Int. J. Appl. Math. Comput. Sci., 2024, Vol. 34, No. 3, 439–451
DOI: 10.61822/amcs-2024-0031

A QUALITY INDEX FOR DETECTION OF ATYPICAL ELEMENTS (OUTLIERS)

PIOTR KULCZYCKI [a,b,*], KRYSTIAN FRANUS [b], MAŁGORZATA CHARYTANOWICZ [b,c]

[a]Faculty of Physics and Applied Computer Science
AGH University of Krakow
Mickiewicza 30, 30-059 Kraków, Poland
e-mail: kulczycki@agh.edu.pl

[b]Systems Research Institute
Polish Academy of Sciences
Newelska 6, 01-447 Warsaw, Poland
e-mail: {kulczycki,krystian.franus,mchmat}@ibspan.waw.pl

[c]Faculty of Electrical Engineering and Computer Science
Lublin University of Technology
Nadbystrzycka 36B, 20-618 Lublin, Poland
e-mail: m.charytanowicz@pollub.pl

Besides clustering and classification, detection of atypical elements (outliers, rare elements) is one of the most fundamental problems in contemporary data analysis. However, contrary to clustering and classification, an atypical element detection task does not possess any natural quality (performance) index. The subject of the research presented here is the creation of one. It will enable not only evaluation of the results of a procedure for atypical element detection, but also optimization of its parameters or other quantities. The investigated quality index works particularly well with frequency types of such procedures, especially in the presence of substantial noise. Using a nonparametric approach in the design of this index practically frees the proposed method from the distribution in the dataset under examination. It may also be successfully applied to multimodal and multidimensional cases.

Keywords: data analysis, atypical elements, outliers, rare elements, quality (performance) index.

1. Introduction

The issue of detecting atypical elements (Aggarwal, 2013; Hodge, 2011; Ranga Suri *et al.*, 2019) constitutes one of the main tasks of modern data analysis (Ott and Longnecker, 2015) and data exploration (Kacprzyk and Pedrycz, 2015; Nisbet *et al.*, 2009; Pedrycz and Chen, 2017). Such elements may be considered in a couple of ways. The most popular one is to connect them with the gross errors hampering these elements of the set under investigation, which can subsequently be corrected or even eliminated. The other, uncommon but more constructive, treats atypical elements as unconventional phenomena, natural talents, and as new trends; they constitute remarkably beneficial information, stimulating exceptional behaviors, and innovative thinking. Investigations concerning atypical elements find a variety of practical usages in many disciplines. In medicine, deviations from norms may indicate illness or pathologies; in technology, faults in a supervised plant; in banking, a fraud attempt; in computer science, hacker attacks. Other possible indicators of atypical elements may also be threats to public order, earthquakes, weather anomalies, climate change or ecological dangers, and many others (Cateni *et al.*, 2008; Kulczycki and Kruszewski, 2019).

A universal definition of atypical elements does not exist. A generic one states that they are created by a different mechanism than the rest. However, such a broadly formulated concept does not help to identify them in a dataset considered and influences a large,

[*]Corresponding author

if not even excessive from an applicational point of view, variety of methods used for this purpose. Most of them are based on the measure of distances between points (a distance-based approach) or the probabilities of particular element occurrences (a frequency approach). This does not exhaust all possibilities; an illustrative example can be an influential approach where elements whose removal from the population changes the model used the most are considered atypical.

Many contemporary methods are based on algorithms of classic data analysis, appropriately adapted for the needs of atypical element detection. The leading concept is here clustering, both in distance-based as well as frequency approaches. Atypical elements can be those which form clusters of low cardinalities or those which do not belong to any cluster at all (e.g., the DBSCAN procedure), or are the most distant from the centers they were assigned to. Another universal tool is the distance-based k-nearest neighbors (kNN) method, commonly used for classification—the decision whether a studied element is typical or not is made on the basis of the distances to the closest points. It is worth mentioning here a related aspect, i.e., in some concepts, only a local neighborhood of the tested element should be taken into account (see, e.g., the algorithm LOF related with kNN). One more group contains methods originating from the classic three-sigma rule, where the elements placed more than three standard deviations away from the mean value are considered atypical. One can include here the Z-score method and its valuable modification MAD-score based on quantiles. The innovative concept concerns encoders—neural networks specialized in atypical element detection. Namely, information about the analyzed set is coded and then encoded to obtain information as close as possible to what was first; the elements encoded the furthest with regards to their initial locations are treated as atypical. Similarly, mutations of the well-known and tested in practice methods, i.e., the support vector machines and the decision trees, or after generalization the decision forests, were created for the atypical element detection task.

The above abbreviated outline does not exhaust all the methods used for atypical element detection (see Aggarwal, 2013; Hodge, 2011; Ranga Suri *et al.*, 2019). However, even that depicts the diversity of the concept applied and, in consequence, indicates the need for constructing an index showing the quality of the results of actions of particular methods. Accordingly, it also becomes possible to change the values of the inner parameters, which were optimized in these adopted procedures according to their main purpose and not for the needs of atypical element detection.

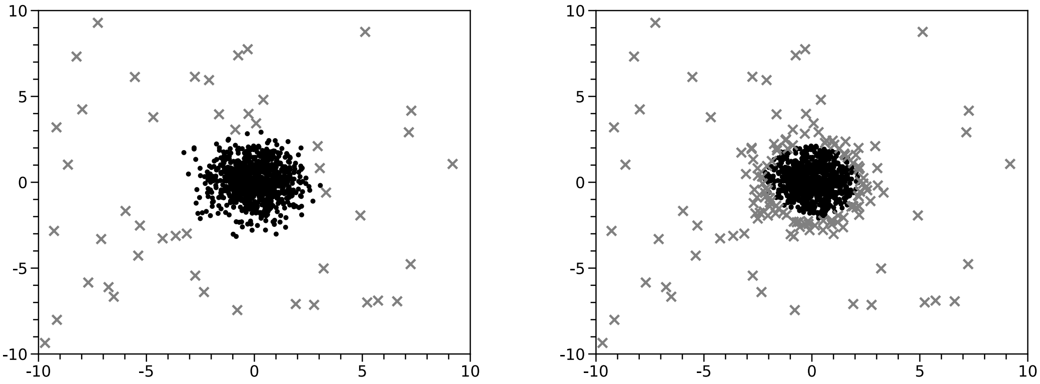
Contrary to clustering and classification, the task of detecting atypical elements does not possess natural quality (performance) indexes like, for example, those based on the sum of intracluster and intercluster distances or the Silhouette index for clustering (Batool and Hennig, 2021; Kłopotek *et al.*, 2020), as well as precision, accuracy or recall in the case of classification (Czmil *et al.*, 2024; Dalianis, 2018). This has a variety of negative repercussions, in particular the inability of automatic improvement in the procedures used for atypical element detection. As an initial illustration, let us take, for instance, a set generated from a two-dimensional standard normal distribution "contaminated" with a uniform noise. Figure 1 shows two example divisions of this set into atypical and typical elements, obtained with different parameters of the atypical element detection procedure. Which of them is better, which to choose? Or maybe something in-between? How can one construct a quality index helpful in decision making? Additionally, can parameters of an atypical element detection procedure be improved? These problems will constitute the subject of investigations presented in this paper. The above example will be made more precise (see Eqn. (23)) and continued at the beginning of Section 5.1. In particular, one may compare Fig. 1 to the right panel of Fig. 2, which indicates the result that is intermediate with respect to both the options from Fig. 1, and optimal in terms of the proposed quality index.

