Skowronek et al. *Journal of Mathematics in Industry* (2024) 14:19
https://doi.org/10.1186/s13362-024-00157-6

Journal of Mathematics in Industry
a SpringerOpen Journal

RESEARCH Open Access

Testing for finite variance with applications to vibration signals from rotating machines

Katarzyna Skowronek[1*], Radosław Zimroz[2] and Agnieszka Wyłomańska[1]

*Correspondence: katarzyna.skowronek@pwr.edu.pl
[1] Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland
Full list of author information is available at the end of the article

Abstract

In this paper we propose an algorithm for testing whether the independent observations come from finite-variance distribution. The preliminary knowledge about the data properties may be crucial for its further analysis and selection of the appropriate model. The idea of the testing procedure is based on the simple observation that the empirical cumulative even moment (ECEM) for data from finite-moments distribution tends to some constant whereas for data coming from heavy-tailed distribution, the ECEM exhibits irregular chaotic behavior. Based on this fact, in this paper we parameterize the regular/irregular behavior of the ECEM and construct a new test statistic. The efficiency of the testing procedure is verified for simulated data from three heavy-tailed distributions with possible finite and infinite variances. The effectiveness is analyzed for data represented in time domain. The simulation study is supported by analysis of real vibration signals from rotating machines. Here, the analyses are provided for data in both the time and time-frequency domains.

Keywords: Testing; Finite variance; Heavy-tailed distribution; Monte Carlo simulations; Condition monitoring

1 Introduction

In this article, we discuss the problem of detection whether the analyzed data come from a distribution with a finite variance. Since in some real-world problems the model describing the data is well-known (and its probabilistic properties are understood), the appearance of infinite variance distributions can be explained theoretically; see e.g., [5, 7]. However, in many complex systems there are no trustworthy theoretical models that can adequately explain the existence of heavy-tail behavior. Thus, in such cases information about specific properties of the corresponding models can only be received based on an analysis of empirical data. Among these, we list vibration-based machine condition monitoring being a motivation for our research. In such a case, the measured vibrations from rotating machines may exhibit impulsive behavior. The impulsiveness may be related to the local damage and in such a case the impulses are expected to be periodic. However, one may also observe the non-cyclic impulses in the data. In that case they are signatures of the heavy-tailed distribution of the background noise. This is the case considered in this paper. According to our research, sources of impulsive (non-cyclic) behavior may be related to specific processes performed by machine (cutting, crushing, milling, drilling, compression, etc.), to completely random external disturbances, disturbances during data transmission, or even to numerical problems during data processing [25].

The considered problem was previously discussed in the literature from different perspectives, see e.g., [31]. Here, we focus on the existence of finite variance of the distribution corresponding to the analyzed data. We note that the diverging variance is strictly related to the heavy-tailed property. However, these two notions are not equivalent (not all distributions with heavy-tailed behavior have infinite variance). We mention that some statistical and signal processing techniques are dedicated only for data coming from finite-variance distributions. The classical examples are methods utilizing the sample autocorrelation function (ACF), see e.g. [4]. The sample ACF is considered as the classical statistic used for periodic/cyclic behavior identification. It is used for data represented in different domains, like time- or time-frequency domain, see e.g. [30]. However, when applying the sample ACF-based methods, one needs to take into account that they are properly defined only for finite-variance distributed data. Thus, utilizing such methods for data coming from diverging-variance distribution may not give expected results and the final conclusion may be unreliable. This problem was discussed in our previous research, see e.g., [19, 20] but also other authors highlighted the small efficiency of the classical ACF-based methods for heavy-tailed distributed data, see e.g. [1, 2]. We mention that in the literature one can find dedicated techniques for data with infinite-variance distributions, see e.g. [21, 22]. Thus, the knowledge about the probabilistic properties of the data (expressed in the means of finite- and infinite-variance of its distribution) can help to avoid inappropriate conclusions resulting from the use of wrong tools.

In the statistical literature there are proposed techniques that may help to distinguish the finite- and infinite-variance distribution describing the data. However, most of the methods test a given specific distribution, see e.g. [9]. In the problem of distinguishing the finite- and infinite-variance distributed data, one can also apply statistical tests dedicated for some specific heavy-tailed distributions, like e.g., for α-stable class of distributions [10, 11]. Although the aforementioned tests are widely used in various applications, their limitations are also discussed in the literature, see e.g. [16].

The problem considered in this paper is much more general than the classical statistical testing, where the null hypothesis is defined such that the data come from given theoretical distribution. We highlight that the analyzed problem was also indicated by some authors, see e.g. [13, 14, 26, 36] where interesting approaches for testing for finite-variance (and other moments) were proposed. This problem is also discussed from different perspectives, i.e., as the problem of testing a power-law behavior which may be also a signature of infinite-variance distribution, see e.g. [12]. In our previous research, we also discussed the issue of heavy-tailed behavior recognition, see e.g. [5, 7]. However, the introduced algorithms were dedicated for specific classes of distributions (like α-stable distributions). In this paper, we present a broader perspective and discuss the problem in the context of any distribution with possible infinite variance. Moreover, the information we get here ("0-1" for finite-infinite variance distribution, respectively) is easy to interpret and can be successfully used in practical applications in contrast to [5, 7, 34], where only visual tests were proposed. We would like to emphasize that the idea of testing finite-variance distribution is not novel, as in [15] the authors introduced approach based on converging variance to discriminate between finite and infinite variance distributions.

We propose a methodology based on the observation discussed in [7], where it was indicated that the statistic defined as the empirical cumulative even moment (ECEM) converges to some constant value for sample of observations from distribution with finite moments, whereas the ECEM exhibits irregular chaotic behavior for data coming from distributions with infinite moments. In [7] we took advantage of this fact and proposed a procedure that was effective in the problem of distinguishing between a Gaussian and close to Gaussian α-stable distribution. In this article, we take a step forward in two aspects. First, we parameterize the regular behavior of the ECEM for finite-variance distributed data and introduce a new statistic (denoted by A) useful in the testing procedure. Second, we propose a test based on A statistic for finite-variance distribution testing that is effective for wide class of distributions. Moreover, we show that the ECEM statistic (denoted by C) can also be useful for the considered problem. The efficiency (expressed by the power of the tests) of the testing procedures based on A and C statistics is verified using Monte Carlo simulations for three classes of distributions (corresponding to alternative hypothesis), namely mixture of Gaussian [32], Student's t [35] and α-stable [33] distributions that are selected to cover a wide class of distributions with possible finite and infinite-variance. The results are compared with the known methods presented in [13, 36]. Finally, the methodology is verified for real vibration signals coming from the rotating machines. The real signals are analyzed in time- and time-frequency domains. The obtained results clearly confirm our research [34], where the same data were examined for condition monitoring purposes. The real-world examples and presented Monte Carlo studies clearly demonstrate the efficiency and universality of the proposed testing methodology.

The rest of the paper is organized as follows. In Sect. 2 we describe in details the construction of A and C test statistics and introduce the testing methodology. In Sect. 3 we discuss the analyzed distributions and demonstrate the behavior of the considered test statistics for examined cases. Next, in Sect. 4, we present the results of the testing procedures based on A and C statistics for simulated data. In this section, we also present a comparison with known methods for finite-variance distribution testing. In Sect. 5 we analyze real vibration signals in the context of the presented methodology. Here, the analyses are provided for data in both the time and time-frequency (spectrogram) domains. The last section concludes the paper.

2 Methodology

In this section, we introduce a new methodology for testing whether a given vector of observations comes from finite-variance distribution. The proposed technique is based on the approach presented in [5, 7], where the problem of discrimination between Gaussian and near-Gaussian (with infinite variance) distributions was discussed. First, we define the statistic called empirical cumulative even moment

$$C(k,N) = \frac{1}{k}\sum_{i=1}^{k} x_i^{2N}, \qquad k = 1,2,\ldots,n, \tag{1}$$

where $x_1, x_2, \ldots, x_n$ is a sample of independent observations from zero-mean distribution and $N \in \mathbb{N} = \{1, 2, \ldots\}$. In [5, 7] the authors used the statistic $C(k, N = 2)$ - called empirical cumulative fourth moment (ECFM) and highlighted that it converges to a constant when the underlying sample is Gaussian. However, this property is also fulfilled for any distribution with finite fourth moment. In practice, for given sample of length n, one observes that ECFM exhibits irregular chaotic behavior only for data from distributions with infinite fourth moment. Similar property can also be observed for other values of N and we expect that the statistic $C(k, N)$ will stabilize if the vector of observations comes from a distribution with finite moment of order $2N$. Otherwise, the statistic will exhibit chaotic behavior. This observation is a starting point for the proposed testing methodology. The specific behavior of the statistic defined in Eq. (1) is also discussed in [5, 23, 34] for samples from selected finite- and infinite-variance distributions.

