DOI 10.1007/s10986-017-9350-7
Lithuanian Mathematical Journal, Vol. 57, No. 2, April, 2017, pp. 155–170

Goodness-of-fit tests based on the empirical characteristic function

Aleksej Bakshaev and Rimantas Rudzkis

Institute of Mathematics and Informatics, Vilnius University, Akademijos str. 4, LT-08663 Vilnius, Lithuania
(e-mail: aleksej.bakshaev@gmail.com; rimantas.rudzkis@mii.vu.lt)

Received May 5, 2016; revised February 17, 2017

Abstract. The paper is devoted to the supremum-type multivariate goodness-of-fit tests based on the empirical characteristic function. Particular attention is devoted to the composite hypothesis of normality and Gaussian distribution mixture model. An analytical way to approximate the null asymptotic distributions of the considered test statistics is discussed applying the theory of large excursions of differentiable Gaussian random fields. The produced comparative Monte Carlo power study shows that the considered tests are powerful competitors to the existing classical criteria, clearly dominating in verification of the goodness-of-fit hypotheses against the specific types of alternatives.

MSC: 62H15

Keywords: goodness-of-fit, empirical characteristic function

1 Introduction

Let $X^n = (X_1, \dots, X_n)$ be a sample of observations of a random vector X with unknown cumulative distribution function (cdf) $F(x)$ and probability density function (pdf) $f(x)$, $x \in \mathbb{R}^d$. In the classical statistical analysis of observations, there often arises a need to check whether the observations are taken from a certain distribution, that is, to test the goodness-of-fit hypothesis (H0) $f = f_0$, where f_0 is the prespecified expected distribution of X. Classical well-studied approaches to solve this problem are based on the empirical processes [1, 2, 3, 19, 20, 33], that is, the differences between the empirical and theoretical distribution functions. These procedures also have analogues related to empirical quantile processes and density estimates (see, e.g., [4, 5, 6, 28]). However, the most powerful (simultaneously for all the hypotheses of this type) test does not exist. Even among the classical tests there is no clear winner. Therefore, the creation of more powerful criteria for different types of densities f_0 and considered classes of alternative hypotheses remains nowadays a topical problem in statistics. Particular attention in this work is devoted to the hypothesis of normality and the mixture model of Gaussian distributions, which are widely applicable in classification problems. This work is a continuation of our research in [4, 5]. The null hypothesis assumes that f_0 is a Gaussian density or a mixture of a known number of Gaussian densities. The alternative assumes the existence of an additional small distribution cluster g, that is,

$$\text{(H1)}\ \ f = (1-\epsilon) f_0 + \epsilon g, \quad \epsilon \leqslant \frac{1}{2}.$$

In this paper, we suggest to apply supremum-type goodness-of-fit tests based on empirical characteristic functions (ecf) for testing the stated type of hypothesis. This idea is not new. Certain probabilistic properties of ecf are in detail discussed in [9], where ecf is also suggested to be a useful tool in verification of statistical hypotheses, for example, symmetry about the origin. During the last decades, applications of ecf attracted plenty of attention in statistical papers adjusting the methodology for different practical purpose. Regarding the goodness-of-fit tests, utilization of ecf became widely spread and popular primarily in the problems of testing the hypothesis of normality (see, e.g., [10, 11, 12, 13, 27, 34]), with the following extension of the methods to testing for Cauchy, Laplace, and other distributions [8, 18, 21, 22]. Supremum-type goodness-of-fit ecf criteria were also considered in [7, 26]. A comprehensive review of testing procedures based on ecf is presented in [14, 15, 25, 35, 36] and references therein.

Additional motivation for the usage of ecf in goodness-of-fit testing is related to the tests for stable distributions; see, for example, [25, 37]. Classical goodness-of-fit procedures such as the Kolmogorov–Smirnov test, the Cramer–von Mises test, etc. generally are not able to handle the stable distributions directly because of the lack of analytical pdf and cdf. Since stable distributions can be fully characterized by their characteristic functions, goodness-of-fit tests based on the ecf can be particularly useful.

The problem of analytical approximation of the null distribution of the proposed test statistics, and therefore establishment of the critical regions of the tests, is discussed in Section 4. The results are obtained using the theory of high excursions of Gaussian (and, in some sense, close to Gaussian) random fields developed in [30, 31]. Produced simulation study shows that the precision of the derived approximations is good enough even for small samples sizes and moderate test significance levels.

Specific algorithms proposed in this work are constructed using the results of both theoretical authors investigations in the theory of large excursions of random fields and produced comparative simulation analysis. One of the main results of this analysis, presented in Section 5, justifies the expediency of application of the described procedures. Provided comparative Monte Carlo power study shows that the considered supremum-type tests are powerful competitors to the existing classical criteria in verification of the normality hypothesis and detection of additional contaminating clusters in Gaussian distribution mixtures. Restriction to consideration of stated null and alternative distributions within the scope of this paper, however, does not limit the generality of the application of proposed test procedures to other goodness-of-fit hypotheses.

2 Statement of the problem. Simple hypothesis

Let $X_1, \dots, X_n$ be a sample of independent observations of a random vector X with an unknown pdf $f(x)$, $x \in \mathbb{R}^d$, $d \geqslant 1$. Based on the given sample, it is required to test a simple hypothesis of goodness-of-fit

$$\text{(H0)}\ \ f(x) = f_0(x), \tag{2.1}$$

where $f_0(x)$ is a prespecified pdf. In view of the one-to-one correspondence between density functions and characteristic functions, their Fourier transforms, the initial null hypothesis can be replaced with

$$\text{(H0)}\ \ c(x) = c_0(x),$$

where $c(t)$ and $c_0(t)$ are the characteristic functions corresponding to $f(t)$ and $f_0(t)$, respectively.

Particular attention in the work, especially in the simulation power study, is devoted to the complex alternative of the form

$$\text{(H1)}\ \ f(x) = (1-\epsilon) f_0(x) + \epsilon g(x), \tag{2.2}$$

where ϵ is a mixing probability, $0 < \epsilon \ll 1/2$, and $g(x)$ is an arbitrary distribution. Different forms of deviations of contaminating cluster g from f_0 are investigated. In particular, we consider differences in scale, for example, $\sigma_i^g \gg \sigma_i^{f_0}$, $i = 1, \dots, d$, where σ_i^f is the standard deviation of the ith component of a random vector with density function f. The form of the alternative is topical in some social and economic studies, for instance, the data clusterization problem in Gaussian mixture models.

