UDC 004.02

S.M. Vovk, V.V. Hnatushenko

CRITERIA AND TECHNIQUES FOR PROCESSING NOISY DATA WITH ANOMALOUS VALUES

Abstract. An approach to solve the problem of processing noisy data with anomalous values when the statistics of noise and anomalies are unknown is proposed. This approach is based on a method of generalized maximum likelihood and a superset of cost functions. The superset allows to tune the data processing onto the current noise environment. This tuning is performed by setting the values of three free parameters related to the scale, the heaviness of tails, and the form of random values distribution, as well as to the presence of anomalous values. In the general case, the proposed approach requires to solve a multimodal optimization problem.

Key words: data processing, criterion, method, noise, anomalous value.

Introduction. Processing of noisy data with anomalous values is one of the important and complex problems [1]. In order to solve this problem, the known criteria and methods for processing noisy data with anomalous values are analyzed under assumption that data model is a constant. Firstly, we analyze the criteria and methods for processing of data distorted by the additive noise with Gaussian and Laplacian distributions, Cauchy distribution, as well as by the generalized Gaussian distribution and by the generalized Cauchy distribution. The general basis of this analysis is the Fischer's criterion of maximum likelihood. Further, the criteria and methods for detecting anomalous values as well as the robust methods for processing of data with contaminated noise are analyzed. It is noted that generalization of the maximum likelihood criterion, performed by P. Huber [2] and complemented by the framework of cost functions with horizontal asymptotes [3], provides a fundamental possibility for solving the considered problem. At the end of this paper, we propose an approach to constructing the generalized maximum likelihood criterion on the base of "superset" of cost functions [4]. This approach improves the generalized maximum likelihood criterion due to the possibility to tune the data processing onto unknown noise environment.

Problem formulation. The problem is to develop the approach for processing noisy data with anomalous values, which generalizes the traditional approaches to data processing in a well-known noise environment and allows to tune the data processing onto unknown noise environment.

Criteria and methods for processing noisy data. For processing of data distorted by additive Gaussian noise, the least squares criterion is traditionally used [2]. In the case when data is represented by a real sequence $x_1,...,x_N$ of length $N \geq 1$, and the unknown value is the constant value θ, the use of this criterion leads to the problem

$$\theta^* = \arg\min_{\theta}[\sum_{i=1}^{N}(x_i - \theta)^2]. \quad (1)$$

From (1) it follows that

$$\theta^* = \frac{1}{N}\sum_{i=1}^{N} x_i \ , \quad (2)$$

i.e., the estimate of θ is the arithmetic mean value. If each quadratic term in (1) is weighted by w_i, then instead of (1) we get the problem:

$$\theta^* = \arg\min_{\theta}[\sum_{i=1}^{N} w_i(x_i - \theta)^2]. \quad (3)$$

The solution of (3) is the weighted arithmetic mean value:

$$\theta^* = \sum_{i=1}^{N} w_i x_i \Bigg/ \sum_{i=1}^{N} w_i \ . \quad (4)$$

The solutions (2) and (4) have a clear probabilistic justification within the framework of maximum likelihood criterion. Indeed, let the data elements (or samples) $x_1,...,x_N$ represent the noisy values of a constant value θ, i.e.

$$x_i = \theta + n_i \ ; \qquad i = 1,...,N, \quad (5)$$

where $n_1,...,n_N$ are the independent random variables. If each of n_i obeying the Gaussian distribution with zero mean and uniform variance σ^2, then the joint probability density function $p(x_1,...,x_N \mid \theta)$ is given by

$$p(x_1,...,x_N \mid \theta) = p(x_1 \mid \theta) \cdot ... \cdot p(x_N \mid \theta) = \frac{1}{(\sqrt{2\pi}\sigma)^N} \cdot \exp\left(-\sum_{i=1}^{N}\frac{(x_i - \theta)^2}{2\sigma^2}\right). \quad (6)$$

Applying the maximum likelihood criterion

$$\theta^* = \arg\max_{\theta} p(x_1,...,x_N \mid \theta) = \arg\max_{\theta} \ln[p(x_1,...,x_N \mid \theta)] \quad (7)$$

to (6), we have the problem (1), the solution of which is given by (2). But if each of n_i has different variance σ_i^2, then

$$p(x_1,...,x_N \mid \theta) = \frac{1}{(\sqrt{2\pi})^N} \cdot \frac{1}{(\sigma_1 \cdot ... \cdot \sigma_N \cdot)} \cdot \exp\left(-\sum_{i=1}^{N} \frac{(x_i - \theta)^2}{2\sigma_i^2} \right). \quad (8)$$

After applying (7) to (8), instead of (1) we get the problem (3), the solution of which is given by (4), where $w_i = 1/\sigma_i^2$.

