mathematics

MDPI

Article

Multiple Outlier Detection Tests for Parametric Models

Vilijandas Bagdonavičius [1] and Linas Petkevičius [2,*]

[1] Institute of Applied Mathematics, Vilnius University, Naugarduko 24, Vilnius LT-03225, Lithuania; vilijandas.bagdonavicius@mif.vu.lt

[2] Institute of Computer Science, Vilnius University, Didlaukio 47, Vilnius LT-08303, Lithuania

* Correspondence: linas.petkevicius@mif.vu.lt

Received: 8 November 2020; Accepted: 30 November 2020; Published: 3 December 2020

Abstract: We propose a simple multiple outlier identification method for parametric location-scale and shape-scale models when the number of possible outliers is not specified. The method is based on a result giving asymptotic properties of extreme z-scores. Robust estimators of model parameters are used defining z-scores. An extensive simulation study was done for comparing of the proposed method with existing methods. For the normal family, the method is compared with the well known Davies-Gather, Rosner's, Hawking's and Bolshev's multiple outlier identification methods. The choice of an upper limit for the number of possible outliers in case of Rosner's test application is discussed. For other families, the proposed method is compared with a method generalizing Gather-Davies method. In most situations, the new method has the highest outlier identification power in terms of masking and swamping values. We also created R package outliersTests for proposed test.

Keywords: location-scale models; outliers identification; unknown number of outliers; outlier region; robust estimators

1. Introduction

The problem of multiple outliers identification received attention of many authors. The majority of outlier identification methods define rules for the rejection of the most extreme observations. The bulk of publications are concentrated on the normal distribution (see [1–6], see surveys in [7,8]. For non-normal case, the most of the literature pertains to the exponential and gamma distributions, see [9–17]. Outliers identification is important analyzing data collected in wide range of areas: pollution [18], IoT [19], medicine [20], fraud [21], smart city applications [22], and many more.

Constructing outlier identification methods, most authors suppose that the number s of observations suspected to be outliers is specified. These methods have a serious drawback: only two possible conclusions are done: exactly s observations are admitted as outliers or it is concluded that outliers are absent. More natural is to consider methods which do not specify the number of suspected observations or at least specify the upper limit s for it. Such methods are not very numerous and they concern normal or exponential samples. These are [1,5,23] methods for normal samples, [15,16,24] methods for exponential samples. The only method which does not specify the upper limit s is the [2] method for normal samples.

We give a competitive and simple method for outlier identification in samples from location-scale and shape-scale families of probability distributions. The upper limit s is not specified, as in the the case of Davies-Gather method. The method is based on a theorem giving asymptotic properties of extreme z-scores. Robust estimators of model parameters are used defining z-scores.

The following investigation showed that the proposed outlier identification method has superior performance as compared to existing methods. The proposed method widens considerably the scope of

models applied in statistical analysis of real data. Differently from the normal probability distribution family many two-parameter families such as Weibull, logistic and loglogistic, extreme values, Cauchy, Laplace and other families can be applied for outlier identification. So it may be useful for models which cannot be symmetrized by simple transformations such as log-transform or others.

An advantage of the new method is that complicated computing is not needed because search of test statistic's critical values by simulation is not needed for each sample size. It allowed to create an R package outliersTests which can be used for outlier search in real datasets. Another advantage is a very good potential for generalizations of the proposed method to regression, time series and other models. To have a competitor, we present not only the new method but also generalize the Davies-Gather method for non-normal data.

In Section 2 we present a short overview of the notion of the outlier region given by [2]. In Section 3 we give asymptotic properties of extreme z-scores based on equivariant estimators of model parameters, and introduce a new outlier identification method for parametric models based on the asymptotic result and robust estimators. In Section 4 we consider rather evident generalizations of Davies-Gather tests for normal data to location-scale families. In Section 5 we give a short overview of known multiple outlier identification methods for normal samples which do not specify an exact number of suspected outliers. In Section 6 we compare performance of the new and existing methods.

2. Outliers and Outlier Regions

Suppose that data are independent random variables $X_1, \ldots, X_n$. Denote by $F_i(x)$ the c.d.f. of X_i.

Let $\mathcal{F}_0 = \{F(x,\theta), \theta \in \Theta \subset \mathbf{R}^m\}$ be a parametric family of absolutely continuous cumulative distribution functions with continuous unimodal densities f on the support $\mathrm{supp}(F)$ of the c.d.f. F.

Suppose that if the data are not contaminated with unusual observations, then the following null hypothesis H_0 is true: there exist $\theta \in \Theta$ such that

$$F_1(x) = \ldots = F_n(x) = F(x,\theta). \tag{1}$$

There are two different definitions of an outlier. In the first case the outlier is an observation which falls into some outlier region $out(X)$. The outlier region is a set such that the probability for at least one observation from a sample to fall into it is small if the hypothesis H_0 is true. In such a case the probability that a specified observation X_i falls into $out(X)$ is very small. If an observation X_i has distribution different from that under H_0 then this probability may be considerably higher.

In the second case, the value x_i of X_i is an outlier if the probability distribution of X_i is different from that under H_0, formally $F_i \neq F(x,\theta)$. In this case, outliers are often called *contaminants*.

Therefore, in the first case, there exists a very small probability to have an outlier under H_0. If the hypothesis H_0 holds, then contaminants are absent and with very small probability some outliers (in the first sense) are possible. If contaminants are present, then the hypothesis H_0 is not true. Nevertheless, contaminants are not necessary outliers (in the first sense) because it is possible that they do not fall into the outlier region. So the two notions are different. Both definitions give approximately the same outliers if the alternative distribution is concentrated in the outlier region. Namely such contaminants can be called outliers in the sense that outliers are anomalous extreme observations. In such a case it is possible to compare outlier and contaminant search methods.

In this paper, we consider location-scale and shape-scale families. Location-scale families have the form $\mathcal{F}_{ls} = \{F_0((x-\mu)/\sigma), \mu \in \mathbf{R}, \sigma > 0\}$ with the completely specified baseline c.d.f F_0 and p.d.f. f_0. Shape-scale families have the form $\mathcal{F}_{ls} = \{G_0(((x/\theta)^\nu), \theta, \nu > 0\}$ with completely specified baseline c.d.f G_0 and p.d.f. g_0. By logarithmic transformation the shape-scale families are transformed to location-scale

family, so we concentrate on location-scale families. Methods for such families are easily modified to methods for shape-scale families.

The right-sided α-outlier region for a location-scale family is

$$out_r(\alpha_n, F) = \{x \in \mathbf{R} : x > \mu + \sigma F_0^{-1}(1-\alpha)\}$$

and the left-sided α-outlier region is

$$out_l(\alpha_n, F) = \{x \in \mathbf{R} : x < \mu + \sigma F_0^{-1}(\alpha)\}.$$

The two-sided α-outlier region has the form

$$out(\alpha, F) = \{x \in \mathbf{R}/[\mu + \sigma F_0^{-1}(\alpha/2), \mu + \sigma F_0^{-1}(1-\alpha/2)]\}. \quad (2)$$

If f_0 is symmetric, then the two-sided outlier region is simpler:

$$out(\alpha, F) = \{x \in \mathbf{R} : |x-\mu| > \sigma F_0^{-1}(1-\alpha/2)\}.$$

The value of α is chosen depending on the size n of a sample: $\alpha = \alpha_n$. The choice is based on assumption that under H_0 for some $\bar{\alpha}$ close to zero

$$\mathbf{P}\{\cap_{i=1}^n \{X_i \notin out(\alpha_n, F)\}\} = (\mathbf{P}\{X_i \notin out(\alpha_n, F)\})^n = 1-\bar{\alpha}. \quad (3)$$

The equality (3) means that under H_0 the probability that *none* of X_i falls into α_n-outlier region is $1-\bar{\alpha}$. It implies that

$$\alpha_n = 1 - (1-\bar{\alpha})^{1/n}. \quad (4)$$

The sequence α_n decreases from $\bar{\alpha}$ to 0 as n goes from 1 to ∞.

The first definition of an outlier is as follows: for a sample size n a realization x_i of X_i is called *outlier* if $x_i \in out(\alpha_n, F)$; x_i is called *right outlier* if $x_i \in out_r(\alpha_n, F)$.

The number of outliers D_n under H_0 has the binomial distribution $B(n, \alpha_n)$ and the expected number of outliers in the sample under H_0 is $\mathbf{E}D_n = n\alpha_n$. Please note that $\mathbf{E}D_n \to -\ln(1-\bar{\alpha}) \approx \bar{\alpha}$ as $n \to \infty$. For example, if $\bar{\alpha} = 0.05$, then $\ln(1-\bar{\alpha}) \approx 0.05129$ and for $n \geq 10$ the expected number of outliers is approximately 0.051, i.e., it practically does not depend on n. So under H_0 the expected number of outliers 0.051 is negligible with respect to the sample size n.

3. New Method

3.1. Preliminary Results

Suppose that a c.d.f. $F \in \mathcal{F}_{ls}$ belongs also to the domain of attraction $\mathcal{G}_\gamma$, $\gamma \geq 0$ (see [25]).

If $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, then there exist normalizing constants $a_n > 0$ and $b_n \in \mathbf{R}$ such that $\lim_{n\to\infty} F_0^n(a_n x + b_n) = e^{-e^{-x}}$. Similarly, if $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, then $\lim_{n\to\infty} F_0^n(a_n x + b_n) = e^{-(-x)^{-1/\gamma}}$, $x < 0$, $\lim_{n\to\infty} F^n(a_n x + b_n) = 1, x \geq 0$.

One of possible choices of the sequences $\{b_n\}$ and $\{a_n\}$ is

$$b_n = F_0^{-1}(1-\frac{1}{n}), \quad a_n = 1/(n f_0(b_n)). \quad (5)$$

In the particular case of the normal distribution equivalent form $a_n = 1/b_n$ can be used. Expressions of b_n and a_n for some most used distributions are given in Table 1.

Table 1. Expressions of b_n and a_n.

Distribution	$F_0(x)$	b_n	a_n
Normal	$\Phi(x)$	$\Phi^{-1}(1-1/n)$	$1/b_n$
Type I extreme value	$1-e^{-e^x}$	$\ln\ln n$	e^{-b_n}
Type II extreme value	$e^{-e^{-x}}$	$\ln(-\ln(1-1/n))$	$e^{b_n}/(n-1)$
Logistic	$\frac{1}{1+e^{-x}}$	$\ln(n-1)$	$n/(n-1)$
Laplace	$\frac{1}{2}+\frac{1}{2}\text{sign}(x)(1-e^{-\lvert x\rvert})$	$\ln(n/2)$	1
Cauchy	$\frac{1}{2}+\frac{1}{\pi}\arctan(x)$	$\cot(\frac{\pi}{n})$	$\frac{\pi}{n}/\sin^2(\frac{\pi}{n})$

Condition A.