Let $c_n(t)$ be the ecf based on the given sample and defined by

$$c_n(t) = \frac{1}{n}\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}t^\top X_j}. \tag{2.3}$$

Denote

$$\xi(t) = c_n(t) - c_0(t), \tag{2.4}$$

$$\xi_1(t) = \Re\bigl(\xi(t)\bigr), \qquad \xi_2(t) = \Im\bigl(\xi(t)\bigr), \tag{2.5}$$

where $\Re(c(t))$ and $\Im(c(t))$ are the real and imaginary parts of the complex-valued function $c(t)$, respectively. Taking into account the form of the alternative hypothesis of interest, the following functionals are considered as the test statistics for (H0):

$$\zeta_1 = \max_{\substack{i=1,2,\\ t\in I}} \left|\frac{\xi_i(t)}{\sigma_i(t)}\right|, \tag{2.6}$$

$$\zeta_2 = \max_{t\in I} \left|\frac{\xi_1(t)}{\sigma_1(t)}\right|, \tag{2.7}$$

$$\zeta_3 = \max_{t\in I} \bigl|\xi(t)\bigr|^2, \tag{2.8}$$

where $\sigma_i^2(t) = \mathrm{Var}(\xi_i(t))$, $i = 1, 2$, are defined in the case of null hypothesis, and I is a fixed d-dimensional interval without zero neighborhood.

Remark. Consideration of three forms of test statistics (2.7)–(2.8) is primarily justified by the fact that $c_0(t)$ is a complex-valued function, and therefore different forms of deviations of empirical and theoretical functions are investigated. Consequently, it is worth noting that that test statistic (2.7) based on the empirical function $\xi_1(t)$ is assumed to be applied for symmetric null distributions when the imaginary part of $c_0(t)$ is zero. In contrast to criteria ζ_i, $i = 1, 3$, the test ζ_2 is not consistent against all fixed alternatives; however, the test is practically useful and powerful relative to other criteria, as it is shown in the subsequent simulation study. Regarding the test (2.8), similar supremum-type statistics applied for testing multivariate normality were primarily considered in [26]. Essentially, ζ_3 seems weaker relative to other statistics and primarily considered in this study for comparison.

Naturally, we should reject the null hypothesis in the case of large values of the test statistics, that is, if $\zeta_i > z_\alpha$, where z_α can be found from the equation

$$\mathbf{P}_0(\zeta_i > z_\alpha) = \alpha, \tag{2.9}$$

where α is a significance level of the test.

3 Composite hypothesis. Normality tests

The modifications of the proposed statistics ζ_i, $i = 1, 2, 3$, for testing the composite hypothesis of goodness-of-fit, that is, based on the sample $X_1, \dots, X_n$ to test whether the unknown characteristic function $c(t)$ belongs to a certain parametric family, are straightforward. Since in this case, $c_0(t) = c_0(t, \theta)$, $\theta \in \Theta \subset \mathbb{R}^p$, and the variances $\sigma_i(t)$, $i = 1, 2$, depend on the unknown parameter θ, we have to replace it by $\widehat{\theta}_n$, where $\widehat{\theta}_n$ is a consistent estimator of θ, the true parameter value under (H0). Finally, the hypothesis can be verified by using statistics (2.7)–(2.8), where $\sigma_i^2(t) = \mathrm{Var}_{\widehat{\theta}_n} \xi_i(t)$, and the characteristic function $c_0(t)$ is replaced by $c_0(t, \widehat{\theta}_n)$ in the definition of the process (2.4).

However, in the general case, similarly to the classical Kolmogorov–Smirnov and Cramer–von Mises statistics with estimated parameters, the asymptotic distributions of ζ_i, $i = 1, 2, 3$, with $c_0(t, \widehat{\theta}_n)$ are not parameter-free, that is, depend on the value of unknown parameter θ. This makes the establishment of the critical region of the test complicated. Nevertheless, in some cases, this parametric dependence problem can be avoided.

Further, in the composite hypothesis case, we restrict to consideration of the hypothesis of normality, where the initial sample $X_1, \dots, X_n$ is first standardized by means of the sample covariance matrix $\widehat{S}_n$ and mean $\bar{X}$, that is,

$$Y_i = \widehat{S}_n^{-1/2}(X_i - \bar{X}), \quad i = 1, \dots, n. \tag{3.1}$$

For testing the hypothesis that X has a nondegenerate multivariate normal distribution, which means that $\widehat{S}_n$ is nonsingular with probability one, we consider the statistics (2.7)–(2.8), where ecf is constructed based on $Y_1, \dots, Y_n$, that is, $c_n(t) = (1/n)\sum_{j=1}^{n} \mathrm{e}^{\mathrm{i}t^\top Y_j}$ and $c_0(t) = \mathrm{e}^{-t^\top t/2}$. Since the joint distribution of the standardized sample $Y_1, \dots, Y_n$ does not depend on the initial distribution parameters of X, the critical regions of ζ_i, $i = 1, 2, 3$, can be obtained by means of Monte Carlo simulations.

4 Analytical approximation of the null distribution of the statistic ζ_i

In practice, the critical regions of the examined tests can be determined by means of Monte Carlo simulations. An alternative approach refers to the establishment of the asymptotic null distribution of the test statistics, which is the objective of this section. Further, we discuss the problem of analytical approximation of the null distribution of statistics (2.6)–(2.7). The problem is investigated using the theory of large excursions of Gaussian (and, in some sense, close to Gaussian) random fields introduced in [30, 31].

Remark. One can see that the empirical process $|\xi(t)|^2$ in the definitions of statistic (2.8) can be rewritten in the form of V-statistics with degenerate kernel

$$\begin{aligned}\left|\xi(t)\right|^2 = \frac{1}{n^2}\sum_{i,j=1}^{n}\Big[&\cos\big(t^\top(X_i - X_j)\big) - 2\Re\big(c_0(t)\big)\cos\big(t^\top X_i\big)\\ &- 2\Im\big(c_0(t)\big)\sin\big(t^\top X_i\big) + \left|c_0(t)\right|^2\Big].\end{aligned}$$

This implies that the finite-dimensional distributions of the process $|\xi(t)|^2$ are not Gaussian. Therefore, the following approximations of the null distribution of test statistics will be derived only for tests (2.6)–(2.7) based on the maxima of empirical random processes $\xi_1(t)$ and $\xi_2(t)$.

For calculation of the thresholds for the critical regions of the tests, we are concerned with the asymptotics of the probabilities

$$P_i(u) = \mathbf{P}_0\{\zeta_i < u\}, \quad n \to \infty,\ i = 1, 2, \tag{4.1}$$

representing the distribution function of ζ_i, $i = 1, 2$. The fact that $\xi_i(t)$, $i = 1, 2$, are close to the Gaussian random field, that is, the multivariate central limit theorem implies the asymptotical normality of the finite-dimensional distributions of $\xi_i(t)$, $i = 1, 2$, which suggests us to apply the mentioned results from the theory of high excursions of Gaussian fields to approximate the probabilities $P_i(u)$.

It has been shown in [30, 31] that if a differentiable (in the mean square sense) Gaussian random field $\{\eta(t),\ t \in T\}$ with $\mathbf{E}\eta(t) \equiv 0$, $\mathrm{Var}(\eta(t)) \equiv 1$, and continuous trajectories defined on a d-dimensional interval $T \subset \mathbb{R}^d$ satisfies certain smoothness and regularity conditions [30, Thm. 1], then $\mathbf{P}\{-v_1(t) < \eta(t) < v_2(t),\ t \in T\} \cong \mathrm{e}^{-Q}$ since for all $t \in T$, $v_1(t), v_2(t) > \chi$, $\chi \to \infty$, where $v_i(\cdot)$, $i = 1, 2$, are smooth enough functions, and Q is a certain constructive functional depending on v_1, v_2, T, and the matrix function $R(t) = \mathrm{cov}(\eta'(t), \eta'(t))$. Here $\eta'(t)$ is the gradient of $\eta(t)$.