Finally, the subject of this paper is the creation of a quality index of the atypical element detection procedure based on the frequency approach, when atypical elements are considered rare, i.e., whose probability is small. Apart from the quality evaluation of the division of the analyzed set into atypical and typical, the presented material can serve as a methodological basis for the task of determining or improving the values of the parameters or other quantities of the decision model used to distinguish atypical elements. The proposed procedure can be applied with respect to the set with practically any distribution, in particular multimodal and incoherent (consisting of many components), and also in the multidimensional cases. The simplicity and illustrativeness of the investigated procedure seems to be especially valuable for a deeper interpretation as well as potential individual modifications. Thus, Section 2 briefly presents a mathematical base—the statistical kernel estimators methodology. The quality index is designed in Section 3 and discussed in Section 4. Next, Section 5 demonstrates the results of the numerical verification using illustrative synthetic data (Section 5.1) and benchmarks (Section 5.2). The summary and bibliography conclude this work.

2. Preliminaries: Kernel estimators

First, assume the n-element set of D-dimensional vectors with continuous attributes:

$$\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \in \mathbb{R}^D. \tag{1}$$



(a) (b)

Fig. 1. Two example divisions of an analyzed set into atypical (grey crosses) and typical (black circles) elements for different parameters of the atypical element detection procedure (note that crosses and circles symbolizing atypical and typical elements, respectively, merge together in dense areas).

The kernel estimator $\hat{f} : \mathbb{R}^D \to [0, \infty)$ of the distribution density (Chacon and Duong, 2020) of the dataset (1) can be defined as

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K\left(\boldsymbol{x}, \boldsymbol{x}_i, \boldsymbol{h}\right), \tag{2}$$

while after separating into coordinates one has

$$\boldsymbol{x}_i = \begin{bmatrix} x_{i,1} \\ x_{i,2} \\ \vdots \\ x_{i,D} \end{bmatrix}, \quad i = 1, 2, \dots, n,$$

$$\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_D \end{bmatrix}, \quad \boldsymbol{h} = \begin{bmatrix} h_1 \\ h_2 \\ \vdots \\ h_D \end{bmatrix}, \tag{3}$$

where the constants $h_d > 0$ (for $d = 1, 2, \dots, D$) are named smoothing parameters; the kernel $K : \mathbb{R}^D \to [0, \infty)$ will hence be used in the product form

$$K\left(\boldsymbol{x}, \boldsymbol{x}_i, \boldsymbol{h}\right) = \prod_{d=1}^{D} \frac{1}{h_d} K_d\left(\frac{x_d - x_{i,d}}{h_d}\right), \tag{4}$$

whilst the one-dimensional kernels $K_d : \mathbb{R} \to [0, \infty)$, for $d = 1, 2, \dots, D$, are measurable with unit integral $\int_{\mathbb{R}} K_d(y) \, \mathrm{d}y = 1$, symmetrical with respect to zero, and have a weak global maximum in this place. (However, nothing prevents from the use of other types of kernels, e.g., radial or asymmetrical, in the procedure presented here.)

In general, the choice of the kernels K_d is insignificant in practice, and one should primarily take into account the advantageous features of the constructed estimator, e.g., continuity, differentiability, or convenient integrability. Thus, the one-dimensional normal kernel $K_d : \mathbb{R} \to [0, \infty)$, i.e.,

$$K_d(x) = K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right), \quad d = 1, 2, \dots, D, \tag{5}$$

is usually considered basic and will be applied hereinafter for each coordinate. (Again, nothing prevents from the employment of other forms of the kernels, especially not the same for each coordinate.)

In contrast, the determination of the values of the smoothing parameters h_d is generally vital for the estimation quality. It is fortunate that many convenient procedures for calculating these values exist. Notably, in the one-dimensional ($D = 1$) case, the concept based on the normal distribution may initially be proposed in practice. Then, one has

$$h = \left(\frac{8\sqrt{\pi}}{3} \frac{W(K)}{U(K)^2} \frac{1}{n}\right)^{1/5} \hat{\sigma}, \tag{6}$$

where the estimator of the standard deviation $\hat{\sigma}$ may be calculated with classic formulas (Lehmann and Casella, 2011), and $W(K) = \int_{-\infty}^{\infty} K(y)^2 \, \mathrm{d}y$ and $U(K) = \int_{-\infty}^{\infty} y^2 K(y) \, \mathrm{d}y$; for the normal kernel (5) we have $W(K) = 1/2\sqrt{\pi}$ and $U(K) = 1$. In the multidimensional case, during the initial phase of the research without excessive performance requirements, the formula (6) can be used to each individual coordinate $d = 1, 2, \dots, D$. In the remainder of the situations, one of the methods collected by Chacon and Duong (2020) may be applied. In some cases, particular concepts possibly matching the model to the reality considered, e.g., the support boundary (Kulczycki, 2005; Silverman, 1986), can be employed. Further information on kernel estimators is available in the classic monographs by Chacon and Duong (2020) or Wand and Jones (1995); example applications can be found in the works of Baszczyńska (2016), Charytanowicz *et al.*, (2018; 2020), or Kulczycki (2020). Kernel estimators in the presence of categorical and discrete attributes are described by Agresti (2002) as well as Rajagopalan and Lall (1995).

3. Quality index

Consider the data set (1) comprising both atypical and typical elements. Using the content of Section 2, one may specify the density estimator $\hat{f}$ of the above data set distribution. Then, from the statistical point of view, the value $\hat{f}(\boldsymbol{x}_i)$ can be interpreted as the occurrence frequency of the element $\boldsymbol{x}_i$ in the population characterized by the analyzed set (1).

Let us introduce a partition of the n-element set (1) into the n_t-element subset of typical elements

$$\boldsymbol{x}_1^t, \boldsymbol{x}_2^t, \ldots, \boldsymbol{x}_{n_t}^t \tag{7}$$

and the n_{at}-element subset of atypical elements

$$\boldsymbol{x}_1^{at}, \boldsymbol{x}_2^{at}, \ldots, \boldsymbol{x}_{n_{at}}^{at} \ , \tag{8}$$

obtained using any available method. Indeed, the sets (7) and (8) are disjoint, whereas their union constitutes the set (1) and

$$n_{at} + n_t = n. \tag{9}$$

The inequality

$$n_{at} < n_t \tag{10}$$

is also justified, otherwise the atypical elements would become typical. In practice, n_{at} is less than n_t between two to 100 times, therefore, the share of the atypical elements is generally from 1% to 30%:

$$0.01 \leq \frac{n_{at}}{n} \leq 0.3 \ . \tag{11}$$

The quality index defined below should grade the quality of the partition of the set (1) into typical (7) and atypical (8).

For the elements from the sets (7) and (8) one can calculate, respectively,

$$\hat{f}(\boldsymbol{x}_1^t), \hat{f}(\boldsymbol{x}_2^t), \ldots, \hat{f}(\boldsymbol{x}_{n_t}^t) \tag{12}$$

and

$$\hat{f}(\boldsymbol{x}_1^{at}), \hat{f}(\boldsymbol{x}_2^{at}), \ldots, \hat{f}(\boldsymbol{x}_{n_{at}}^{at}). \tag{13}$$

Additionally, we can sort the set (12) non-decreasingly so that

$$\hat{f}(\boldsymbol{x}_1^t) \leq \hat{f}(\boldsymbol{x}_2^t) \leq \hat{f}(\boldsymbol{x}_{n_t}^t). \tag{14}$$

In the case of the frequency approach for the quality index being designed herein, the requirement is natural for the value

$$\frac{1}{n_{at}} \sum_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^{at}) \ , \tag{15}$$

characterizing atypical elements, to be the lowest possible. On the contrary, the values $\hat{f}(\boldsymbol{x}_i^t)$ corresponding to typical elements should be generally as high as possible. However, these elements for which $\hat{f}(\boldsymbol{x}_i^t)$ is large (e.g., located near the global mode) should not influence the choice of atypical elements since they are too distinct and their role in the population is completely different. Let us limit our discussion to these n_{at} among typical elements[1] which have the lowest value $\hat{f}(\boldsymbol{x}_i^t)$, i.e., those being as though "frequently the closest" to atypical elements. They are those that should have a real impact on the grading of the partition into atypical versus typical elements. Taking into account the order of the set (12) after sorting (14), we therefore require the average

$$\frac{1}{n_{at}} \sum_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^t) \tag{16}$$

to be as large as possible. Joining the conditions (15) and (16), we obtain the following quality index:

$$\mathrm{QI}_{\mathrm{KFC}} = \frac{\sum\limits_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^{at})}{\sum\limits_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^t)} \ . \tag{17}$$

The smaller the value of $\mathrm{QI}_{\mathrm{KFC}}$, the sharper the division into atypical and typical elements. This is more firm, and in the majority of applications simply better, "stronger". The use of the kernel K with the positive values, e.g., normal (5), to construct the estimator $\hat{f}$ guarantees the denominator to be nonzero while the quality index $\mathrm{QI}_{\mathrm{KFC}}$ is then positive.

