-\infty,-1)\cup(1,\infty)\end{cases} \tag{3}$$

is the most effective. In the multidimensional case, two natural generalizations for the above concept are used: the radial kernel

$$K(x)=C\,\mathscr{K}(\sqrt{x^{\mathrm{T}}x}) \tag{4}$$

and the product kernel

$$K(x)=K([\,\mathsf{x}_1,\,\mathsf{x}_2,\ldots,\mathsf{x}_n]^{\mathrm{T}})=\mathsf{K}\,(\mathsf{x}_1)\cdot\,\mathsf{K}\,(\mathsf{x}_2)\cdot\ldots\cdot\,\mathsf{K}\,(\mathsf{x}_n), \tag{5}$$

where $\mathscr{K}$ denotes the aforementioned one-dimensional kernel, while C is a positive constant, calculated so that $\int_{\mathbb{R}^n} K(x)\,\mathrm{d}x=1$. For any assumed one-dimensional kernel $\mathscr{K}$, radial kernel (4) appears to be more effective than product kernel (5), although the difference – from an applicational point of view – seems to be negligible. Due to this fact, the product kernel is often preferred in practical problems. Apart from specific statistical tasks, it proves to be significantly more suitable in further analysis – for example procedures of integration and differentiation of a product kernel are not much different from the one-dimensional case. Among radial kernels the most effective is the radial Epanechnikov kernel, i.e. defined by dependence (4) when $\mathscr{K}$ is given by formula (3). Likewise in the family of product kernels, the most effective proves to be the product Epanechnikov kernel, described by equalities (5) and (3).

To summarize: the possibility of significant flexibility in choosing the form of the kernel K constitutes a great practical advantage which becomes more obvious as the specific problem under investigation is more complex.

2.2 Calculation of Smoothing Parameter Value

Unlike the kernel's form, the value of the smoothing parameter has a strong influence on the quality of the kernel estimator obtained. Thankfully, suitable algorithms have been developed, allowing the calculation of the value h based on random sample (1), which is close to optimal in a mean-square sense.

So, an approximation can be made – in the primary phase of the analysis – for the value obtained for the normal distribution

$$h=\left(\frac{V(K)}{U(K)^2}\,\pi^{n/2}\,\frac{2^{n+2}}{n+2}\,\frac{1}{m}\right)^{1/(n+4)}\bar{\sigma}\ , \tag{6}$$

where $\bar{\sigma}$ denotes the geometric mean of standard deviations for particular coordinates of the variable X, while $V(K)=\int_{\mathbb{R}^n} K(y)^2\,\mathrm{d}y$ and

$U(K) = \int_{\mathbb{R}^n} y^{\mathrm{T}} y\, K(y)\, \mathrm{d}y$. The cross-validation method is universal, whereby one calculates the value minimizing the real function $g : (0, \infty) \to \mathbb{R}$ defined as

$$g(h) = \frac{1}{m^2 h^n} \sum_{i=1}^{m} \sum_{j=1}^{m} \tilde{K}\left(\frac{x_j - x_i}{h} \right) + \frac{2}{m h^n} K(0) \quad , \tag{7}$$

when $\tilde{K}(x) = K^{*2}(x) - 2K(x)$, where K^{*2} denotes the convolution square of the function *K*. Applying iterative procedures to find the minimum, the initial value can be taken from formula (6).

For particular cases a number of appropriate algorithms have been worked out, such as the simple and effective plug-in method, used in the one-dimensional case.

2.3 Additional Procedures

The basic form of kernel estimator (2) can be generalized for overall improvement of its features as well as possibly extended to include additional aspects adapting the model to reality. As examples of such generalizations, the following procedures will be described:

- modification of the smoothing parameter (Section 2.4),
- linear transformation (Section 2.5),

while sample extensions will be presented in

- support boundary (Section 2.6),
- binary coordinates (Section 2.7).

2.4 Modification of Smoothing Parameter

The value for the smoothing parameter *h* introduced in definition (2) is the same for all kernels related to particular values of the random sample (1). Generally, a small value for this parameter causes a "thinning" of the kernel, whereas a large one, its "flattening". If therefore it is possible to individualize the influence of the smoothing parameter on specific kernels, then for areas "denser" with elements of a random sample, this value should be decreased (thereby better showing specific features of a distribution), as opposed to "sparser" areas, where it should be increased (which results in additional flattening of "tails"). With this aim modification of the smoothing parameter realizes the above according to the following algorithm.

Firstly one calculates the basic form of kernel estimator (2), then its values for particular elements of the random sample, i.e. $\hat{f}(x_1)$, $\hat{f}(x_2), \ldots, \hat{f}(x_m)$, as well as their geometrical mean $\bar{s}$. The modifying parameters $s_i > 0 \quad (i = 1,2,\ldots,m)$ are given by

$$s_i = \left(\frac{\hat{f}(x_i)}{\bar{s}} \right)^{-c} \quad , \tag{8}$$

while $c \geq 0$, and finally, the kernel estimator with modified smoothing parameter is defined as

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} \frac{1}{s_i^n} K\left(\frac{x - x_i}{hs_i} \right) \quad . \tag{9}$$

This formula constitutes the generalization of definition (2). If $c = 0$, then $s_i \equiv 1$ and consequently these dependences are equivalent. The parameter c refers to intensity of the modification procedure. Corollaries arising from the integrated mean-square error criterion show the value $c = 0{,}5$.

Apart from an obvious improvement in quality of estimation, kernel estimators with a modified smoothing parameter maintain a number of additional advantages in practical applications. Above all they are more robust with respect to imprecise calculation of a smoothing parameter value. What is more, application of this procedure decreases the difference in effectiveness of particular kernel types with respect to optimal Epanechnikov kernel (3), as well as lowering the difference in effectiveness between product kernel (5) with respect to radial kernel (5) in the multidimensional case. These features are especially valuable in practice, since additionally, they increase the possibility of selecting more suitable kernels for further analysis in concrete applicational tasks.

2.5 Linear Transformation

In the multidimensional case, when the radial kernel is used, one can apply a simple procedure fitting a kernel shape to the form of distribution of a random variable under investigation, based on the idea of linear transformation. The definition of radial kernel (4) then generalizes to the equality

$$K(x) = \frac{C}{\sqrt{\det(R)}} \mathscr{K}(\sqrt{x^{\mathrm{T}} R^{-1} x}) \quad , \tag{10}$$

where R is a positively defined matrix, while the meaning and value of the constant C remain unchanged. If R is a unique matrix, then the above formula is equivalent to dependence (4). Particularly useful results can be obtained by taking

$$R = \hat{\mathrm{Cov}} \quad , \tag{11}$$

where $\hat{\mathrm{Cov}}$ denotes an estimator of covariance of the random variable X.