For the processing of data distorted by additive Laplacian noise, the method of median filtering is used. It is based on the criterion of the minimum sum of absolute values, which leads to the problem [5]:

$$\theta^* = \arg\min_{\theta}[\sum_{i=1}^{N} | x_i - \theta |] \ . \quad (9)$$

The solution of the problem (9) corresponds to the median $med\{x_i |_{i=1}^{N}\}$ of data sequence, which for odd value of N is equal to the value of the middle element of ordered sequence. However, if N is even, then the median is equal to the arithmetic mean of the two middle elements in ordered data sequence. If each absolute term in (9) is weighted by w_i, then we have the problem:

$$\theta^* = \arg\min_{\theta}[\sum_{i=1}^{N} w_i | x_i - \theta |] \ . \quad (10)$$

The solution of (10) is given by:

$$\theta^* = med\{w_i \circ x_i |_{i=1}^{N}\} \ , \quad (11)$$

where $\circ$ is the replication operator which replicates the data element x_i by the number of times w_i [6].

The criterion of maximum likelihood provides such a justification for the problems (9) and (10). Let the independent random variables n_i obeying the Laplacian distribution with zero mean and uniform variance $\sigma^2 = 2\lambda^2$. Then

$$p(x_1,...,x_N \mid \theta) = \frac{1}{(2\lambda)^N} \cdot \exp\left(-\frac{1}{\lambda} \sum_{i=1}^{N} | x_i - \theta | \right). \quad (12)$$

Applying (7) to (12), we have the problem (9), the solution of which is the median. But if each of n_i has different variance $\sigma_i^2 = 2\lambda_i^2$, then

$$p(x_1,...,x_N \mid \theta) = \frac{1}{2^N(\lambda_1 \cdot ... \cdot \lambda_N)} \cdot \exp\left(-\sum_{i=1}^{N} \frac{| x_i - \theta |}{\lambda_i} \right). \quad (13)$$

Assuming $w_i = 1/\lambda_i$ and applying (7) to (13), we obtain the problem (10), the solution of which is given by (11).

The Gaussian and Laplacian distributions are special cases of the generalized Gaussian distributions family [7]:

 ISSN 1562-9945

$$p(x) = \frac{\rho}{2s\Gamma(1/\rho)} \cdot \exp\left[-\left(\frac{|x-\theta|}{s}\right)^{\rho}\right] = \frac{1}{2s\Gamma(1+1/\rho)} \cdot \exp\left[-\left(\frac{|x-\theta|}{s}\right)^{\rho}\right], \quad (14)$$

where $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ is a gamma function, θ is a location parameter, $s > 0$ is a scale parameter, $\rho > 0$ is a shape parameter, which determines, in particular, the tail decay rate. The mean, the median and the mode of generalized Gaussian distribution is equal to θ, and the variance is $\sigma^2 = s^2[\Gamma(3/\rho)/\Gamma(1/\rho)]$. From (14) it follows that the values of $\rho = 1$ and $\rho = 2$ correspond to the Laplacian and Gaussian distributions, respectively, and the boundary value of $\rho \to \infty$ corresponds to the uniform distribution for the interval $(\theta - s, \theta + s)$.

Applying (7) to the joint probability density function of independent noisy samples (5), each of which obeying the law (14) with the same scale s, leads to the problem:

$$\theta^* = \arg\max_{\theta}\{-\sum_{i=1}^{N}(|x_i - \theta|/s)^{\rho} - N\ln[2s\Gamma(1+1/\rho)]\} = \arg\min_{\theta}[\sum_{i=1}^{N}|x_i - \theta|^{\rho}], \quad (15)$$

where $0 < \rho < \infty$. If $\rho \to \infty$, then from (15) we have the problem:

$$\theta^* = \arg\min_{\theta}\max_{1 \le i \le N}|x_i - \theta|. \quad (16)$$

The problem (16) has an obvious analytical solution [9]

$$\theta^* = [\min_{1 \le i \le N}(x_i) + \max_{1 \le i \le N}(x_i)]/2 = [x_{(1)} + x_{(N)}]/2, \quad (17)$$

where the lower indices in parentheses indicate the ordered sequence elements. Equation (17) gives an analytical solution of the problem of location parameter estimation for the noise with a uniform distribution.