Finally, if the studied aspect of the procedure of the division into atypical and typical elements can be modified, for example, by changing the values of the parameters existing there, then the index $\mathrm{QI}_{\mathrm{KFC}}$ (17) should be minimized. One can thus note informally

$$\mathrm{QI}_{\mathrm{KFC}} = \frac{\sum\limits_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^{at})}{\sum\limits_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i^t)} \to \text{min.} \tag{18}$$

(The lower index KFC originates from the authors' surnames, which coincides by chance with the name of a chain of fast food restaurants.)

[1] Note the use of the parameter n_{at} instead of n_t natural for the typical elements.

4. Comments and practical suggestions

The procedure for evaluating the index $\mathrm{QI}_{\mathrm{KFC}}$ (17) value can be synthetically expressed in the form of Algorithm 1.

All the procedures applied in Section 3 have linear computational complexity with respect to n and D, apart from the method for calculating the smoothing parameter value and the sorting algorithm, with the complexity from linear to quadratic depending on the employed procedure (Chacon and Duong, 2020; Knuth, 1988). Current computer systems enable effortless computations for n up to a range from $1,000$ to $100,000$ when the computing time does not exceed a few seconds. If, apart from the continuous attributes, categorical and/or discrete ones occur, the kernel estimator can be generalized according to the investigations published in the subject literature, (e.g., Agresti, 2002; Rajagopalan and Lall, 1995). The material presented here is also convenient in such situations.

The very nature of the above described issue leads to the natural procedure for atypical element detection. Let us calculate for all elements of the set (1) the values of the kernel estimator which, after sorting in nondecreasing order, can be noted as

$$\hat{f}(\boldsymbol{x}_1) \le \hat{f}(\boldsymbol{x}_2) \le \ldots \le \hat{f}(\boldsymbol{x}_n) \,. \tag{19}$$

The division of the set (1) into subsets of typical (7) and atypical (8) elements, under this procedure, consists in finding the index at which the quality index (17) is minimized, subject to the condition (11), belonging to the set

$$\{\lceil 0.01n\rceil, \lceil 0.01n\rceil + 1, \ldots, \lceil 0.3n\rceil\} \,, \tag{20}$$

where $\lceil\cdot\rceil$ means rounding up to the nearest integer. Finally, we are searching for the value n_{at}^* minimizing the expression

$$\min_{n_{at}\in\{\lceil 0.01n\rceil,\lceil 0.01n\rceil+1,\ldots,\lceil 0.3n\rceil\}} \frac{\sum_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i)}{\sum_{i=n_{at}+1}^{2n_{at}} \hat{f}(\boldsymbol{x}_i)} \,, \tag{21}$$

therefore

$$n_{at}^* = \arg \min_{\{\lceil 0.01n\rceil,\lceil 0.01n\rceil+1,\ldots,\lceil 0.3n\rceil\}} \frac{\sum_{i=1}^{n_{at}} \hat{f}(\boldsymbol{x}_i)}{\sum_{i=n_{at}+1}^{2n_{at}} \hat{f}(\boldsymbol{x}_i)} \,. \tag{22}$$

Elements with the indexes $i \le n_{at}^*$ (after renumbering (19)) are treated as atypical, while the remainder of elements, i.e., those for which $n_{at}^* < i$, are considered typical. Such a division is optimal in the sense of the index (17).

The procedure presented as Algorithm 2 constitutes an extension of the algorithm presented in the paper by Kulczycki and Kruszewski (2017), supplemented with the optimal choice of the parameter r value, which arbitrarily defined there the share of atypical elements in the analyzed set (1). This optimization is based on the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) proposed in this paper. In consequence, Algorithm 2 does not require fixing any parameter or other quantity value, which is a beneficial feature in practice.

As hereby shown, the concept of the quality index investigated in this paper can be used to improve the atypical element detection procedure through fixing or modifying the values of its parameters. In particular, applying the above procedure to detect atypical elements, we can change the smoothing parameter $\boldsymbol{h}$ value used in the construction of the kernel estimator (2) in order to improve the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) value. Another

Algorithm 1. Evaluation of the quality index $\mathrm{QI}_{\mathrm{KFC}}$.

```
1: Load datasets $\{\boldsymbol{x}_1^t, \boldsymbol{x}_2^t, \ldots, \boldsymbol{x}_{n_t}^t\}$ and $\{\boldsymbol{x}_1^{at}, \boldsymbol{x}_2^{at}, \ldots, \boldsymbol{x}_{n_{at}}^{at}\}$.
2: Based on the union of the above sets, calculate kernel estimator $\hat{f}$ (Section 2).
3: Sort set $\left\{\hat{f}(\boldsymbol{x}_1^t), \hat{f}(\boldsymbol{x}_2^t), \ldots, \hat{f}(\boldsymbol{x}_{n_t}^t)\right\}$.
4: Using $n_{at}$ smallest elements of above set and set $\{\hat{f}(\boldsymbol{x}_1^{at}), \hat{f}(\boldsymbol{x}_2^{at}), \ldots, \hat{f}(\boldsymbol{x}_{n_{at}}^{at})\}$, evaluate index $\mathrm{QI}_{\mathrm{KFC}}$ (17).
```

Algorithm 2. Atypical element detection based on the quality index $\mathrm{QI}_{\mathrm{KFC}}$.

```
1: Load analyzed set $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$.
2: Using the above set, calculate kernel estimator $\hat{f}$ (Section 2).
3: Sort set $\left\{\hat{f}(\boldsymbol{x}_1), \hat{f}(\boldsymbol{x}_2), \ldots, \hat{f}(\boldsymbol{x}_n)\right\}$.
4: for $n_{at}$ from $\lceil 0.01n\rceil$ to $\lceil 0.3n\rceil$ do
5:     Treating $\{\hat{f}(\boldsymbol{x}_1^{at}), \hat{f}(\boldsymbol{x}_2^{at}), \ldots, \hat{f}(\boldsymbol{x}_{n_{at}}^{at})\}$ as typical set, calculate quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17).
6: end for
7: Specify $n_{at}^*$, i.e., a value of $n_{at}$ for which index $\mathrm{QI}_{\mathrm{KFC}}$ is smallest.
8: Provide a graph of index $\mathrm{QI}_{\mathrm{KFC}}$ as a function of $n_{at}$ and partition of the investigated set into atypical and typical elements for $n_{at}^*$.
9: If $n_{at}^* = \lceil 0.3n\rceil$ carry out an individual case study.
```

example follows: in the case of using the k-nearest neighbors method (Yang *et al.*, 2023), the index proposed above can be used to fix the value of the parameter k. It should, however, be remarked that connecting the distance-based k-nearest neighbors method with the frequency index $\mathrm{QI}_{\mathrm{KFC}}$ (17) may potentially introduce interpretational ambiguity.