2.6 Support Boundary

In practical problems, particular coordinates of the random variable X may represent various quantities. Many of these, including those dealing with distance or time, if

they are to be correctly and strictly interpreted, should belong to bounded subsets exclusively, for example the set of nonnegative numbers. In order to avoid calculational errors and misinterpretations arising from this, one can apply a simple procedure bounding a kernel estimator's support.

First, the case of a one-dimensional random variable and its left boundary – i.e. when the condition $\hat{f}(x) = 0$ for $x < x_*$, with $x_* \in \mathbb{R}$ fixed, is required – will be presented below. A part of any *i*-th kernel which finds itself beyond the interval $[x_*, \infty)$ becomes symmetrically "reflected" with respect to the boundary x_*, and is treated as a part of the kernel "grounded" in the symmetrical "reflection" of the element x_i with respect to the boundary x_*, so in the point $2x_* - x_i$. Therefore if one defines the function $K_{x_*} : \mathbb{R} \to [0, \infty)$ as

$$K_{x_*}(x) = \begin{cases} K(x) & \text{when} \quad x \geq x_* \\ 0 & \text{when} \quad x < x_* \end{cases}, \tag{12}$$

then the basic form of kernel estimator (2) takes the shape

$$\hat{f}(x) = \frac{1}{mh^n} \sum_{i=1}^{m} \left[K_{x_*}\left(\frac{x - x_i}{h} \right) + K_{x_*}\left(\frac{x + x_i - 2x_*}{h} \right) \right] . \tag{13}$$

Parts of particular kernels "cut" beyond an assumed support are thus "filled in" inside the support directly beside its boundary, therefore within the most commonly accepted – in practice – margin of error.

The case of a right boundary for a kernel estimator's support can be considered similarly. In the multidimensional instance, the above concept can be applied separately for each particular coordinate of the random variable under investigation.

2.7 Binary Coordinates

In many problems of contemporary engineering binary quantities appear, that is taking only two values, denoted symbolically hereinafter by 0 and 1. Besides quantities which are binary in nature, simplifications can often be applied to this form of description of even complicated phenomena, which however are of little influence to the final result. The methodology of kernel estimators enables binary coordinates to be taken into account.

Consider first the *k*-dimensional binary random variable $Y : \Omega \to \{0,1\}^k$. Its distribution is characterized by 2^k probabilities of the appearance of each possible *k*-dimensional vector of binary values. In many practical tasks, one can also infer with respect to values for the probability of the appearance of a given vector with the help of observations of vectors "similar" to it. The greater the "similarity", the likelier the inference becomes. Let therefore the function $p : \{0,1\}^k \to [0,1]$ be given, mapping to every *k*-dimensional vector of binary values the probability of its appearance. The

kernel estimator of this function $\hat{p}:\{0,1\}^k \to [0,1]$ is calculated based on values of the *m*-elements random sample

$$y_1, y_2, \ldots, y_m \tag{14}$$

obtained from the variable *Y*, and is given as

$$\hat{p}(y) = \frac{1}{m}\sum_{i=1}^{m} L(y, y_i) \quad , \tag{15}$$

while the function $L:\{0,1\}^k \times \{0,1\}^k \to [0,1]$ defines the dependence

$$L(y, y_i) = \delta^{k-d(y,y_i)}(1-\delta)^{d(y,y_i)} \quad , \tag{16}$$

where $0{,}5 \le \delta \le 1$ and $0^0 = 1$ can be assumed, whereas the function $d:\{0,1\}^k \times \{0,1\}^k \to \mathbb{N}$ is given by $d(y, y_i) = (y - y_i)^{\mathrm{T}}(y - y_i)$. The value $k - d(y, y_i)$ shows the number of coordinates, to which the vectors *y* and y_i are equal, and represents the aforementioned "similarity" of binary vectors. The function *L* takes the role filled in definition (2) by the kernel *K*, and is called a binary kernel, while the parameter δ is termed a binary smoothing parameter. In practice its value can be calculated by minimizing the function $g:[0{,}5;1] \to \mathbb{R}$ given by formula

$$g(\delta) = -\sum_{i=1}^{m} \log[\hat{p}_{-i}(y_i)] \quad , \tag{17}$$

where $\hat{p}_{-i}$ denotes estimator (15) obtained using the random sample $y_1, y_2, \ldots, y_{i-1}, y_{i+1}, \ldots, y_m$, therefore, omitting the *i*-th element of sample (14).

And finally, the (*n*+*k*)-dimensional random variable $Z \equiv [X, Y]^{\mathrm{T}}$ will be considered, being a composition of the *n*-dimensional continuous variable *X* investigated so far and the above defined *k*-dimensional binary variable *Y*. If the kernel estimator $\hat{f}:\mathbb{R}^n \times \{0,1\}^k \to [0,\infty)$ for the variable *Z* is calculated using the *m*-elements random sample

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix}, \begin{bmatrix} x_2 \\ y_2 \end{bmatrix}, \ldots, \begin{bmatrix} x_m \\ y_m \end{bmatrix} \quad , \tag{18}$$

then the basic form of kernel estimator (2) becomes

$$\hat{f}(x, y) = \frac{1}{mh^n}\sum_{i=1}^{m} K\left(\frac{x - x_i}{h}\right) L(y, y_i) \quad . \tag{19}$$

2.8 Sample Size

The last parameter requiring discussion is the crucial sample size, especially its dependence on the dimension of the random variable under investigation. The sample size m_* necessary to ensure a 10% accuracy at point zero for the *n*-dimensional standard normal distribution can be more or less taken as $m_* = 4^n$. Because of the exceptional regularity of the above distribution, and the significant looseness of the above criterion, such values seem to constitute an absolute minimum (suggested by e.g. $m_* = 4$ for $n = 1$). In the presence of binary coordinates these values should be multiplied by $(3/2)^k$. Next, the obtained result is multiplied by heuristically determined coefficients arising from the necessity to improve the quality of estimation, multimodality and asymmetry of the distribution, and also the correlation of random sample elements – in practice the product of these coefficients is most often equal to 3-10, in extreme cases even 100. For the one-dimensional random variable, the desired sample size in reality amounts to 20-50, increasing accordingly with the growth of a variable dimension. However, thanks to modern computer technology, even where multidimensional and complex tasks with adverse distribution features appear, this does not necessarily prove to be a significant applicational obstacle nowadays. One must always take into account the considerable advantages resulting from the application of kernel estimators, as they enable the identification of practically any distribution appearing in applicational problems, although they require a sample size adequate for immensity and universality of the information contained within.