Applying (7) to the joint probability density function of independent noisy samples (5), each of which obeying the law (14) with different scale s_i, leads to the problem:

$$\theta^* = \arg\min_{\theta}\{\sum_{i=1}^{N} w_i\,|x_i - \theta|^{\rho}\} \quad (18)$$

where $w_i = (1/s_i)^{\rho}$. In the general case, the problem (18) has no analytical solution. However, one can note the following. If $0 < \rho < 1$, then the solution of the problem (18) must occur at one of the samples x_i. Hence, it can be obtained by a search among them. For this case, the objective function of the problem (18) is no convex neither unimodal, and its derivative has infinite derivative discontinuity at the points x_i. This property makes it possible to apply just zero-order optimization methods based on using values of the objective function. If $\rho = 1$, then the solution of the problem (18) is the weighted median (11). Although in this case the objective function of the problem (18) is convex, but it has finite derivative discontinuity at the points x_i. Such a behavior of the objective function also makes it possible to apply just zero-order optimization methods. If $1 < \rho < \infty$, then the objective function of the problem (18) becomes strictly convex, and the solution of the problem (18) is achieved at a single point. First and second order optimization methods based on the use of the first (for $1 < \rho < \infty$) and the second (for $2 \le \rho < \infty$) derivatives of the objective function can be used to search this point.

Let the independent n_i obeying the Cauchy distribution:

$$p(x) = \frac{s}{\pi} \cdot \frac{1}{s^2 + |x - \theta|^2} = \frac{1}{\pi s} \cdot \frac{1}{(1 + |x - \theta|^2 / s^2)}, \quad (19)$$

where θ is the location parameter, and $s > 0$ is the scale parameter. Then the joint probability density function is

$$p(x_1, ..., x_N \mid \theta) = \left(\frac{s}{\pi}\right)^N \cdot \prod_{i=1}^{N} \frac{1}{s^2 + |x_i - \theta|^2}. \quad (20)$$

Applying (7) to (20), we have the following equivalent problems:

$$\begin{aligned} \theta^* &= \arg\max_{\theta} \prod_{i=1}^{N} (s^2 + |x_i - \theta|^2)^{-1} = \arg\max_{\theta} \ln[\prod_{i=1}^{N} (s^2 + |x_i - \theta|^2)^{-1}] = \\ &= \arg\min_{\theta} \sum_{i=1}^{N} \ln(s^2 + |x_i - \theta|^2) = \arg\min_{\theta} \sum_{i=1}^{N} \ln(1 + |x_i - \theta|^2 / s^2) \end{aligned}. \quad (21)$$

The penultimate record in (21) is used as a mathematical statement of the "myriad" filtering problem, the solution of which received the designation $myriad\{x_i |_{i=1}^N; s\}$ and the name "myriad" [10].

In the case of different values of scale s_i, we obtain:

$$\theta^* = \arg\max_{\theta} \prod_{i=1}^{N} (s_i^2 + |x_i - \theta|^2)^{-1} = \arg\min_{\theta} \sum_{i=1}^{N} \ln(s_i^2 + |x_i - \theta|^2). \quad (22)$$

Assuming $h_i = (\gamma / s_i)^2$, where γ is a free parameter, from (22) we obtain [10]

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \ln(\gamma^2 + h_i |x_i - \theta|^2) = myriad\{h_i \circ x_i |_{i=1}^N; \gamma\}, \quad (23)$$

where $\circ$ denotes the weighting operation. Note that the myriad does not have to coincide with any x_i, but its value does not go beyond the range of these values. Limiting cases of the myriad are the arithmetic mean (when $\gamma \to \infty$) and the mode (when $\gamma \to 0$) [10].

Let the independent n_i obeying the distribution [6]:

$$p(x) = \frac{s}{2} \cdot \frac{1}{(s + |x - \theta|)^2}, \tag{24}$$

where θ is the location parameter, and $s > 0$ is the scale parameter. Then

$$p(x_1, \ldots, x_N \mid \theta) = \left(\frac{s}{\pi}\right)^N \cdot \prod_{i=1}^{N} \frac{1}{(s + |x_i - \theta|)^2}. \tag{25}$$

Applying (7) to (25), we have the following equivalent problems:

$$\theta^* = \arg\max_{\theta} \prod_{i=1}^{N} (s + |x_i - \theta|)^{-2} = \arg\min_{\theta} \sum_{i=1}^{N} \ln(s + |x_i - \theta|). \tag{26}$$

The solution of (26) is called the "meridian" and denoted as $meridian\{x_i |_{i=1}^{N}; s\}$.

In the case of different values of scale s_i, we obtain

$$\theta^* = \arg\max_{\theta} \prod_{i=1}^{N} (s_i + |x_i - \theta|)^{-2} = \arg\min_{\theta} \sum_{i=1}^{N} \ln(s_i + |x_i - \theta|). \tag{27}$$

Assuming $h_i = \gamma / s_i$, where γ is the free parameter, from (27) we obtain [6]

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \ln(\gamma + h_i |x_i - \theta|) = meridian\{h_i \circ x_i |_{i=1}^{N}; \gamma\}, \tag{28}$$

where $\circ$ denotes the weighting operation. Note that the meridian coincides with the some x_i. Limiting cases of the meridian are the median (when $\gamma \to \infty$) and the mode (when $\gamma \to 0$) [6].