(a) *$\hat{\mu}$ and $\hat{\sigma}$ are consistent estimators of μ and σ;*

(b) *the limit distribution of $(\sqrt{n}(\hat{\mu}-\mu), \sqrt{n}(\hat{\sigma}-\sigma))$ is non-degenerate;*

(c)

$$\lim_{x\to\infty}\frac{xf_0(x)}{\sqrt{1-F_0(x)}}=0.$$

Condition A (c) is satisfied for many location-scale models including the normal, type I extreme value, type II extreme value, logistic, Laplace ($F\in\mathcal{G}_0$), Cauchy ($F\in\mathcal{G}_1$).

Set $Y_i=(X_i-\mu)/\sigma$, $\hat{Y}_i=(X_i-\hat{\mu})/\hat{\sigma}$. The random variables $\hat{Y}_i$ are called z-scores. Denote by $Y_{(1)}\le Y_{(2)}\le\ldots\le Y_{(n)}$ and $\hat{Y}_{(1)}\le\ldots\le\hat{Y}_{(n)}$ the respective order statistics

The following theorem is useful for right outliers detection test construction.

Theorem 1. *If $F\in\mathcal{G}_0\cap\mathcal{F}_{ls}$ and Conditions A hold, then for fixed s*

$$((\hat{Y}_{(n)}-b_n)/a_n,(\hat{Y}_{(n-1)}-b_n)/a_n,\ldots,(\hat{Y}_{(n-s+1)}-b_n)/a_n)\xrightarrow{d}$$

$$L_0=(-\ln E_1,-\ln(E_1+E_2),\ldots,-\ln(E_1+\ldots+E_s))$$

as $n\to\infty$, where $E_1,\ldots,E_s$ are i.i.d. standard exponential random variables.

If $F\in\mathcal{G}_\gamma\cap\mathcal{F}_{ls}$, $\gamma>0$ and Conditions A hold, then the limit random vector is

$$L_\gamma=(E_1^{-1}-1,(E_1+E_2)^{-1}-1,\ldots,(E_1+\ldots+E_s)^{-1}-1).$$

Proof of Theorem 1. Please note that

$$\frac{\hat{Y}_{(n-i+1)}-b_n}{a_n}=\frac{Y_{(n-i+1)}-b_n}{a_n}\frac{\sigma}{\hat{\sigma}}+\frac{(\mu-\hat{\mu})}{\hat{\sigma}a_n}+\frac{b_n}{a_n}\frac{\sigma-\hat{\sigma}}{\hat{\sigma}}$$

The s-dimensional random vector such that its ith component is the first term of the right side converges in distribution to the random vector given in the formulation of the theorem. It follows from Theorem 2.1.1 of [25] and Condition A (a). So it is sufficient to show that the second and the third terms converge to zero in probability. The second term is

$$-\sqrt{n}f_0(F_0^{-1}(1-\frac{1}{n}))\frac{\sqrt{n}(\hat{\mu}-\mu)}{\hat{\sigma}},$$

the third term is

$$-\sqrt{n}F_0^{-1}(1-\frac{1}{n})f_0(F_0^{-1}(1-\frac{1}{n}))\frac{\sqrt{n}(\hat{\sigma}-\sigma)}{\hat{\sigma}}.$$

By Condition A (c)

$$\lim_{n\to\infty}\sqrt{n}F_0^{-1}(1-\frac{1}{n})f_0(F_0^{-1}(1-\frac{1}{n}))) = \lim_{x\to\infty}\frac{xf_0(x)}{\sqrt{1-F_0(x)}} = 0.$$

It also implies $\lim_{n\to\infty}\sqrt{n}f_0(F_0^{-1}(1-\frac{1}{n})) = 0$ because $\lim_{n\to\infty}F_0^{-1}(1-\frac{1}{n}) = \infty$. Proof completed. □

Remark 1. *Please note that* $2(E_1+\ldots+E_i) \sim \chi^2(2i)$. *It implies that if* $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, *then for fixed* i, $i = 1,\ldots,s$,

$$\mathbf{P}\{(\hat{Y}_{(n-i+1)} - b_n)/a_n \le x\} \to 1 - F_{\chi^2_{2i}}(2e^{-x}) \quad as \quad n \to \infty. \tag{6}$$

Similarly, if $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, *then for fixed* i, $i = 1,\ldots,s$,

$$\mathbf{P}\{(\hat{Y}_{(n-i+1)} - b_n)/a_n \le x\} \to 1 - F_{\chi^2_{2i}}(\frac{2}{1+x}) \quad as \quad n \to \infty. \tag{7}$$

The following theorem is useful for construction of outlier detection tests in two-sided case when f_0 is symmetric. For any sequence $\zeta_1,\ldots,\zeta_n$ denote by $|\zeta|_{(1)} \le \ldots \le |\zeta|_{(n)}$ the ordered absolute values $|\zeta_1|,\ldots,|\zeta_n|$.

Theorem 2. *Suppose that the function* f_0 *is symmetric. If* $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma \ge 0$ *and Conditions A hold, then for fixed* s

$$((|\hat{Y}|_{(n)} - b_{2n})/a_{2n}, (|\hat{Y}|_{(n-1)} - b_{2n})/a_{2n}, \ldots, (|\hat{Y}|_{(n-s+1)} - b_{2n})/a_{2n}) \xrightarrow{d} L_\gamma$$

as $n \to \infty$.

Proof of Theorem 2. For any $i = 1,\ldots,s$ the following equality holds:

$$\frac{|\hat{Y}|_{(n-i+1)} - b_{2n}}{a_{2n}} = \frac{|\hat{Y}|_{(n-i+1)} - |Y|_{(n-i+1)}}{a_{2n}} + \frac{|Y|_{(n-i+1)} - b_{2n}}{a_{2n}}. \tag{8}$$

The c.d.f. of the random variables $|Y_i|$ is $2F_0(x) - 1$, so if $F_0 \in \mathcal{G}_\gamma$, $\gamma \ge 0$ then $2F_0 - 1 \in \mathcal{G}_\gamma$, and for the sequence $|Y_n|$ the normalizing sequences are a_{2n}, b_{2n}. So the s-dimensional random vector such that its ith component is the second term of the right side converges in distribution to the random vector given in the formulation of the theorem. It follows from Theorem 2.1.1 of [25]. So it is sufficient to show that the first term converges in probability to zero.

Please note that $|\hat{Y}_i| \le |Y_i| + |\hat{Y}_i - Y_i|$, and

$$|\hat{Y}_i - Y_i| = \frac{1}{\hat{\sigma}}|\mu - \hat{\mu} + (\sigma - \hat{\sigma})Y_i| \le \frac{|\hat{\mu}-\mu|}{\hat{\sigma}} + \frac{|\sqrt{n}(\hat{\sigma}-\sigma)|}{\hat{\sigma}}\frac{1}{\sqrt{n}}|Y|_{(n)}.$$

So $|\hat{Y}|_{(n-j+1)} \le |Y|_{(n-j+1)} + \frac{|\hat{\mu}-\mu|}{\hat{\sigma}} + \frac{|\sqrt{n}(\hat{\sigma}-\sigma)|}{\hat{\sigma}}\frac{1}{\sqrt{n}}|Y|_{(n)}$. Analogously, the inequality $Y_i| \le ||\hat{Y}_i| + |\hat{Y}_i - Y_i|$ implies that $|Y|_{(n-j+1)} \le |\hat{Y}|_{(n-j+1)} + \frac{|\hat{\mu}-\mu|}{\hat{\sigma}} + \frac{|\sqrt{n}(\hat{\sigma}-\sigma)|}{\hat{\sigma}}\frac{1}{\sqrt{n}}|Y|_{(n)}$.

Theorem 2.1.1 in [25] applied to the random variables $|Y_i|$ implies that there exist a random variable V_1 with the c.d.f. $G(x) = e^{-e^{-x}}$ ($\gamma = 0$) or $G(x) = e^{-(-x)^{-1/\gamma}}, x < 0, G(x) = 1, x \geq 0$ ($\gamma > 0$), such that

$$\frac{1}{\sqrt{n}}|Y|_{(n)} = (b_{2n} + a_{2n}(V_1 + o_P(1)))/\sqrt{n}. \tag{9}$$

$$|\frac{|\hat{Y}|_{(n-i+1)} - |Y|_{(n-i+1)}}{a_{2n}}| \leq \frac{|\sqrt{n}(\hat{\mu} - \mu)|}{\hat{\sigma}\sqrt{n}a_{2n}} + \frac{|\sqrt{n}(\hat{\sigma} - \sigma)|}{\hat{\sigma}}(\frac{b_{2n}}{\sqrt{n}a_{2n}} + \frac{V_1 + o_p(1)}{\sqrt{n}}).$$

The convergence $b_n \to \infty$ and Condition A (c) imply:

$$\lim_{n\to\infty} \frac{b_{2n}}{\sqrt{n}a_{2n}} = \lim_{n\to\infty} \sqrt{n}F_0^{-1}(1 - \frac{1}{2n})f_0F_0^{-1}(1 - \frac{1}{2n})) =$$

$$\frac{1}{\sqrt{2}} \lim_{x\to\infty} \frac{xf_0(x)}{\sqrt{1 - F_0(x)}} = 0, \quad \lim_{n\to\infty} \frac{1}{\sqrt{n}a_{2n}} = 0.$$

These results and Conditions A (a), (b) imply that the first term at the right of (8) converges in probability to zero. Proof completed. □

Remark 2. *Theorem 2 implies that if $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, $n \to \infty$, then for fixed i, $i = 1, \ldots, s$,*

$$\mathbf{P}\{(|\hat{Y}|_{(n-i+1)} - b_{2n})/a_{2n} \leq x\} \to 1 - F_{\chi^2_{2i}}(2e^{-x}), \tag{10}$$

and if $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, then

$$\mathbf{P}\{(|\hat{Y}|_{(n-i+1)} - b_{2n})/a_{2n} \leq x\} \to 1 - F_{\chi^2_{2i}}(2/(1 + x)). \tag{11}$$

Suppose now that the function f_0 is not symmetric. Set $Y_i^* = -(X_i - \mu)/\sigma$. The c.d.f. and p.d.f. of Y_i^* are $1 - F_0(-x)$ and $f_0(-x)$, respectively. Set

$$b_n^* = -F_0^{-1}(1/n), \quad a_n^* = 1/(nf_0(-b_n^*)). \tag{12}$$

For example, if type I extreme value distribution is considered, then

$$b_n = \ln\ln n, \quad a_n = \frac{1}{\ln n}, \quad b_n^* = -\ln(-\ln(1 - \frac{1}{n})), \quad a_n^* = -\frac{1}{(n-1)\ln(1 - \frac{1}{n})}.$$

For the type II extreme value distribution a_n, b_n, a_n^*, b_n^* have the same expressions as a_n^*, b_n^*, a_n, b_n for the Type I extreme value distribution, respectively.