The exact definition of the functional Q in the general case is presented in [4, 5, 30, 31]. In the case $d = 2$ and $v_1(t) = v_2(t) = u$, the functional $Q(u)$ has the form

$$Q(u) = \frac{1}{\pi}\big(1 - \Phi(u) + u\phi(u)\big) \int_I |R|^{1/2}\,\mathrm{d}t_1\,\mathrm{d}t_2 + \frac{\phi(u)}{\sqrt{2\pi}} \int_{I_1} \big(R_{1,1}^{1/2}(t_1, a_2) + R_{1,1}^{1/2}(t_1, b_2)\big)\,\mathrm{d}t_1 + \frac{\phi(u)}{\sqrt{2\pi}} \int_{I_2} \big(R_{2,2}^{1/2}(a_1, t_2) + R_{2,2}^{1/2}(b_1, t_2)\big)\,\mathrm{d}t_2,$$

where $t = (t_1, t_2) \in \mathbb{R}^2$, R is the covariance matrix with the elements $R_{k,l} = R_{k,l}(t_1, t_2)$, $k, l = 1, 2$, and $I = I_1 \times I_2 = [a_1, b_1] \times [a_2, b_2]$. Regarding the probabilistic interpretation of $Q(u)$, it equals to the average number of local maxima of the considered Gaussian random field above the level u within the interval I.

The stated result leads to the following approximation of the probabilities $P_i(u)$, $i = 1, 2$:

$$P_1(u) \cong \mathrm{e}^{-Q_1(u)-Q_2(u)} =: \widehat{P}_1(u), \tag{4.2}$$

$$P_2(u) \cong \mathrm{e}^{-Q_1(u)} =: \widehat{P}_2(u), \tag{4.3}$$

where Q_i depends on u, I, and the matrix function $R_i(t)$,

$$R_i(t) = \mathrm{Var}\left(\left(\frac{\xi_i(t)}{\sigma_i(t)}\right)'\right) = \frac{\mathrm{Var}(\xi_i'(t))}{\sigma_i^2(t)} - \frac{[(\sigma_i^2(t))'][(\sigma_i^2(t))']^\top}{4\sigma_i^4(t)}, \tag{4.4}$$

where $\sigma_i^2(t) = \mathrm{Var}(\xi_i(t))$, $i = 1, 2$.

Remark. Regarding the statistic ζ_1, we are interested in the large excursions of the bivariate random process, that is, the real and imaginary parts of the empirical random process $\xi(t)$. The asymptotic results stated in [30, 31] imply that, as $u \to \infty$, the events $\{\max_{t\in I} |\xi_i(t)/\sigma_i(t)| > u\}$, $i = 1, 2$, can be treated independently. This implies (4.2).

In the univariate case, the proposed approximations (4.3) and (4.2) can be improved by applying the large excursion results for Gaussian processes presented in [29]. In this case,

$$P_1(u) \cong \big[2\Phi(u) - 1\big]^2 \exp\left\{-\frac{\exp(-u^2/2)}{\pi} \int_I \big(\beta_1(t) + \beta_2(t)\big)\,\mathrm{d}t\right\},$$

$$P_2(u) \cong \big[2\Phi(u) - 1\big] \exp\left\{-\frac{\exp(-u^2/2)}{\pi} \int_I \beta_1(t)\,\mathrm{d}t\right\},$$

where $\beta_i^2(t) = \mathrm{Var}((\xi_i(t)/\sigma_i(t))')$, and Φ is the standard normal distribution function.

4.1 Simple goodness-of-fit hypothesis

Recall that, in this case, $X_1, \dots, X_n$ are iid random vectors, and $c_0(t)$ is a known hypothesized characteristic function.

Let us rewrite the empirical processes (2.5) in the form

$$\xi_1(t) = \frac{1}{n}\sum_i \cos\bigl(t^\top X_i\bigr) - \Re\bigl(c_0(t)\bigr), \qquad \xi_2(t) = \frac{1}{n}\sum_i \sin\bigl(t^\top X_i\bigr) - \Im\bigl(c_0(t)\bigr). \tag{4.5}$$

Further, from the above representation of $\xi_i(t)$, $i = 1, 2$, by means of straightforward but rather lengthy calculations we obtain the following exact expressions for computation of the corresponding covariance matrices $R_i(t)$ defined in (4.4):

$$\mathrm{Var}\,\bigl(\xi_1'(t)\bigr) = \frac{1}{2n}\bigl(\mathbf{E}XX^\top + \Re\bigl(c_0''(2t)\bigr)\bigr) - \frac{1}{n}\bigl(\Re\bigl(c_0'(t)\bigr)\bigr)\bigl(\Re\bigl(c_0'(t)\bigr)\bigr)^\top, \tag{4.6}$$

$$\sigma_1^2(t) = \frac{1}{n}\biggl[\frac{\Re(c_0(2t)) + 1}{2} - \bigl(\Re\bigl(c_0(t)\bigr)\bigr)^2\biggr], \tag{4.7}$$

$$\mathrm{Var}(\xi_2'(t)) = \frac{1}{2n}\bigl(\mathbf{E}XX^\top - \Re\bigl(c_0''(2t)\bigr)\bigr) - \frac{1}{n}\bigl(\Im\bigl(c_0'(t)\bigr)\bigr)\bigl(\Im\bigl(c_0'(t)\bigr)\bigr)^\top, \tag{4.8}$$

$$\sigma_2^2(t) = \frac{1}{n}\biggl[\frac{1 - \Re(c_0(2t))}{2} - \bigl(\Im\bigl(c_0(t)\bigr)\bigr)^2\biggr], \tag{4.9}$$

where X is a random vector with charactersitic function $c_0(t)$.

4.2 Composite hypothesis of normality

As discussed in Section 3, in this case, statistics (2.6)–(2.7) for testing the normality are constructed based on the standardized sample $Y_1, \dots, Y_n$. Unfortunately, the produced transformation (3.1) leads to the dependency among the observations within the sample, which makes a direct application of the mentioned large excursions results and therefore calculation of the critical regions of ζ_i, $i = 1, 2$, complicated. These difficulties can be overcome using the approach suggested in the proof of Theorem 2.1 in [12] considering the approximations for the processes $\xi_i(t)$, $i = 1, 2$, constructed based on the sample of independent observations.