2.9 Notes and Comments

In Section 2 a compendium of a kernel estimators methodology has been described. The basic form of definition of this type of estimator found in equality (2) was formulated, as were procedures for fixing quantities appearing therein. This form was generalized by the concept of smoothing parameter modification (9) and linear transformation (10), improving the properties of the estimator obtained. Additional procedures enabling the boundary of its support (13) and including binary coordinates (19) were presented, which allows the obtained model to be better fitted to the examined reality many times over. The above generalizations and supplementations were formulated with respect to the basic form of the kernel estimator, which makes it possible to combine them according to the demands of the concrete applicational problem. Because of clear interpretation, the introduction for additional generalizations is also possible. As an example – for the needs of further investigations in Section 3 – basic form (2) can be enhanced by nonnegative coefficients w_i with $i = 1,2,...,m$, and not all equal zero, which can be interpreted as the "meaning" of particular elements of random sample (1). Then

$$\hat{f}(x) = \frac{1}{h^n \sum_{i=1}^{m} w_i} \sum_{i=1}^{m} w_i K\left(\frac{x - x_i}{h}\right) \quad . \tag{20}$$

If $w_i \equiv 1$, the above formula is equivalent to definition (2).

Kernel estimators allow modeling of the distribution density – a basic functional characteristic of one- and multidimensional random variables – practically independent of its form and features. Consequently this is fundamental to obtain other functional characteristics and parameters. For example, if in a one-dimensional case one chooses the kernel K, such that its primitive $I(x) = \int_{-\infty}^{x} K(y)\,\mathrm{d}x$ may be analytically obtained, then the estimator of the distribution function

$$\hat{F}(x) = \frac{1}{m}\sum_{i=1}^{m} I\left(\frac{x - x_i}{h}\right) \tag{21}$$

can be defined. Next, if the kernel K has positive values, the solution for the equation

$$\hat{F}(x) = r \tag{22}$$

constitutes the kernel estimator of quantile of order $r \in (0,1)$.

Similar to the estimator of density of random variable distribution, the concept of the kernel estimator of spectral density can be formulated, as can be a particularly interesting notion – from an applicational point of view – the kernel estimator of regression function. According to the general nature of kernel estimators, this function is obtained without arbitrary assumptions fixing its shape, for example as linear or logarithmical.

3 Sample Industrial Applications

The universal character of kernel estimators allows their broad application in a variety of contemporary problems in science and practice. This process is possible thanks to today's modern computer systems, ubiquitous and ever more powerful, as well as the automation of procedures for both measuring and data gathering. Below are shown sample applications of kernel estimators in the following contemporary industrial tasks:

- data analysis and exploration – recognition of atypical elements (outliers), clustering, and classification – applied to the detection and diagnosis of devices working in real time, and planning of strategy for mobile phone operators (Section 3.1);
- parametric identification illustrated in automatic control applications and by tasks of sharpening of imprecise information (Section 3.2);
- definition of spatial distribution of demand based on fuzzy data for the needs of a problem from the telecommunications field (Section 3.3).

3.1 Data Analysis and Exploration

Here, the application of kernel estimators in basic tasks of data analysis and exploration will be considered, as will the recognition of atypical elements (outliers), clustering, and classification, and also the action of these procedures in the fault detection and diagnosis of devices working in real-time, and planning of strategy for mobile phone operators.

First, in many problems of data analysis the task of recognizing atypical elements (outliers) – those which differ greatly from the general population – arises. This enables the elimination of such elements from the available set of data, which increases its homogeneity (uniformity), and facilitates analysis, especially in complex and unusual cases. In practice, the recognition process for outliers is most often carried out using procedures of testing for statistical hypotheses (Barnett and Lewis 1994). The significance test based on the kernel estimators methodology will now be described.

Let therefore the random sample $x_1,\ x_2,\ldots,x_m$ treated as representative, therefore including a set of elements as typical as possible, be given. Furthermore, let $r\in(0,1)$ denote an assumed significance level. The hypothesis that $\tilde{x}\in\mathbb{R}^n$ is a typical element will be tested against the hypothesis that it is not, i.e. one should denote it as an outlier. The statistic $S:\mathbb{R}^n\to[0,\infty)$, used here, can be defined as

$$S(\tilde{x})=\hat{f}(\tilde{x})\quad , \tag{23}$$

where $\hat{f}$ denotes a kernel estimator of density, obtained for the random sample x_1, $x_2,\ldots,x_m$ mentioned above, while the critical set takes the left-sided form

$$A=(-\infty,a]\quad , \tag{24}$$

when a constitutes the kernel estimator of quantile of order r, calculated for the sample $\hat{f}(x_1),\ \hat{f}(x_2),\ldots,\hat{f}(x_m)$, with the assumption that random variable support is bounded to nonnegative numbers.

Secondly, the aim of clustering is the division of a data set – for example given in the form of the random sample $x_1,\ x_2,\ldots,x_m$ – into subgroups (clusters), with every one including elements "similar" to each other, but with significant differences between particular subgroups (Hand et al. 2001, Larose 2005). In practice this often allows the decomposition of large data sets with differing characteristics of elements, into subsets containing elements of similar properties, which considerably facilitates further analysis, or even makes it possible at all. The following clustering procedure based on kernel estimators, taking advantage of the gradient methods concept (Fukunaga and Hostetler 1975) will be presented now.

Here the natural assumption is made, that clusters are prescribed to modes (local maximums) of the density kernel estimator $\hat{f}$, calculated for the considered random sample $x_1,\ x_2,\ldots,x_m$. Within this procedure, particular elements are moved in a direction defined by a gradient, according to the following iterative algorithm:

$$x_j^0=x_j\quad \text{for}\quad j=1,2,\ldots,m \tag{25}$$

$$x_j^{k+1}=x_j^k+b\frac{\nabla\hat{f}(x_j^k)}{\hat{f}(x_j^k)}\quad \text{for}\quad j=1,2,\ldots,m\ \text{ and }\ k=0,1,\ldots\ , \tag{26}$$

where $b>0$ and ∇ denotes a gradient. Thanks to the proper choice of form of the kernel K, a suitable analytical formula for the gradient ∇ becomes possible. In practice the value $b=h^2/(n+2)$ can be recommended.