2.1 Construction of the test statistic

In this part, we present how to construct the statistic that may be used to parameterize the chaotic behavior of the ECEM statistic for a given N for infinite-variance distributed samples. Let us assume that $x_1, x_2, \ldots, x_n$ is a sample of independent observations from zero-mean distribution. The procedure consists of the following steps:

1. First, we calculate $C(k, N)$ statistic for vector $x_1, x_2, \ldots, x_n$ according to Eq. (1) for the selected value of N and all $k = 1, 2 \cdots, n$.
2. We identify the segments of the ECEM statistic between the jumps. To identify the segments, first we calculate the increments of the statistic and then identify their peaks,[1] considering them as the points separating the segments.[2]
3. We select the segments of the ECEM statistic that are long enough. In our analysis, we selected the segments of minimum $10\%n$. For further analysis, we take the last long segment as with increasing value of k for samples from finite variance distributions the statistic $C(k, N)$ tends to theoretical value, and for samples from infinite variance distributions statistic $C(k, n)$ diverges to infinity. Thus, the last segment showcases the limiting behaviour of the statistic. The data in this segment are denoted as $c(t_1, N), c(t_2, N), \ldots, c(t_m, N)$, where m is the length of that segment. If there are no peaks identified (i.e. there are no identified segments), we take the last 10% of the points of the $C(k, N)$ statistic.
4. For the vector $c(t_1, N), c(t_2, N), \ldots, c(t_m, N)$ we fit the straight line using the least-squares method. As a consequence we use the linear regression model where the values $c(t_1, N), c(t_2, N), \cdot, c(t_m, N)$ are random observations while the points $t_1, t_2, \ldots, t_m$ are covariates. The estimated value of the slope in the linear regression model we denote as $A(N)$. The $A(N)$ we consider further as a test statistic. Let us note that $C(N)$ statistic is calculated for the entire dataset, while the slope $A(N)$ is determined for the selected segment. The rationale for choosing $A(N)$ statistic for testing whether the vector comes from finite variance distribution is as follows. As it was previously mentioned, if the distribution generating the data has finite

[1] To obtain the increments of the statistic, we use MATLAB function *diff*.

[2] To identify the peaks, we utilize MATLAB function *findpeaks*.

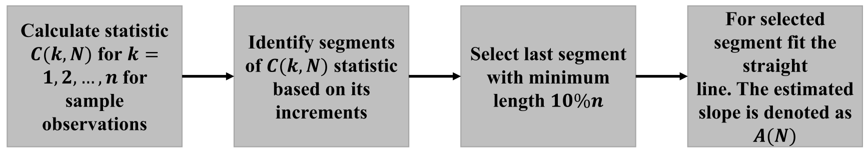

Figure 1 The flowchart of the derivation of statistic $A(N)$

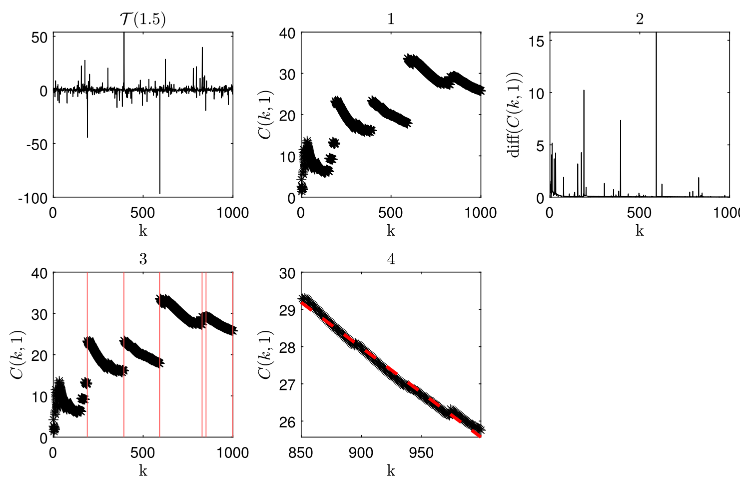

Figure 2 The procedure presenting derivation of $A(1)$ statistic for Student's t distribution with $\nu = 1.5$ (infinite variance case, see Eq. (5)). At first, the trajectory of $C(k,1)$ statistic is obtained (see panel 1). Next, the differences between neighbouring values of statistic are calculated (see panel 2). Based on the peaks, we segment the trajectory of $C(k,1)$ statistic (see panel 3)- red lines represent points separating segments. Lastly, we fit straight line (red dashed line) to the last segment (see panel 4) - the obtained slope is our statistic $A(1)$

variance, we expect that the $C(k,N)$ statistic stabilizes. Thus, in this case, the $A(N)$ statistic is close to zero. More precisely, the distribution of $A(N)$ is concentrated around zero. On the other side, if the distribution of the random sample has infinite variance, then the ECEM statistic exhibits chaotic behavior and $A(N)$ is significantly lower than zero. In summary, the distribution of $A(N)$ statistic is significantly different for the finite- and infinite-variance cases.

A flowchart of the above procedure for calculating the statistic $A(N)$ is presented in Fig. 1. Moreover, we include a step-by-step scheme of the proposed procedure in Fig. 2 (example of infinite variance case) and Fig. 16 (example of finite variance case) for random samples drawn from Student's t distribution. The numbers above each subplots correspond to the steps of the above procedure. For infinite-variance case we assumed $\nu = 1.5$ while for finite-variance case - $\nu = 15$, see next section for more details.

Let us note that similarly as it was proposed in [15], the assumption about zero-mean sample $x_1, x_2, \ldots, x_n$ does not neccesarily need to hold. Namely, one can normalize the random sample using robust algorithms, such as substracting mean (or median) and dividing by the conditional variance (see [29]), thus enabling the application of proposed methodology in various scenarios when the given data are not centered around 0. This procedure was applied in the simulated and real data analysis. See Sects. 4 and 5 for more details.

2.2 Testing procedure

We assume that the null and alternative hypotheses are defined as follows:

$$H_0: \text{ the given vector of observations comes from finite variance distribution}$$

$$H_1: \text{ the vector of observations comes from infinite variance distribution.}$$

In the testing procedure we use the classical approach based on the acceptance region calculated for the test statistic under H_0 hypothesis. We proceed as follows:

- For sample of observations $x_1, x_2, \ldots, x_n$ and given N we calculate the value of $A(N)$ statistic.
- We calculate the acceptance region

  $$[Q_{d/2}(n,N), Q_{1-d/2}(n,N)], \tag{2}$$

  where $Q_p(n,N)$ is the quantile of order p of the distribution corresponding to the test statistic $A(N)$ under H_0 hypothesis.
- We reject the H_0 hypothesis if the test statistic calculated for vector $x_1, x_2, \ldots, x_n$ is extreme, either larger than an upper critical value or smaller than a lower critical value with a given significance level d (with probability d it is inside the critical region or equivalently outside the acceptance region). If the value of the statistic $A(N)$ calculated for vector $x_1, x_2, \ldots, x_n$ falls into the acceptance region with a given significance level d, we may conclude that there is no evidence to reject the H_0 hypothesis at this significance level.

Since the theoretical distribution of the test statistic under the H_0 hypothesis is not known, the acceptance region (2) is calculated based on Monte Carlo simulations. Thus, there is need to assume a specific distribution corresponding to H_0 hypothesis. However, as demonstrated in the next section, the distribution of the $A(N)$ statistic exhibits similar behavior for finite-variance distributed samples. Therefore, any distribution with this property can be assumed in the H_0 hypothesis. In this paper, we use Monte Carlo simulated samples of length n from the Gaussian distribution to obtain the empirical distribution of the test statistic and construct the acceptance region (2). Let us note that the values of statistic $A(N)$ are closer to 0 for samples from finite-variance distributions, and for infinite-variance case the values of this statistic are significantly lower than 0, thus it is possible to construct one-sided acceptance region $[Q_d(n,N), \infty]$ instead of the one proposed in (2).

Instead of analyzing the $A(N)$ statistic obtained by using the above procedure, one may also test the finite variance distribution utilizing directly the ECEM statistic for given k and N. The similar methodology was proposed in [23] to distinguish the Gaussian and close to Gaussian α-stable processes (with stationary increments). As mentioned, the ECEM statistic given in (1) exhibits different behavior for samples from finite and infinite variance distributions, especially for larger values of N, see the next section for more details. Thus, the ECEM can be also considered as the test statistic for identification of finite variance. In this paper, the testing methodology is used for $k = n$ and in this case the statistic defined in Eq. (1) we denote as $C(N)$. The testing procedure based on ECEM statistic is similar to that described above. However, in this case the quantiles in (2) are calculated based on the values of $C(N)$ under H_0 hypothesis. In this case under H_0 hypothesis we also assume Gaussian distribution.

3 Analyzed distributions

In this section, we present three distributions that are further analyzed in the context of the proposed methodology, namely mixture of Gaussian distributions, Student's t distribution and the α-stable distribution. A random variable with mixture of Gaussian distributions has finite variance for any set of parameters. However, depending on the values of the parameters, this distribution may belong to platykurtic (when excess kurtosis is smaller than 0, for definition of excess kurtosis see [39]) or leptokurtic class of distributions (i.e. when the excess kurtosis is greater than zero or it is infinite), see e.g. [39]. For particular values of the parameters this distribution reduces to the Gaussian one. The Student's t distribution belongs to the leptokurtic class of distributions for which the excess kurtosis is greater than 0 or is infinite. The variance of Student's t distribution is finite when the number of degrees of freedom is greater than 2, otherwise we have the infinite-variance case. For large number of degrees of freedom the Student's t distribution tends to Gaussian one. The last considered distribution, α-stable one, belongs to the leptokurtic class of distributions. In this case the variance is infinite for all values of stability index (see Eq. (6)). The only exception is when the stability index is equal to 2. In that case the α-stable distribution reduces to the Gaussian one.