5. Applications, experimental verification

5.1. Synthetic data. Let us discuss the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) applied to synthetic data with easy-to-illustrate features. First, consider the two-dimensional case ($D = 2$) and the dataset with the distribution

$$(1-a)n\,\mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) + a\,n\,\mathcal{U}([-10,10])\,, \tag{23}$$

where $\mathcal{N}(\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes the two-dimensional normal distribution with the vector of expected values $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$, while $\mathcal{U}(A)$ is the two-dimensional uniform distribution on the interval A with independent coordinates, whereas the parameter $a \in [0,1]$ defines the share of the above factors. The first of them represents typical elements and the second, which plays the role of noise, atypical ones. For clarity of the figures, the relative small size $n = 1,000$ was assumed. The classic plug-in method was applied for calculation of the smoothing parameter $\boldsymbol{h}$ value of the kernel estimator (2)–(5). The large robustness of the investigated procedure with respect to this parameter should be underlined. The impacts of the possible changes in the parameter $\boldsymbol{h}$ generally reduce one another in the nominator and denominator in the definition of the quality index (17). Moreover, the possibility of optimization of the parameters appearing in the procedure used for atypical element detection frees the result from the smoothing parameter $\boldsymbol{h}$ value obtained using classic criterions based on the L^2 norm; in this case, such a value takes on the role of only a reference state, as an initial point for further iterations (compare Figs. 8 and 9). From a practical point of view, these are valuable features.

Three cases of the distribution (23) will be considered, with $a = 0.05$ being an example of a small number of atypical elements, $a = 0.1$ medium, and $a = 0.2$ large.

The left panel of Fig. 2 shows the values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for the particular quantities n^*_{at}/n with the range from 0.01 to 0.3 (see the condition (11)). The minimum occurs for $n^*_{at}/n = 0.084$ (see the notation (22) with (9)). The numbers of selected atypical and typical elements correctly amounted to 84 and 916, respectively. The division of the set (23) into atypical and typical elements for the optimal value n^*_{at}/n is illustrated in the right panel of Fig. 2.

Let us return to the initial example from Section 1. The left panel of Fig. 1 was obtained with the arbitrarily assumed value $n^*_{at}/n = 0.047$, whereas the right one with $n^*_{at}/n = 0.140$. In the former, there are too few atypical elements, and in the latter there seem to be too many. The above result, $n^*_{at}/n = 0.084$, minimizing the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17), constitutes a valuable compromise, found "automatically", without the subjective judgment of an analyst. Note that, in higher dimensionality, such evaluation may be impossible, yet apart from naturally increased requirements concerning the size of the analyzed set (1) the procedure optimizing the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) does not change.

Similar results obtained for medium and large noise are shown in Figs. 3 and 4. The values of the minimum of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) increased to $n^*_{at}/n = 0.132$ and $n^*_{at}/n = 0.232$, respectively, roughly proportional to the growth in the value of a. In the first case, the numbers of the atypical and typical elements in consequence amounted to 132 and 868, whereas in the second to 232 and 768.

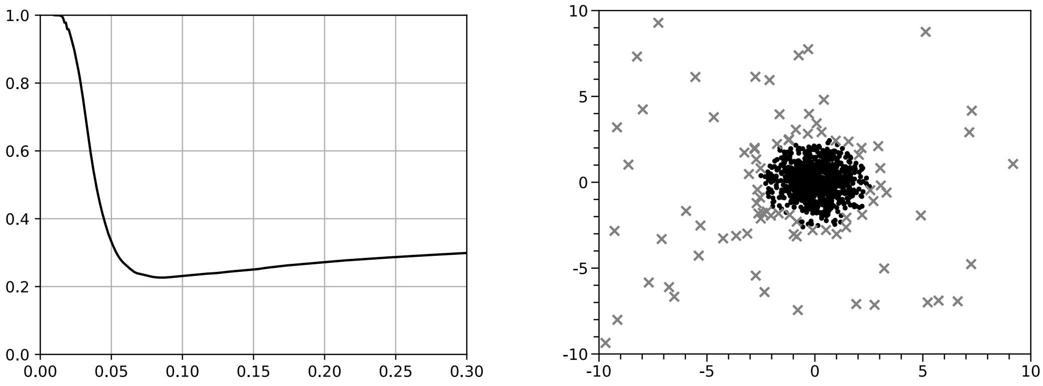
Consider now the set $\mathcal{HM}$ with the distribution density in the form of two half-moons partially overlapping one another, with the noise of the uniform distribution

$$(1-a)n\,\mathcal{HM} + a\,n\,\mathcal{U}([-4,4]), \tag{24}$$

where the notation is compliant with that introduced in the formula (23), while the two-dimensional set with the distribution $\mathcal{HM}$ was generated using the method `make_circles` of `scikit-learn` (scikit-learn, 2004), additionally transferring the coordinates so that the arithmetic means of both attributes equal zero. The shape of this distribution density is shown in the valid parts (b) of Figs. 5–7.

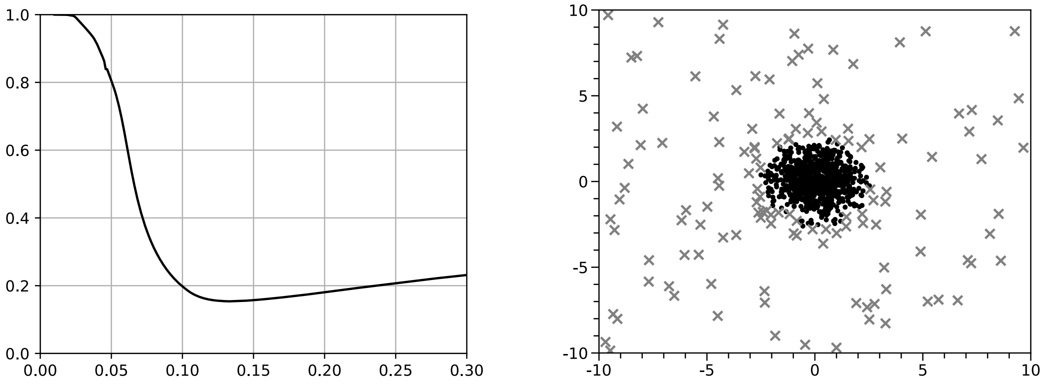
All corollaries formulated earlier for the normal distribution (23) also remain valid in the case of the set (24), which seems to be fairly troublesome to analyze. Special attention should be paid to the correctly detected elements placed in the valley between the half-moons. For the small ($a = 0.05$), medium ($a = 0.1$), and large ($a = 0.2$) noise, the minimum quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) occurred at $n^*_{at}/n = 0.043, 0.089$, and 0.182, respectively; it therefore properly increased, approximately proportionally to the value of a. The number of atypical and typical elements amounted, in consequence, to 43 and 89, 182 and 957, 911 and 818.

In both of the above examples, the number of atypical elements correctly increased as the noise share grew. Such a relation occurred successively for all examined values of the parameter a. Similar results were obtained in many cases studied for a variety of distributions in the presence of noise. The index $\mathrm{QI}_{\mathrm{KFC}}$ (17) rightly judged the quality of the division into atypical and typical element subsets, while the dependency of its value on the quantity n_{at}/n mostly had one proper (i.e., placed in the interior of a domain) minimum. The results of the



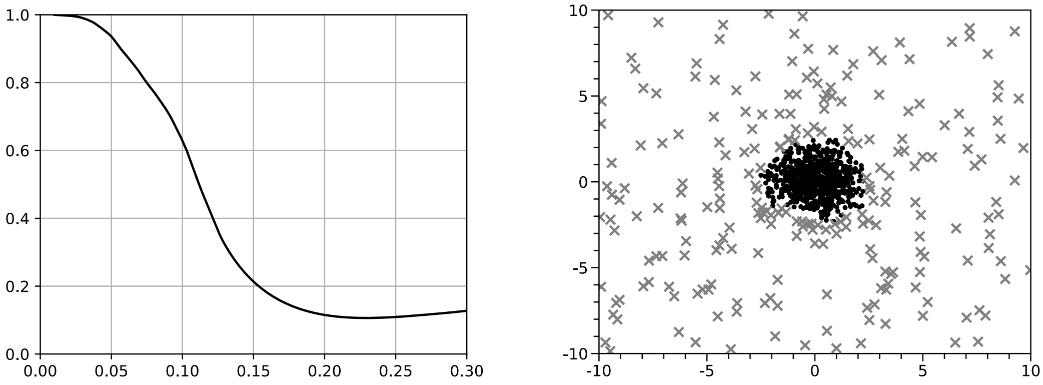
(a) (b)

Fig. 2. Results for the distribution (23) with small noise ($a = 0.05$): values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (23) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n_{at}^{*}/n = 0.084$ (b) (note that circles symbolizing typical elements merge together in the central area).