Let $X_1, \dots, X_n$ be independent standard normal random vectors. Define the auxiliary process

$$\widehat{\xi}_1(t) = \frac{1}{n}\sum_{j=1}^{n}\biggl[\cos\bigl(t^\top X_j\bigr) - \mathrm{e}^{-t^\top t/2} + \mathrm{e}^{-t^\top t/2}\frac{(t^\top X_j)^2 - t^\top t}{2}\biggr], \tag{4.10}$$

$$\widehat{\xi}_2(t) = \frac{1}{n}\sum_{j=1}^{n}\bigl[\sin\bigl(t^\top X_j\bigr) - t^\top X_j \mathrm{e}^{-t^\top t/2}\bigr]. \tag{4.11}$$

In a way analogous to that in [12], we can obtain that, for any fixed interval I,

$$\sup_{t\in I}\bigl|\xi_i(t) - \widehat{\xi}_i(t)\bigr| \xrightarrow{\mathrm{P}} 0, \quad n \to \infty,\ i = 1, 2,$$

where $\xi_i(t)$, $i = 1, 2$, are previously defined random fields constructed based on the sample $Y_i = \widehat{S}_n^{-1/2}(X_i – \bar{X})$, $i = 1, \dots, n$.

This fact gives us a possibility to approximate the probabilities (4.1) of large excursions of the random fields $\xi_i(t)$, $i = 1, 2$, by the excursions of the fields $\widehat{\xi}_i(t)$, $i = 1, 2$, based on the sample $X_1, \dots, X_n$ of independent observations. Moreover, difficulties with calculation of the variances $\sigma_i^2(t)$, $i = 1, 2$, in the definitions of (2.6)–(2.7) can be solved by replacing them with $\widehat{\sigma}_i^2(t) = \mathrm{Var}\,\widehat{\xi}_i(t)$, $i = 1, 2$.

Finally, to obtain the constructive way for calculation of the large excursion probabilities of the processes $\widehat{\xi}_i(t)$ and therefore null distribution approximations (4.3), (4.2) of ζ_i, $i = 1, 2$, we further provide exact expressions for computation of the covariance matrices $\widehat{R}_i(x) = \mathrm{Var}(\widehat{\xi}_i(t)/\widehat{\sigma}_i(t))'$ (4.4) in this case:

$$\begin{aligned}
\mathrm{Var}\,\big(\widehat{\xi}_1'(t)\big) &= \frac{1}{n}\bigg[\frac{1}{2}\big(I - \mathrm{e}^{-2t^\top t}\big(I - 4tt^\top\big)\big) - \frac{1}{2}\big(t^\top t\big)^2 \mathrm{e}^{-t^\top t} tt^\top \\
&\quad - \big(t^\top t\big)\mathrm{e}^{-t^\top t} I + 2\big(t^\top t\big)\mathrm{e}^{-t^\top t} tt^\top - 2\mathrm{e}^{-t^\top t} tt^\top\bigg], \\
\widehat{\sigma}_1^2(t) &= \frac{1}{2n}\big[\mathrm{e}^{-2t^\top t} - 2\mathrm{e}^{-t^\top t} - \big(t^\top t\big)^2 \mathrm{e}^{-t^\top t} + 1\big],
\end{aligned} \tag{4.12}$$

$$\begin{aligned}
\mathrm{Var}\big(\widehat{\xi}_2'(t)\big) &= \frac{1}{n}\bigg[\frac{1}{2}\big(I + \mathrm{e}^{-2t^\top t}\big(I - 4tt^\top\big)\big) - \big(t^\top t\big)\mathrm{e}^{-t^\top t} tt^\top - \mathrm{e}^{-t^\top t} I + 2\mathrm{e}^{-t^\top t} tt^\top\bigg], \\
\widehat{\sigma}_2^2(t) &= \frac{1}{2n}\big[1 - \mathrm{e}^{-2t^\top t} - 2\big(t^\top t\big)\mathrm{e}^{-t^\top t}\big].
\end{aligned} \tag{4.13}$$

4.3 Accuracy of proposed approximations

For a graphical assessment of the precision of the proposed approximations, we performed a simulation study where the empirical distribution functions of statistics ζ_i, $i = 1, 2$, were compared with the corresponding asymptotic distribution functions (4.2) and (4.3) in the cases $d = 1, 2$. Simple and composite hypotheses of normality were investigated. The empirical distributions of ζ_i, $i = 1, 2$, were obtained by generating 5000 samples of sizes 20–1000 from the standard normal distribution. In the case of composite hypothesis, the considered statistics were calculated using the procedure described in Section 3. Different variants of intervals I in the definitions of considered statistics were investigated.

The experimental results show that, in the general case, the precision of obtained approximations strongly depends on the size of the available sample and chosen maximization interval I, especially its length and deviation from zero. In the univariate case, for a simple goodness-of-fit hypothesis, obtained approximations for the distributions of statistics ζ_i, $i = 1, 2$, are sufficiently precise even for small sample sizes $n \geqslant 20$ and moderate significance levels $\alpha \leqslant 0.2$ for arbitrary intervals I without zero neighborhood. However, a similar precision level for the same sample sizes in the case of composite hypothesis could be obtained only for $I = [a, b]$, $a \geqslant 1$. Achieving the same accuracy level dealing with the closer to zero intervals I requires the increase of sample size. For example, if $I = [0.5, 2]$, then the sample size should be $n \geqslant 500$. A similar situation could be also seen in the case $d = 2$, where in general an adequate approximation precision was obtained for larger sample sizes relative to univariate case, that is, $n \geqslant 50, 100$. Some simulation results for the statistics ζ_i, $i = 1, 2$, composite hypothesis of normality, $d = 1, 2$, and sample sizes $n = 25, 50$ are presented in Fig. 1.

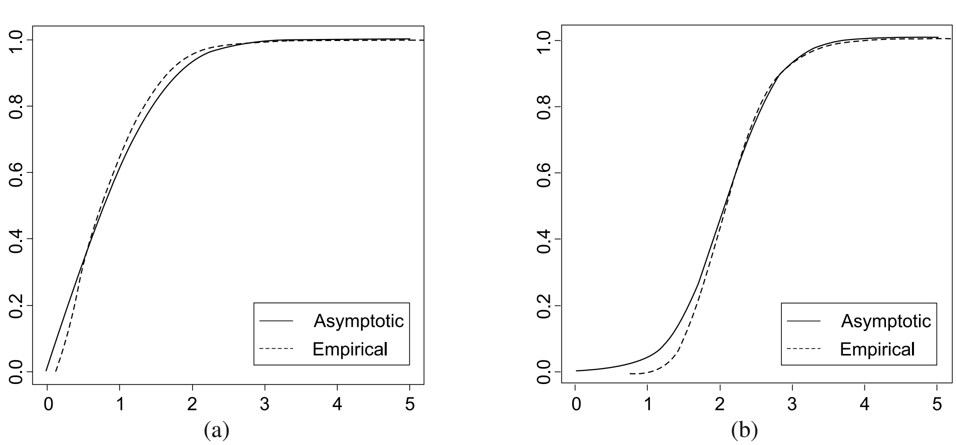


Figure 1. Composite hypothesis of normality, empirical and asymptotic distributions of ζ_i, $i = 1, 2$: (a) ζ_2, $d = 1$, $n = 25$, $I = [1, 2]$; (b) ζ_1, $d = 2$, $n = 50$, $I = [1, 2] \times [-2, 2]$.