3.1 Mixture of Gaussian distributions

Let $a_1, a_2, \dots, a_l$ denote a series of non-negative weights satisfying $\sum_{i=1}^{l} a_i = 1$. Let $F_1(\cdot), F_2(\cdot), \dots, F_l(\cdot)$ denote an arbitrary sequence of Gaussian cumulative distribution functions (CDFs) with means $\mu_1, \mu_2, \dots, \mu_l$ and variances $\sigma_1^2, \sigma_2^2, \dots, \sigma_l^2$. Let $f_1(\cdot), f_2(\cdot), \dots, f_l(\cdot)$ be the corresponding probability density functions (PDFs). A random variable with the following CDF and PDF

$$F(z) = \sum_{i=1}^{l} a_i F_i(z),\ f(z) = \sum_{i=1}^{l} a_i f_i(z),\ z \in \mathbb{R}, \tag{3}$$

$$\text{where } f_i(z) = \frac{1}{\sqrt{2\pi}\sigma_i} \exp\left(-\frac{1}{2}\left(\frac{z-\mu_i}{\sigma_i}\right)^2\right) \tag{4}$$

is called a mixture of Gaussian distributions [3, 32, 39] (denoted further as $\mathcal{MG}$). For $l = 1$ this distribution reduces to the Gaussian one (denoted further as $\mathcal{G}$). In that case, the excess kurtosis is equal to zero. As was mentioned, this distribution may belong to the leptokurtic class of distributions, but the variance is finite for any set of parameters. In this paper we assume $l = 2$, $\mu_1 = \mu_2 = 0$. Moreover we assume $\sigma_1 = 1$, $\sigma_2 = \sqrt{5}$ and the analysis is provided with respect to the $a_1 \in [0.7, 1]$ parameter. In that case, the excess kurtosis is greater than 0, the only exception is when $a_1 = 1$. In that case, the mixture of Gaussian distributions reduces to the Gaussian one.

In Fig. 3, we demonstrate the boxplots of the analyzed statistics $A(N)$ and $C(N)$ defined in the previous section for mixture of Gaussian distributions and $N = 1, 2, 3$. The boxplots are obtained for samples from a mixture of Gaussian distributions for sample

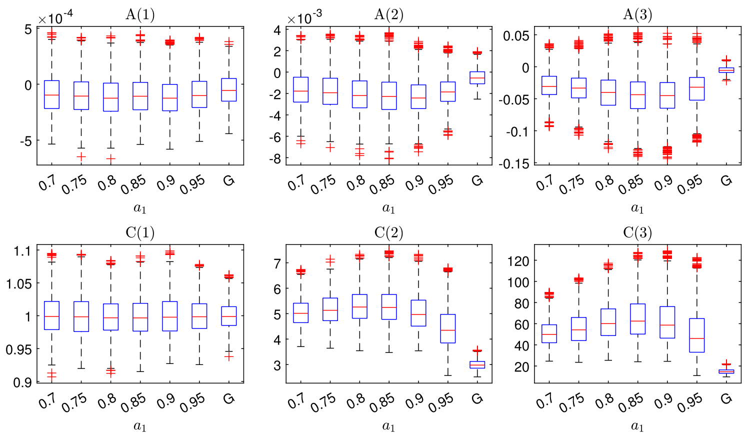

Figure 3 The boxplots of the statistics $A(1)$, $A(2)$, $A(3)$ (top panels) and $C(1)$, $C(2)$, $C(3)$ (bottom panels) for samples of length $n = 1000$ from mixed Gaussian distributions with $m = 2$, $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$ and $\sigma_2 = \sqrt{5}$. The boxplots are presented with respect to a_1 parameter and are obtained based on $M = 1000$ Monte Carlo simulations

length $n = 1000$ based on $M = 1000$ Monte Carlo simulations. As can be seen, for $N = 1$ the difference between the statistics considered is not visible for all cases (including the Gaussian case, that is, when $a_1 = 1$). For higher N there is a significant difference between the cases $a_1 \in [0.7, 1)$ and $a_1 = 1$. This is visible in the interquartile ranges and medians of the sample statistics. Thus, in the case considered, one may expect that the test for finite variance distribution will be more efficient for $N = 1$. Note that for a mixture of Gaussian distributions, the test should not reject the H_0 hypothesis for any values of a_1.

3.2 Student's t distribution

Student's t distributed random variable is defined through the following PDF [35, 37]

$$f(z) = \frac{1}{\sqrt{\nu\pi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{z^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad z \in \mathbb{R}, \tag{5}$$

where the number of degrees of freedom $\nu > 0$ is a parameter responsible for heavy-tailed properties. In the following parts, we denote this distribution by $\mathcal{T}$. The lower the ν parameter, the more impulsive behavior that occurs in the corresponding random sample. For $\nu \in (0, 2]$ the variance is not defined. Moreover, as $\nu \to \infty$, the Student's t distribution tends to the Gaussian one.

In Fig. 4, we show boxplots of the statistics $A(N)$ (top panels) and $C(N)$ (bottom panels) for $N = 1, 2, 3$ obtained for samples from Student's t distribution of length $n = 1000$ for different numbers of degrees of freedom. As before, the boxplots are obtained with $M = 1000$ Monte Carlo simulations. One can see the significant differences between finite- and infinite-variance cases we receive for higher N for both types of statistics.

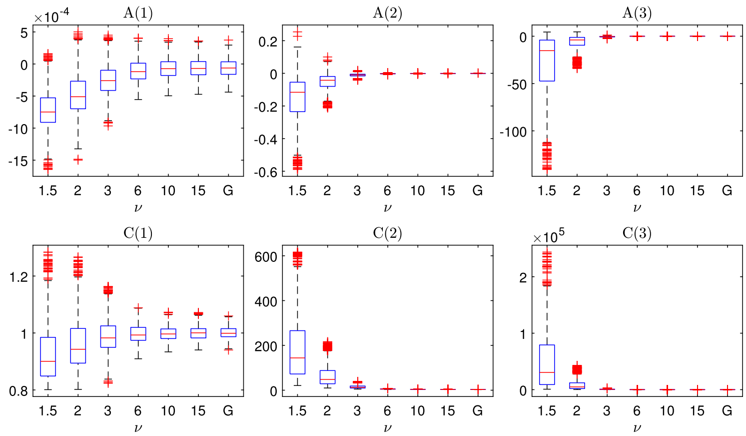

Figure 4 The boxplots of the statistics $A(1)$, $A(2)$, $A(3)$ (top panels) and $C(1)$, $C(2)$, $C(3)$ (bottom panels) for samples of length $n = 1000$ from Student's t distribution with different v parameters. The boxplots are obtained based on $M = 1000$ Monte Carlo simulations

3.3 The α-stable distribution

The α-stable distributed random variable is defined by the following characteristic function [27, 33, 38]

$$\Phi(z) = \begin{cases} \exp\left(-|z|^{\alpha}[1 - i\beta\tan\frac{\pi\alpha}{2}\operatorname{sign}(z)]\right), & \alpha \neq 1 \\ \exp\left(-|z|^{\alpha}[1 + i\beta\frac{2}{\pi}\operatorname{sign}(x)\log|z|]\right), & \alpha = 1, \end{cases} \tag{6}$$

where $z \in \mathbb{R}$, $\alpha \in (0,2]$ is the stability index responsible for heavy-tailed behavior, $\beta \in [-1,1]$ is the skewness parameter, $\gamma \in \mathbb{R}_{+}$ is the scale parameter, and $\mu \in \mathbb{R}$ is the shift parameter. The tail of the α-stable distribution decays as the power-law function, that is, $1 - F(z) \sim z^{-\alpha}$, where $F(z)$ is the corresponding CDF. Therefore, one can conclude that the lower the stability index α, the more impulsive the behavior that occurs in the sample. When $\alpha = 2$, the α-stable distribution reduces to the Gaussian one, hence the variance, as well as higher order moments, such as skewness or kurtosis, exist. For $\alpha < 2$, the variance is infinite. In this paper, we consider a symmetric $\alpha-$ stable distribution, that is, when $\beta = 0$ and $\mu = 0$. Moreover, we assume $\sigma = 1$. In the further analysis, we denote this distribution by $\mathcal{S}$.

In Fig. 5, we provide boxplots of the statistics $A(N)$ and $C(N)$ for $N = 1,2,3$ obtained for the samples of the α-stable distribution for the length of the sample $n = 1000$ with different parameters α. The boxplots are obtained with $M = 1000$ Monte Carlo simulations. Similar to the case of Student's t distribution, there is significant difference between distributions of the statistics for finite- (i.e. Gaussian) and infinite-variance random samples for all values of N.

4 Simulated data analysis

In this section, we present the efficiency of the procedures for finite variance testing for simulated data. Here, we analyze three distributions described in the previous section. The effectiveness of the testing procedures is demonstrated based on the powers of the tests.