Remark 3. *Similarly as in Theorem 1 we have that if s is fixed and $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, then for fixed i, $i = 1, \ldots, s$,*

$$\mathbf{P}\{(Y_{(i)} + b_n^*)/(-a_n^*) \leq x\} = \mathbf{P}\{(\hat{Y}^*_{(n-i+1)} - b_n^*)/a_n^* \leq x\} \to 1 - F_{\chi^2_{2i}}(2e^{-x}), \tag{13}$$

and if $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, then for fixed i, $i = 1, \ldots, s$,

$$\mathbf{P}\{(Y_{(i)} + b_n^*)/(-a_n^*) \leq x\} = \mathbf{P}\{(\hat{Y}^*_{(n-i+1)} - b_n^*)/a_n^* \leq x\} \to 1 - F_{\chi^2_{2i}}(2/(1 + x)). \tag{14}$$

3.2. Robust Estimators for Location-Shape Distributions

The choice of the estimators $\hat{\mu}$ and $\hat{\sigma}$ is important when outlier detection problem is considered. The ML estimators from the complete sample are not stable when outliers exist.

In the case of location-scale families highly efficient robust estimators of the location and scale parameters μ and σ are (see [26])

$$\hat{\mu} = MED - \hat{\sigma}F_0^{-1}(0.5), \quad \hat{\sigma} = Q_n = d\, W_{([0.25n(n-1)/2])}, \tag{15}$$

where MED is the empirical median, $W_{ij} = |X_i - X_j|, \quad 1 \le i < j \le n$ are $C_n^2 = n(n-1)/2$ absolute values of the differences $X_i - X_j$ and $W_{(l)}$ is the lth order statistic from W_{ij}.

The constant d has the form $d = 1/K_0^{-1}(5/8)$, where $K_0^{-1}(x)$ is the inverse of the c.d.f. of $Y_1 - Y_2$, $Y_i = (X_i - \mu)/\sigma \sim F_0(x)$.

Expressions of $K_0^{-1}(x)$ and values d for some well-known location-scale families are given in Table 2.

Table 2. Values of d for various probability distributions.

Distribution	$K_0(x)$	d
Normal	$\Phi(x/\sqrt{2})$	2.2219
Type I extr.val.	$1/(1+e^{-x})$	1.9576
Type II extr.val.	$1/(1+e^{-x})$	1.9576
Logistic	$1 - \frac{(x-1)e^x+1}{(e^x-1)^2}$	1.3079
Laplace	$1 - \frac{1}{2}(1+\frac{x}{2})e^{-x}$	1.9306
Cauchy	$\frac{1}{2} + \frac{1}{\pi}\arctan(x/2)$	1.2071

The above considered estimators are equivariant under H_0, i.e. for any $e \in \mathbf{R}$, $f > 0$, the following equalities hold:

$$\hat{\mu}((X_1 - e)/f, \ldots, (X_n - e)/f) = (\hat{\mu}(X_1, \ldots, X_n) - e)/f,$$
$$\hat{\sigma}((X_1 - e)/f, \ldots, (X_n - e)/f) = \hat{\sigma}(X_1, \ldots, X_n)/f.$$

Equivariant estimators have the following property: the distribution of $(\hat{\mu} - \mu)/\sigma$, $\hat{\sigma}/\sigma$ and $(\hat{\mu} - \mu)/\hat{\sigma}$ does not depend on the values of the parameters μ and σ.

3.3. Right Outliers Identification Method for Location-Scale Families

Suppose that $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma \ge 0$. Let a_n, b_n be defined by (5). Set

$$U^+_{(n-i+1)}(n) = 1 - F_{\chi^2_{2i}}(2e^{-(\hat{Y}_{(n-i+1)} - b_n)/a_n}), \ \gamma = 0,$$

$$U^+_{(n-i+1)}(n) = 1 - F_{\chi^2_{2i}}(2/(1 + (\hat{Y}_{(n-i+1)} - b_n)/a_n), \ \gamma > 0,$$

$$U^+(n,s) = \max_{1 \le i \le s} U^+_{(n-i+1)}(n). \tag{16}$$

Theorem 3. *The distribution of the statistic $U^+(n,s)$ is parameter-free for any fixed n.*

Proof of Theorem 3. The result follows from the equality

$$\frac{\hat{Y}_{(n-i+1)} - b_n}{a_n} = \frac{Y_{(n-i+1)} - b_n}{a_n}\frac{\sigma}{\hat{\sigma}} + \frac{b_n}{a_n}(\frac{\sigma}{\hat{\sigma}} - 1) + \frac{1}{a_n}\frac{\mu - \hat{\mu}}{\hat{\sigma}},$$

equivariance of the estimators $\hat{\mu}$, $\hat{\sigma}$ and the fact that the distribution of the random vector $(Y_1, \ldots, Y_n)^T$ does not depend on the values of the parameters μ and σ. □

Denote by $u_\alpha^+(n,s)$ the α critical value of the statistic $U^+(n,s)$. Please note that it is exact, not asymptotic α critical value: $\mathbf{P}\{U^+(n,s) \geq u_\alpha^+(n,s)\} = \alpha$ under H_0.

Theorem 1 implies that the limit distribution (as $n \to \infty$) of the random variable $U^+(n,s)$ coincides with the distribution of the random variable $V^+(s) = \max_{1\leq i\leq s} V_i^+$, where $V_i^+ = 1 - F_{\chi^2_{2i}}(2(E_1 + \ldots + E_i))$, $E_1, \ldots, E_s$ are i.i.d. standard exponential random variables. The random variables $V_1^+, \ldots, V_s^+$ are dependent identically distributed and the distribution of each V_i^+ is uniform: $V_i^+ \sim U(0,1)$.

Denote by $v_\alpha^+(s)$ the α critical values of the random variable $V^+(s)$. They are easily found by simulation many times generating s i.i.d. standard exponential random variables and computing values of the random variables $V^+(s)$.

Our simulations showed that the below proposed outlier identification methods based on exact and approximate critical values of the statistic $U^+(n,s)$ give practically the same results, so for samples of size $n \geq 20$ we recommend to approximate the α-critical level of the statistic $U^+(n,s)$ by the critical values $v_\alpha^+(s)$ which depend only on s. We shall see that for the purpose of outlier identification only the critical values $v_\alpha^+(5)$ are needed. We found that the critical values $v_\alpha^+(5)$ are: $v_{0.1}^+(5) = 0.9677$, $v_{0.05}^+(5) = 0.9853$, $v_{0.01}^+(5) = 0.9975$.

Our simulations showed that the performance of the below proposed outlier identification method based on exact and approximate critical values of the statistic $U^+(n,5)$ is similar for samples of size $n \geq 20$.

We write shortly *BP*-method for the below considered method.

BP method for right outliers. Begin outlier search using observations corresponding to the largest values of $\hat{Y}_i$. We recommend begin with five largest. So take $s = 5$ and compute the values of the statistics

$$U^+(n,5) = \max_{1\leq i\leq 5} U_{(n-i+1)}^+(n).$$

If $U^+(n,5) \leq v_\alpha^+(5)$, then it is concluded that outliers are absent and no further investigation is done. Under H_0 the probability of such event is approximately $1 - \alpha$.

If $U^+(n,5) > v_\alpha^+(5)$, then it is concluded that outliers exist.

Please note that (see the classification scheme below) that if $U^+(n,5) > v_\alpha^+(5)$, then minimum one observation is declared as an outlier. So the probability to declare absence of outliers does not depend on the following classification scheme.

If it is concluded that outliers exist then search of outliers is done using the following steps.

Step 1. Set $d_1 = \max\{i \in \{1, \ldots, 5\} : U_{(n-i+1)}^+(n) > v_\alpha^+(5)\}$. Please note that the maximum $d_1 > 0$ exists because $U^+(n,5) > v_\alpha^+(5)$.

If $d_1 < 5$, then classification is finished at this step: d_1 observations are declared as right outliers because if the value of $X_{(n-d_1)}$ is declared as an outlier, then it is natural to declare values of $X_{(n)}, \ldots, X_{(n-d_1+1)}$ as outliers, too.

If $d_1 = 5$, then it is possible that the number of outliers is higher than 5. Then the observation corresponding to $i = 1$ (i.e., corresponding to $X_{(n)}$) is declared as an outlier and we proceed to the step 2.

Step 2. The above written procedure is repeated taking $U^+(n-1,5) = \max_{1\leq i\leq 5} U_{(n-i)}^+(n-1)$ instead of $U^+(n,5)$; here

$$U_{(n-i)}^+(n-1) = 1 - F_{\chi^2_{2i}}(2e^{-(\hat{Y}_{(n-i)} - b_{n-1})/a_{n-1}}), \quad i = 1, \ldots, 5,$$

Set $d_2 = \max\{i \in \{1, \ldots, 5\} : U_{(n-i)}^+(n-1) > v_\alpha^+(5)\}$. If $d_2 < 5$, the classification is finished and $d_2 + 1$ observations are declared as outliers.

If $d_2 = 5$, then it is possible that the number of outliers is higher than 6. Then the observation corresponding to the largest $\hat{Y}_{(n-1)}$ is declared as an outlier, in total 2 observations (i.e., the observations corresponding to $i = 1, 2$ (i.e., corresponding to $X_{(n)}$ and $X_{(n-1)}$) are declared as outliers and we proceed to the *Step 3*, and so on. Classification finishes at the lth step when $d_l < 5$. So we declare $(l-1)$ outliers in the previous steps and d_l outliers in the last one. The total number of observations declared as outliers is $l - 1 + d_l$. These observations are values of $X_{(n)}, \ldots, X_{(n-d_l-l+2)}$.

3.4. Left Outliers Identification Method for Location-Scale Families

Let a_n^*, b_n^* be the normalizing constants defined by (12). If $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, $i = 1, \ldots, s$, then set

$$U_{(i)}^-(n) = 1 - F_{\chi_{2i}^2}(2e^{(\hat{Y}_{(i)}+b_n^*)/a_n^*}), \quad U^-(n,s) = \max_{1\le i\le s} U_{(i)}^-(n).$$

If $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, then replace $e^{(\hat{Y}_{(i)}+b_n^*)/a_n^*}$ by $1/(1 + (\hat{Y}_{(i)} + b_n^*)/a_n^*)$. Denote by $u_\alpha^-(n,s)$ the α critical value of the statistic $U^-(n,s)$.

Theorem 1 and Remark 3 imply that the limit distribution (as $n \to \infty$) of the random variable $U^-(n,s)$ coincides with the distribution of the random variable $V^+(s)$. So the critical values $u_\alpha^-(n,s)$ are approximated by the critical values $v_\alpha^-(s) = v_\alpha^+(s)$.

The left outliers search method coincides with the right outliers search method replacing $+$ to $-$ in all formulas.