As a result of the following iterative steps, the elements of the random sample move successively, focusing more and more clearly on a certain number of clusters. Their final shape can be defined after completing the k^*-th step. To this end one may calculate the kernel estimator for mutual distances of the elements $x_1^{k^*}$, $x_2^{k^*}$,…, $x_m^{k^*}$ (under the assumption of nonnegative support of the random variable), after which the value is found where this estimator takes on the lowest local minimum, omitting a possible minimum in zero point. Particular clusters are assigned those elements whose distance is not greater than the above value. Thanks to the possibility of change in the smoothing parameter value, it becomes possible to affect the range of a number of obtained clusters, albeit without arbitrary assumptions concerning the strict value of this number, which allows it to be suited to a true data structure. Moreover, possible changes in intensity of the smoothing parameter modification – defined in formula (8) by the constant c – allows influence on the proportion of clusters located in dense areas of random sample elements to the number of clusters on the "tails" of the distribution under investigation.

Thirdly, the application of kernel estimators in a classification task is considered. Let the number $J\in\mathbb{N}\setminus\{0,1\}$ be given. Assume also, that the sample $x_1,\ x_2,\ldots,x_m$ obtained from the n-dimensional random variable has been divided into J separate subsets

$$x_1,\ x_2,\ldots,x_{m_1} \tag{27}$$

$$x_1,\ x_2,\ldots,x_{m_2} \tag{28}$$

$$\vdots$$

$$x_1,\ x_2,\ldots,x_{m_J}\quad , \tag{29}$$

while $m_1,m_2,\ldots,m_J\in\mathbb{N}\setminus\{0\}$ and $\sum_{j=1}^{J}m_j=m$, representing classes with features as mutually different as possible. The classification task requires deciding into which of them the given element $\tilde{x}\in\mathbb{R}^n$ should be reckoned (Hand et al. 2001, Larose 2005).

The kernel estimators methodology provides a natural mathematical tool for solving the above problem in the optimal Bayes approach. Let thus $\hat{f}_1,\ \hat{f}_2,\ldots,\hat{f}_J$ denote kernel estimators of density calculated for samples (27)-(29), respectively. If sizes $m_1,\ m_2,\ldots,m_J$ are proportional to the "frequency" of appearance of elements

from particular classes, the considered element $\tilde{x}$ should be reckoned into the class for which the value $m_1\hat{f}_1(\tilde{x})$, $m_2\hat{f}_2(\tilde{x}),\ldots,m_J\hat{f}_J(\tilde{x})$ is the greatest.

The applicational possibilities of the above-presented procedures will now be illustrated in examples from the areas of fault detection and diagnosis for devices working in real-time, and a mobile phone operator's strategy.

Thus, the fault detection and diagnosis problem has lately become one of the most important challenges in modern control engineering. Early discovery of anomalies appearing in the operation of a controlled system, from an industrial robot to a nuclear reactor (i.e. detection), most often allows serious incidents and even catastrophes to be avoided, which could save material damage, or loss of human life. Secondly, confirmation of kind and location of these anomalies (i.e. diagnosis) is of fundamental meaning, especially when supervising large systems like complex chemical installations, as well as modern ships and airplanes. The importance of the above actions is multiplied by a psychological factor expressed by an increased feeling of safety, as well as – for the producer – prestige and commercial reputation. Finally, economic reasons often translate into a significant decrease in running costs, above all by ensuring the proper technological conditions as well as rationalizing overhauls and reducing repairs. Among the many different procedures used with this aim, the most universal are statistical methods. These very often consist in generating a certain group of variables that characterize the state of technical performance of the device, and then making a statistical inference, on the basis of their current values, as to whether or not the device is working properly, and in the event of a negative response, on the nature of the anomalies.

The procedures presented in this section, of recognition of atypical elements (outliers), clustering, and classification provides an appropriate tool for constructing an effective and suitable algorithm for use. If therefore a random sample represents conditions regarded as typical, signifying the correct operation of a device, and $\tilde{x}$ its current state, then, using the procedure of recognizing outliers it can be confirmed whether this state also should be considered as typical, or rather showing the appearance of anomalies (detection). In this case then, with random samples (27)-(29) characterizing particular kinds of typical faults, applying the classification procedure, one can tell which of them is being dealt with. If the partition of elements describing different kinds of anomalies is ambiguous, the appropriate division into classes can be obtained by clustering.

The next example of a direct application of data analysis and exploration procedures using kernel estimators is the planning of the strategy for mobile phone operators.

So, the continuously high rate of growth on the global mobile phone market forces the development of analytical methods, which serve to precisely specify the needs of an ever increasing group of users of this service. All mobile phone operators now have offers which differ less and less, both in services provided to clients and their obligatory tariffs. The market is consequently dividing with the aim of considering the various customer requirements. This forces effective realization of a company's strategy for satisfying these needs, while at the same time maximizing profits. However, continuation of such a process could lead to excessive segmentation through characterizing subscriber groups, which results in a loss of coherence in strategy with respect to clients. To avoid this one should find applications for new solutions of a

systematic form. These allow a given group of mobile phone users to be divided in such a way that specifying a strategy with respect to the newly-formed subgroups does not require the whole procedure to be carried out again, yet uniquely defines the type of action for a given subgroup which maximizes an operator's profits, while on the other hand satisfying the client. In summary, among many factors which characterize every client one should select a set of such quantities which exclusively assign a subscriber to a specific, previously defined group of mobile service users, and specify a coherent strategy for each of them.