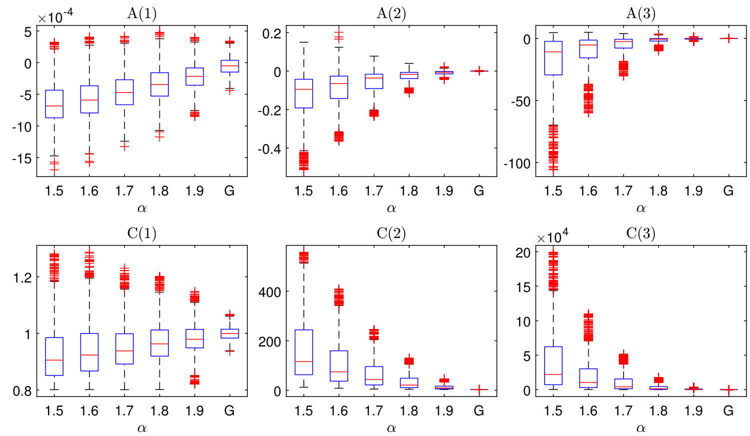

Figure 5 The boxplots of the statistics $A(1)$, $A(2)$, $A(3)$ (top panels) and $C(1)$, $C(2)$, $C(3)$ (bottom panels) for samples of length $n = 1000$ from α-stable distribution with different α parameters. The boxplots are obtained based on $M = 1000$ Monte Carlo simulations

In the performed simulation studies the test procedures based on the statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$, $C(2)$, $C(3)$ are applied to the samples represented in the time domain. In our analysis, we investigate the performance of the proposed tests based on 1000 Monte Carlo simulations of samples from each distribution with lengths $n = 500, 1000$. The empirical distribution of test statistic is obtained based on 1000 simulated trajectories from Gaussian distribution (corresponding to the H_0 hypothesis - finite variance case) and the power of the test is calculated based on 1000 simulated trajectories from mixture of Gaussian distribution, Student's t distribution and α-stable distribution. More precisely, for each simulated trajectory from each distribution we calculate the test statistics and we compare if they fall in the obtained acceptance region (as described in proposed testing procedure in Sect. 2.2).

In the case of mixture of Gaussian distributions, we fixed parameters $m = 2$, $\mu_1 = \mu_2 = 0$, $\sigma_1 = 1$ and $\sigma_2 = \sqrt{5}$ and we analyze the performance of the tests with respect to the parameter $a_1 \in \{0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1\}$. In the case of the Student's t distribution, we apply the test to samples from the distribution with parameter $\nu \in \{1.5, 2, 3, 6, 10, 15\}$. Finally, in case of the α-stable distribution, we verify the results of the tests by the assumption that the scale parameter $\sigma = 1$ and the stability index $\alpha \in \{1.5, 1.6, 1.7, 1.8, 1.9, 2\}$. As it was mentioned, the empirical distributions of the analyzed test statistics under H_0 hypothesis are calculated for samples from the standard Gaussian distribution.

Due to the possible different scales of the samples considered, we propose applying a normalization procedure based on conditional variance [29]. A similar approach was used in [23]. The procedure is as follows: first, the sample median is calculated for given data. Then the conditional standard deviation is calculated, that is, the sample is trimmed at an arbitrarily chosen quantile levels (in our case $q_1 = 0.1$ and $q_2 = 0.9$), and then the sample standard deviation is calculated for the trimmed data. Finally, the vector is normalized by subtracting the obtained sample median and dividing by the conditional sample standard deviation.

The powers of the tests (for $d = 5\%$) obtained for the mixture of Gaussian, Student's t, and α-stable distribution are presented in Fig. 6, Fig. 7 and Fig. 8, respectively. For a mixture of Gaussian distributions, we expect the power of tests based on proposed statistics

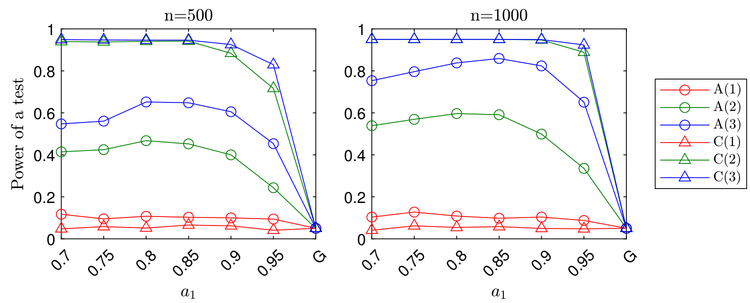

Figure 6 Power curves of the introduced tests obtained in 1000 Monte Carlo simulations of samples from mixture of Gaussian distributions with $n = 500, 1000$ observations. The results are presented with respect to parameter a_1 responsible for heavy-tailed properties of this distribution. We assume $d = 5\%$

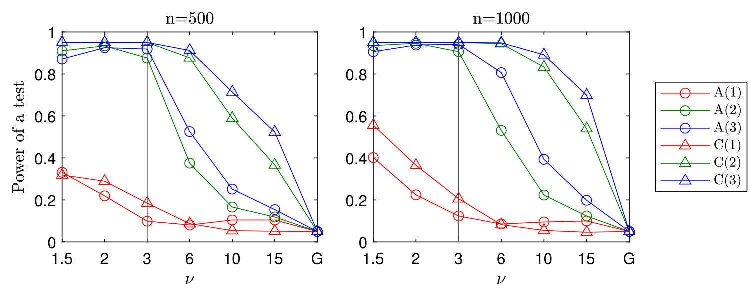

Figure 7 Power curves of the introduced tests obtained in 1000 Monte Carlo simulations of samples from Student's t distribution with $n = 500, 1000$ observations. The results are presented with respect to parameter ν responsible for heavy-tailed properties of the distribution. We assume $d = 5\%$

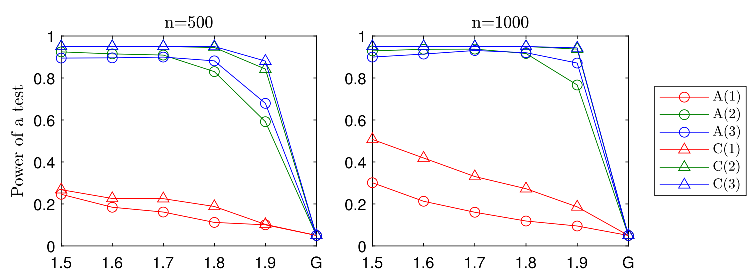

Figure 8 Power curves of the introduced tests obtained in 1000 Monte Carlo simulations of samples drawn from α-stable distributions with $n = 500, 1000$ observations. The results are presented with respect to parameter α responsible for heavy-tailed properties of the distribution. We assume $d = 5\%$

to be close to 0.05 reflecting the chosen significance level, as for any value of parameter a_1 the variance of the distribution exists. Based on the results obtained for a mixture of Gaussian distributions, it can be observed that the test statistics $A(2)$, $A(3)$, $C(2)$ and $C(3)$ are very sensitive to any violation from the Gaussian distribution, that is, when $a_1 = 1$. All tests based on the statistics $A(2)$, $A(3)$, $C(2)$ and $C(3)$ falsely reject the H_0 hypothesis for a mixture of Gaussian distributions. In contrast, tests based on $A(1)$ and $C(1)$ do not reject the H_0 hypothesis and correctly identify a mixture of Gaussian distribution as a finite-variance case for any a_1.

In the case of the Student's t distribution, the power of all tests decreases with respect to the increasing value of ν, see Fig. 7. This means that the statistics are sensitive to changes in the variance of this distribution. Let us note that for $\nu > 2$ we expect the tests to not reject the H_0 hypothesis. Therefore, for $\nu > 2$, we expect the power of the tests to be close to 0.05, as selected significance level, otherwise the power of the tests should be significantly higher than the significance level. In Fig. 7 the cases corresponding to finite and infinite variance are separated by the vertical line. Based on the results presented in Fig. 7 it can be concluded that the tests based on statistics $A(1)$ and $C(1)$ favor not rejecting the H_0 hypothesis even for $\nu = 1.5$ and $\nu = 2$. At the same time, the other tests tend to reject the H_0 hypothesis more often even if it is true, which can be observed for $\nu = 3$ and $\nu = 6$. For $\nu = 10$ and $\nu = 15$, tests based on statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$ favor not rejecting the H_0 hypothesis and statistics $C(2)$ and $C(3)$ reject it in most of the cases. The power of the tests based on $A(1)$ and $C(1)$ is lower than that of any other test. Furthermore, tests $A(2)$, $A(3)$, $C(2)$ and $C(3)$ incorrectly reject the H_0 hypothesis for $\nu = 6$ and $\nu = 3$.

For the α-stable distribution, the results are presented in Fig. 8. The tests based on the statistics $A(2)$, $A(3)$, $C(2)$, and $C(3)$ discriminate in favor of the infinite variance of the distribution for all the values of the stability index, as expected. Furthermore, the introduced test statistics can distinguish between finite and infinite variance (strictly α-stable) distributions for samples of length $n = 500, 1000$ for the stability index close to 2. Considering all statistics, the least efficient of the proposed tests for α-stable distribution are tests based on the statistics $A(1)$ and $C(1)$. In these cases, the power of the test is the lowest. Moreover, for sample length $n = 1000$ the values of the power of the tests based on $A(2)$, $A(3)$, $C(2)$, and $C(3)$ are above 0.8 for stability index $\alpha \in \{1.5, 1.6, 1.7, 1.8, 1.9\}$ which means that all these tests accurately reject H_0 hypothesis. Hence, for $\alpha \in \{1.5, 1.6, 1.7, 1.8, 1.9\}$ tests based on $A(2)$, $A(3)$, $C(2)$ and $C(3)$ rightfully reject the H_0 hypothesis.

In the Appendix we include the comparison of the power of tests based on the $A(N)$ statistic with two-sided and one-sided acceptance region for all analyzed distributions (see Fig. 17, Fig. 18 and Fig. 19). In case of Student's t distribution and α-stable distribution the results of a test with two-sided acceptance region are similar to those obtained for a test with one-sided acceptance region (see Fig. 18 and Fig. 19. For mixture of Gaussian distributions (see Fig. 17), the power of a test with one-sided acceptance region is lower than for a test with two-sided acceptance region. Thus, for mixture of Gaussian distributions the test with two-sided acceptance region is more restrictive as it rejects the H_0 hypothesis more often.