3.5. Outlier Detection Tests for Location-Scale Families: Two-Sided Alternative, Symmetric Distributions

Let a_n, b_n be defined by (5). If $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$, $i = 1, \ldots, s$, then set

$$U_{(n-i+1)}(n) = 1 - F_{\chi_{2i}^2}(2e^{-(|\hat{Y}|_{(n-i+1)}-b_{2n})/a_{2n}}), \quad U(n,s) = \max_{1\le i\le s} U_{(n-i+1)}(n).$$

If $F \in \mathcal{G}_\gamma \cap \mathcal{F}_{ls}$, $\gamma > 0$, then replace $e^{(\hat{Y}_{(i)}+b_n^*)/a_n^*}$ by $1/(1 + (\hat{Y}_{(i)} + b_n^*)/a_n^*)$. Denote by $u_\alpha(n,s)$ the α critical value of the statistic $U(n,s)$.

Theorem 1 and Remark 2 imply that the limit distribution (as $n \to \infty$) of the random variable $U(n,s)$ coincides with the distribution of the random variable $V^+(s)$. So the critical values $u_\alpha(n,s)$ are approximated by the critical values $v_\alpha(s) = v_\alpha^+(s)$.

The outliers search method coincides with the right outliers search method skipping upper index $+$ in all formulas.

3.6. Outlier Detection Tests for Location-Scale Families: Two-Sided Alternative, Non-Symmetric Distributions

Suppose now that the function f_0 is not symmetric. Let a_n, b_n, a_n^*, b_n^* be defined by (12).

Begin outlier search using observations corresponding to the largest and the smallest values of $\hat{Y}_i$. We recommend begin with five smallest and five largest. So compute the values of the statistics $U^-(n,5)$ and $U^+(n,5)$. If $U^-(n,5) \le v_{\alpha/2}(5)$ and $U^+(n,5) \le v_{\alpha/2}(5)$, then it is concluded that outliers are absent and no further investigation is done.

If $U^-(n,5) > v_{\alpha/2}(5)$ or $U^-(n,5) > v_{\alpha/2}(5)$, then it is concluded that outliers exist. If $U^-(n,5) > v_{\alpha/2}(5)$, then left outliers are searched as in Section 3.3. If $U^+(n,5) > v_{\alpha/2}(5)$, then right outliers are searched as in Section 3.2. The only difference is that α is replaced by $\alpha/2$ in all formulas.

3.7. Outlier Identification Method for Shape-Scale Families

If shape-scale families of the form $\{F(t;\theta,\nu) = G_0((t/\theta)^\nu),\ \theta,\nu > 0\}$ with specified G_0 are considered then the above given tests for location-scale families could be used because if $X_1,\ldots,X_n$ is a sample from shape scale family then $Z_1,\ldots,Z_n$, $Z_i = \ln X_i$, is a sample from location-scale family $\{F_0((x-\mu)/\sigma, \mu \in \mathbf{R}, \sigma > 0\})$ with $\mu = \ln\theta$, $\sigma = 1/\nu$, $F_0(x) = G_0(e^x)$.

3.8. Illustrative Example

To illustrate simplicity of the *BP*-method, let us consider an illustrative example of its application (sample size $n = 20$, $r = 7$ outliers). The sample of size $n = 20$ from standard normal distribution was generated. The 1st-3rd and 17th-20th observations were replaced by outliers. The observations x_i, the absolute values $|\hat{Y}_i|$ of the z-scores $\hat{Y}_i$, and the ranks (i) of $|\hat{Y}_i|$ are presented in Table 3.

Table 3. Illustrative sample ($n = 20, r = 7$).

i	x_i	$\lvert\hat{Y}_i\rvert$	(i)	i	x_i	$\lvert\hat{Y}_i\rvert$	(i)
1	6.10	3.18	16	11	−0.69	0.28	9
2	10	5.17	18	12	−0	0.07	5
3	6.20	3.23	17	13	0.05	0.10	6
4	−0.08	0.03	2	14	−0.20	0.03	1
5	0.63	0.39	11	15	−0.25	0.06	4
6	−0.54	0.21	7	16	−0.64	0.25	8
7	1.37	0.77	13	17	−6.30	3.14	15
8	0.46	0.30	10	18	−5.50	2.73	14
9	−0.22	0.04	3	19	−12.10	6.10	19
10	0.94	0.55	12	20	−20	10.13	20

In Table 4 we present steps of the classification procedure by the *BP* method. First, we compute (see line 1 of Table 4) value of the statistic $U(20,5) = \max_{1\le i\le 5} U_{(20-i+1)}(20) = 1$. Since $U(20,5) = 1 > 0.9853 = v_{0.05}(5)$, we reject the null hypothesis, conclude that outliers exist and begin the search of outliers.

Step 1. The inequality $U_{(16)}(20) = 1.0000 > 0.9853 = v_{0.05}(5)$ (note that $U_{(16)}(20)$ corresponds to the fifth largest observation in absolute value) implies that $d_1 = 5$. So it is possible that the number of outliers might be greater than 5. We reject the largest in absolute value 20th observation as an outlier and continue the search of outliers.

Step 2. The inequality $U_{(15)}(19) = 1.0000 > 0.9853 = v_{0.05}(5)$ (note that $U_{(15)}(19)$ corresponds to the fifth largest observation in absolute value from the remaining 19 observations) implies that $d_2 = 5$. So it is possible that the number of outliers might be greater than 6. We declare the second largest in absolute value observation as an outlier. So two observations (19th and 20th) are declared as outliers. We continue the search of outliers.

Step 3. The inequality $U_{(14)}(18) = 0.999997 > 0.9853 = v_{0.05}(5)$ implies that $d_3 = 5$. We declare the third largest in absolute value observation as an outlier. So three observations (2nd, 19th and 20th) are declared as outliers. We continue the search of outliers.

Step 4. The inequalities $U_{(13)}(17) = 0.084290 < 0.9853 = v_{0.05}(5)$ and $U_{(14)}(17) = 0.999940 > 0.9853 = v_{0.05}(5)$ imply that $d_4 = 4$. So four additional observations (the fourth, fifth, sixth and seventh largest in absolute value observations), namely the 3d, 1st, 17th, and 7th are declared as outliers, The outlier search is finished. In all, 7 observations were declared as outliers: 1–3,17–20, as was expected. Please note that since the outlier search procedure was done after rejection of the null hypothesis, the significance level did not change.

Table 4. Illustrative example of *BP* test observations classification.

$U_{(20)}(20)$	$U_{(19)}(20)$	$U_{(18)}(20)$	$U_{(17)}(20)$	$U_{(16)}(20)$	$U(20,5)$
1.000000	1.000000	1.000000	0.999998	1.000000	1.000000
$U_{(19)}(19)$	$U_{(18)}(19)$	$U_{(17)}(19)$	$U_{(16)}(19)$	$U_{(15)}(19)$	$U(19,5)$
0.999685	0.999998	0.999916	0.999998	1.000000	1.000000
$U_{(18)}(18)$	$U_{(17)}(18)$	$U_{(16)}(18)$	$U_{(15)}(18)$	$U_{(14)}(18)$	$U(18,5)$
0.998046	0.996970	0.999893	0.999997	0.999997	0.999997
$U_{(17)}(17)$	$U_{(16)}(17)$	$U_{(15)}(17)$	$U_{(14)}(17)$	$U_{(13)}(17)$	$U(17,5)$
0.924219	0.996446	0.999871	0.999940	0.084290	0.999940

3.9. Practical Example

Let's consider the stent fatigue testing dataset from reliability control [27]. The dataset contains 100 observations. Let us consider the Weibull, lollogistic and lognormal models. These are the most applied models for analysis of reliability data. For preliminary choice of suitable model we compare the values of various goodness-of-fit statistics and information criteria (see Table 5). The Weibull model is obviously the most suited because values of all five statistics are smallest for this model.

Table 5. Values of goodness-of-fit statistics and information criteria (initial sample).

Goodness-of-Fit Statistics	**Weibull**	**Logistic**	**Log-Normal**
Kolmogorov-Smirnov statistic	0.05	0.09	0.07
Cramer-von Mises statistic	0.03	0.23	0.127
Anderson-Darling statistic	0.21	1.36	1.08
Goodness-of-fit criteria			
Akaike's Information Criterion	1056.515	1074.783	1073.13
Bayesian Information Criterion	1061.725	1079.993	1078.34

Using the function WEDF.test from the R package EWGoF we applied the following goodness-of-fit tests for Weibull distribution : Anderson-Darling (p-value = 0.86), Kolmogorov-Smirnov (p-value = 0.82), Cramer-von-Mises (p-value = 0.795), Watson (p-value = 0.795). So all tests do not contradict to the Weibull model.

The logarithms $X_1, \ldots, X_{100}$ of observations have type I extreme value distribution. Minimal and maximal values are $X_{(1)} = 1.609$ and $X_{(100} = 5.670$. Let us consider the situation, where fatigue data contain two outliers $X_3 = 6.5$ and $X_5 = 6.5$. All goodness-of-fit tests applied to the data with outliers reject the Weibull model: Anderson-Darling (p-value $< 10^{-15}$), Kolmogorov-Smirnov (p-value 0.005), Cramer-von-Mises (p-value $< 10^{-15}$), Watson (p-value $< 10^{-15}$).

Let us apply the *BP* method for outlier identification. Values of the statistics U_i are: $U_{(100)}(100) = 0.92, U_{(99)}(100) = 0.997, U_{(98)}(100) = 0.96, U_{(97)}(100) = 0.92, U_{(96)}(100) = 0.96$. Since $U(100,5) = 0.997 > 0.9853$, we reject the null hypothesis.

Step 1. Since $d_1 = \max_{\{i\in\{1,\ldots,5\}\}} : U_{(101-i)} > 0.9853\} = 2 < 5$. the search procedure is finished and the observations $X_{(99)}$ and $X_{(100)}$, namely X_3 and X_5, are declared as outliers. We see that our method did not allow masking other equal observations $X_3 = X_5 = 6.5$. It is a very important advantage of the *BP* method.

After outliers removal, we repeated goodness-of-fit procedure. All tests did not reject the Weibull model: Anderson-Darling (p-value = 0.88), Kolmogorov-Smirnov (p-value = 0.8), Cramer-von-Mises

(p-value = 0.93), Watson (p-value = 0.895). Once more, we compared values of goodness-of-fit statistics and information criteria for above considered models using data without removed outliers, see Table 6.

Table 6. Values of goodness-of-fit statistics and information criteria (sample without removed outliers).

Goodness-of-Fit Statistics	**Weibull**	**Logistic**	**Log-Normal**
Kolmogorov-Smirnov statistic	0.048	0.09	0.07
Cramer-von Mises statistic	0.027	0.21	0.11
Anderson-Darling statistic	0.18	1.25	1.01
Goodness-of-fit criteria			
Akaike's Information Criterion	1037.09	1054.76	1053.49
Bayesian Information Criterion	1042.26	1059.93	1058.66

The Weibull distribution gives clearly the best fit.

Values of ML estimators from the initial non-contaminated data and from the final cleared from outliers data are similar: shape practically did not change: 1.83 $\rightarrow$ 1.83, scale changed slightly: 100.8 $\rightarrow$ 101.4.