The above tasks, to which the methodology presented above will be applied, concern the business market of a particular mobile phone operator. The aim of the research is to define the appropriate strategy for a given client, taking into account such factors as – for example – mean monthly income from each SIM-card, length of subscription, and number of active SIM-cards. Based on the model constructed with the statistical kernel estimators methodology, the selected group of firms undergoes division into subgroups (clusters) characterized by the above factors. Both before and after clustering, those elements which are noted to be atypical (outliers) are excluded from the sample. The obtained division of clusters will result in the possibility of defining a concrete, preferred strategy with respect to each of them. Next, the algorithm enables the classification of any one firm involved in negotiations into the closest cluster with respect to its given characteristics, while also allowing the specification of an appropriate strategy. This will consist of the proper use of discounts which the operator can apply with regard to the client, all the while monitoring the costs incurred. Depending on which solution is applied concerning the subscriber, the risk of their leaving the given operator changes. On the other hand, it is important to find the threshold, where it is still profitable to continue providing a discount with the goal of keeping the client. In the conditions of fierce competition on the mobile phone market, the proper analysis and exploration of information contained in a client database not only allows for it to be effectively maintained, with the appropriate steering of the operator's development, but also becomes a source of information and solutions crucial to the acquisition of new clients.

It is worth mentioning the possibility of applying data compensation and dimensionality reduction procedures for the methodology described in this section. These among others efficiently lower calculation time in the above presented tasks.

In practice some elements of samples representing particular classes may be of little importance – from the point of view of propriety of statistical inference – or even contribute to mistakes. Their removal should therefore result in a reduction in the number of inappropriate decisions and also an increase in calculation speed. In order to realize this task one can apply the sensitivity method, patterned on the artificial neural networks theory.

For similar reasons the suggestion arises to reduce the dimensionality of the considered random variable X by the linear transformation

$$X^* = AX \tag{30}$$

mapping the space $\mathbb{R}^n$ onto the space $\mathbb{R}^{n^*}$ with significantly smaller dimension $n^* < n$. The elements of the matrix A should be selected such that the distances of particular random sample elements in both spaces are as mutually close as possible.

Since the number of these elements amounts to $n \cdot n^*$, which in practice makes using classical optimization methods impossible, then evolutionary algorithms can be proposed for the above task.

The above section contains results presented in the book (Kulczycki 1998), as well as material from research carried out together with Cyprian Prochot and Karina Daniel, published in the common papers (Kulczycki and Prochot 2004) and (Kulczycki and Daniel 2006), and also recently commenced works with Piotr Andrzej Kowalski and Szymon Lukasik.

3.2 Parameter Identification

One of the main problems of systems engineering is parameter identification – the specification of values of parameters existing in an investigated model. In a typical practical task, *m* independent measurements of the parameter, although suffering from errors of different origin, are available. On this basis one should define the value which, from an overall point of view of the problem to be worked out, would best represent phenomena described by this parameter. Usual estimation procedures, based on minimum integrated mean-square error or maximum likelihood methods, are applied mostly because of their popularity and availability in subject literature, however they do not allow differing causes of estimation errors to be taken into account.

This problem will be illustrated for the example of optimal control. The performance index, fundamental for the above task, may be used for testing not only the quality of a control, but also the procedure of identifying model parameters. As an example let the system, whose dynamic is described by the following differential equation

$$\dot{x}(t) = \begin{bmatrix} v & 1 \\ 0 & v \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ v \end{bmatrix} u(t) \quad , \tag{31}$$

while $v \in \mathbb{R} \setminus \{0\}$, be given. If the optimal feedback controller with quadratic performance index has been constructed with the value $V \in \mathbb{R} \setminus \{0\}$ not necessarily equal to the parameter v existing in object (31), then the obtained graph for this index can be approximated with reasonable precision by a quadratic function where coefficients differ for negative and positive errors (Fig. 2). Treating obtained values of an examined parameter as realizations of a random variable with the distribution density f, one can calculate the value of the optimal – in the sense of minimum expectation value of performance index for the control – estimator using Bayes' decision rule (Berger 1980). If the distribution of the above random variable is obtained with the aid of kernel estimators, the algorithm worked out is suitable for calculational procedures and in consequence for practical use.

First, a basic case will be investigated, considering the single parameter *v*. As a set of possible decisions $D = \mathbb{R}$ can be assumed, while the loss function is given as the asymmetrical quadratic form:

$$l(\hat{v}, v) = \begin{cases} p(\hat{v} - v)^2 & \text{for} \quad \hat{v} - v \le 0 \\ q(\hat{v} - v)^2 & \text{for} \quad \hat{v} - v \ge 0 \end{cases} , \tag{32}$$

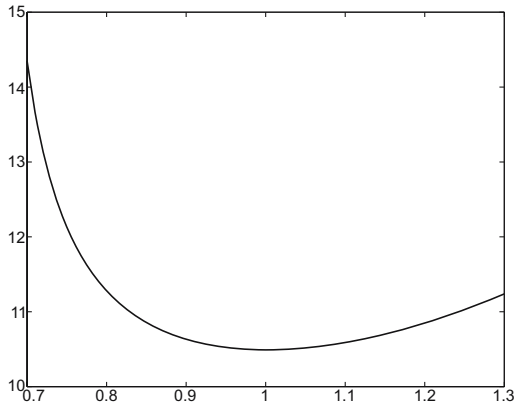

Fig. 2. Performance index value as a function of the parameter V ; ($v=1$)

where $p, q > 0$ (note that these coefficients can be different), and $\hat{v}$ denotes the desired value of the Bayes decision which here fulfills the role of optimal estimator. This value is then given as the solution of the following equation with the argument $\hat{v}$:

$$(p-q)\int_{-\infty}^{\hat{v}}(\hat{v}-v)f(v)\,\mathrm{d}v = p\int_{-\infty}^{\infty}(\hat{v}-v)f(v)\,\mathrm{d}v \quad . \tag{33}$$

Solving the above criterion is generally no easy task. If, however, the kernel estimators methodology is used in specifying the density *f*, then, thanks to the proper choice of the kernel form, the effective numerical algorithm can be obtained. Let therefore *m* measurements for the examined parameter be given, treated as random sample (1). For the chosen kernel *K* one may define the following real mappings:

$$I(x) = \int_{-\infty}^{x} K(y)\,\mathrm{d}y \tag{34}$$

$$J(x) = \int_{-\infty}^{x} yK(y)\,\mathrm{d}y \quad , \tag{35}$$

and (for the basic form of kernel estimator (2)):

$$\bar{I}(x) = \frac{1}{mh}\sum_{i=1}^{m} I\left(\frac{x-x_i}{h}\right) \tag{36}$$

$$\bar{J}(x) = \frac{1}{mh}\sum_{i=1}^{m} J\left(\frac{x-x_i}{h}\right) \quad . \tag{37}$$