To compare the introduced methods, the power of each test for every distribution is included in Table 1 (see Appendix) for the significance level $d = 5\%$ for different values of the sample lengths. Moreover, we demonstrate the results of two tests known from the literature to compare efficiency with the proposed methods. In the first method introduced in [36], denoted as T1, the H_0 hypothesis assumes an infinite second moment of a random sample. Thus, in Table 1 in the column "T1", the results do not present the power of a test, but rather 1 – power of a test. In this method, one first calculates the empirical second moment s^2 of the tested sample $x_1, x_2, \ldots, x_n$, and then generates a vector of independent random variables $\xi_1, \xi_2 \cdots, \xi_r$, where $\xi_j \sim \mathcal{N}(0, \sqrt{e^{s^2}})$, where $r = \lfloor n^{\frac{4}{5}} \rfloor$ and $\lfloor \cdot \rfloor$ is the floor function. Secondly, two sequences $\zeta_1(u), \zeta_2(u), \ldots, \zeta_r(u)$ for $u \in \{-1, 1\}$ are generated, where $\zeta_j(u) = \mathrm{I}[\xi_j \leq u]$ and $\mathrm{I}[\cdot]$ is an indicator function. Lastly, one defines $\upsilon_r(u)$ as $\upsilon_r(u) = \frac{2}{\sqrt{r}}\sum_{j=1}^{r}[\zeta_j(u) - 0.5]$ and the test statistic is constructed as $\theta_1 = \sum_u 0.5\upsilon_r^2(u)$. The test statistic θ_1 under H_0 follows χ^2 distribution with one degree of freedom, which enables one to create the acceptance region of the test at a given significance level.

In the second method [13], denoted as T2, the H_0 hypothesis is the same as in our case, i.e. that the considered random sample comes from the distribution with finite second moment. The method incorporates the bootstrap approach and is based on randomly generating $M = 10000$ sub-samples of the original sample with a fixed length $\lceil 0.4 \log n \rceil$, where $\lceil\cdot\rceil$ is the ceiling function. Then, for each sub-sample, the empirical second moment is calculated $s_1^2, s_2^2, \dots s_M^2$. As a test statistic θ_2 the authors selected the ratio of the number of sub-samples with a higher empirical second moment than the scaled empirical second moment of the tested initial sample to the number of all simulated sub-samples, i.e. $\theta_2 = \frac{\sum_{i=1}^{M} I[s_i^2 > 0.999 s^2]}{M}$, where s^2 is the empirical second moment of the initial sample. If the defined ratio exceeds the chosen significance level, the H_0 hypothesis is not rejected, otherwise the test rejects H_0 in favor of the H_1 hypothesis. For both tests T1 and T2, we apply the same normalization of the data as in the tests introduced in this paper. The results obtained for T1 and T2 for $d = 5\%$ are presented in Table 1 (see columns "T1" and "T2").

It can be observed that especially in the case of α-stable distribution, where for stability index $\alpha \in \{1.7, 1.8, 1.9\}$ tests T1 and T2 identify the distribution as finite variance, i.e. test T1 rejects the H_0 hypothesis that the distribution has an infinite second moment, while test T2 does not reject H_0 hypothesis, assuming finite second moment. As it was previously noted, tests based on $A(1)$ and $C(1)$ favor not rejecting the H_0 hypothesis and their performance is at least as good or better than the performance of the tests T1 and T2. The only case where T1 and T2 are more efficient than proposed tests is observed for Student's t distribution with $\nu = 3$. When applying the tests to infinite-variance samples, tests based on statistics $A(2)$, $A(3)$, $C(2)$ and $C(3)$ have higher power than test T2. The methods proposed in this paper are also more efficient than T1 for infinite-variance identification, comparing the power of the tests based on $A(2)$, $A(3)$, $C(2)$ and $C(3)$ and 1 – power of a test for T1. Thus, the tests $A(2)$, $A(3)$, $C(2)$ and $C(3)$ are more efficient to reject the H_0 hypothesis than tests T1 and T2.

5 Real data analysis

In this section, we demonstrate the results of proposed tests, based on statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$, $C(2)$, $C(3)$, applied to real data. We analyze two vibration signals collected from healthy machines. Thus, we do not expect here the components related to local damage (i.e. cyclic impulses). In the next subsections, we study the signals in the time and time-frequency (spectrogram) domains, respectively. The analyzed vibration signals come from a rolling element bearing used in belt a conveyor system (signal 1) and a hammer crusher used for fragmentation of hard solid material (signal 2). The signals are presented in Fig. 9 in the time domain (see the top panels) and in the time-frequency domain (see the bottom panels). The sampling frequency of signal 1 is 19200, and for signal 2 - it is 25000. In case of signal 1, we have in total 48000 number of observations corresponding to 2.5 seconds of measurement. For signal 2,

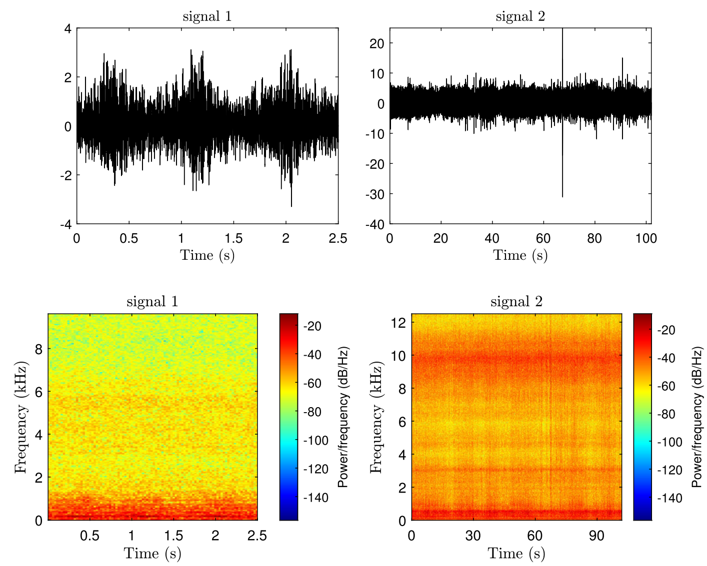

Figure 9 Top panels: analyzed signals in the time domain. Bottom panels: analyzed signals in the time-frequency domain (spectrograms)

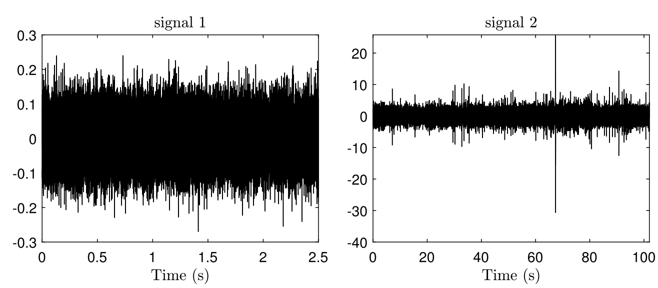

Figure 10 Analyzed signals after applying high-pass filtering

we obtained 2550000 sampling points measured during 102 seconds. The same signals were examined in the time-frequency domain in the article [34], which addressed the issue of infinite-variance distributed signals in the context of condition monitoring.

5.1 Analysis in time domain

Real vibration signals are often considered as a mixture of deterministic and random components. These components should be separated by decomposition methods. For simplicity, assuming that deterministic components are low frequency signals, we used basic high-pass filter to remove low-frequency, high energy deterministic components. Selection of cutoff frequency is not critical here, it was set to 2 kHz for signal 1, and 1 kHz for signal 2, as the energy is located in the range up to those frequencies. Let us note that in general the deterministic components may also appear for high frequencies. Thus, the pre-processing proposed here (high-pass filtration) is designed for such specific signals. In Fig. 10 we present the examined signals after applying high-pass filtering. In the analysis, we cut the signals into non-overlapping windows of length corresponding to 1 second. In

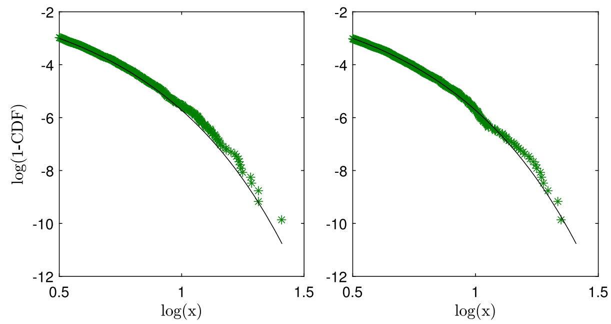

Figure 11 Comparison of the empirical tails (green stars) and theoretical tail of Gaussian distribution (black solid line) in logarithmic scale. Left panel: results obtained for the first segment of length 1 second, right panel: result obtained for the second segment of length 1 second