We created R package outliersTests (https://github.com/linas-p/outliersTests) to be able to use the proposed BP test in practice within R package.

4. Generalization of Davies-Gather Outlier Identification Method

Let us consider location-scale families. Following the idea of Davies-Gather [2] define an empirical analogue of the right outlier region as a random region

$$OR_r(\alpha_n) = \{x : x > \hat{\mu} + \hat{\sigma} g_{n.\alpha}\}, \tag{17}$$

where $g_{n.\alpha}$ is found using the condition

$$\mathbf{P}\{X_i \bar{\in} OR_r(\alpha_n), i = 1, \ldots, n | \mathrm{H}_0\} = 1 - \alpha, \tag{18}$$

and $\hat{\mu}, \hat{\sigma}$ are robust equivariant estimators of the parameters μ, σ.

Set

$$\hat{Y}_{(n)} = (X_{(n)} - \hat{\mu})/\hat{\sigma}.$$

The distribution of $\hat{Y}_{(n)}$ is parameter-free under H_0.

The Equation (18) is equivalent to the equation equation

$$\mathbf{P}\{\hat{Y}_{(n)} \leq g_{n,\alpha}\} | \mathrm{H}_0\} = 1 - \alpha.$$

So $g_{n,\alpha}$ is the upper α critical value of the random variable $\hat{Y}_{(n)}$. It is easily computed by simulation.

Generalized Davies-Gather method for right outliers identification: if $\hat{Y}_{(n)} \leq g_{n,\alpha}$, then it is concluded that right outliers are absent. The probability of such event is α. If $\hat{Y}_{(n)} > g_{n,\alpha}$, then it is concluded that right outliers exist. The value x_i of the random variable X_i is admitted as an outlier if $x_i \in OR_r(\alpha_n)$, i.e., if $x_i > \hat{\mu} + \hat{\sigma} g_{n,\alpha}$. Otherwise it is admitted as a non-outlier.

An empirical analogue of the left outlier region as a random region

$$OR_l(\alpha_n) = \{x : x < \hat{\mu} + \hat{\sigma} h_{n.1-\alpha}\}, \tag{19}$$

where $h_{n.1-\alpha}$ is found using the condition

$$\mathbf{P}\{X_i \bar{\in} OR_l(\alpha_n), i = 1, \ldots, n|\mathrm{H}_0\} = 1 - \alpha, \tag{20}$$

Set

$$\hat{Y}_{(1)} = (X_{(1)} - \hat{\mu})/\hat{\sigma}.$$

The distribution of $\hat{Y}_{(1)}$ is parameter-free under H_0.

The Equation (20) is equivalent to the equation equation

$$\mathbf{P}\{\hat{Y}_{(1)} \geq h_{n,1-\alpha}|\mathrm{H}_0\} = 1 - \alpha.$$

So $h_{n,\alpha}$ is the upper $1 - \alpha$ critical value of the random variable $\hat{Y}_{(1)}$. It is easily computed by simulation.

Generalized Davies-Gather method for left outliers identification: if $\hat{Y}_{(1)} \geq h_{n,1-\alpha}$, then it is concluded that left outliers are absent. The probability of such event is α. If $\hat{Y}_{(1)} < h_{n,\alpha}$, then it is concluded that left outliers exist. The value x_i of the random variable X_i is admitted as an outlier if $x_i \in OR_l(\alpha_n)$, i.e., if $x_i < \hat{\mu} + \hat{\sigma} h_{n,1-\alpha}$. Otherwise it is admitted as a non-outlier.

Let us consider two-sided case.

If the distribution of X_i is symmetric, then the empirical analogue of the outlier region is the random region

$$OR(\alpha_n) = \{x : |x - \hat{\mu}| > \hat{\sigma} g_{n.\alpha/2}\}. \tag{21}$$

In this case

$$1 - \alpha = \mathbf{P}\{X_i \in OR(\alpha_n), i = 1, \ldots, n|\mathrm{H}_0\} = \mathbf{P}\{|\hat{Y}|_{(n)} \leq g_{n.\alpha/2}\}.$$

Generalized Davies-Gather method for left and right outliers identification (symmetric distributions): if $|\hat{Y}|_{(n)} \leq g_{n.\alpha/2}$, then it is concluded that outliers are absent. The probability of such event is α. If $|\hat{Y}|_{(n)} > g_{n.\alpha/2}$, then it is concluded that outliers exist. The value x_i of the random variable X_i is admitted as a left outlier if $x_i < \hat{\mu} - \hat{\sigma} g_{n,\alpha/2}$, it is admitted as a right outlier if $x_i > \hat{\mu} + \hat{\sigma} g_{n,\alpha/2}$. Otherwise it is admitted as a non-outlier.

If distribution of X_i is non-symmetric, then the empirical analogue of the outlier region is defined as follows:

$$OR(\alpha_n) = \{x \in \mathbf{R}/[\hat{\mu} + \hat{\sigma} g_{n,1-\alpha/2}, \hat{\mu} + \hat{\sigma} g_{n,\alpha/2})]\},$$

In this case

$$1 - \alpha = \mathbf{P}\{X_i \bar{\in} [\hat{\mu} + \hat{\sigma} h_{n,1-\alpha/2},\ \hat{\mu} + \hat{\sigma} g_{n,\alpha/2}], i = 1, \ldots, n|\mathrm{H}_0\} =$$

$$\mathbf{P}\{h_{n,1-\alpha/2} \leq \hat{Y}_{(1)} \leq \hat{Y}_{(n)} \leq g_{n,\alpha/2})|\mathrm{H}_0\}.$$

Generalized Davies-Gather method for left and right outliers identification (non-symmetric distributions): if $\hat{Y}_{(1)} \geq h_{n,1-\alpha/2}$ and $\hat{Y}_{(n)} \leq g_{n,\alpha/2}$, then it is concluded that outliers are absent. The probability of such event is α. If $\hat{Y}_{(1)} < h_{n,1-\alpha/2}$ or $\hat{Y}_{(n)} > g_{n,\alpha/2}$, then it is concluded that outliers exist. The value x_i of the random variable X_i is admitted as a left outlier if $x_i < \hat{\mu} + \hat{\sigma} h_{n,1-\alpha/2}$, it is admitted as a right outlier if $x_i > \hat{\mu} + \hat{\sigma} g_{n,\alpha/2}$. Otherwise it is admitted as a non-outlier.

5. Short Survey of Multiple Outlier Identification Methods for Normal Data

5.1. Rosner's Method

Let us formulate Rosner's method in the form mostly used in practice. Suppose that the number of outliers does not exceed s and the two-sided alternative is considered. Set (see [5,28])

$$R_1 = \max_{1\leqslant j\leqslant n} |\tilde{Y}_j| = \max_{1\leqslant j\leqslant n} |X_j - \bar{X}|/S_X, \quad S_X^2 = \sum_{j=1}^{n} (X_{(j)} - \bar{X})^2/(n-1).$$

$|\tilde{Y}_j| = |(X_j - \bar{X})/S_X|$ may be interpreted as a distance between X_j and $\bar{X}$. Remove the observation X_{j_1} which is most distant from $\bar{X}$. This maximal distance is R_1. The value of X_{j_1} is a possible candidate for contaminant.

Recompute the statistic using $n-1$ remaining observations and denote by R_2 the obtained statistic. Remove the observation X_{j_2} which is most distant from the new empirical mean. The value of X_{j_2} is also possible candidate for contaminant. Repeat the procedure until the statistics $R_1, \cdots, R_s$ are computed. So we obtain all possible candidates for contaminants. They are values of $X_{j_1}, \ldots, X_{j_s}$

Fix α and find λ_{in} such that

$$\mathbf{P}\{R_1 > \lambda_{in}|\mathrm{H}_0\} = \ldots = \mathbf{P}\{R_s > \lambda_{in}|\mathrm{H}_0\}, \quad \mathbf{P}\{\cup_{i=1}^{s}\{R_i > \lambda_{in}\}|\mathrm{H}_0\} = \alpha.$$

If $n > 25$, then the approximations

$$\lambda_{in} \approx t_{\frac{\alpha}{2(n-i-1)}}(n-i+1)\sqrt{\frac{n-i}{n-i-1+t^2_{\frac{\alpha}{2(n-i-1)}}(n-i+1)}}\sqrt{1-\frac{1}{n-i+1}},$$

are recommended (see [5]); here $t_p(\nu)$ is the p critical value of the Student distribution with ν degrees of freedom.

Rosner's method for left and right outliers identification: if $R_i \leqslant \lambda_{in}$ for all $i = 1, \cdots, s$, then it is concluded that outliers are absent. If there exists $i_0 \in \{1, \ldots, s\}$ such that $R_{i_0} > \lambda_{i_0 n}$, i.e., the event $\cup_{i=1}^{s}\{R_i > \lambda_{in}\}$ occurs, then it is concluded that outliers exist. In this case, classification of observations to outliers and non-outliers is done in the following way: if $R_s > \lambda_{sn}$, then it is concluded that there are s outliers and they are values of $X_{j_1}, \ldots, X_{j_s}$. If $R_j \leqslant \lambda_{jn}$ for $j = s, s-1, \ldots, i+1$, and $R_i > \lambda_{in}$, then it is concluded that there are i outliers and they are values of $X_{j_1}, \ldots, X_{j_i}$.

If right outliers are searched, then define $R_1^+ = \max_{1\leqslant i\leqslant n} \tilde{Y}_i$, and repeat the above procedure taking approximations

$$\lambda_{in}^+ \approx t_{\frac{\alpha}{n-i-1}}(n-i+1)\sqrt{\frac{n-i}{n-i-1+t^2_{\frac{\alpha}{n-i-1}}(n-i+1)}}\sqrt{1-\frac{1}{n-i+1}}.$$

Denote by R_s the Rosner's test with a fixed upper limit s. Our simulation results confirm that the true significance level is different from the level α suggested by the approximation when n is not large. Nevertheless, it is approaching α as n increases, see Figure 1. The true significance value of the BP test, which uses asymptotic values of the test statistic are also presented in Figure 1.

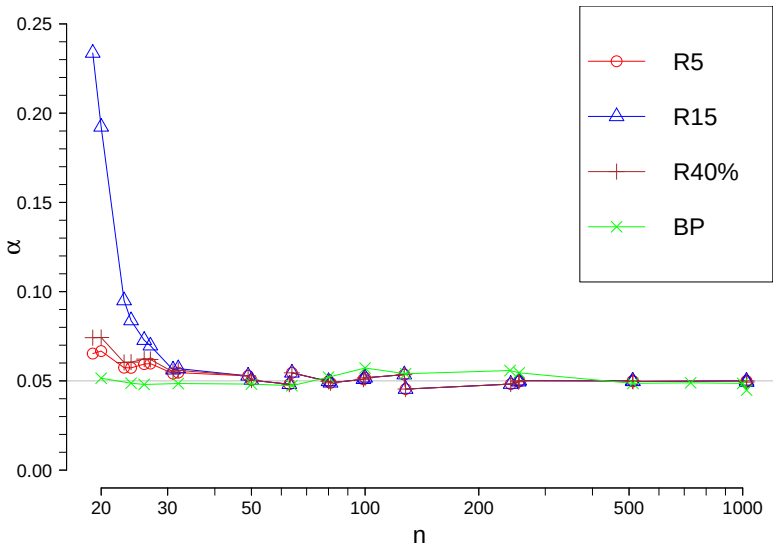


Figure 1. The true values of the significance level of Rosner's and *BP* tests in function of n for different values of s ($\alpha = 0.05$ is used in approximations).