Then, criterion (32) takes the form of the equation

$$(p-q)[\hat{v}\bar{I}(\hat{v})-\bar{J}(\hat{v})]-p\hat{v}=-p\frac{1}{m}\sum_{i=1}^{m}x_i \quad . \tag{38}$$

If the kernel K assumes (strictly) positive values, the above solution exists and is unique. Denoting the left and right sides of the above equation as follows

$$L(\hat{v})=(p-q)[\hat{v}\bar{J}(\hat{v})-\bar{I}(\hat{v})]-p\hat{v} \tag{39}$$

$$P=-p\frac{1}{m}\sum_{i=1}^{m}x_i \quad , \tag{40}$$

and calculating the derivative of function (39):

$$L'(\hat{v})=(p-q)\bar{J}(\hat{v})-p \quad , \tag{41}$$

one can then – using Newton's iterative method – effectively obtain a solution for criterion (32) as a limit of the sequence $\{v^{(i)}\}_{i=1}^{\infty}$ given by the formulas

$$v^{(0)}=\frac{1}{m}\sum_{i=1}^{m}x_i \tag{42}$$

$$v^{(k+1)}=v^{(k)}+\frac{P-L(v^{(k)})}{L'(v^{(k)})} \qquad \text{for } k=0,1,\ldots \quad . \tag{43}$$

The properties of kernel estimators allow generalizations of the above concept to be made for the multidimensional (parameters' vector), conditional (dependence on conditional factors), and polynomial (loss function to a power greater than two) cases.

Thus, in the multidimensional case, i.e. the task of a parameters' vector estimation, it is possible to carry out the above-proposed procedure with respect to a multidimensional random variable. As an example, if for the two-dimensional parameters' vector $v=[v_1,v_2]^{\mathrm{T}}$ one assumes the loss function in the following asymmetrical quadratic form:

$$l\left(\begin{bmatrix}\hat{v}_1\\ \hat{v}_2\end{bmatrix}\begin{bmatrix}v_1\\ v_2\end{bmatrix}\right)=\begin{cases}c_l(\hat{v}_1-v_1)^2+c_{ld}(\hat{v}_1-v_1)(\hat{v}_2-v_2)+c_d(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\le 0 \text{ i } \hat{v}_2-v_2\le 0 \\ c_p(\hat{v}_1-v_1)^2+c_{pd}(\hat{v}_1-v_1)(\hat{v}_2-v_2)+c_d(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\ge 0 \text{ i } \hat{v}_2-v_2\le 0 \\ c_l(\hat{v}_1-v_1)^2+c_{lg}(\hat{v}_1-v_1)(\hat{v}_2-v_2)+c_g(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\le 0 \text{ i } \hat{v}_2-v_2\ge 0 \\ c_p(\hat{v}_1-v_1)^2+c_{pg}(\hat{v}_1-v_1)(\hat{v}_2-v_2)+c_g(\hat{v}_2-v_2)^2 \\ \qquad\qquad \text{when } \hat{v}_1-v_1\ge 0 \text{ i } \hat{v}_2-v_2\ge 0\end{cases} , \tag{44}$$

where $c_l, c_p, c_g, c_d > 0$, $c_{ld}, c_{pg} \geq 0$ and $c_{pd}, c_{lg} \leq 0$, then here criterion (32) takes on the form of two equations, defining Bayes' decisions $\hat{v}_1$ and $\hat{v}_2$ of similar, albeit slightly more complex form.

In the case where the examined parameter depends heavily on a conditioning factor, the proposed procedures can be generalized with the purpose of including its influence. Let therefore the conditioning factor be given in the form of the random variable Z, taking the form of the composition of an *n*-dimensional continuous variable, whose distribution has a density, and a *k*-dimensional binary variable. Statistical kernel estimators can be used to estimate the density of total distribution. During application of the procedure, after fixing the concrete value of the conditional factor $Z(\omega)$, criterion (33) can be applied to a "cross-section" defined by this value. When significant conditioning for factors represented by variable Z occurs to an object under consideration, such an approach can considerably improve the quality of received results.

The form of the asymmetrical quadratic loss function (32) may be generalized to a polynomial case:

$$l(\hat{v}, v) = \begin{cases} (-1)^k \, p(\hat{v} - v)^k & \text{for} \quad \hat{v} - v \leq 0 \\ q(\hat{v} - v)^k & \text{for} \quad \hat{v} - v \geq 0 \end{cases}, \tag{45}$$

while $p,q > 0$ and $k = 2,3,...$. For instance when $k = 3$, a criterion analogical to equation (33) becomes

$$(p+q) \int_{-\infty}^{\hat{v}} (\hat{v} - v)^2 f(v) \, \mathrm{d}v = \mathrm{q} \int_{-\infty}^{\infty} (\hat{v} - v)^2 f(v) \, \mathrm{d}v \quad . \tag{46}$$

The above text contains material from research carried out together with Aleksander Mazgaj, published in the common paper (Kulczycki and Mazgaj 2005).

Similar investigations concerning an asymmetrical linear loss function, i.e. in the case when formula (32) is replaced with the following dependence

$$l(\hat{v}, v) = \begin{cases} -p(\hat{v} - v) & \text{for} \quad \hat{v} - v \leq 0 \\ q(\hat{v} - v) & \text{for} \quad \hat{v} - v \geq 0 \end{cases}, \tag{47}$$

where $p, q > 0$ (while these coefficients can differ) were used to sharpen imprecise information in research carried out together with Malgorzata Charytanowicz, and published in the common paper (Kulczycki and Charytanowicz 2005).

3.3 Definition of Spatial Distribution of Demand

The Local Multipoint Distribution System (LMDS) is applied by telecommunications operators for wireless broadband data transmission purposes, which to a large degree results from a radical rise in the demand for Internet access. This system allows to connect the operator's network node to the buildings in which customers are located, without the necessity of constructing an expensive cable infrastructure. Thus, data is transmitted between base-stations, distributed across a metropolitan area, and providing regular

connections with subscriber-stations located within the effective coverage of transceivers belonging to base-stations. Subscriber-stations installed on building roofs or facades then transmit data to customers through local (e.g. cable) networks. An essential factor which often decides about the economic justification of the LMDS system implementation is to determine base-station locations so that the highest profit can be achieved within the available investment funds. In this section an algorithm for designing an optimal system of LMDS base-stations will be presented.