4.4 Considered tests application algorithm

We further provide an algorithm for application of the proposed statistics in practice, where the critical region of the tests is established using the obtained approximations. For simplicity, we consider the simple hypothesis case. The modification of the procedure for the case of composite hypothesis of normality is straightforward and requires the replacement of the further stated formulas by their analogues from Section 4.2.

1. For a given sample $X_1, \dots, X_n$, evaluate the statistic of interest by calculating the maximum with respect to $t \in I$ using a small enough partition of the interval I. The processes $\xi_i(t)/\sigma_i(t)$, $i = 1, 2$, are computed by means of formulas (4.5), (4.7), and (4.9).
2. For a prespecified significance level α, determine the critical test level c_α using the equation $\widehat{P}_i(c_\alpha) = 1 - \alpha$, $i = 1, 2$, where $\widehat{P}_i$ are defined in (4.2) and (4.3), respectively.
3. Reject the null hypothesis if the calculated value of statistics is greater than c_α.

5 Simulation study

To evaluate the relative efficiency of the proposed criteria, a simulation power study was performed, where analysed tests (2.7)–(2.8) were compared with some popular criteria in the following problems:

- Verification of the composite hypothesis of normality in the cases $d = 1, 2, 5$.
- A simple goodness-of-fit hypothesis, where the null distribution is a mixture of two Gaussian densities. The following cases were investigated:
  - (H0) $f_0 = 0.6N(0, 1) + 0.4N(0, 3)$,
  - (H0) $f_0 = 0.6N(0, 3) + 0.4N(0, 1)$.

The following classical general and specific normality criteria for comparison were considered:

- Tests based on empirical distribution function: Anderson–Darling (AD) [1,2], Cramer–von Mises (CM) [3,24], Kolmogorov–Smirnov (KS) [20,33], and Mahalanobis[1] ($d > 1$);
- Tests based on distribution skewness and kurtosis: Jarque–Bera (JB, $d = 1$) [16,17] and Mardia ($d > 1$) [23];
- L_2-type ecf test: BHEP [11,13,27];
- Regression and correlation tests: Shapiro–Wilk (SW, $d = 1$) [32].

The tests were studied for a wide range of alternative distributions. Particular attention was devoted to the mixtures of Gaussian distributions (with different location and scale parameters) and mixtures of Gaussian and non-Gaussian symmetric, asymmetric, as well as short- and long-tailed distributions. Recall that the power of a statistical test is the probability that the test will reject the null hypothesis when the alternative hypothesis is true.

In all the cases, the behavior of the above-mentioned tests was investigated for sample sizes $n = 20, 50, 100, 200, 500$[2] and the significance level $\alpha = 0.05$. The simulations were carried out in R. A brief description and specifications for calculation of the considered classical test statistics are presented in [5] or the corresponding references. In the univariate case, the Kolmogorov–Smirnov, Anderson–Darling, Cramer–von Mises, and Shapiro–Wilk criteria were applied using the corresponding procedures from the statistical package *nortest*. Regarding the considered ecf statistics ζ_i, $i = 1, 2, 3$, they were calculated applying the procedure described in Section 4.4. In the case of composite hypothesis of normality, the unknown parameters were

[1] The test is based on the Mahalanobis transformation of initial sample $X_1, \dots, X_n$ using the sample mean and covariance matrix, i.e., $Y_i = (X_i - \bar{X})^\top \widehat{S}^{-1}(X_i - \bar{X})$, $i = 1, \dots, n$. Transformed univariate sample has chi-squared limit distribution with d degrees of freedom. After that, the null hypothesis should be rejected in the case of large values of one-sample Kolmogorov–Smirnov statistic.

[2] To shorten the presentation of the comparative study, in the tables below, the simulation results are provided only for sample sizes $n = 20, 100, 500$.

calculated using the ML-estimates of the mean and covariance matrix parameters. Since, according to the methodology in this case, the initial sample $X_1, \dots, X_n$ is first standardized by means of (3.1), which leads to the sample $Y_1, \dots, Y_n$ with dependent observations, the variances $\sigma_i^2(t)$, $i = 1, 2$, in the definitions of statistics ζ_i, $i = 1, 2$, were replaced by $\widehat{\sigma}_i^2(t)$, $i = 1, 2$, calculated using formulas (4.12) and (4.13).

The maxima with respect to t in all proposed test statistics ζ_i, $i = 1, 2, 3$, were calculated using the following intervals without zero neighborhood:

$$I = [-2, 2]^k \setminus [-0.05, 0.05]^k \quad (d = k,\ k = 1, 2, 5).$$

Remark. Empirically investigated dependence of the power of ecf tests on the choice of the interval I shows that the extension of the internal I outside the outer thresholds does not improve the results of the tests; however, it increases the calculation time of the considered statistics.

For unification of the calculation approaches and comparability of the results, the critical regions of all the investigated tests were established using the finite sample null distribution of the corresponding test statistics obtained by the Monte Carlo method applying the following procedure:

1. Generate repeatedly i.i.d. random samples $X_1, \dots, X_n$ from the null distribution. In our study, 5000 samples were simulated.
2. For each sample, evaluate all the statistics under consideration, that is, calculate the maxima with respect to $t \in I$ of the corresponding processes $\xi_i(t)/\sigma_i(t)$, $i = 1, 2$, and $|\xi(t)|^2$. The processes $\xi_i(t)$, $i = 1, 2$, were calculated using formulas (4.5), where the corresponding variances σ_i, $i = 1, 2$, were obtained from (4.7), (4.9), (4.12), (4.13), respectively.
3. For each statistic, calculate the empirical distribution function $F_n(x)$ on the basis of previously computed values.
4. For each statistic and chosen significance level $\alpha > 0$, find c_α from the equation

$$c_\alpha = \inf\{c_\alpha\colon F_n(c_\alpha) > 1 - \alpha\}.$$

The power of the tests was estimated by the proportion of the 5000 samples of alternative distributions for which the values of the corresponding statistics fall into the critical region of the tests.

5.1 Simulation results

The main simulation results of the power study are summarized in Tables 1–8. Tables 1–6 present the empirical powers, that is, the percentage of the rejected null hypothesis in the case of composite hypothesis of normality with $d = 1, 2, 5$, whereas Tables 7 and 8 contain the results for the Gaussian mixture model.

5.1.1 Normality hypothesis

It can be observed from the simulations that, in the univariate case, the best ecf tests behave similarly to the most powerful classical tests chosen for comparison.

The tests ζ_i, $i = 1, 2$, were slightly more powerful than the majority of other criteria testing the null hypothesis against the scale contaminated mixtures of normal distributions (see Table 1), where the variance of the contaminating distribution was significantly greater than the variance of the main one. This is especially seen in the case of unbalanced distribution mixtures with different mixing probabilities and can be explained by the usage of the uniform metric as the loss function for ecf in (2.6) and (2.7). Among the other criteria close to ζ_i, $i = 1, 2$, results were shown by the Jarque–Bera and Shapiro–Wilk tests; see, for example, Fig. 2(a).