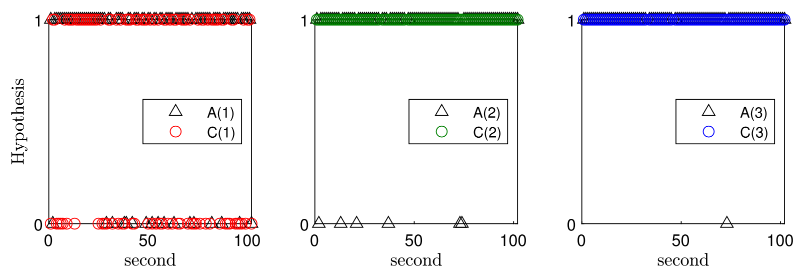

Figure 12 The results of proposed tests obtained for signal 2 in time domain for $d = 5\%$. For segments where the tests rejected the H_0 hypothesis we assigned the value 1, whereas for segments where the tests did not reject H_0 we assigned 0

case of signal 1 it is 19200 observations, and in case of signal 2 - 25000 observations. For signal 1 there are 2 segments (as the signal was measured during 2.5 seconds), whereas for signal 2 we obtain 102 segments. We apply the testing procedures based on the $A(1)$, $A(2)$, $A(3)$ and $C(1)$-$C(3)$ statistics to each segment separately. In the analysis, we used 1000 Monte Carlo simulations of Gaussian samples. More precisely, in the case of signal 1 to calculate the acceptance regions (2) under the H_0 hypothesis we simulate Gaussian samples of length 19200 while for signal 2 we used samples of length 25000. Moreover, we assumed significance levels $d = 5\%$. In case of signal 1 there is no evidence to reject the H_0 hypothesis (all tests do not reject the hypothesis). Moreover, this signal is confirmed to be Gaussian distributed by the Kolmogorov-Smirnov test [10] with $p_{value} = 0.27$ and $p_{value} = 0.48$ for the first and second segment, respectively. In Fig. 11 we present a comparison of the theoretical tail (1- cumulative distribution function) of the Gaussian distribution with the empirical tail of the data corresponding to the first two seconds of signal 1. In case of signal 2, tests based on statistics $A(2) - A(3)$ and $C(2) - C(3)$ reject the H_0 hypothesis for both significance levels. The tests based on statistics $A(1)$ and $C(1)$ do not reject H_0 hypothesis in some cases. However, there are more segments classified as infinite variance distributed. The results of the tests obtained for signal 2 in the time domain are presented in Fig. 12. The x axis presents the seconds of the signal and the y axis presents the result of the hypothesis testing, which is 0 (not rejecting the H_0 hypothesis) or 1 (rejecting the H_0 hypothesis).

5.2 Analysis in time-frequency domain

In this part, we present the results of the tests applied to the examined signals in their spectrogram representations. The spectrogram is defined as a square of the short time Fourier transform (STFT) [17][3]

$$S(t,f) = |STFT(t,f)|^2 = \left|\sum_{m=1}^{n} x_m w(t-m) e^{-i2\pi f \frac{m}{n}}\right|^2, \tag{7}$$

where $x_1, x_2, \dots, x_n$ is the considered vector of observations, $w(\cdot)$ is a window, $t \in T$ is a time point and $f \in \mathcal{F}$ is the frequency.

Let us note, that the probabilistic properties of a given random sample change after transformation via spectrogram. More precisely, in [18] it was shown that a Gaussian random sample after transformation into the time-frequency domain (spectrogram) has the so-called generalized χ^2 distribution [24]. In that case, the finite variance distribution property is valid for samples in the time and time-frequency domains. The situation is much more complicated when the random sample in the time domain has infinite-variance distribution. In that case, after the transformation into time-frequency domain considered here, we have also the infinite-variance distribution. However, there are also cases where the finite-variance distributed samples are transformed into the infinite-variance case. This point was extensively discussed in [34].

As mentioned, the same real signals were analyzed in our previous work in the time-frequency domain. In [34] it was concluded that signal 1 has a generalized χ^2 distribution. This thesis was confirmed in Sect. 5.1, where we have shown that this signal in the time domain can be considered as Gaussian distributed. Signal 2 in [34] was classified as infinite variance distributed in the time-frequency domain. In this work, we apply tests based on the statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$, $C(2)$, $C(3)$ to confirm this result.

In the time-frequency analysis, we calculated the spectrograms of the given signals after high-pass filtering. For signal 1, we selected $kaiser(500,5)$ windowing (see [28]), with 512 points to calculate the Fourier transform. The spectrogram of signal 2 was obtained using the $kaiser(2000,5)$ window and 2048 points to calculate the Fourier transform. As it was previously mentioned, to ensure that the observations in spectrogram representations are independent, in both cases the overlap parameter was set to 0. The spectrograms of signal 1 and signal 2 calculated after high-pass filtering are presented in Fig. 13.

It should be noted that selecting a single sub-signal associated with one frequency to determine the behavior of the background noise (i.e. independent observations) is not allowed due to the significant differences occurring for different frequencies. Moreover, to ensure that there are no auto-dependencies within the data, for each signal, we selected only the frequencies for which the mean of the robust autocorrelation measure (robust ACF), defined as in [40], was the lowest (in our case, we selected frequencies for which robust ACF was lower than 0.05). In our analysis, for signal 1 in such a way we selected 100 sub-signals, and for signal 2- 200 sub-signals. In each case, we analyzed test statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$, $C(2)$, $C(3)$ separately for all selected sub-signals (with the smallest

[3]In our analysis, we selected Kaiser window kaiser(L,b), where L represents window length and b is the shape factor.

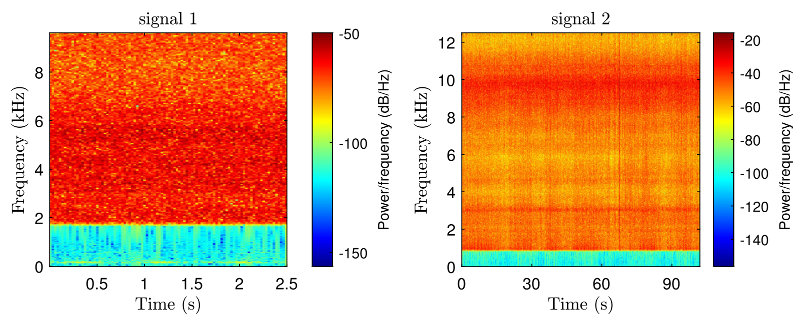

Figure 13 Spectrograms of analyzed signals after applying high-pass filtering

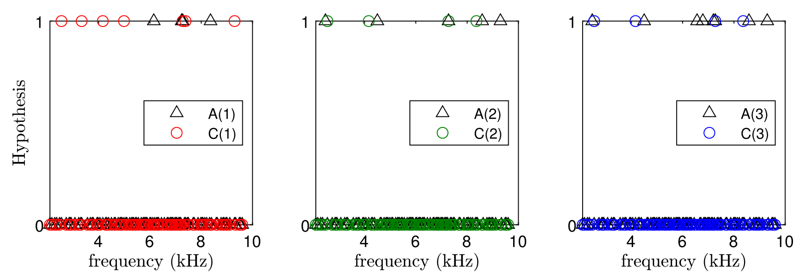

Figure 14 The results of proposed tests obtained for signal 1 in time-frequency domain for $d = 5\%$. For frequencies where the test rejected the H_0 hypothesis we assigned the value 1, whereas for frequencies where the test did not reject H_0 we assigned 0

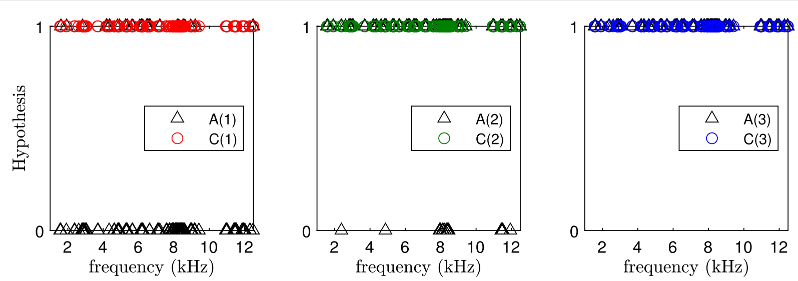

Figure 15 The results of proposed tests obtained for signal 2 in time-frequency domain for $d = 5\%$. For frequencies where the test rejected the H_0 hypothesis we assigned the value 1, whereas for frequencies where the test did not reject H_0 we assigned 0

robust ACF). Namely, for each sub-signal from the spectrogram, we calculate test statistics $A(1)$, $A(2)$, $A(3)$ and $C(1)$, $C(2)$, $C(3)$ and verify if they are contained in the respective acceptance region constructed based on 1000 Monte Carlo simulations. If the value of the statistic calculated for the analyzed sub-signal falls into the acceptance region, there is no evidence to reject the H_0 hypothesis at given significance levels, and for the selected sub-signal we classify it as finite variance distributed. The procedure is then repeated for all sub-signals within the selected frequency range. In Figs. 14 and 15 we demonstrate the results obtained for signal 1 and signal 2, respectively. The y-axis corresponds to the hypothesis testing result, i.e. if the H_0 was rejected, the result is 1, otherwise it is 0. In order to be consistent with the results presented in the simulation study, we assume $d = 5\%$.

In case of the signal originating from the rolling bearings, i.e. signal 1, the tests did not reject the H_0 hypothesis for the majority of sub-signals for the selected frequencies. In case of signal 2 obtained from the crusher, the tests based on statistics $A(2)$-$A(3)$ and $C(1)$, $C(2)$, $C(3)$ reject the H_0 hypothesis for most of the sub-signals, that confirms the results presented in [34]. The only exception is the test based on statistics $A(1)$ which does not reject H_0 hypothesis. The results obtained in the time-frequency domain correspond to the results obtained in time domain. Signal 1 in the time and time-frequency domain is classified as finite variance distributed. Signal 2 is classified as infinite variance distributed by most of the tests, both in the time and time-frequency domain.