The Cauchy distribution (19) and the "meridian" distribution (24) are special cases of the generalized Cauchy distribution [6]:

$$p(x) = \frac{\rho \Gamma(2/\rho) s}{2[\Gamma(1/\rho)]^2} \cdot \frac{1}{[s^\rho + |x - \theta|^\rho]^{2/\rho}} = \frac{\rho \Gamma(2/\rho)}{2s[\Gamma(1/\rho)]^2} \cdot \left[1 + \left(\frac{|x - \theta|}{s}\right)^\rho\right]^{-2/\rho}, \tag{29}$$

where $\Gamma(x)$ is a gamma function, θ is a location parameter, $s > 0$ is a scale parameter, $\rho > 0$ is a shape parameter, which is also called a tail constant [8]. From (29) it follows that in the case of $0 < \rho \le 2$ the mean and the variance are not exist because the corresponding integrals are divergent.

Applying (7) to the joint probability density function of independent noisy samples (5), each of which obeying the law (29) with the different scale s_i, leads to the problem:

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \ln(\gamma^\rho + w_i |x_i - \theta|^\rho), \tag{30}$$

where $w_i = (\gamma / s_i)^{\rho}$ and $0 < \rho < \infty$. The problem (30) generalizes the problems given by (23) and (28). Moreover, if $\gamma \to \infty$, then from the problem (30) we obtain the problem (18) formulated for the generalized Gaussian distribution (because $\ln(1+x) \approx x$ for $x << 1$). Thus, the method for solving the problem (30) is more general than the solving method of the problem (18). In addition, if $0 < \rho \le 1$ then the solution of the problem (30) is achieved in one of the points x_i. Therefore, it can be obtained by checking these points. In this case, the objective function of the problem (30) is not unimodal, and its derivative has a discontinuity at these points. This makes it possible to apply just zero-order optimization methods. However, if $1 < \rho \le 2$ then the objective function of the problem (30) is also not unimodal, and the solution is obtained in one of the local minima, the number of which is limited by the total number of points. At the same time, solving the problem of finding a global minimum is a rather complicated problem [6].

Criteria and methods for detecting anomalous values. The anomalous values are the values that are outside the acceptable range, i.e. they are the outliers. The traditional approach to the problem of removing outliers is to use the method of median filtering. For successful median filtering it is necessary that the amount of outliers within the aperture of median filter was less than half of the total number of samples in aperture. Otherwise, the quality of median filtering can be bad.

An important class of methods to remove outliers is the class of robust non-parametric methods based on trimming and winsorizing procedures. Trimming means throwing out the "bad" values, while winsorizing means replacing them with the certain "good" ones. In particular, such a method is the method of constructing an α-trimmed mean value:

$$m(\alpha) = \frac{1}{(N-2q)} \sum_{i=q+1}^{N-q} x_{(i)} , \qquad (31)$$

where α is chosen such that $q = \lfloor N\alpha \rfloor$, where $\lfloor ... \rfloor$ means the integer part of the number, and $x_{(i)}$ denotes the element of ordered sequence of x_i. From (31) it follows if $\alpha = 1/4$ then $m(1/4)$ is the mean value of samples in the interquartile interval. Winzorizing consists in replacing with $x_{(q+1)}$ the q number of the first elements of ordered sequence, in replacing with $x_{(N-q)}$ the q number of the last elements of ordered sequence, where $1 \le q < N/2$, and then in calculating the mean [12].

Classical criteria of anomalous values detection when a large number of samples have the Gaussian statistics are considered in [13]. In particularly, the statistical methods of outlier's detection, which are based on the generalization of the Grabbs criterion [11], are considered in [14]. These methods can be used when the number of outliers is small, and there are the appropriate tables of percentage points of statistics for the given data samples size. In particular, the Grabbs criterion for detecting the one outlier, which can be the maximum value $x_{(N)}$, is based on statistics [14],