(a) (b)

Fig. 3. Results for the distribution (23) with medium noise ($a = 0.1$): values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (23) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n_{at}^{*}/n = 0.132$ (b) (note that circles symbolizing typical elements merge together in the central area).



(a) (b)

Fig. 4. Results for the distribution (23) with large noise ($a = 0.2$): values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (23) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n_{at}^{*}/n = 0.232$ (b) (note that circles symbolizing typical elements merge together in the central area).

Table 1. Numbers of elements indicated as atypical by Algorithm 2 and correct classification of elements from Class 2 for the distribution (23), using the notation "mean value ± standard deviation", calculated on the basis of 100 measurements.

	$a = 0.05$	$a = 0.1$	$a = 0.2$
A2	47.25 ± 1.75	94.56 ± 2.56	189.08 ± 2.90
NB	45.07 ± 2.32	90.65 ± 2.77	184.08 ± 4.22
kNN	38.98 ± 3.04	84.0 ± 3.82	176.85 ± 5.14
DT	43.68 ± 2.72	88.81 ± 3.28	180.86 ± 5.08

Table 2. Numbers of elements indicated as atypical by Algorithm 2 and correct classification of elements from Class 2 for the distribution (24), using the notation "mean value ± standard deviation", calculated on the basis of 100 measurements.

	$a = 0.05$	$a = 0.1$	$a = 0.2$
A2	44.87 ± 2.15	88.59 ± 2.88	173.46 ± 5.43
NB	22.11 ± 3.36	54.29 ± 5.36	127.46 ± 7.07
kNN	32.16 ± 3.53	75.57 ± 4.48	163.49 ± 5.46
DT	40.59 ± 2.75	85.34 ± 3.60	175.32 ± 5.33

natural procedure for the division into atypical and typical elements presented in Section 4 were readily interpretable and the procedure itself complete (i.e., no fixing of any parameter or other quantity value is necessary).

The above examples will be used to conduct a comparative study. In the literature, there is a lack of methods which can be directly compared with the concept proposed here. Let us then treat the first factor existing in the formulas (23) and (24) as an element of Class 1, and the second factor, representing noise, as an element of Class 2. The share of the latter n_{at}/n is specified by the parameter a with values, subsequently, $0.05, 0.1$, and 0.2. The sizes of both the learning and testing sets again equal $n = 1,000$. The classic methods: the naive Bayes (NB), k-nearest neighbors (kNN), and the decision tree (DT) with standard parameters (James *et al.*, 2023) were subsequently used for classification. The results are shown in Tables 1 and 2. Placed there were the numbers of elements indicated as atypical by Algorithm 2 (A2) described in Section 4, and the correct classifications of the elements from Class 2 were considered to be noise.

The results obtained using Algorithm 2 are more favorable in both the greater average of the number of detected atypical elements (or correctly classified elements of Class 2) and the lower standard deviation. Only in the case of the distribution (24) with $a = 0.2$ (right column of Table 2) were slightly better results achieved by the decision tree method. This took place despite Algorithm 2 being an unsupervised method, whereas the supervised classification procedures benefited from additional information in the form of patterns of both classes. Such valuable results of Algorithm 2 were reached to a large degree thanks to the optimization of the parameter n^*_{at}/n provided based on the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) investigated in this paper. Similar profits can be obtained by using it with respect to other aspects of atypical element detection methods.

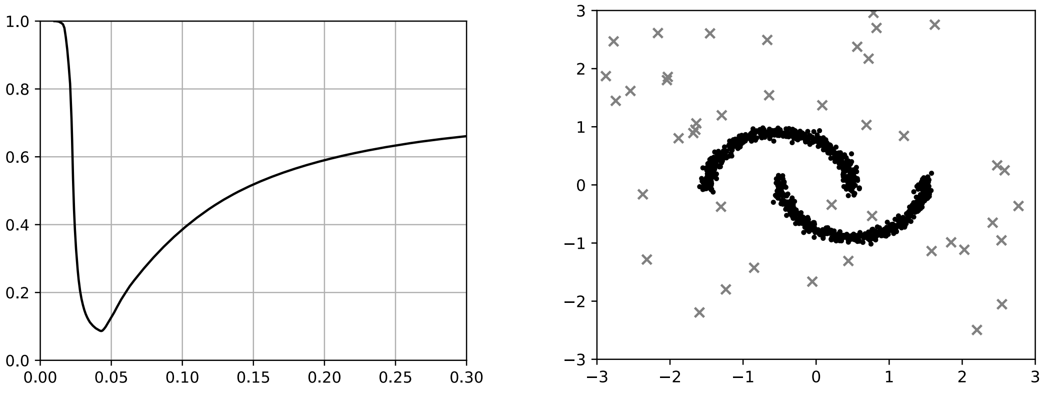
It should be stressed that, in the case of the lack of noise, the results were mainly correct, although, it is harder to judge their accuracy by nature. In particular, in the case when the minimum occurred on the borders of the admissible arguments interval (11), i.e, for 0.01 or 0.3, an additional individual analysis is worth recommending, especially considering possible local minima of the function $\mathrm{QI}_{\mathrm{KFC}}$ placed in the interior (see Fig. 10). In such cases, one should also consider modifying the values of the parameters or other quantities of an atypical element detection procedure. These concepts will be illustrated in the next section (see Figs. 8 and 9).

5.2. Experimental data.

The results obtained with synthetic data will be additionally illustrated using real experimental data accessible on the Internet, from the fields of medicine and sociology (Kaggle, 2024) as well as astronomy (Caltech, 2024).

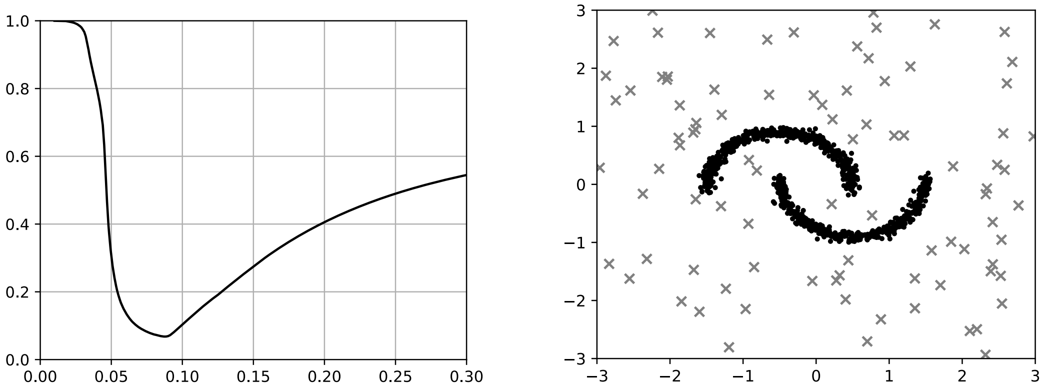
Consider the number of suicides in 36 European countries (excluding ministates and Moldovia, North Macedonia, as well as those partially in Asia) and in the years between 1985 and 2016 (from two surveys in the case of Bosnia and Hercegovina, to 32 for Austria and Iceland) (Kaggle, 2024). Special attention will be paid to the midlife crisis of 35–54 year-olds. At this critical time of life, people become aware of the missed opportunities and unrealized plans, with an escalation of comparing oneself most often to those who are exceptionally successful. This process, because of cultural and environmental factors, affects mostly men. The size of the analyzed set is $n = 923$.