5.2. Bolshev's Method

Suppose that the number of contaminants does not exceed s. For $i = 1, \cdots, n$ set

$$\hat{Y}_i = (X_i - \bar{X})/s, \quad \tau_i^+ = n \cdot (1 - T_{n-2}(\hat{Y}_i)), \quad \tau_i = n \cdot (1 - T_{n-2}(|\hat{Y}_i|)),$$

where $\bar{X}$ and s are the empirical mean and standard deviation, $T_{n-2}(x)$ is the c.d.f. of Thompson's distribution with $n-2$ degrees of freedom.

Let us consider search for right outliers. Please note that the largest s observations $X_{(n-s+1)}, \ldots, X_{(n)}$ define the smallest s order statistics $\tau^+_{(1)} \leq \ldots \leq \tau^+_{(n)}$. Possible candidates for outliers are namely the values of $X_{(n-s+1)}, \ldots, X_{(n)}$.

Set $\tau^+ = \min_{1 \leqslant i \leqslant s} \tau^+_{(i)}/i$.

Bolshev's method for right outliers search. If $\tau^+ \geq \tau^+_{1-\alpha}(n,s)$, then it is concluded that outliers are absent; here $\tau^+_{1-\alpha}(n,s)$ is the $1-\alpha$ critical value of the test statistic under H_0. If $\tau^+ < \tau^+_{1-\alpha}(n,s)$, then it is concluded that outliers exist. In such a case outliers are selected in the following way: if $\tau^+_i/i < \tau^+_{1-\alpha}(n,s)$ then the value of the order statistic $X_{(n-i+1)}$ is admitted as an outlier, $i = 1, \ldots, s$.

In the case of left and right outliers search Bolshev's method uses $\tau_{(i)}$ instead of $\tau^+_{(i)}$, defining the statistic $\tau = \min_{1 \leqslant i \leqslant s} \tau_{(i)}/i$.

Bolshev's method for left and right outliers search. If $\tau \geq \tau_{1-\alpha}(n,s)$, then it is concluded that outliers are absent; here $\tau_{1-\alpha}(n,s)$ is the $1-\alpha$ critical value of the statistic τ under H_0. If $\tau < \tau_{1-\alpha}(n,s)$, then it is concluded that outliers exist. In such a case they are selected in the following way: if $\tau_i/i < \tau_{1-\alpha}(n,s)$ then the observation corresponding to τ_i is admitted as an outlier, $i = 1, \ldots, s$.

5.3. Hawking's Method

Suppose that the number of contaminants does not exceed s. Let us consider the search for right outliers. For $k = 1, \ldots, s$ set

$$b_k^+ = \frac{1}{\sqrt{k(n-k)}} \sum_{i=1}^{k} \tilde{Y}_{(n-i+1)} = \frac{1}{\sqrt{k(n-k)}} \sum_{i=1}^{k} (X_{(n-i+1)} - \bar{X})/S_X.$$

b_k^+ proportional to the sum of k largest $\tilde{Y}_{(n-i+1)}$. Set $B^+ = \max_{1\leqslant k\leqslant s} b_k^+$.

Hawking's method. If $B^+ \leq B_\alpha^+(n,s)$ then it is concluded that outliers are absent; here $B_\alpha^+(n,s)$ is the α critical value of the statistic under H_0. If $B^+ > B_\alpha^+(n,s)$, then it is concluded that outliers exist. In such a case outliers are selected in the following way: if $b_i^+ > B_\alpha^+(n,s)$, then the value of the order statistic $X_{(n-i+1)}$ is admitted as an outlier, $i = 1,\ldots,s$.

6. Comparative Analysis of Outlier Identification Methods by Simulation

In the case of location-scale classes probability distribution of all considered test statistics does not depend on μ and σ, so we generated samples of various sizes n with $n-r$ observations with the c.d.f. F_0 and r observations with various alternative distributions concentrated in the outlier region. We shall call such observations "contaminant outliers", shortly *c*-outliers. As was mentioned, outliers which are not c-outliers, i.e., outliers from regular observations with the c.d.f. F_0, are very rare.

We repeated simulations $M = 100{,}000$ times and using various methods we classified observations to outliers and non-outliers and computed the mean number D_{O_cO} of correctly identified c-outliers, the mean number D_{ON} of c-outliers which were not identified, the mean number D_{NO} of non c-outliers admitted as outliers, and the mean number D_{NN} of non c-outliers admitted as non-outliers.

An outlier identification method is ideal if each outlier is detected and each non-outlier is declared as a non-outlier. In practice it is impossible to do with the probability one. Two errors are possible: (a) an outlier is not declared as such (masking effect); (b) a non-outlier is declared as an outlier (swamping effect). We shall write shortly "masking value" for the mean number of non-detected c-outliers and "swamping value" for the mean number of "normal" observations declared as outliers in the simulated samples.

If swamping is small for two tests then a test with smaller masking effect should be preferred because in this case the distribution of the data remaining after excluding of suspected outliers should be closer to the distribution of non-outlier data.

From the other side, if swamping for Method 1 is considerably bigger than swamping of Method 2 and masking is smaller for Method 1, then it does not mean that Method 1 is better because this method rejects many extreme non-outliers from the tails of the regular distribution F_0 and the sample remaining after classification may be not treated as a sample from this regular distribution even if all c-outliers are eliminated.

For various families of distributions, sample sizes n, and alternatives we compared Davies-Gather (DG) and new (BP) methods performance. In the case of normal distribution we also compared them with Rosner's, Bolshev's and Hawking's methods.

We used two different classes of alternatives: in the first case c-outliers are spread widely in the outlier region around the mean, in the second case c-outliers are concentrated in a very short interval laying in the outlier region. More precisely, if right outliers were searched, then we simulated r observations concentrated in in the right outlier region $out_r(\alpha_n, F_0) = \{x : x > x_{\alpha_n}\}$ using the following alternative families of distribution:

(1) Two parameter exponential distribution $\mathcal{E}(\theta, x_{\alpha_n})$ with the scale parameter θ. If θ is small, then outliers are concentrated near the border of the outlier region. If θ is large then outliers are widely spread in the outlier region. If θ increases, then the mean of outlier distribution increases. Please note that even if θ is very near 0 and the true number of outliers r is large, these outliers may corrupt strongly the data making tails of histogram two heavy.

(2) Truncated normal distribution $\mathcal{TN}(x_{\alpha_n}, \mu, \rho)$ with the location and scale parameters μ, ρ ($\mu > x_{\alpha_n}$). If ρ is small then this distribution is concentrated in a small interval around μ. If μ increases, then the mean of outlier distribution increases.

For lack of place we present a small part of our investigations. Please note that the results are very similar for all sample sizes $n \geq 20$. Multiple outlier problem is not very relevant for smaller sample sizes.

6.1. Investigation of Outlier Identification Methods for Normal Data

We use notation B, H, R, DG, and BP for the Bolshev's, Hawking's, Rosner's, Davies-Gather's, and the new methods, respectively. If DG method is based on maximum likelihood estimators, then we write DG_{ml} method, if it is based on robust estimators, we write DH_{rob} method.

For comparison of above considered methods we fixed the significance level $\alpha = 0.05$. We remind that the significance level α is the probability to reject minimum one observation as an outlier under the hypothesis H_0 which means that all observations are realizations of i.i.d. with the same normal distribution. The only test, namely R method uses approximate critical values of the test statistic, so the significance values for this test is only approximately 0.05 and depends on s and n. In Figure 1 the true significance level value for $s = 5, 15$ and $[0.4n]$ in function of n are given.

The B, H, and R tests methods have a drawback that the upper bound for the possible number of outliers s must be fixed. The BP and DG tests have an advantage that they do not require it.

Our investigations showed that H,B and DG_{ml} methods have other serious drawbacks. So firstly let us look closer at these methods.

If the true number of c-outliers r exceeds s, then the B and H methods cannot find them even if they are very far from the limits of the outlier region. Nevertheless, suppose that r does not exceed s and look at the performance of the H method. Set $n = 100$, $s = 5$, and suppose that c-outliers are generated by right-truncated normal distribution $\mathcal{TN}(x_{\alpha_n}, \mu, \rho)$ with fixed ρ and increasing μ. Note that the true number of c-outliers is supposed to be unknown but do not exceed $s = 5$. In Figure 2 the mean numbers of rejected non-c-outliers D_{NO} are given in function of the parameter μ (the value of the parameter $\rho = 0.1^2$ is fixed) for fixed values of r see Figure 2. In Table 7 the values of D_{NO} plus the values of the mean numbers of truly rejected c-outliers are given. Table 7 shows that if $r = 1$, then if μ is sufficiently large, the c-outlier is found but the number of rejected non-c-outliers D_{NO} increases to 4, so swamping is very large. Similarly, if $r = 2$, then D_{NO} increases to 3, so swamping is large. Beginning from $r = 3$ not all c-outliers are found even for large μ. Swamping is smallest if the true value r coincides with s but even in this case one c-outlier is not found even for large μ. Taking into account that the true number r of c-outliers is not known in real data, the performance of the H methos is very poor. Results are similar for other values of n, s, and distributions of c-outliers. As a rule, H mehod finds rather well the c-outliers but swamping is very large because this method has a tendency to reject a number near s of observations for remote alternatives. which is good if $r = s$ but is bad if r is different from s.

Table 7. Hawkin's method: the values of $D_{NO} + D_{OO}$ in function of μ and r ($n = 100$, $s = 5$).

r \ μ	**0.1**	**1**	**6.3**	**10**
1	0.31 + 0.00	0.66 + 0.00	3.93 + 1.00	3.99 + 1.00
2	0.87 + 0.00	2.15 + 0.06	3.00 + 1.21	3.00 + 2.00
3	1.33 + 0.08	1.99 + 0.84	2.00 + 2.00	2.00 + 2.00
4	0.89 + 0.58	1.00 + 1.42	1.00 + 3.00	1.00 + 3.00
5	0.01 + 1.15	0.00 + 2.03	0.00 + 3.02	0.00 + 3.96

The B and DG_{ml} tests have a drawback that they use maximum likelihood estimators which are not robust and estimate parameters badly in presence of outliers. Once more, set $n = 100$, $s = 5$, and suppose that c-outliers are generated by two-parameters exponential distribution $\mathcal{TE}(x_{\alpha_n}, \theta)$ with increasing θ. Swamping values are negligible in, so only masking values(mean numbers of non-rejected c-outliers D_{ON}) are important. In Figure 3 the masking values in function of the parameter θ are given for fixed values of r.