$$G_N = (x_{(N)} - \bar{x}) / S, \tag{32}$$

where $\bar{x} = \sum_{i=1}^{N} x_i$ and $S = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$. The value $x_{(N)}$ is considered to be the outlier if the value of G_N exceeds the critical value $G_{N,1-\alpha}$, where α is the significance level [15]. Similarly to (32), the Grabbs criterion is also used to test the smallest sample value $x_{(1)}$. In this case, the statistics $G_1 = (\bar{x} - x_{(1)}) / S$ are compared with the critical value $G_{1,1-\alpha}$. If $G_1 > G_{1,1-\alpha}$, then the decision that the $x_{(1)}$ is the outlier is made. It is noted in [14] that the Grabbs criterion can actually be used to detect the "abnormal" data elements, that is, "to detect the abnormal element only in the case of a normal law". If the distribution law differs from normal (Gaussian) law, then the use of corresponding tables can lead both to the omission of outlier as well as to the erroneous assignment of a good data element to the class of outliers. The applicability of the Grabbs type criteria faces at least with the two following problems. Firstly, the data may have more number of outliers than number of outliers was given a priori. This leads to the errors in the estimates of the mean and the variance, since these estimates are not robust. Therefore, to use the Grabbs criterion it is necessary to check the data for a different number of outliers consistently. Secondly, the number of possible "true" data distribution is too great. Therefore, the identification of "true" distribution, even within the framework of parametric approach (when it is necessary to determine the values of few parameters) also faces with the stability problem, since outliers can lead to the significant errors in estimating the parameters of "true" distribution.

Criteria and methods for processing of data with contaminated noise. If noise distribution is "contaminated" by another distribution, the robust methods are used for data processing [16]. The model of contaminated noise is usually described by ratio [9]:

$$p_\varepsilon(x) = (1-\varepsilon)g(x) + \varepsilon h(x), \tag{33}$$

where $g(x)$ is a known probability density function, $h(x)$ is an unknown probability density function, and the parameter ε specifies the value of contamination in the range from 0 to 1 [16]. The well-known model of (33) is the Tukey model, in which $g(x)=\mathbf{N}(x;\mu,\sigma^2)$ and $h(x)=\mathbf{N}(x;\mu,9\sigma^2)$, where $\mathbf{N}(...)$ denotes the density of normal distribution. Considering this model, Huber noted [2] that editing data with abnormal values by cutting the values that are allocated in the general picture, and the subsequent application of classical criteria and evaluation procedures is inferior to the application of robust methods.

The most common method for constructing robust estimation is the estimation method of "maximum-likelihood type" or, in short, the M-estimation method [2]. In literature, this method name is associated with the criterion of "generalized maximum likelihood" [16], although sometimes the method of M-estimation is also called a generalized method of maximum likelihood [9]. Using this method, the problem of estimating the location parameter value is formulated as a minimization problem:

$$\theta^* = \arg\min_{\theta} \sum_{i=1}^{N} \psi(x_i - \theta), \quad (34)$$

where $\psi(x)$ is a cost function which may not have probabilistic interpretation [16]. Although the function $\psi(x)$ is sometimes called a loss function or a weight function [3], hereinafter it will be referred to as the cost function. Note that the designation $\psi(x)$ is taken from [3], and it corresponds to the designation $\rho(x)$ that is given in [17]. According to (34), this function forms the objective function of the minimization problem. List of some known cost functions with normalization coefficients that improve the graphic comparison of these functions is given in Table 1.

A visual comparison of the plots of cost functions given in Table 1 indicates that Tukey's, Hampel's, Andrew's and Meshalkin's cost functions are very similar. In particular, they have quadratic behavior in the neighborhood of zero and quickly go to horizontal asymptotes, thus implementing a strong filter against contamination [3]. In addition, the Geman-McClure-Demidenko function is close to these cost functions, while the cost functions of Huber, Cauchy-Gonzalez-Arce and Aysal-Barner differ significantly from them since they do not have the horizontal asymptotes.

Table 1

Cost functions and normalization coefficients

function name	$\psi(x)$	k
Huber [18]	$k\cdot\begin{cases}x^2; & \lvert x\rvert\le c\\ 2c\lvert x\rvert-c^2; & \lvert x\rvert>c\end{cases}$	$1/c^2$
Tukey [2]	$k\cdot\begin{cases}1-[1-(x/c)^2]^3; & \lvert x\rvert\le c\\ 1; & \lvert x\rvert>c\end{cases}$	1
Hampel [16]	$k\cdot\begin{cases}x^2; & 0\le\lvert x\rvert\le a\\ 2a\lvert x\rvert-a^2; & a<\lvert x\rvert\le b\\ -\dfrac{a(c-\lvert x\rvert)^2}{c-b}+a(b+c-a); & b<\lvert x\rvert\le c\\ a(b+c-a); & \lvert x\rvert>c\end{cases}$	$1/[a(b+c-a)]$
Andrews [19]	$k\cdot\begin{cases}c(1-\cos(x/c)); & \lvert x\rvert\le\pi c\\ 2c; & \lvert x\rvert>\pi c\end{cases}$	$1/(2c)$
Meshalkin [20]	$k\cdot\left(1-\exp\left(-\dfrac{\lambda x^2}{2}\right)\right),\quad \lambda>0$	$(1-e^{-\lambda/2})^{-1}$, $\lambda=10$
Geman-McClure-Demidenko [3]	$k\cdot\dfrac{x^2}{x^2+c},\quad c>0$	$1+c$, $c=0.16$
Cauchy-Gonzalez-Arce [21]	$k\cdot\ln(1+x^2/c^2)$	$1/\ln(1+1/c^2)$, $c=0.16$
Aysal-Barner [6]	$k\cdot\ln(1+\lvert x\rvert/c),\quad c>0$	$1/\ln(1+1/c)$, $c=0.4$