Investigating the alternatives in the form of the mixtures of Gaussian and non-Gaussian distributions (see Table 2), the best results among ecf tests were shown by the criteria ζ_i, $i = 1, 2$, except for the case (H1) $(1-\epsilon) \times N(0, 1) + \epsilon U(0, 1)$. Their empirical power was significantly better, especially for moderate and large sample sizes $n \geqslant 100$, than the power of the majority of classical criteria, comparable only to the Jarque–Bera and Shapiro–Wilk tests results.

Table 1. Normality hypothesis, $d = 1$. Power against scale contaminated Gaussian distribution mixtures. Percentage of the rejected (H0), $\sigma_1 \sim U(1,5)$, $\sigma_2 \sim U(1,10)$, $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	JB	SW
$0.5N(0,1) + 0.5N(0,\sigma_1)$	20	28	28	28	30	29	24	29	27	25
	100	71	73	68	67	68	62	68	64	67
	500	86	86	83	82	83	79	82	85	83
$(1-\epsilon_1)N(0,1) + \epsilon_1 N(0,\sigma_1)$	20	35	37	33	31	32	26	33	37	33
	100	70	71	62	61	63	56	61	69	66
	500	85	85	78	77	79	75	77	85	84
$(1-\epsilon_2)N(0,1) + \epsilon_2 N(0,\sigma_1)$	20	21	21	17	17	18	14	18	23	21
	100	50	52	35	36	40	31	36	54	49
	500	74	75	58	58	61	54	56	77	73
$(1-\epsilon_1)N(0,1) + \epsilon_1 N(0,\sigma_2)$	20	57	58	56	57	58	53	57	55	56
	100	85	86	80	80	81	77	80	86	84
	500	91	91	89	88	89	87	87	91	91

Table 2. Normality hypothesis, $d = 1$. Power against non-Gaussian mixtures. Percentage of the rejected (H0), $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	JB	SW
$(1-\epsilon_1)N(0,1) + \epsilon_1 Logistic(0,1)$	20	17	18	14	12	13	10	15	18	15
	100	46	50	25	25	29	18	25	50	41
	500	87	88	87	67	73	56	64	89	86
$(1-\epsilon_1)N(0,1) + \epsilon_1 t(10)^*$	20	7	6	6	6	5	5	6	7	6
	100	12	13	6	6	7	6	6	15	11
	500	24	23	21	9	10	7	8	28	23
$(1-\epsilon_1)N(0,1) + \epsilon_1 t(4)$	20	13	14	12	11	12	9	12	15	13
	100	35	36	17	20	23	16	19	38	32
	500	75	76	73	49	55	39	43	78	74
$(1-\epsilon_1)N(0,1) + \epsilon_1 U(0,1)^\dagger$	20	18	12	18	18	18	15	19	14	17
	100	49	31	52	56	55	53	55	39	48
	500	70	57	75	81	80	78	80	67	76
$(1-\epsilon_2)N(0,1) + \epsilon_2 \chi^2(2)^\ddagger$	20	17	17	15	14	15	12	16	19	17
	100	48	49	28	33	37	27	30	51	48
	500	82	83	82	68	73	62	62	86	85
$(1-\epsilon_2)N(0,1) + \epsilon_2 Cauchy(0,1)$	20	20	19	18	18	18	17	19	20	19
	100	54	55	41	43	46	41	43	56	53
	500	87	87	87	80	81	77	77	88	88

Considering scale contaminated Gaussian distribution mixtures as an alternative in multivariate cases, that is, $d = 2, 5$ (Tables 3, 5), two different cases of independent and dependent marginal components were investigated. In the first case, that is, (H1) $(1-\epsilon)N(0, I_d) + \epsilon N(0, \sigma I_d)$, the tests ζ_i, $i = 1, 2$, were better than the majority of classical criteria, except the Mardia test with much the same results. Here I_d is the $d \times d$ unit matrix. Dealing with dependent marginal components, that is, $(1-\epsilon)N(0, I_d) + \epsilon N(0, \sigma R)$, $R \neq I_d$, we observed an absolute superiority of the tests ζ_i, $i = 1, 2$, which is especially evident for $n \geqslant 50$. The behavior of the mentioned tests in this case is characterized by an obvious tendency. We observe the increasing superiority of the tests if the correlation between the components of the multivariate distributions becomes greater. The graphical comparison of the power functions of ζ_i, $i = 1, 2, 3$, and the Mardia tests for different sample sizes is presented in Fig. 2(b).

* $t(k)$ denotes the Student t-distribution with k degrees of freedom.
† $U(a, b)$ denotes the uniform distribution in the interval $[a, b]$.
‡ $\chi^2(k)$ denotes the chi-squared distribution with k degrees of freedom.

Table 3. Normality hypothesis, $d = 2$. Power against scale contaminated Gaussian distribution mixtures. Percentage of the rejected (H0), $\sigma_1 \sim U(1,5)$, $\sigma_2 \sim U(1,10)$, $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	Mardia	Mahalanobis
$0.5N(0, I_d) + 0.5N(0, \sigma_1 I_d)$	20	41	39	34	41	36	33	48	37	44
	100	79	79	74	75	74	72	78	79	80
	500	88	89	85	86	85	84	86	89	89
$(1-\epsilon_1)N(0, I_d) + \epsilon_1 N(0, \sigma_1 I_d)$	20	48	46	37	39	38	31	46	47	41
	100	75	75	64	67	68	64	70	78	72
	500	87	88	81	83	83	81	82	90	86
$(1-\epsilon_1)N(0, I_d) + \epsilon_1 N(0, \sigma_1 R_1)^\dagger$	20	13	13	8	10	8	8	12	16	8
	100	44	44	24	23	21	21	30	42	30
	500	74	75	60	55	52	52	60	72	61
$(1-\epsilon_1)N(0, I_d) + \epsilon_1 N(0, \sigma_1 R_2)^\dagger$	20	17	15	10	9	8	7	11	15	7
	100	45	48	29	21	19	18	30	38	24
	500	80	82	66	54	54	51	64	75	61
$(1-\epsilon_1)N(0, I_d) + \epsilon_1 N(0, \sigma_1 R_3)^\dagger$	20	17	15	10	9	8	8	11	15	8
	100	55	57	36	20	23	17	34	41	27
	500	89	90	79	56	69	47	77	82	75
$(1-\epsilon_2)N(0, I_d) + \epsilon_2 N(0, \sigma_1 I_d)$	20	23	25	18	20	19	13	24	28	17
	100	57	57	38	43	45	39	44	62	48
	500	73	74	58	62	63	60	61	79	68
$(1-\epsilon_1)N(0, I_d) + \epsilon_1 N(0, \sigma_2 I_d)$	20	68	69	65	66	63	45	71	69	66
	100	90	90	86	87	87	84	87	92	89
	500	95	95	91	92	92	91	92	95	93