The distribution of the spatial demand for data transmission services in the area under consideration has strict point structure regarding particular potential customers. Such a model has, however, limited applicational significance, since it is practically unidentifiable in a metropolitan area. Here, statistical kernel estimators will be applied for this purpose. The variable X existing in their definition is therefore two-dimensional, with its particular coordinates representing latitude and longitude. Thus, once in possession of a data base composed of m potential locations of subscriber buildings, with each of them characterized by its geographical position $x_i = [x_{i1}, x_{i2}]^{\mathrm{T}}$ and the coefficient w_i representing potential demand for data transmission services corresponding to this location (see formula (20)), one can calculate the kernel estimator describing the density spatial distribution of demand in a given area. Such identified distribution becomes continuous due to the properties of statistical kernel estimators. Moreover, thanks to the averaging aspects of such estimators, it is possible to use a simplified data base, including only the locations of main subscriber buildings, and also taking smaller objects in their neighborhood into account in the corresponding coefficients w_i. This considerably facilities the most difficult and expensive phase of the procedure of planning optimal locations of LMDS base-stations.

In practice, it is not difficult to point out a limited number of k sites for installing base-stations, including e.g. tall buildings and telecommunications towers. When one possesses the kernel estimator $\hat{f}$ characterizing the spatial distribution of demand in the area under investigation, for any such location x_j ($j = 1,2,...,k$) the performance index can be defined as

$$E_j = \int_{C_j} \hat{f}(x)\,\mathrm{d}x \quad , \tag{48}$$

where C_j denotes a circle with the center at x_j and the positive radius r_j representing maximum range of the transceiver mapped to this location. Thanks to the proper form of the kernel it is possible to describe this value with an analytical formula. It should be underlined that, from the point of view of the optimization problem investigated here, the positive constant $1/(h^2 \sum_{i=1}^{m} w_i)$ present in definition (20) may be omitted.

Based on the above dependence one can calculate the performance index for any system of base-stations with specific types of transceiver (or lack thereof, where installation of such equipment is not foreseen) mapped to particular locations. The

proposed procedure also allows "shaded" areas (where transmission is not possible because of uneven terrain or obstacles, e.g. tall buildings) to be taken into account, as well as limited bitrate of equipment. Namely, in very attractive districts, demand for coverage by particular transceivers can not be satisfied by one base-station and so some of the unsatisfied demand should be met by another base-station within range – the optimal division is calculated based on the classic linear optimization methods, in particular the simplex algorithm.

With possession of the performance index for a fixed base-stations system created above, one can calculate – using operational research procedures, in particular the branch and bound method – their optimal system within the available investment budget. It is also possible to expand the problem to the task of planning over a few years, with changing conditions.

The coefficients w_i for $i = 1,2,...,m$ represent the demand for teletransmission services assigned to particular subscriber-station locations. Their identification is in practice conducted on the basis of an expert opinion expressed verbally, often based on intuitional premise. Consequently, the description of the predicted demand for teletransmission services by a subscriber-station will require fuzzy logic elements. The task with fuzzy character of the coefficients w_i introduced in definition (20) is commented below in detail, as such a problem can arise in various applications of kernel estimators in engineering challenges. What should be taken into account here is the specific nature of the task under consideration: a lot of fuzzy numbers (equal to the sample size *m*) necessary to identify and to use in subsequent analysis, as well as the fact that incidentally, the coefficients w_i may be deterministic owing to previously executed agreements.

In this situation, especially suitable are fuzzy numbers of the type *L-R*, whose membership function is denoted here in the following form:

$$\mu_{(w_i,\alpha_i,\beta_i)}(x) = \begin{cases} L\left(\dfrac{w_i - x}{\alpha_i}\right) & \text{for} \quad x \le w_i \\ R\left(\dfrac{x - w_i}{\beta_i}\right) & \text{for} \quad x \ge w_i \end{cases}, \tag{49}$$

where $w_i \in \mathbb{R}$ and $\alpha_i, \beta_i > 0$, while the real functions *L* and *R* are symmetrical with respect to zero, assume here the value 1 and are nondecreasing within the interval $(-\infty, 0]$. The parameter w_i may therefore be interpreted as a modal value, while α_i and β_i describe left- and right-hand concentration around that value, respectively. The fuzzy number $\mathcal{A}$ of the type *L-R* may, therefore, be identified by three parameters, which will be denoted as $\mathcal{A} = (w, \alpha, \beta)$, and, consequently, the process of identification requires only to determine the values which are close to intuitional interpretation. Algebraic operations on fuzzy numbers of the type *L-R* are defined as follows:

$$\mathcal{A} + \mathcal{B} = (w_{\mathcal{A}}, \alpha_{\mathcal{A}}, \beta_{\mathcal{A}}) + (w_{\mathcal{B}}, \alpha_{\mathcal{B}}, \beta_{\mathcal{B}}) = (w_{\mathcal{A}} + w_{\mathcal{B}}, \alpha_{\mathcal{A}} + \alpha_{\mathcal{B}}, \beta_{\mathcal{A}} + \beta_{\mathcal{B}}) \tag{50}$$

$$\mathcal{A}-\mathcal{B}=(w_{\mathcal{A}},\alpha_{\mathcal{A}},\beta_{\mathcal{A}})-(w_{\mathcal{B}},\alpha_{\mathcal{B}},\beta_{\mathcal{B}})=(w_{\mathcal{A}}-w_{\mathcal{B}},\alpha_{\mathcal{A}}+\alpha_{\mathcal{B}},\beta_{\mathcal{A}}+\beta_{\mathcal{B}}) \tag{51}$$

$$c\cdot\mathcal{A}=(cw_{\mathcal{A}},c\alpha_{\mathcal{A}},c\beta_{\mathcal{A}}) \quad , \tag{52}$$

where $\mathcal{A}$ and $\mathcal{B}$ denote fuzzy numbers, while c is a positive real number. If one adopts the notation in which the real number a is described in the form of three parameters $a=(a,0,0)$, those operations may be generalized to addition and subtraction of the fuzzy and real numbers. Moreover, formulas (50)-(52) also correctly express the operations on two real numbers. Finally, the result is that the fuzzy number of the type *L-R* in the above range is a generalization of the real number. In the end, each coefficient w_i introduced in formula (20), was generalized to the three-parameter fuzzy number suitable for identification and calculation in practice, denoted below as $\mathcal{W}_i=(w_i,\alpha_i,\beta_i)$. In a special case, $\mathcal{W}_i=(w_i,0,0)$ may represent the real (nonfuzzy) number w_i.