From (34) it follows that the estimate of the location parameter can be obtained by solving the equation

$$\sum_{i=1}^{N}\psi'(x_i-\theta)=0, \tag{35}$$

where $\psi'(x)$ is a derivative of cost function $\psi(x)$. Since (35) follows from (34), the problem (35) may be more complicated than the problem (34). Indeed, if the cost function is not convex, then the corresponding object function will have several minima and maxima. Hence, the roots of equation (35) will be both the points of the minima and the points of the maxima.

Superset of cost functions and generalized criterion. An aggregation of several cost functions into a general "superset" [4] with the corresponding technique of their transformations [22] allows eliminating redundancy of known cost functions and provide a wide range of possible solutions including mean, median, mode, myriad and meridian as limiting cases. Advantage of such a superset is due to ability to tune a processing method onto unknown noise environment.

The superset of cost functions is defined by the function [4]:

$$\psi_S^{(\alpha,\beta,q)}(x) = k_S^{(\alpha,\beta,q)}[(1+|x/\alpha|^q)^{\beta/q} - 1], \qquad (36)$$

where α is a smoothing parameter; q is a smoothing degree parameter ($0 < q < \infty$); β is a parameter of the form of cost function ($-\infty < \beta \le 1$), where $\beta < q$; $k_S^{(\alpha,\beta,q)}(x) = 1/[(1+|x_0/\alpha|^q)^{\beta/q} - 1]$ is a constant, which is necessary to perform the transformations of cost functions one to another within the superset framework; x_0 is a normalization point, in which the equality: $\psi_S^{(\alpha,\beta,q)}(x_0) = 1$ is ensured (usually $x_0 = 1$). The parameters α, β and q have the sense of free parameters needed to change the behavior of cost function (36).

The optimal settings for these free parameters are the following. For Gaussian noise we have $\alpha \to \infty$, $\beta = const$ and $q = 2$. For Laplacian noise we have $\alpha \to \infty$, $\beta = const$ and $q = 1$. For Cauchy noise with scale s we have $\alpha = s$, $\beta \to 0$ and $q = 2$. For "Meridian" noise with scale s we have $\alpha = s$, $\beta \to 0$ and $q = 1$. For generalized Cauchy noise with scale s and tail constant $0 < \rho \le 2$ we have $\alpha = s$, $\beta \to 0$ and $q = \rho$. In the absence of noise, but in presence of anomalous values we have several options, in particularly, $\alpha = 0$, $0 < \beta < 1$ and $q = const$ as well as $\alpha \ne 0$, $\beta \to 0$ and $q \to 0$.

Thus, the superset of cost functions allows to form various criteria for data processing by changing the values of its free parameters. Therefore, we can formulate the following generalized criterion for data processing:

$$\min_{\theta} \sum_{i=1}^{N} \psi_S^{(\alpha,\beta,q)}(x_i;\, \theta), \qquad (37)$$

where θ is a vector of unknown parameters, and where the free parameters α, β, and q should be tuned onto the current noise environment. Note, if the noise environment is known then the free parameters should be equal to the corresponding optimal values.

In practice, the tuning of free parameters onto the current noise environment can be performed on a certain grid of their values by minimizing the error received for a priori known solution, which may be a constant. Therefore, in the case of unknown noise environment, instead of obtaining the statistics necessary to determine the data distribution and then choosing the cost function corresponding to this

 ISSN 1562-9945

noise environment, one can tune the values of free parameters and use the proposed method. Taking into account that in the general case the obtained object function is not convex and is not unimodal, the implementation of the proposed approach may be laborious.

Conclusions. To solve the problem of processing noisy data with anomalous values when the statistics of noise and/or anomalies are unknown, it is expedient to use the proposed approach based on the method of generalized maximum likelihood with the superset of cost functions. The superset of cost functions allows forming various data processing criteria by changing the values of free parameters. It provides the ability to tune the data processing onto the current noise environment.

REFERENCES

1. Rousseeuw P. J. Anomaly detection by robust statistics / P. J. Rousseeuw, M. Hubert // WIREs Data Mining Knowledge Discovery. – 2018. – V. 8. – N. 2. – P. 1-14. DOI= https://doi.org/10.1002/widm.1236.

2. Huber P. Robust statistics / P. Huber, E. M. Ronchetti. – Hoboken: Wiley. – 2009. – 370 p. DOI: 10.1002/9780470434697.