Table 4. Normality hypothesis, $d = 2$. Power against non-Gaussian mixtures. Percentage of the rejected (H0), $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	Mardia	Mahalanobis
$(1-\epsilon_1)N(0, I_d) + \epsilon_1[\mathit{Logistic}(0,1)]^{2\ddagger}$	20	16	21	12	14	15	13	18	22	13
	100	57	62	22	35	30	27	39	66	51
	500	91	92	76	80	81	75	81	95	88
$(1-\epsilon_1)N(0, I_d) + \epsilon_1[t(10)]^2$	20	7	6	6	7	7	7	6	5	5
	100	11	11	6	7	6	8	7	14	9
	500	21	22	10	12	12	11	10	34	16
$(1-\epsilon_1)N(0, I_d) + \epsilon_1[t(4)]^2$	20	14	15	11	13	13	12	14	15	10
	100	40	44	17	25	24	20	27	50	32
	500	76	77	49	57	60	50	57	86	70
$(1-\epsilon_1)N(0, I_d) + \epsilon_1[U(0,1)]^2$	20	22	17	21	27	25	18	33	17	15
	100	59	46	60	64	59	56	70	50	49
	500	83	74	84	86	85	84	89	77	74
$(1-\epsilon_2)N(0, I_d) + \epsilon_2[\chi^2(2)]^2$	20	21	22	15	16	16	14	19	24	11
	100	59	60	29	41	37	35	39	67	43
	500	88	89	74	79	78	77	75	93	82
$(1-\epsilon_2)N(0, I_d) + \epsilon_2[\mathit{Cauchy}(0,1)]^2$	20	24	26	20	22	23	21	25	28	18
	100	67	68	53	59	60	56	59	72	62
	500	92	92	87	89	89	87	88	93	88

Exploring the alternatives, where the null distribution is contaminated with some non-Gaussian cluster (Tables 4, 6), the relative results of the ecf tests in this case on average reflect the corresponding situation when $d = 1$: the tests ζ_i, $i = 1, 2$, were better than most of the classical criteria but slightly inferior to the performance of the Mardia test, the multivariate analogue of the Jarque–Bera univariate normality criterion. It is worth noting that for additionally examined alternatives in the form of Gaussian distribution mixtures with different location parameters, the dominance of the ecf tests was not so convincing, except probably for

Table 5. Normality hypothesis, $d = 5$. Power against scale contaminated Gaussian distribution mixtures. Percentage of the rejected (H0), $\sigma_1 \sim U(1,5)$, $\sigma_2 \sim U(1,10)$, $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	Mardia	Mahalanobis
$0.5N(0,I_d)+0.5N(0,\sigma_1 I_d)$	20	40	39	39	35	30	20	71	72	62
	100	83	84	76	78	75	65	84	87	87
	500	91	92	87	88	85	87	92	93	94
$(1-\epsilon_1)N(0,I_d)+\epsilon_1 N(0,\sigma_1 I_d)$	20	51	45	33	28	27	20	58	68	47
	100	82	83	68	72	71	57	78	85	82
	500	89	90	80	83	84	83	86	92	90
$(1-\epsilon_1)N(0,I_d)+\epsilon_1 N(0,\sigma_1 R_1)^\dagger$	20	11	10	6	6	5	5	12	20	6
	100	58	60	31	23	22	19	41	51	37
	500	86	87	73	55	51	46	69	73	71
$(1-\epsilon_1)N(0,I_d)+\epsilon_1 N(0,\sigma_1 R_2)^\dagger$	20	11	9	7	6	6	6	10	14	6
	100	65	67	38	19	17	15	34	39	19
	500	93	93	83	54	54	40	81	70	72
$(1-\epsilon_1)N(0,I_d)+\epsilon_1 N(0,\sigma_1 R_3)^\dagger$	20	12	8	6	7	7	5	13	19	5
	100	73	74	46	22	24	16	49	48	32
	500	95	95	88	66	73	45	88	85	87
$(1-\epsilon_2)N(0,I_d)+\epsilon_2 N(0,\sigma_1 I_d)$	20	32	29	10	8	7	6	21	43	11
	100	70	71	38	47	50	29	53	73	63
	500	78	78	59	62	63	61	68	83	77
$(1-\epsilon_1)N(0,I_d)+\epsilon_1 N(0,\sigma_2 I_d)$	20	74	70	62	52	50	18	77	85	66
	100	90	90	85	85	85	73	89	91	90
	500	96	96	91	93	93	92	94	96	96

Table 6. Normality hypothesis, $d = 5$. Power against non-Gaussian mixtures. Percentage of the rejected (H0), $\epsilon_1 \sim U(0,0.5)$, $\epsilon_2 \sim U(0,0.1)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP	Mardia	Mahalanobis
$(1-\epsilon_1)N(0,I_d)+\epsilon_1[\mathit{Logistic}(0,1)]^{5\ddagger}$	20	16	13	9	8	8	7	19	36	9
	100	79	81	21	39	36	27	64	88	77
	500	94	95	73	79	80	78	87	96	95
$(1-\epsilon_1)N(0,I_d)+\epsilon_1[t(10)]^5$	20	6	5	7	5	5	5	5	8	5
	100	13	13	5	8	7	8	8	16	6
	500	23	23	6	7	8	9	10	39	25
$(1-\epsilon_1)N(0,I_d)+\epsilon_1[t(4)]^5$	20	13	11	6	5	5	6	11	22	7
	100	57	59	13	22	21	17	34	66	45
	500	85	85	41	53	55	51	66	89	83
$(1-\epsilon_1)N(0,I_d)+\epsilon_1[U(0,1)]^5$	20	17	13	25	17	18	10	62	40	25
	100	71	64	73	68	64	45	86	65	69
	500	88	85	89	84	81	83	95	83	86
$(1-\epsilon_2)N(0,I_d)+\epsilon_2[\chi^2(2)]^5$	20	24	22	8	8	6	6	17	35	7
	100	81	80	40	44	39	24	56	84	59
	500	96	96	83	83	81	75	86	97	90
$(1-\epsilon_2)N(0,I_d)+\epsilon_2[\mathit{Cauchy}(0,1)]^5$	20	40	39	21	13	15	8	29	44	12
	100	86	87	73	73	73	51	78	87	81
	500	97	97	93	94	93	92	94	97	96

the case of balanced mixtures, that is, (H1) $0.5N(0,I_d)+0.5N(a,I_d)$, especially for small and moderate sample sizes. In the general case, dealing with arbitrary distribution mixtures, their results were similar to the performance of the best examined classical criteria.

† $R_i = (1-\alpha_i)I_d + \alpha_i J_d$, $i = 1,2,3$, where I_d is the $d \times d$ unit matrix, J_d is the $d \times d$ all-ones matrix, and $\alpha_1 = 0.5$, $\alpha_2 = 0.75$, $\alpha_3 = 0.9$.