6 Conclusions

In this paper we have introduced procedures to test whether the sample of observations comes from a finite-variance distribution. Preliminary knowledge about the probabilistic properties of the data (here expressed in terms of the finite variance of the corresponding distribution) is extremely important for selection of appropriate tools for its further analysis. The proposed methodology is based on the ECEM statistic and its specific behavior for data coming from finite- and infinite-variance distributions. We have parameterized this specific behavior and introduced a new test statistic. The efficiency of the testing procedures was verified for three broad classes of heavy-tailed distributions with possible finite and infinite variances. The presented real-world examples confirm the universality and the broad spectrum of possible applications of the introduced methodology. We believe that developed procedures could be useful for various applications and the proposed algorithms can be considered as the tools useful at the pre-processing step where the preliminary knowledge about the data properties is extremely important for further analysis. As the final conclusions we draw the main advantages of the proposed testing methodology:

- It is based on simple observation related to specific behavior of the ECEM statistic for finite- and infinite-variance distributed data.
- The A statistic used for testing is easy to calculate. It requires only a regression method, which is pretty standard and available in various mathematical packages.
- The testing procedure is quite standard and easy to implement.
- The proposed methodology is universal. It can be applied to any data (assuming it represents independent observations). Moreover, one may apply the testing procedure for data represented in any domains (like time, time-frequency domains).
- The presented simulation study clearly confirms the efficiency of the testing procedures for a broad class of distributions.

We highlight also main limitations of the introduced methodology:

- Theoretical distributions of test statistics are not know, thus, the acceptance region is obtained based on Monte Carlo simulations.
- In consequence in H_0 hypothesis there is a need to assume some specific distribution (here Gaussian) to obtain the acceptance region for the tests.
- The proposed methodology based on A statistic is dedicated to vectors of independent observations, and the possible dependencies in the data may reduce the efficiency of the testing procedures. Therefore, it is necessary to perform some pre-processing of the initial data (like high-pass filtration in the presented analysis for real data in time domain) to be sure that the examined data are independent.

However, as it is shown in [23], the testing methodology based on C statistic is also useful for data with a given dependence structure.

Appendix

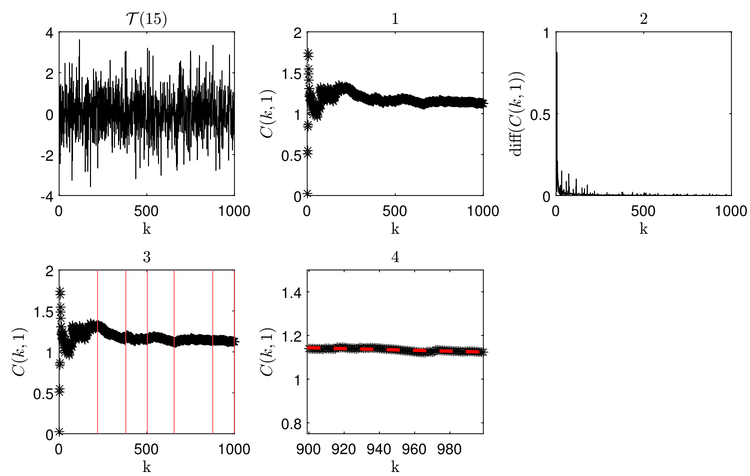

Figure 16 The procedure presenting derivation of $A(1)$ statistic for Student's t distribution with $\nu = 15$ (finite variance case, see Eq. (5)). At first, $C(1)$ statistic is calculated (see panel 1). Next, the differences between neighbouring values of statistic are calculated (see panel 2). Based on the peaks, we segment the trajectory of $C(1)$ statistic (see panel 3) - red lines represent points separating segments. Lastly, we fit straight line (red dashed line) to the last segment (see panel 4) - the obtained slope is our statistic $A(1)$

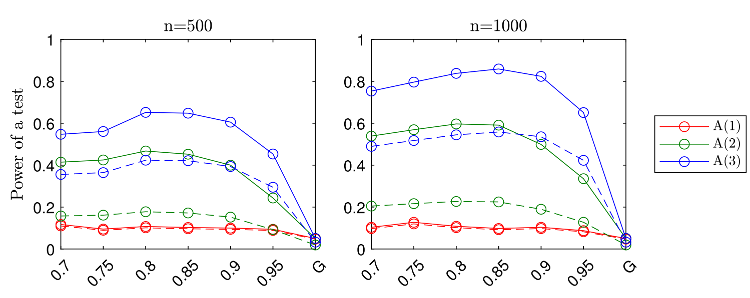

Figure 17 The comparison of power curves of the tests based on the $A(N)$ statistics with two-sided (solid lines) and one-sided (dashed lines) acceptance regions. Power curves of the tests are obtained based on 1000 Monte Carlo simulations of samples from mixture of Gaussian distributions with $n = 500, 1000$. The results are presented with respect to parameter a_1 responsible for heavy-tailed properties of this distribution. We assume $d = 5\%$ in case of both two-sided and one-sided acceptance regions

Table 1 Power of a tests based on statistics *A*(1), *A*(2), *A*(3) and *C*(1), *C*(2), *C*(3), and for T2 for a random sample of length $n = 500$ (top table) and $n = 1000$ (bottom table) for significance level $d = 5\%$. In the case of T1, results present 1 – power of the test

	A(1)	A(2)	A(3)	C(1)	C(2)	C(3)	T1	T2	
$n = 500$									
$\mathcal{MG}(a_1)$	0.7	0.12	0.41	0.55	**0.05**	0.94	0.95	**0.05**	0.06
	0.75	0.10	0.42	0.56	**0.06**	0.94	0.95	**0.06**	0.07
	0.8	0.11	0.47	0.65	**0.05**	0.94	0.95	0.06	0.07
	0.85	0.10	0.45	0.65	0.07	0.94	0.95	**0.05**	0.06
	0.9	0.10	0.40	0.61	0.06	0.88	0.93	**0.05**	0.06
	0.95	0.09	0.24	0.45	0.05	0.72	0.83	**0.04**	0.05
$\mathcal{T}(\nu)$	1.5	0.33	0.91	0.87	0.29	**0.95**	**0.95**	0.85	0.23
	2	0.22	0.93	0.93	0.26	**0.95**	**0.95**	0.72	0.16
	3	**0.10**	0.88	0.92	0.15	0.95	0.95	0.12	**0.10**
	6	0.08	0.38	0.53	**0.06**	0.88	0.91	0.07	0.08
	10	0.10	0.17	0.25	**0.06**	0.59	0.71	**0.06**	**0.06**
	15	0.10	0.12	0.15	0.06	0.37	0.52	**0.05**	**0.05**
$\mathcal{S}(\alpha)$	1.5	0.25	0.92	0.89	0.27	**0.95**	**0.95**	0.87	0.22
	1.6	0.18	0.91	0.90	0.23	**0.95**	**0.95**	0.51	0.14
	1.7	0.16	0.91	0.90	0.23	**0.95**	**0.95**	0.09	0.07
	1.8	0.11	0.83	0.88	0.19	**0.95**	**0.95**	0.05	0.05
	1.9	0.10	0.59	0.68	0.10	0.84	**0.88**	0.05	0.04
Gauss		0.05	0.05	0.05	0.05	0.05	0.05	0.02	0.02
$n = 1000$									
$\mathcal{MG}(a_1)$	0.7	0.10	0.54	0.75	**0.06**	0.95	0.95	**0.06**	**0.06**
	0.75	0.13	0.57	0.80	**0.06**	0.95	0.95	**0.06**	**0.06**
	0.8	0.11	0.60	0.84	**0.05**	0.95	0.95	0.06	0.07
	0.85	0.10	0.59	0.86	0.06	0.95	0.95	**0.05**	0.06
	0.9	0.10	0.50	0.82	**0.05**	0.95	0.95	**0.05**	**0.05**
	0.95	0.09	0.34	0.65	0.05	0.89	0.92	0.05	**0.04**
$\mathcal{T}(\nu)$	1.5	0.40	0.93	0.91	0.53	**0.95**	**0.95**	0.88	0.25
	2	0.22	0.95	0.94	0.34	**0.95**	**0.95**	0.41	0.15
	3	0.12	0.91	0.94	0.18	0.95	0.95	0.08	**0.06**
	6	0.09	0.53	0.81	**0.06**	0.94	0.95	**0.06**	**0.06**
	10	0.10	0.22	0.39	**0.05**	0.83	0.89	**0.05**	**0.05**
	15	0.10	0.12	0.20	0.05	0.54	0.70	**0.04**	**0.04**
$\mathcal{S}(\alpha)$	1.5	0.30	0.93	0.90	0.51	**0.95**	**0.95**	0.88	0.26
	1.6	0.21	0.94	0.91	0.42	**0.95**	**0.95**	0.45	0.17
	1.7	0.16	0.94	0.93	0.33	**0.95**	**0.95**	0.10	0.10
	1.8	0.12	0.92	0.92	0.27	**0.94**	**0.94**	0.07	0.08
	1.9	0.10	0.77	0.87	0.19	**0.94**	**0.94**	0.05	0.07
Gauss		0.05	0.05	0.05	0.05	0.05	0.05	0.02	0.02