3. Демиденко Е. З. Оптимизация и регрессия / Е. З. Демиденко. – М.: Наука. – 1989. – 296 с.

4. Borulko V.F. Minimum-duration filtering / V. F. Borulko, S. M. Vovk // Радіоелектроніка, інформатика, управління. — 2016. — N. 1. — С.7-14. DOI: 10.15588/1607-3274-2016-1-1.

5. Two-dimensional digital signal processing II: transforms and median filters // J.-O. Eklundh, T. S. Huang, B. I. Justusson, H. J. Nussbaumer, S. G. Tyan, S. Zohar. Edited by T. S. Huang. – Springer-Verlag. – 1981. – 222 p.

6. Aysal T. C. Meridian filtering for robust signal processing / T. C. Aysal, K. E. Barner // IEEE Trans. on Signal Processing. – 2007. – V. 55. – N. 8. – P. 3949–3962.

7. Nadarajah S. A generalized normal distribution / S. Nadarajah // Journal of Applied Statistics. – 2005.– V.32.– N.7.– P.685–694. DOI:10.1080/02664760500079464.

8. Carrillo R. E. Generalized Cauchy distribution based robust estimation / R. E. Carrillo, T.C. Aysal, K.E. Barner // Acoustic, Speech and Signal Processing : IEEE International Conference ICASSP 2008, Las Vegas, 31 March – 4 April 2008: proceedings. – IEEE. – 2008. – P. 3389–3392. DOI: 10.1109/ICASSP.2008.4518378.

9. Грановский В.А. Методы обработки экспериментальных данных при измерениях / В. А. Грановский, Т. Н. Сирая – Л.: Энергоатомиздат. – 1990. – 288 с.

10. Gonzalez J.G. Optimality of the myriad filter in practical impulsive-noise environments / J. G. Gonzalez, G. R. Arce // IEEE Trans. on Signal Processing. – 2001. – V. 49. – N. 2. – P. 438–441. DOI: 10.1109/78.902126.

11. Grubbs F.E. Sample Criteria for Testing Outlying observations / F. E. Grubbs // Ann. Math. Statist. – 1950. – V. 21. – N. 1. – P.27-58.

12. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ – М.: Мир, 1982. – 488 с.

13. David H. A. Order Statistics // H. A. David, H. N. Nagaraja. – John Wiley & Sons. – 2003. – 465 p. DOI:10.1002/0471722162.

14. Лемешко Б.Ю. Расширение области применения критериев типа Граббса, используемых при отбраковке аномальных измерений / Б. Ю. Лемешко, С. Б. Лемешко // Измерительная техника. – 2005. – N. 6. – С.13-19.

15. Grubbs F. E. Extension of sample sizes and percentage points for significance tests of outlying observations / F. E. Grubbs, G. Beck // Technometrics. – 1972. – Vol. 14. – No. 4. – P.847-854.

16. Hampel F. R. Robust statistics: the approach based on influence functions / F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. – Hoboken: Wiley. – 2011. – 536 p. DOI=10.1002/9781118186435.

17. Huber P. J. Finite sample breakdown of M- and P-estimators / P. J. Huber // The Annals of Statistics. – 1984. – Vol. 12. – N 1. – P. 119-126.

18. Huber P. J. Robust estimation of a location parameter / P. J. Huber // Annals of Mathematical Statistics. – 1964. – Vol. 35. – Issue 1. – P. 73-101.

19. Andrews D. F. Robust Estimates of Location / D. F. Andrews, P. J. Bickel, F. R. Hampel, P. J. Huber, W. H. Rogers, J. W. Tukey. – Princeton Univ. Press, Princeton. – 1972.

20. Meshalkin L. D. Some mathematical methods for the study of non-communicable diseases / L. D. Meshalkin // Uses of epidemiology in planning of health services. Proc. of the Sixth International Scientific Meeting, Primosten, August 29—September 3, 1971. – Belgrade. – 1973. – P. 250–256.

21. Gonzalez J. G. Weighted myriad filters: a robust filtering framework derived from alpha-stable distributions / J. G. Gonzalez, G. R. Arce // Proc. of Int. Conf. Acoustic, Speech and Signal Processing, ICASSP'96, Atlanta, 1996. – P. 2833–2836.

22. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle / S. M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7 – P. 281-292. DOI: 10.3103/S0735272716070013.

REFERENCES

1. Rousseeuw P. J. Anomaly detection by robust statistics / P. J. Rousseeuw, M. Hubert // WIREs Data Mining Knowledge Discovery. – 2018. – V. 8. – N. 2. – P. 1-14. DOI= https://doi.org/10.1002/widm.1236.

2. Huber P. Robust statistics / P. Huber, E. M. Ronchetti. – Hoboken: Wiley. – 2009. – 370 p. DOI: 10.1002/9780470434697.