Figure 8 illustrates the joint distribution of the number of suicides and the GDP income in the case of men, for particular countries and years. Panel (a) shows the values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for specific values of n_{at}/n in the range from 0.01 to 0.3. The minimum appears for $n^*_{at}/n = 0.088$. In Panel (b), the atypical elements obtained for this value are presented; 82 such elements were selected. They are aggregated into four groups denoted in Fig. 8 as A, B, C, and D. The first is made up of Iceland, Luxembourg, Norway, and Switzerland, i.e., the countries with the highest GDP per person. It is clear that, in this group, the number of suicides clearly oscillates around 20 per one million inhabitants, which becomes a sort of "sociological noise" of a magnitude independent of the level of wealth of the state in group A. Group D characterizes European countries with the lowest income,



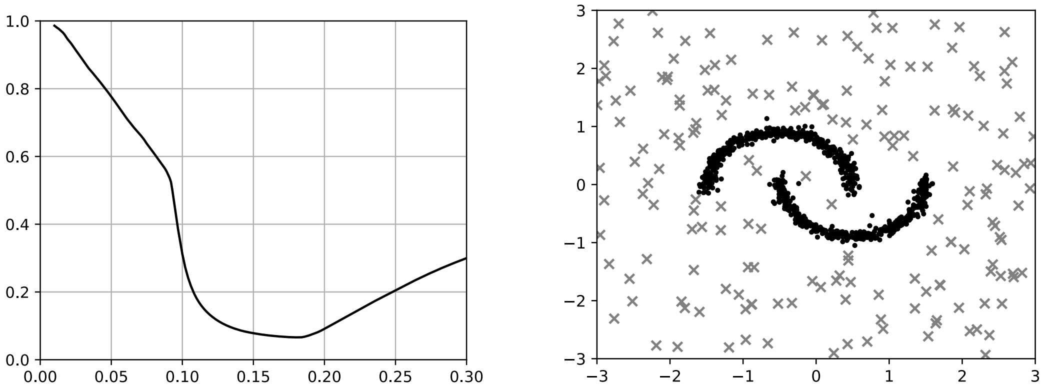
(a) (b)

Fig. 5. Results for the distribution (24) with small noise $(a = 0.05)$: values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (24) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.043$ (b) (note that circles symbolizing typical elements merge together in the central area).



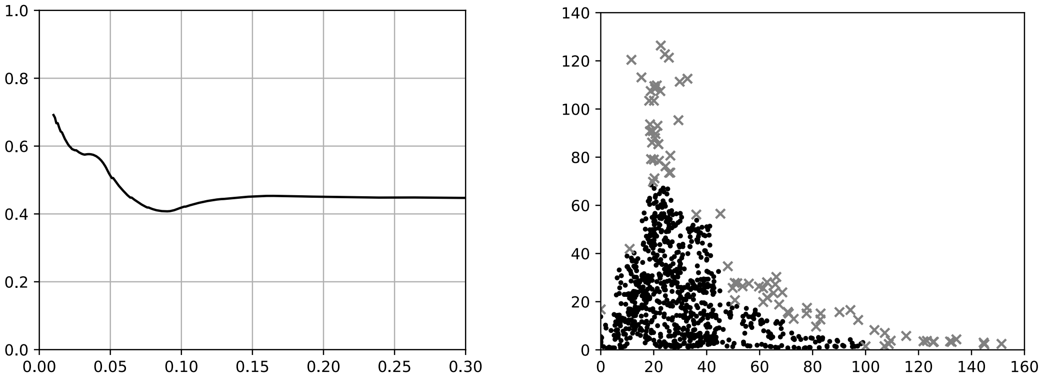
(a) (b)

Fig. 6. Results for the distribution (24) with medium noise $(a = 0.1)$: values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (24) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.089$ (b) (note that circles symbolizing typical elements merge together in the central area).



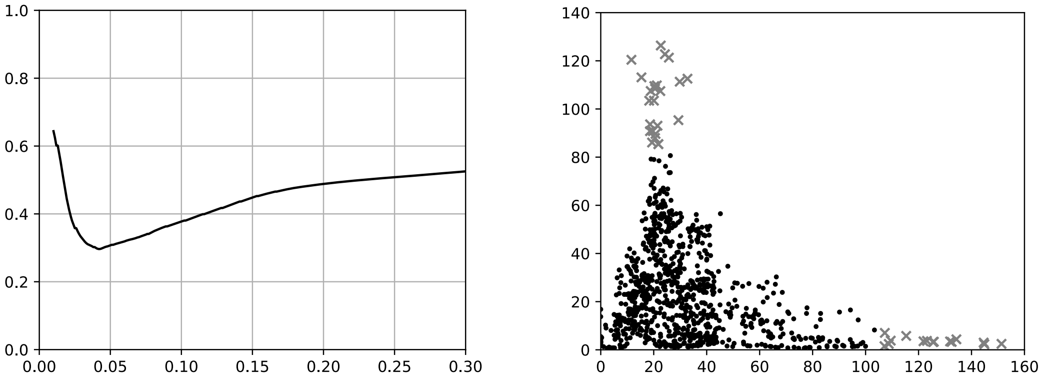
(a) (b)

Fig. 7. Results for the distribution (24) with large noise $(a = 0.2)$: values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division of the set (24) into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.182$ (b) (note that circles symbolizing typical elements merge together in the central area).



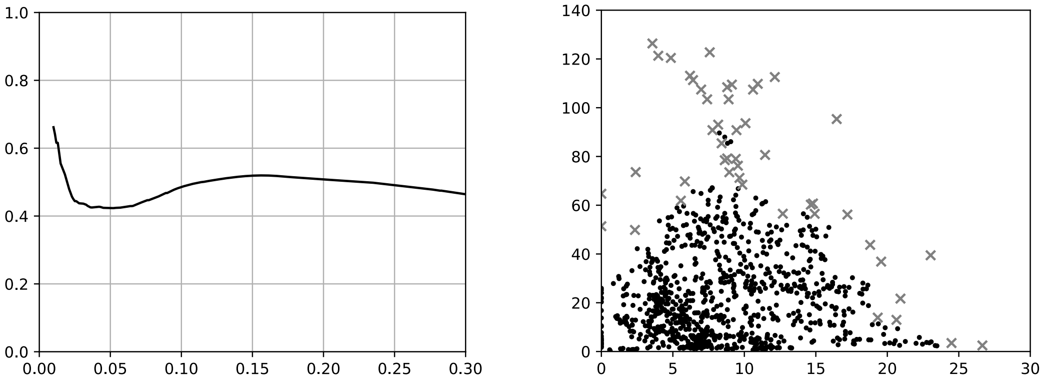
(a) (b)

Fig. 8. Results for the joint distribution of suicides of men aged 35–54 (per one million inhabitants—horizontal axis) and GDP (in 1,000 USD per person—vertical axis): values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.088$ (b).