Both methods perform very similarly. The masking values are large for every value of $r > 1$. If r increases, then masking values increase, too. For example, if $r = 5$, then almost 3 c-outliers from 5 are not rejected on average even for large values of θ.

Similar results hold taking other values of n, s and various distributions of c-outliers.

The above analysis shows that the B, H, DG_{ml} methods have serious drawbacks, so we exclude these methods from further consideration.

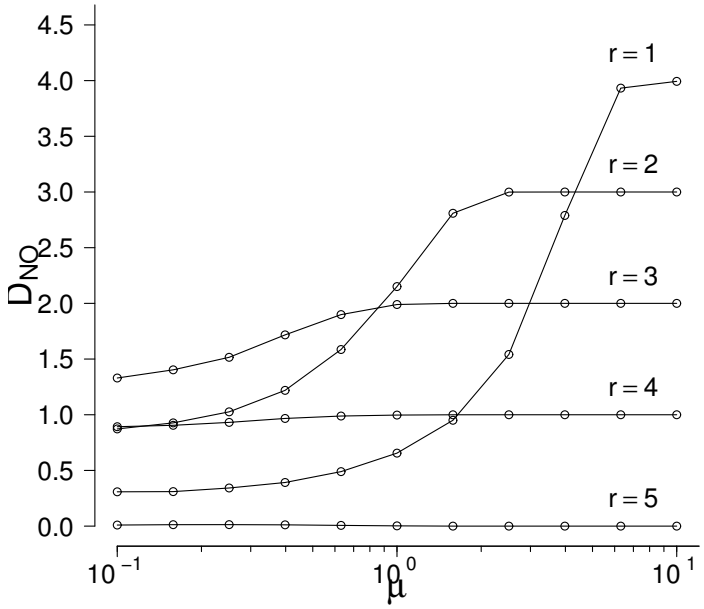


Figure 2. Hawkin's method: the values of $D_{NO} + D_{OO}$ in function of μ and r ($n = 100$, $s = 5$).

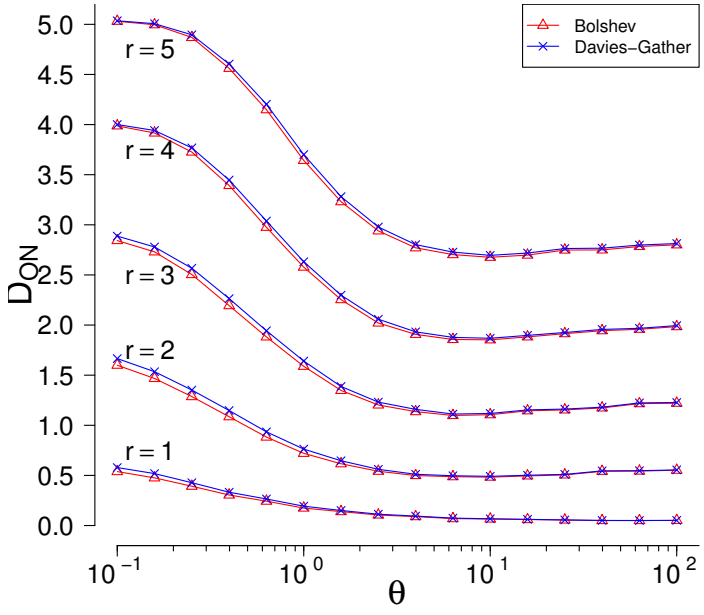


Figure 3. The number of outliers rejected as non-outliers (D_{ON}). The alternative: two-sided, the outliers generated by two-parameters exponential distribution on both sides.

Let us consider the remaining three methods: R, DG, and BP. For small n the true significance level of Rosner's test differ considerably from the suggested, so we present comparisons of tests performance for $n = 50, 100, 1000$ (see Tables 8 and 9). Truncated exponential distribution was used for outliers simulation. Remoteness of the mean of outliers from the border of the outlier region is characterized by the parameter θ.

Table 8. The masking values D_{ON} ($n = 50$ and $n = 100$).

	$n = 50$						$n = 100$				
r	**Method**$\backslash\theta$	**0.1**	**0.4**	**1**	**4**	**10**	r	**0.1**	**0.4**	**1**	**4**	**10**
2	$Rosner_5$	1.36	0.95	0.51	0.15	0.06	2	1.19	0.71	0.33	0.09	0.04
	$Rosner_{15}$	1.36	0.95	0.51	0.15	0.06		1.19	0.71	0.33	0.09	0.04
	$Rosner_{[0.4n]}$	1.36	0.95	0.51	0.15	0.06		1.19	0.71	0.33	0.09	0.04
	DGr_{rob}	1.56	1.17	0.71	0.24	0.10		1.31	0.84	0.44	0.13	0.06
	BP	0.92	0.66	0.37	0.10	0.04		0.50	0.32	0.15	0.04	0.02
5	$Rosner_5$	3.79	3.31	2.11	0.48	0.16	5	3.52	2.57	1.27	0.27	0.10
	$Rosner_{15}$	3.66	3.21	2.04	0.46	0.16		3.43	2.52	1.24	0.26	0.10
	$Rosner_{[0.4n]}$	3.66	3.21	2.04	0.46	0.16		3.43	2.52	1.24	0.26	0.10
	DG_{rob}	4.70	4.10	2.90	1.09	0.48		4.23	3.01	1.81	0.57	0.25
	BP	2.00	1.68	1.18	0.40	0.15		0.78	0.60	0.43	0.15	0.07
8	$Rosner_5$	8.00	7.97	7.54	3.70	3.06	10	10.0	9.90	8.21	5.10	5.00
	$Rosner_{15}$	5.70	5.48	4.52	1.00	0.29		6.88	6.54	4.36	0.69	0.22
	$Rosner_{[0.4n]}$	5.70	5.48	4.52	1.00	0.29		6.88	6.54	4.36	0.69	0.22
	DG_{rob}	7.90	7.49	6.10	2.67	1.24		9.74	8.38	5.78	2.12	0.92
	BP	4.27	3.84	3.25	1.47	0.57		2.21	1.90	1.73	0.74	0.30

Table 9. The masking values D_{ON} ($n = 1000$).

r	**Method**$\backslash\theta$	**0.1**	**0.4**	**1**	**4**	**1000**
5	$Rosner_5$	2.15	0.69	0.29	0.07	0.00
	$Rosner_{15}$	2.12	0.66	0.27	0.07	0.00
	$Rosner_{[0.4n]}$	2.12	0.66	0.27	0.07	0.00
	DG_{rob}	1.99	0.78	0.35	0.09	0.00
	BP	0.25	0.23	0.22	0.11	0.00
20	$Rosner_5$	19.0	15.8	15.0	15.0	15.0
	$Rosner_{15}$	19.2	10.9	5.52	5.00	5.00
	$Rosner_{[0.4n]}$	12.7	6.94	1.76	0.30	0.00
	DG_{rob}	14.8	6.97	3.32	1.93	0.00
	BP	0.29	0.26	0.23	0.18	0.00
100	$Rosner_5$	100	99.9	96.7	95.0	95.0
	$Rosner_{15}$	100	99.92	96.4	85.0	85.0
	$Rosner_{[0.4n]}$	55.8	56.8	50.4	4.43	0.01
	DG_{rob}	100	89.9	61.6	22.2	0.1
	BP	4.72	4.00	3.95	3.58	0.04

Swamping values D_{NO} (the mean numbers of non-c-outliers declared as outliers) are very small for all tests. For example, even if $n = 1000$, the R and DG methods reject on average as outliers only 0.05 from $n - r = 995, 980, 900$ non-c-outliers. For the BP method this number is $0.25, 0.19, 0.05$ from $995, 980$, and 900 non-c-outliers, respectively. So only masking values D_{ON} (the mean numbers of c-outliers declared as non-outliers) are important for outlier identification methods comparison.

Necessity to guess the upper limit s for a possible number of outliers is considered as a drawback of the Rosner's method. Indeed, if the true number of outliers r is greater than the chosen upper limit s, then $r - s$ outliers are not identified with the probability one. In addition, even if $r \leq s$, it is not clear how important is closeness of r to s. So first we investigated the problem of the upper limit choice.

Here we present masking values D_{ON} of the Rosner's tests for $s = 5, 15$ and $[0.4n]$. Similar results are obtained for other values of s.

Our investigations show that it is sufficient to fix $s = [0.4n]$, which is clearly larger than it can be expected in real data. Indeed, Tables 8 and 9 show that for $r > s$ $Rosner_5$ and $Rosner_{15}$ do not find $r - s$ outliers even if they are very remote, as it should be. Nevertheless, we see that even if the true number of outliers r is much smaller than $[0.4n]$, for any considered n, $r \leq s = 5, 15$ the masking values of the $Rosner_{[0.4n]}$ test are approximately the same (even a little smaller) as the masking values of the tests $Rosner_5$ and $Rosner_{15}$, for $r > s$ they are clearly smaller.

Hence, $s = [0.4n]$ should be recommended for Rosner's test application, and performance of $Rosner_{[0.4]}$, Davies-Gather robust (DG_{rob}) and the proposed BP methods should be compared.

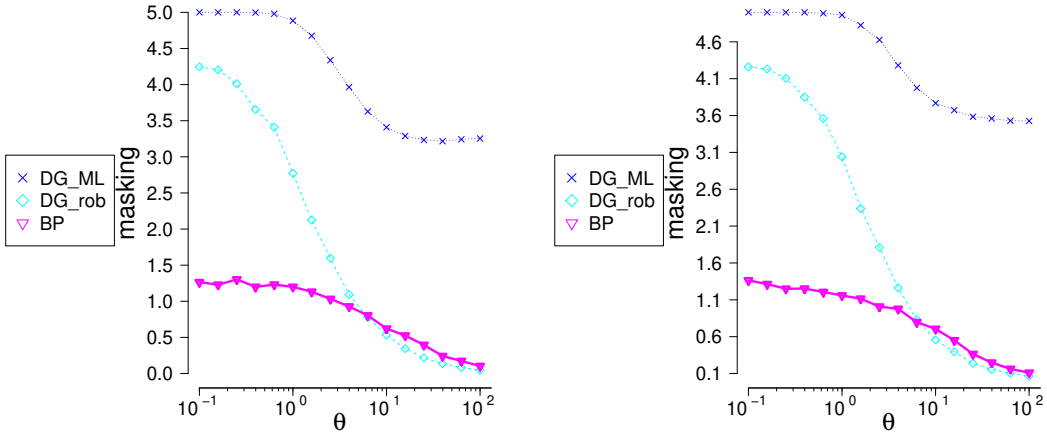
All three methods find all c-outliers if they are sufficiently remote. For $n = 50$ the BP method gives uniformly smallest masking values and the DG method gives uniformly largest masking values for any considered r in all diapason of alternatives. For $n = 100$ and $r = 2, 5$ the result is the same. For $n = 100$ and $r = 10$ (it means that even for very small θ the data is seriously corrupted) the BP method is also the best except that for the most remote alternatives the $Rosner_{[0.4n]}$ method slightly outperforms the BP method. For $n = 1000$ and the most of alternatives the BP method strongly outperforms other methods, except the most remote alternatives.