3. Demidenko E. Z. Optimizatsiya i regressiya / E. Z. Demidenko. – M.: Nauka. – 1989. – 296 s.

4. Borulko V.F. Minimum-duration filtering / V. F. Borulko, S. M. Vovk // Radio Electronics, Computer Science, Control.—2016.—N.1.— C.7-14.DOI: 10.15588/1607-3274-2016-1-1.

5. Two-dimensional digital signal processing II: transforms and median filters // J.-O. Eklundh, T. S. Huang, B. I. Justusson, H. J. Nussbaumer, S. G. Tyan, S. Zohar. Edited by T. S. Huang. – Springer-Verlag. – 1981. – 222 p.

6. Aysal T. C. Meridian filtering for robust signal processing / T. C. Aysal, K. E. Barner // IEEE Trans. on Signal Processing. – 2007. – V. 55. – N. 8. – P. 3949–3962.

7. Nadarajah S. A generalized normal distribution / S. Nadarajah // Journal of Applied Statistics. – 2005.–V.32.–N.7.– P. 685–694. DOI:10.1080/02664760500079464.

8. Carrillo R. E. Generalized Cauchy distribution based robust estimation / R. E. Carrillo, T.C. Aysal, K.E. Barner // Acoustic, Speech and Signal Processing : IEEE International Conference ICASSP 2008, Las Vegas, 31 March – 4 April 2008: proceedings. – IEEE. – 2008. – P. 3389–3392. DOI: 10.1109/ICASSP.2008.4518378.

9. Granovskiy V.A. Metodyi obrabotki eksperimentalnyih dannyih pri izmereniyah / V.A. Granovskiy, T. N. Siraya – L.: Energoatomizdat. – 1990. – 288 s.

10. Gonzalez J.G. Optimality of the myriad filter in practical impulsive-noise environments / J. G. Gonzalez, G. R. Arce // IEEE Trans. on Signal Processing. – 2001. – V. 49. – N. 2. – P. 438–441. DOI: 10.1109/78.902126.

11. Grubbs F.E. Sample Criteria for Testing Outlying observations / F. E. Grubbs // Ann. Math. Statist. – 1950. – V. 21. – N. 1. – P.27-58.

12. Afifi A., Eyzen S. Statisticheskiy analiz. Podhod s ispolzovaniem EVM – M.: Mir, 1982. – 488 s.

13. David H. A. Order Statistics // H. A. David, H. N. Nagaraja. – John Wiley & Sons. – 2003. – 465 p. DOI:10.1002/0471722162.

14. Lemeshko B.Yu. Rasshirenie oblasti primeneniya kriteriev tipa Grabbsa, ispolzuemyih pri otbrakovke anomalnyih izmereniy / B. Yu. Lemeshko, S. B. Lemeshko // Izmeritelnaya tehnika. – 2005. – N. 6. – S.13-19.

15. Grubbs F. E. Extension of sample sizes and percentage points for significance tests of outlying observations / F. E. Grubbs, G. Beck // Technometrics. – 1972. – Vol. 14. – No. 4. – P.847-854.

16. Hampel F. R. Robust statistics: the approach based on influence functions / F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, W. A. Stahel. – Hoboken: Wiley. – 2011. – 536 p. DOI=10.1002/9781118186435.

17. Huber P.J. Finite sample breakdown of M- and P-estimators / P. J. Huber // The Annals of Statistics. – 1984. – Vol. 12. – N 1. – P. 119-126.

18. Huber P.J. Robust estimation of a location parameter / P. J. Huber // Annals of Mathematical Statistics. – 1964. – Vol. 35. – Issue 1. – P. 73-101.

19. Andrews D. F. Robust Estimates of Location / D. F. Andrews, P. J. Bickel, F. R. Hampel, P.J. Huber, W. H. Rogers, J. W. Tukey. – Princeton Univ. Press, Princeton. – 1972.

20. Meshalkin L.D. Some mathematical methods for the study of non-communicable diseases / L. D. Meshalkin // Uses of epidemiology in planning of health services. Proc. of the Sixth International Scientific Meeting, Primosten, August 29—September 3, 1971. – Belgrade. – 1973. – P. 250–256.

21. Gonzalez J.G. Weighted myriad filters: a robust filtering framework derived from alpha-stable distributions / J. G. Gonzalez, G. R. Arce // Proc. of Int. Conf. Acoustic, Speech and Signal Processing, ICASSP’96, Atlanta, 1996. – P. 2833–2836.

22. Vovk S.M. General approach to building the methods of filtering based on the minimum duration principle / S. M. Vovk // Radioelectronics and Communications Systems. – 2016. – V. 59. – N. 7 – P. 281-292. DOI: 10.3103/S0735272716070013

 ISSN 1562-9945