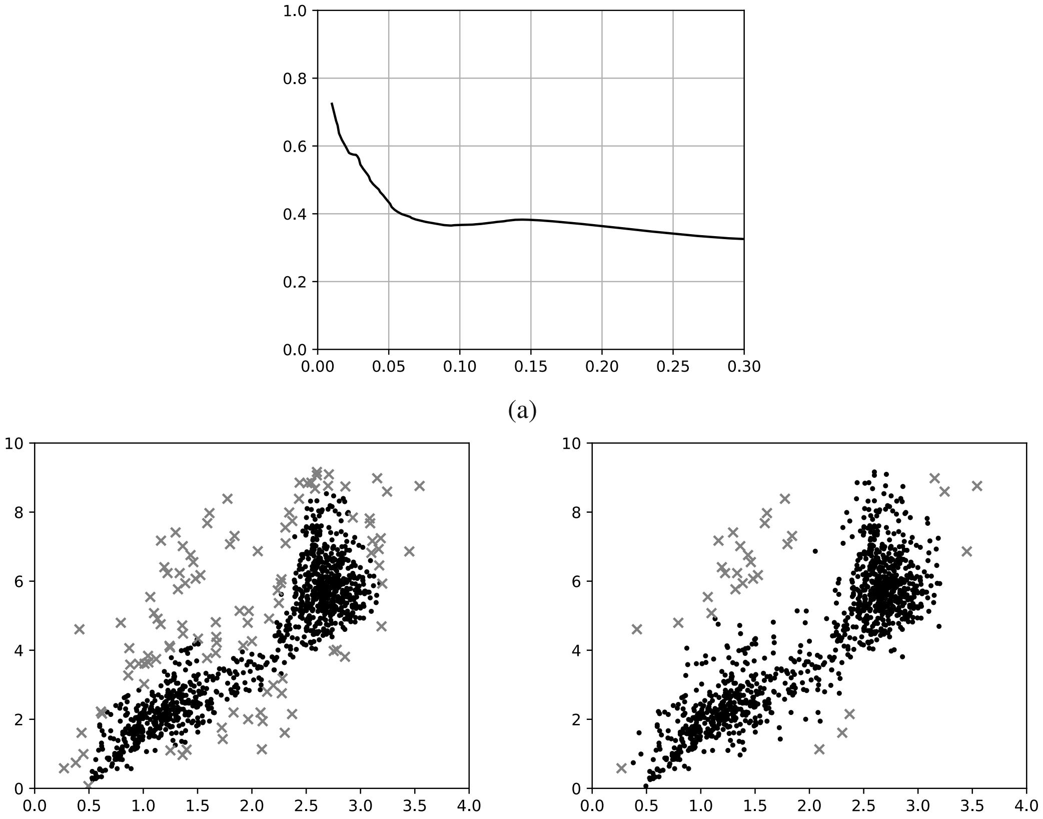
(a) (b)

Fig. 9. Results for the joint distribution of suicides of men aged 35–54 (per one million inhabitants—horizontal axis) and GDP (in 1,000 USD per person—vertical axis), with the value of smoothing parameter $\boldsymbol{h}$ multiplied by 7.1: values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.042$ (b).



(a) (b)

Fig. 10. Results for the joint distribution of suicides of women aged 35–54 (per one million inhabitants—horizontal axis) and GDP (in 1,000 USD per person—vertical axis): values of the quality index $\mathrm{QI}_{\mathrm{KFC}}$ (17) for particular n_{at}/n values (a), division into atypical (grey crosses) and typical (black circles) elements for the optimal value $n^*_{at}/n = 0.050$ (b).



(a)

(b)

(c)

Fig. 11. Results for the joint distribution of radius (in relation to the Earth's radius—horizontal axis) and mass (in relation to the Earth's mass—vertical axis) for exoplanets: values of the quality index QI_{KFC} (17) for particular n_{at}/n values (a), division into atypical (grey crosses) and typical (black circles) elements for the global optimal value $n^*_{at}/n = 0.300$ (b), division into atypical (grey crosses) and typical (black circles) elements for the local optimal value $n^*_{at}/n = 0.094$ (c) (note that circles symbolizing typical elements merge together in dense areas).

i.e., Belarus, Estonia, Latvia, and Lithuania, where the economic reality clashes with the culturally conditioned responsibility of men for the material status of the family. Finally, group C contains intermediate cases (Finland, Lithuania, Luxemburg, Slovenia) and B those on the borders of the typical element sets (Luxemburg, Norway, Slovakia, Italy). The indications of the quality index QI_{KFC} (17) allowed a clear partition into atypical and typical elements.

The results revealed in Fig. 8 can be additionally modified by optimizing the quality index QI_{KFC} (17) with respect to the parameters of the procedure of atypical element detection. In the case being considered, this role is played by the smoothing parameter $\boldsymbol{h}$ of the kernel estimator introduced in the formula (4). Figure 9 shows the result for the parameter $\boldsymbol{h}$ obtained with the classic mean-square criterion with both coordinates multiplied by 7.1, which constitutes the optimal value for the quality index QI_{KFC} (17). Visible are the sharpened results of the division into atypical and typical elements—only groups A (excluding Iceland) and C (no change) remain, making the interpretation much more distinct. The number of indicated atypical elements decreased to 39.

Out of scientific curiosity, let us consider the joint distribution of the number of suicides and the GDP income in the case of women. The equivalent of Fig. 8 is now Fig. 10. Above all, one should note the changed scale of the horizontal axis, representing the number of suicides, being five-times smaller for women (Fig. 10) than for men (Fig. 8). The minimum of the function QI_{KFC} occurs for $n^*_{at}/n = 0.050$. Then, 47 elements were indicated as atypical. Group A, containing wealthy countries, is similar to the case of men (Denmark, Luxemburg, Norway, Switzerland). Group D, including the poorest states, does not practically exist—the elements potentially belonging here (in the bottom-right corner, corresponding to Lithuania) become part of the general noise, previously represented by groups B and C.

After the dark subject of suicides, let us illustrate the different operation aspect of the quality index QI_{KFC} (17) in regard to exoplanets, i.e., the planets orbiting around a star or stars other than the Sun (Caltech, 2024). The size of this set is $n = 1,138$. The attributes are the logarithms of the radius (in units of the Earth's radius—horizontal axis) and mass (in units of the Earth's mass—vertical axis), after adding 1 to both of the arguments, i.e., $x \mapsto \log(x + 1)$ in units related to the Earth. The left panel of Fig. 11 shows the value of the index $\mathrm{QI}_{\mathrm{KFC}}$ (17) as a function of n_{at}/n. Its minimum on the interval $[0.01, 0.3]$ occurs on the right border 0.3, so the situation requires special consideration. In fact, the number of atypical elements is in consequence significant and amounts to 341 with 797 typical. The partition into atypical and typical elements, presented in Panel (b) of Fig. 11, does not seem to be fully justified. In such a case, one should pay attention to the local minimum appearing for $n^*_{at}/n = 0.094$; the obtained division is shown in Panel (c) of Fig. 11. It is more distinct, and suitable for convenient interpretation.

6. Additional comments and summary

This paper presented the concept of a quality index enabling the judgement of the results of atypical element detection in an analyzed set, especially rare elements in the frequency approach (for which this method is particularly predisposed) as well as outliers in the distance approach and other concepts. In consequence, the subject of evaluation constitutes the quality of division of the set under investigation into atypical and typical elements. Notably convenient for interpretation are the indications being obtained in the case of the existence of noise superimposed on the conglomeration of typical elements. It may be any shape, especially incoherent (consisting of isolated components).

Thanks to having the possibility to evaluate the results of any procedure of atypical element detection, this method also allows the optimization of the values of parameters or other quantities occurring in such a procedure. In the presented material, this aspect was applied to the natural frequency algorithm based on the probability of appearance (Kulczycki and Kruszewski, 2017). The subject of the optimization was the quantity n^*_{at}/n characterizing the share of elements recognized as atypical with respect to the size of the set being analyzed, in the range of $[0.01, 0.3]$. In the case when the global minimum of the quality index occurs on the borders of this interval, individual analysis is recommended, especially with the consideration of local minima of the quality index. Positive effects can also be obtained by optimizing the values of the parameters or other quantities existing in the procedure used (e.g., the smoothing parameter $\boldsymbol{h}$ of the kernel estimator in the presented example).