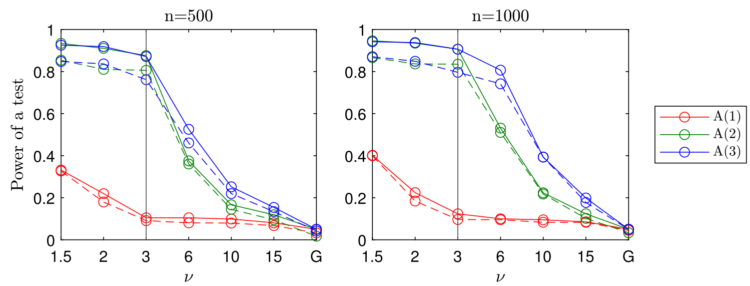

Figure 18 The comparison of power curves of the tests based on the *A*(*N*) statistics with two-sided (solid lines) and one-sided (dashed lines) acceptance regions. Power curves of the tests are obtained based on 1000 Monte Carlo simulations of samples from Student's t distribution with $n = 500, 1000$. The results are presented with respect to parameter ν responsible for heavy-tailed properties of the distribution. We assume $d = 5\%$ in case of both two-sided and one-sided acceptance regions

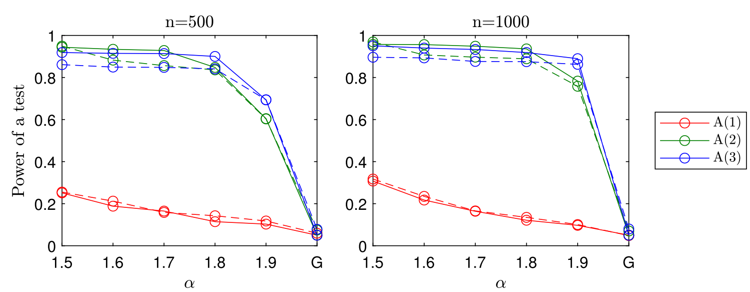

Figure 19 The comparison of power curves of the tests based on the *A*(*N*) statistics with two-sided (solid lines) and one-sided (dashed lines) acceptance regions. Power curves of the introduced tests are obtained based on 1000 Monte Carlo simulations of samples drawn from α-stable distributions with $n = 500, 1000$. The results are presented with respect to parameter α responsible for heavy-tailed properties of the distribution. We assume $d = 5\%$ in case of both two-sided and one-sided acceptance regions

Author contributions

KS developed and validated proposed statistical tests for finitness of the variance. RZ supported supervision from industrial and applicative perspective, and provided the real data from mining machines. AW supervised the mathematical part of this work. All authors read and approved the final manuscript.

Funding

The work of KS, RZ and AW is supported by the National Center of Science under Sheng2 project No. UMO-2021/40/Q/ST8/00024 "NonGauMech - New methods of processing non-stationary signals (identification, segmentation, extraction, modeling) with non-Gaussian characteristics for the purpose of monitoring complex mechanical structures".

Data availability

The real data is not available due to confidentiality agreement between the parties.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Faculty of Pure and Applied Mathematics, Wrocław University of Science and Technology, Wrocław, Poland. [2]Faculty of Geoengineering, Mining and Geology, Wrocław University of Science and Technology, Wrocław, Poland.

Received: 29 September 2023 Accepted: 10 August 2024 Published online: 27 August 2024

References

1. Barszcz T, Jabłoński A. A novel method for the optimal band selection for vibration signal demodulation and comparison with the kurtogram. Mech Syst Signal Process. 2011;25(1):431–51.

2. Barszcz T, Randall RB. Application of spectral kurtosis for detection of a tooth crack in the planetary gear of a wind turbine. Mech Syst Signal Process. 2009;23(4):1352–65.
3. Behboodian J. On the modes of a mixture of two normal distributions. Technometrics. 1970;12(1):131–9.
4. Brockwell PJ, Davis RA, Yang Y. Estimation for non-negative Lévy-driven CARMA processes. J Bus Econ Stat. 2011;29(2):250–9.
5. Burnecki K, Wyłomańska A, Beletskii A, Gonchar V, Chechkin A. Recognition of stable distribution with Lévy index α close to 2. Phys Rev E. 2012;85:056711.
6. Burnecki K, Wyłomańska A, Beletskii A, Gonchar V, Chechkin A. Recognition of stable distribution with Lévy index α close to 2. Phys Rev E. 2012;85:056711.
7. Burnecki K, Wyłomańska A, Chechkin A. Discriminating between light- and heavy-tailed distributions with limit theorem. PLoS ONE. 2015;10:e0145604.
8. Burnecki K, Wyłomańska A, Chechkin A. Discriminating between light- and heavy-tailed distributions with limit theorem. PLoS ONE. 2015;10(12):1–23.
9. Chakravarti I, Laha RG, Roy J. Handbook of methods of applied statistics, Volume I. Hoboken: Wiley; 1967.
10. Chakravarti IM, Laha R, Roy J. Handbook of methods of applied statistics. New York: Wiley; 1967.
11. Cizek P, Haerdle W, Weron R. Statistical tools for finance and insurance. Berlin: Springer; 2005.
12. Clauset A, Shalizi CR, Newman MEJ. Power-law distributions in empirical data. SIAM Rev. 2009;51(4):661–703.
13. Fedotenkov I. A bootstrap method to test for the existence of finite moments. J Nonparametr Stat. 2013;25(2):315–22.
14. Fedotenkov I. A simple nonparametric test for the existence of finite moments. Germany. 2015. MPRA Paper 66089.
15. Granger CWJ, Orr D. "Infinite variance" and research strategy in time series analysis. J Am Stat Assoc. 1972;67:275–85.
16. Hájek J, Šidák Z. Theory of rank tests. Probability and mathematical statistics. San Diego: Academic Press; 1967.
17. Hebda-Sobkowicz J, Nowicki J, Zimroz R, Wyłomańska A. Alternative measures of dependence for cyclic behaviour identification in the signal with impulsive noise—application to the local damage detection. Electronics. 2021;**10**(15).
18. Huillery J, Millioz F, Martin N. On the description of spectrogram probabilities with a chi-squared law. IEEE Trans Signal Process. 2008;56(6):2249–58.
19. Kruczek P, Zimroz R, Antoni J, Wyłomańska A. Generalized spectral coherence for cyclostationary signals with α-stable distribution. Mech Syst Signal Process. 2021;159:107737.
20. Kruczek P, Zimroz R, Wyłomańska A. How to detect the cyclostationarity in heavy-tailed distributed signals. Signal Process. 2020;172:107514.
21. Liu T, Qiu T, Luan S. Cyclic frequency estimation by compressed cyclic correntropy spectrum in impulsive noise. IEEE Signal Process Lett. 2019;26(6):888–92.
22. Liu Y, Zhang Y, Qiu T, Gao J, Na S. Improved time difference of arrival estimation algorithms for cyclostationary signals in alpha-stable impulsive noise. Digit Signal Process. 2018;76:94–105.
23. Maraj-Zygmąt K, Sikora G, Pitera M, Wyłomańska A. Goodness-of-fit test for stochastic processes using even empirical moments statistic. Chaos, Interdiscip J Nonlinear Sci. 2023;33(1):013128.
24. Mathai AM, Provost SB. Quadratic forms in random variables: theory and applications. New York: Dekker; 1992.
25. Mauricio A, Qi J, Smith W, Sarazin M, Randall R, Janssens K, Gryllias K. Bearing diagnostics under strong electromagnetic interference based on integrated spectral coherence. Mech Syst Signal Process. 2020;140:106673.
26. Ng WL, Yau CY. Test for the existence of finite moments via bootstrap. J Nonparametr Stat. 2018;30(1):28–48.
27. Nolan JP. Numerical calculation of stable densities and distribution functions. Commun Stat, Stoch Models. 1997;13(4):759–74.
28. Nuttall A. Some windows with very good sidelobe behavior. IEEE Trans Acoust Speech Signal Process. 1981;**29**(1).
29. Pitera M, Chechkin A, Wyłomańska A. Goodness-of-fit test for alpha-stable distribution based on the quantile conditional variance statistics. Stat Methods Appl. 2021. 387–424.
30. Randall RB, Antoni J, Chobsaard S. The relationship between spectral correlation and envelope analysis in the diagnostics of bearing faults and other cyclostationary machine signals. Mech Syst Signal Process. 2001;15(5):945–62.
31. Resnick S. Heavy-tail phenomena: probabilistic and statistical modeling. Springer series in operations research and financial engineering. New York: Springer; 2007.
32. Robertson CA, Fryer JG. Some descriptive properties of normal mixtures. Scand Actuar J. 1969;1969(3–4):137–46.
33. Samoradnitsky G, Taqqu MS. Stable non-Gaussian random processes: stochastic models with infinite variance. London: Routledge; 1994.
34. Skowronek K, Barszcz T, Antoni J, Zimroz R, Wyłomańska A. Assessment of background noise properties in time and time–frequency domains in the context of vibration-based local damage detection in real environment. Mech Syst Signal Process. 2023;199:110465.
35. Student. The probable error of a mean. Biometrika. 1908;6(1):1–25.
36. Trapani L. Testing for (in)finite moments. J Econom. 2016;191(1):57–68.
37. Welch BL. 'Student' and small sample theory. J Am Stat Assoc. 1958;53(284):777–88.
38. Weron A, Weron R. Computer simulation of Lévy α-stable variables and processes. In: Garbaczewski P, Wolf M, Weron A, editors. Chaos – the interplay between stochastic and deterministic behaviour. Berlin: Springer; 1995. p. 379–92.
39. Wyłomańska A, Iskander R, Burnecki K. Omnibus test for normality based on the Edgeworth expansion. PLoS ONE. 2020;15(6):e0233901.
40. Żuławiński W, Maraj-Zygmąt K, Shiri H, Wyłomańska A, Zimroz R. Framework for stochastic modelling of long-term non-homogeneous data with non-Gaussian characteristics for machine condition prognosis. Mech Syst Signal Process. 2023;184:109677.

Publisher's Note