The DG and Rosner's methods have very large masking if many outliers are concentrated near the outlier region border. In this case data is seriously corrupted; however, these methods do not see outliers.

Conclusion: in most considered situations the BP method is the best outlier identification method. The second is Rosner's method with $s = [0.4]$, and the third is the Davies-Gather method based on robust estimation. Other methods have poor performance.

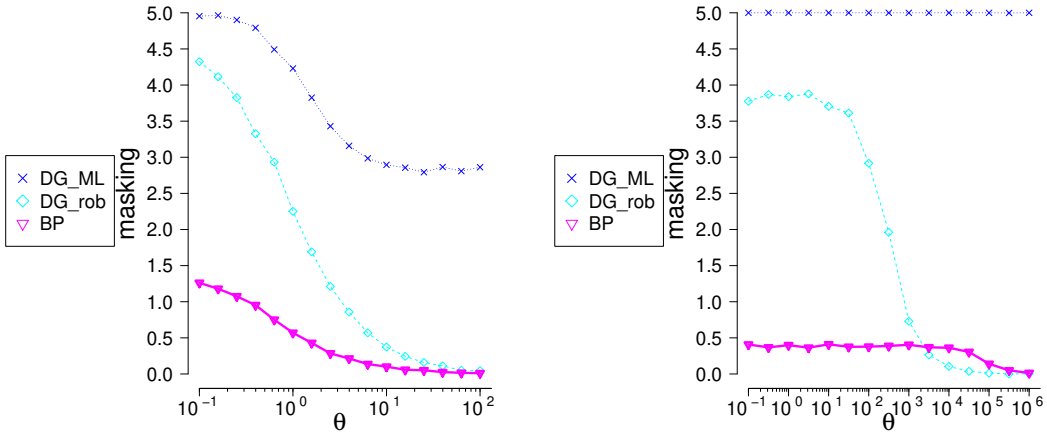
6.2. Investigation of Outlier Identification Methods for Other Location-Scale Models

We investigated performance of the new method for location-scale families different from normal. We compare the BP method with the generalized Davies-Gather method for logistic, Laplace (symmetric, $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$), extreme values (non-symmetric $F \in \mathcal{G}_0 \cap \mathcal{F}_{ls}$), and Cauchy (symmetric, $F \in \mathcal{G}_1 \cap \mathcal{F}_{ls}$) families. C-outliers were generating using truncated exponential distribution concentrated in two-sided outlier region. Swamping values being small, masking value, see Table 10 and differences between the true number of c-outliers and the number of rejected observations, see Figures 4 and 5, were compared. The BP and DG_{rob} methods find very well the most remote outliers; meanwhile, the BP method identifies much better closer outliers. The DG_{rob} method identifies badly multiple outliers concentrated near the border of the outlier region, whereas the BP method does well. The DG_{ML} is not appropriate for multiple outlier search.



(**a**) Logistic distribution. (**b**) Laplace distribution.

Figure 4. The difference between number outliers and rejected observations given that sample size $n = 100$ and $r = 10$ outliers.



(**a**) Extreme value II distribution. (**b**) Cauchy distribution.

Figure 5. The difference between number outliers and rejected observations given that sample size $n = 100$ and $r = 10$ outliers.

Table 10. Masking values for logistic, Laplace, extreme value II and Cauchy distribution, when $n = 100$, $r = 5$.

	Logistic				**Laplace**			
Method\θ	**0.1**	**1**	**6.3**	**10**	**0.1**	**1**	**6.3**	**10**
DG_{ML}	5	4.89	3.64	3.42	5	4.96	3.98	3.78
DG_{rob}	4.21	2.69	0.76	0.51	4.27	2.98	0.87	0.59
BP	1.3	1.13	0.78	0.64	1.31	1.21	0.8	0.66
	Extreme Value II				**Cauchy**			
Method\θ	**0.1**	**1**	**6.3**	**10**	**1**	**100**	**1000**	**10^5**
DG_{ML}	4.96	4.19	3	2.9	5	5	5	5
DG_{rob}	4.29	2.25	0.59	0.4	3.81	2.89	0.8	0.01
BP	1.25	0.56	0.14	0.11	0.38	0.4	0.39	0.13

7. Conclusions

We compared by simulation outlier identification results of the new method and methods given in previous studies. Even in the case of the normal model, which is investigated by many authors, the new method shows excellent identification power. In many situations, it has superior performance as compared to existing methods.

The obtained results widened considerably the spectre of most used non-regression models needing outlier identification methods. Many two-parameter models such as Weibull, logistic and loglogistic, extreme values, Cauchy, Laplace, and others can be investigated applying the new method.

The advantage of the proposed outlier identification method is that it has very good potential for generalizations. The authors are at the completion stage of research on outlier identification methods for accelerated failure time regression models and generalized linear models, gamma regression model in particular. Outlier identification methods for time series is another direction of the future work. Possible direction is investigation of Gaussian mixture regression models (see [29]).

Limitation of the new method is that it cannot be applied for analysis of discreet models. Taking into consideration that the method is based on asymptotic results, we recommend not applying it to samples of very small size $n \leq 15$.

The R package outliersTests was created for the practical usage of proposed test.

Author Contributions: Investigation, V.B. and L.P.; Methodology, V.B. and L.P.; Supervision, V.B.; Writing—original draft, V.B. and L.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Bol'shev, L.; Ubaidullaeva, M. Chauvenet's Test in the Classical Theory of Errors. *Theory Probab. Appl.* **1975**, *19*, 683–692.
2. Davies, L.; Gather, U. The Identification of Multiple Outliers. *J. Am. Stat. Assoc.* **1993**, *88*, 782–792.
3. Dixon, W.J. Analysis of Extreme Values. *Ann. Math. Stat.* **1950**, *21*, 488–506.
4. Grubbs, F.E. Sample Criteria for Testing Outlying Observations. *Ann. Math. Stat.* **1950**, *21*, 27–58.
5. Rosner, B. On the Detection of Many Outliers. *Technometrics* **1975**, *17*, 221–227.
6. Tietjen, G.L.; Moore, R.H. Some Grubbs-Type Statistics for the Detection of Several Outliers. *Technometrics* **1972**, *14*, 583–597.
7. Barnett, V.; Lewis, T. *Outliers in Statistical Data*; John Wiley & Sons: Hoboken, NJ, USA, 1974.

8. Zerbet, A. Statistical Tests for Normal Family in Presence of Outlying Observations. In *Goodness-of-Fit Tests and Model Validity*; Huber-Carol, C., Balakrishnan, N., Nikulin, M.S., Mesbah, M., Eds.; Birkhäuser Boston: Basel, Switzerland, 2002; pp. 57–64.
9. Chikkagoudar, M.; Kunchur, S.H. Distributions of test statistics for multiple outliers in exponential samples. *Commun. Stat. Theory Methods* **1983**, *12*, 2127–2142.
10. Kabe, D.G. Testing outliers from an exponential population. *Metrika* **1970**, *15*, 15–18.
11. Kimber, A. Testing upper and lower outlier paris in gamma samples. *Commun. Stat. Simul. Comput.* **1988**, *17*, 1055–1072.
12. Lalitha, S.; Kumar, N. Multiple outlier test for upper outliers in an exponential sample. *J. Appl. Stat.* **2012**, *39*, 1323–1330.
13. Lewis, T.; Fieller, N.R.J. A Recursive Algorithm for Null Distributions for Outliers: I. Gamma Samples. *Technometrics* **1979**, *21*, 371–376.
14. Likeš, I.J. Distribution of Dixon's statistics in the case of an exponential population. *Metrika* **1967**, *11*, 46–54.
15. Lin, C.T.; Balakrishnan, N. Exact computation of the null distribution of a test for multiple outliers in an exponential sample. *Comput. Stat. Data Anal.* **2009**, *53*, 3281–3290.
16. Lin, C.T.; Balakrishnan, N. Tests for Multiple Outliers in an Exponential Sample. *Commun. Stat. Simul. Comput.* **2014**, *43*, 706–722.
17. Zerbet, A.; Nikulin, M. A new statistic for detecting outliers in exponential case. *Commun. Stat. Theory Methods* **2003**, *32*, 573–583.
18. Torres, J.M.; Pastor Pérez, J.; Sancho Val, J.; McNabola, A.; Martínez Comesaña, M.; Gallagher, J. A functional data analysis approach for the detection of air pollution episodes and outliers: A case study in Dublin, Ireland. *Mathematics* **2020**, *8*, 225.
19. Gaddam, A.; Wilkin, T.; Angelova, M.; Gaddam, J. Detecting Sensor Faults, Anomalies and Outliers in the Internet of Things: A Survey on the Challenges and Solutions. *Electronics* **2020**, *9*, 511.
20. Ferrari, E.; Bosco, P.; Calderoni, S.; Oliva, P.; Palumbo, L.; Spera, G.; Fantacci, M.E.; Retico, A. Dealing with confounders and outliers in classification medical studies: The Autism Spectrum Disorders case study. *Artif. Intell. Med.* **2020**, *108*, 101926.
21. Zhang, C.; Xiao, X.; Wu, C. Medical Fraud and Abuse Detection System Based on Machine Learning. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7265.
22. Souza, T.I.; Aquino, A.L.; Gomes, D.G. A method to detect data outliers from smart urban spaces via tensor analysis. *Future Gener. Comput. Syst.* **2019**, *92*, 290–301.
23. Hawkins, D.M. *Identification of Outliers*; Springer: Dordrecht, The Netherlands, 1980; Volume 11.
24. Kimber, A.C. Tests for Many Outliers in an Exponential Sample. *J. R. Stat. Soc.* **1982**, *31*, 263–271.
25. De Haan, L.; Ferreira, A. *Extreme Value Theory: An Introduction*; Springer: New York, NY, USA, 2007.
26. Rousseeuw, P.J.; Croux, C. Alternatives to the median absolute deviation. *J. Am. Stat. Assoc.* **1993**, *88*, 1273–1283.
27. Liu, Y.; Abeyratne, A.I. *Practical Applications of Bayesian Reliability*; John Wiley & Sons: Hoboken, NJ, USA, 2019.
28. Rosner, B. Percentage points for the RST many outlier procedure. *Technometrics* **1977**, *19*, 307–312.
29. Su, H.; Hu, Y.; Karimi, H.R.; Knoll, A.; Ferrigno, G.; De Momi, E. Improved recurrent neural network-based manipulator control with remote center of motion constraints: Experimental results. *Neural Netw.* **2020**, *131*, 291–299.