If the coefficients w_i are fuzzy, then the performance index of the base-station system under consideration has a form of linear combination of three-parameter fuzzy numbers $\mathcal{W}_i$, and, therefore, due to formulas (50)-(52), it also becomes a three-parameter fuzzy number, denoted below as $\mathcal{E}$. To allow for the comparison of qualities of particular base-station systems, the methodology of fuzzy preference theory (Fodor and Rubens 1994) will be applied. The preference function $\mathcal{P}$ of the fuzzy number $\mathcal{E}$, with the bounded support of the membership function, will be adopted in the form resulting from the decision-making practice (Berger 1980):

$$\mathcal{P}(\mathcal{E})=\delta\frac{\int_{\min\operatorname{supp}\mu_{\mathcal{E}}}^{\max\operatorname{supp}\mu_{\mathcal{E}}}x\mu_{\mathcal{E}}(x)\,dx}{\int_{\min\operatorname{supp}\mu_{\mathcal{E}}}^{\max\operatorname{supp}\mu_{\mathcal{E}}}\mu_{\mathcal{E}}(x)\,dx}+(1-\delta)\min\operatorname{supp}\mu_{\mathcal{E}} \quad , \tag{53}$$

where $\delta\in[0,1]$, $\mu_{\mathcal{E}}$ means the membership function of the fuzzy number $\mathcal{E}$, while $\operatorname{supp}\mu_{\mathcal{E}}$ denotes its support. The value of the membership function is therefore a linear combination with weights δ and $1-\delta$ of the average value of the fuzzy number and the minimum value of its support. The average number corresponds to the Bayes decision rule and expresses a "realistic" operation, while the minimum value of the membership function support results from the minimax rule and represents the "pessimistic" point of view. The parameter δ determines therefore the company's strategy in the range from realistic (assuming average predicted demand) for $\delta=1$, to pessimistic (assuming the lowest level of predicted demand) for $\delta=0$. When clear preferences are missing, the value $\delta=0{,}5$ can be proposed.

Finally, when two base-station systems characterized by fuzzy performance indexes are considered, the one for which the preference function (53) is larger should be treated as the "better".

The above section contains material from research carried out together with Jacek Waglowski, published in the common paper (Kulczycki and Waglowski 2005).

References

Barnett, V., Lewis, T.: Outliers in Statistical Data. Wiley, Chichester (1994)

Berger, J.O.: Statistical Decision Theory. Springer, New York (1980)

Fodor, J., Roubens, M.: Fuzzy Preference Modelling and Multicriteria Decision Support. Kluwer, Dordrecht (1994)

Fukunaga, K., Hostetler, L.D.: The estimation of the gradient of a density function, with applications in Pattern Recognition. IEEE Transactions on Information Theory 21, 32–40 (1975)

Hand, J.H., Mannila, H., Smyth, P.: Principles of Data Mining. MIT Press, Cambridge (2001)

Kulczycki, P.: Wykrywanie uszkodzen w systemach zautomatyzowanych metodami statystycznymi. Alfa, Warsaw (1998)

Kulczycki, P.: Fuzzy Controller for Mechanical Systems. IEEE Transactions on Fuzzy Systems 8, 645–652 (2000a)

Kulczycki, P.: A Statistical Fault Detection System. In: 16th IMACS World Congress on Scientific Computation. Applied Mathematics and Simulation, Lausanne, August 21-25, 2000, pp. CD: 216–214 (2000b)

Kulczycki, P.: An Algorithm for Bayes Parameter Identification. Journal of Dynamic Systems, Measurement, and Control 123, 611–614 (2001a)

Kulczycki, P.: A Fuzzy Approach for Mechanical System with Uncertain Load. In: European Control Conference, Porto, September 4-7, 2001, pp. 2658–2663, CD: ecc4561 (2001b)

Kulczycki, P.: Statistical Inference for Fault Detection: A Complete Algorithm Based on Kernel Estimators. Kybernetika 38, 141–168 (2002a)

Kulczycki, P.: A test for comparing distribution functions with strongly unbalanced samples. Statistica 62, 39–49 (2002b)

Kulczycki, P.: Bayes Parameter Identification with Reference to Nonlinear Optimal Control. In: 15th World Congress IFAC, Barcelona, July 21-26, 2002, CD: 1329 (2002c)

Kulczycki, P.: Estymatory jadrowe w analizie systemowej. WNT, Warsaw (2005)

Kulczycki, P.: Estymatory jadrowe w badaniach systemowych. In: Kulczycki, P., Hryniewicz, O., Kacprzyk, J. (eds.) Techniki informacyjne w badaniach systemowych, pp. 79–105. WNT, Warsaw (2007)

Kulczycki, P., Charytanowicz, M.: Bayes sharpening of imprecise information. Applied Mathematics and Computer Science 15, 393–404 (2005)

Kulczycki, P., Daniel, K.: Algorytm wspomagania strategii marketingowej operatora telefonii komorkowej. In: Kacprzyk, J., Budzinski, R. (eds.) Badania operacyjne i systemowe; metody i techniki, EXIT, Warsaw, pp. 245–256 (2006)

Kulczycki, P., Mazgaj, A.: An algorithm for Bayes parameter identification with quadratic asymmetrical loss function. Control and Cybenetics 34, 1127–1148 (2005)

Kulczycki, P., Prochot, C.: Wykrywanie elementow odosobnionych za pomoca metod estymacji nieparametrycznej. In: Kulikowski, R., Kacprzyk, J., Slowinski, R. (eds.) Badania operacyjne i systemowe; podejmowanie decyzji – podstawy teoretyczne i zastosowania, EXIT, Warsaw, pp. 313–328 (2004)

Kulczycki, P., Waglowski, J.: On the application of statistical kernel estimators for the demand-based design of a wireless data transmission system. Control and Cybernetics 34, 1149–1167 (2005)

Kulczycki, P., Wisniewski, R.: Fuzzy Controller for a System with Uncertain Load. Fuzzy Sets and Systems 131, 185–195 (2002)

Larose, D.T.: Discovering Knowledge in Data. In: An Introduction to Data Mining, Wiley, New York (2005)

Silverman, B.W.: Density Estimation for Statistics and Data Analysis. Chapman and Hall, London (1986)

Wand, M.P., Jones, M.C.: Kernel Smoothing. Chapman and Hall, London (1994)