The index is intended for one- and multi-dimensional data, in particular with continuous attributes, although the presented concept may be easily generalized with any attribute types for which the construction of the kernel estimator is possible, i.e., apart from continuous also categorical, discrete, and their combinations. Convenient interpretability of the proposed method can be facilitated by the primary application of the reduction of size (Wasserman, 2004) and dimensionality (Sorzano *et al.*, 2014) algorithms. In practice, it can also be beneficial to employ the conditional density, making the model of the studied reality more precise by adding the current value of the conditional factor (Kulczycki and Franus, 2021). These aspects will be the subjects of further detailed research.

Acknowledgment

This work was partially supported by the Faculty of Physics and Applied Computer Science of the AGH University of Krakow as well as the Systems Research Institute of the Polish Academy of Sciences, and also by the Faculty of Electrical Engineering and Computer Science of the Lublin University of Technology.

References

Aggarwal, C.C. (2013). *Outlier Analysis*, Springer, Cham.

Agresti, A. (2002). *Categorical Data Analysis*, Wiley, Hoboken.

Baszczyńska, A. (2016). *Smoothing Parameter of the Density Functions for Random Variables in Economic Research*, Lodz University Press, Łódź, (in Polish).

Batool, F. and Hennig, C. (2021). Clustering with the average silhouette width, *Computational Statistics and Data Analysis* **158**(6): 107190.

Cateni, S., Colla, V. and Vannucci, M. (2008). Outlier detection methods for industrial applications, *in* J. Aramburo and A.R. Trevino (Eds), *Advances in Robotics, Automation and Control*, I-Tech, Vienna, pp. 265–282.

Caltech (2024). *NASA Exoplanet Archive*, https://exoplanetarchive.ipac.caltech.edu/.

Chacon, J.E. and Duong, T. (2020). *Multivariate Kernel Smoothing and Its Applications*, Chapman and Hall/CRC, Boca Raton.

Charytanowicz, M., Kulczycki, P., Kowalski, P.A., Lukasik, S. and Czabak-Garbacz, R. (2018). An evaluation of utilizing geometric features for wheat grain classification using x-ray images, *Computers and Electronics in Agriculture* **144**(1): 260–268.

Charytanowicz, M., Perzanowski, K., Januszczak, M., Wołoszyn-Gałęza, A. and Kulczycki, P. (2020). Application of complete gradient clustering algorithm for analysis of wildlife spatial distribution, *Ecological Indicators* **113**(6): 106216.

Czmil, S., Kluska, J. and Czmil, A. (2024). An empirical study of a simple incremental classifier based on vector quantizzation and adaptive resonance theory, *International Journal of Applied Mathematics and Computer Science* **34**(1): 149–165, DOI: 10.61822/amcs-2024-0011.

Dalianis, H. (2018). *Clinical Text Mining*, Springer, Cham.

Hodge, V. (2011). *Outlier and Anomaly Detection: A Survey of Outlier and Anomaly Detection Methods*, Lambert Academic Publishing, Saarbrucken.

James, G., Witten, D., Hastie, T., Tibshirani, R. and Taylor, J. (2023). *An Introduction to Statistical Learning*, Springer, Cham.

Kacprzyk, J. and Pedrycz, W. (2015). *Springer Handbook of Computational Intelligence*, Springer, Berlin.

Kaggle (2024). Suicide rates overview 1985 to 2016, Dataset, http://www.kaggle.com/datasets/russellyates88/suicide-rates-overview-1985-to-2016.

Kłopotek, R., Kłopotek, M. and Wierzchoń, S. (2020). A feasible k-means kernel trick under non-Euclidean feature space, *International Journal of Applied Mathematics and Computer Science* **30**(4): 703–715, DOI: 10.34768/amcs-2020-0052.

Knuth, D.E. (1988). *Art of Computer Programming. Vol. 3: Sorting and Searching*, Addison-Wesley, Upper Saddle River.

Kulczycki, P. (2005). *Kernel Estimators in Systems Analysis*, Scientific and Engineering Publishers, Warsaw, (in Polish).

Kulczycki, P. (2020). Methodically unified procedures for outlier detection, clustering and classification, *in* K. Arai (Ed.), *Proceedings of the Future Technologies Conference (FTC)*, Springer, Cham, pp. 460–474.

Kulczycki, P. and Franus, K. (2021). Methodically unified procedures for a conditional approach to outlier detection, clustering, and classification, *Information Sciences* **560**: 504–527.

Kulczycki, P. and Kruszewski, D. (2017). Identification of atypical elements by transforming task to supervised form with fuzzy and intuitionistic fuzzy evaluations, *Applied Soft Computing* **60**(11): 623–633.

Kulczycki, P. and Kruszewski, D. (2019). Detection of rare elements in investigation of medical problems, *in* N.T. Nguen *et al.*, (Eds), *Intelligent Information and Database Systems*, Springer, Singapore, pp. 257–268.

Lehmann, E.L. and Casella, G. (2011). *Theory of Point Estimation*, Springer, New York.

Nisbet, R., Miner, G. and Yale, K. (2009). *Handbook of Statistical Analysis and Data Mining Applications*, Elsevier, London.

Ott, R.L. and Longnecker, M.T. (2015). *An Introduction to Statistical Methods and Data Analysis*, Cengage, Boston.

Pedrycz, W. and Chen, S.-M. (2017). *Data Science and Big Data: An Environment of Computational Intelligence*, Springer, Cham.

Rajagopalan, B. and Lall, U. (1995). A kernel estimator for discrete distributions, *Journal of Nonparametric Statistics* **4**(1): 409–426.

Ranga Suri, N.N.R., Narasimha-Murty, M. and Athithan, G. (2019). *Outlier Detection: Techniques and Applications*, Springer, Cham.

scikit-learn (2004). make_circles, Dataset, https://scikit-learn.org/stable/modules/generated/sklearn.datasets.make_circles.html.

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

Sorzano, C., Vargas, J. and Pascual-Montano, A. (2014). A survey of dimensionality reduction techniques, *arXiv:* 1403.2877v1.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*, Chapman and Hall, New York.

Wasserman, L. (2004). *All of Statistics: A Concise Course in Statistical Inference*, Springer, New York.

Yang, J., Tan, X. and Rahardja, S. (2023). Outlier detection: How to select k for k-nearest-neighbors-based outlier detectors, *Pattern Recognition Letter* **174**: 112–117.

Piotr Kulczycki holds professorial positions at the Systems Research Institute of the Polish Academy of Sciences, where he is the head of the Centre of Information Technology for Data Analysis Methods, as well as at the AGH University of Krakow, Faculty of Physics and Applied Computer Science, where he is the head of the Division for Information Technology and Systems Research. In the past he had also held the position of a visiting professor at Aalborg University. Professor Kulczycki has supervised 10 doctoral students, and has published 13 books and more than 200 scientific works in reputable journals, edited volumes, and conference proceedings. The field of his scientific activity to date is the applicational aspects of information technology, particularly data analysis and mining, in diverse issues of contemporary systems research and control engineering.

Krystian Franus holds an MSc degree in applied computer science from the AGH University of Krakow. He is currently pursuing a PhD at the Systems Research Institute of the Polish Academy of Sciences, and works as a data scientist at Samsung, focusing on data analysis, machine learning, and enhancing recommender systems. His scientific interests include leveraging data to drive innovation, combining academic research with practical industry experience to advance the field of intelligent systems.

Małgorzata Charytanowicz is a research scientist at the Systems Research Institute, Polish Academy of Sciences, and at the Department of Computer Science, Lublin University of Technology. She holds an MSc degree in mathematics from Maria Curie-Skłodowska University in Lublin, as well as a PhD and a DSc (habilitation) from the Systems Research Institute of the Polish Academy of Sciences, both in the area of computer science. Her research interests are image processing methods, data analysis, artificial intelligence and their applications especially in medicine. She has published and presented papers in journals as well as at national and international conferences.

Received: 3 February 2024
Revised: 21 April 2024
Re-revised: 8 June 2024
Accepted: 26 June 2024