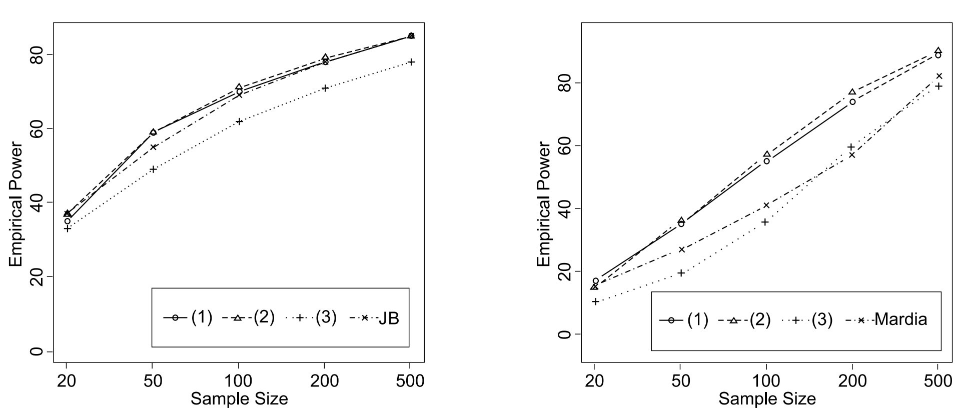
‡ $[P]^k := P \times \cdots \times P$ denotes the k-variate distribution having independent marginals P.

Table 7. Gaussian mixture model. Power against location and scale contaminated Gaussian distribution mixtures. Percentage of the rejected (H0) $f_0 = 0.6N(0,1) + 0.4N(0,3)$, $a \sim U(1,10)$, $\sigma_1 \sim U(1,5)$, $\sigma_2 \sim U(1,10)$, $\sigma_3 \sim U(0,1)$, $\epsilon \sim U(0,0.2)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP
$(1-\epsilon)f_0 + \epsilon N(a,1)$	20	33	34	12	12	25	9	14
	100	60	58	43	46	59	40	44
	500	79	76	69	71	78	70	67
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_1)$	20	9	8	7	6	7	5	6
	100	14	16	9	7	10	5	6
	500	35	41	20	14	24	12	24
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_2)$	20	25	26	7	7	12	6	7
	100	42	45	12	9	24	7	13
	500	63	67	38	29	49	27	43
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_3)$	20	5	5	6	5	5	5	7
	100	14	18	16	8	8	9	14
	500	41	50	39	29	30	30	40

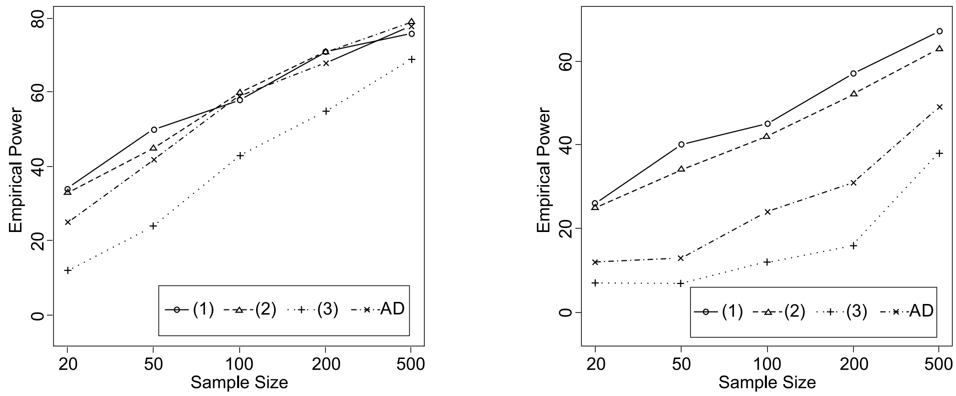
Table 8. Gaussian mixture model. Power against location and scale contaminated Gaussian distribution mixtures. Percentage of the rejected (H0) $f_0 = 0.6N(0,3) + 0.4N(0,1)$, $\sigma_1 \sim U(1,5)$, $\sigma_2 \sim U(1,10)$, $\sigma_3 \sim U(0,1)$, $\epsilon \sim U(0,0.2)$.

Alternative	n	ζ_1	ζ_2	ζ_3	CvM	AD	KS	BHEP
$(1-\epsilon)f_0 + \epsilon N(a,1)$	20	25	28	8	13	21	11	14
	100	55	53	36	40	54	37	45
	500	80	75	71	73	78	72	73
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_1)$	20	6	9	5	5	5	5	7
	100	10	12	7	6	7	5	10
	500	24	30	13	11	16	8	28
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_2)$	20	18	21	6	5	9	5	6
	100	39	40	11	9	23	7	20
	500	61	63	31	26	45	21	53
$(1-\epsilon)f_0 + \epsilon N(0,\sigma_3)$	20	5	6	5	5	5	5	5
	100	13	19	16	7	7	9	4
	500	50	57	48	33	35	34	11



(a) $d = 1$, (H1) $(1-\epsilon)N(0,1) + \epsilon N(0,\sigma)$ (b) $d = 2$, (H1) $(1-\epsilon)N(0,I_d) + \epsilon N(0,\sigma R_3)$

Figure 2. Empirical power of the tests, where $(i) \sim \zeta_i$, $i = 1,2,3$. Composite hypothesis of normality, $a \sim U(1,5)$, $\sigma \sim U(1,5)$, $\epsilon \sim U(0,0.5)$, $R_3 = \binom{1\ \ 0.9}{0.9\ \ 1}$.



(a) $(1-\epsilon)f_0+\epsilon N(a,1)$ (b) $(1-\epsilon)f_0+\epsilon N(0,\sigma)$

Figure 3. Empirical power of the tests, where $(i) \sim \zeta_i, i = 1,2,3$. (H0) $f_0 = 0.6N(0,1)+0.4N(0,3)$, $a \sim U(1,10)$, $\sigma \sim U(1,10)$, $\epsilon \sim U(0,0.2)$.

5.1.2 Gaussian mixture model

Considering the hypothesis devoted to the detection of additional cluster with different location in Gaussian distribution mixtures, for example, (H1) $(1-\epsilon)f_0+\epsilon N(a,1)$, one can see that the results of the tests ζ_i, $i=1,2$, were in general similar to the performance of the Anderson–Darling criterion and slightly superior to the other classical examined tests (see, e.g., Fig. 3(a)). However, absolute dominance of ζ_i, $i=1,2$, over the other criteria is observed in the case where the additional distribution cluster has the same location but different scale parameters. This is especially evident in the case where the variance of the contaminating cluster is significantly different from the variance of the null distribution, for example, (H0) $f_0=0.6N(0,1)+0.4N(0,3)$ and (H1) $(1-\epsilon)f_0+\epsilon N(0,\sigma_2)$, where $\sigma_2 \sim U(1,10)$ and $\epsilon \leqslant 0.2$ (see Fig. 3(b)).

5.2 Conclusion

In this paper, we have considered supremum-type statistics for comparing distributions based on empirical characteristic functions. All examined tests ζ_i, $i=1,2,3$, are practical to apply for moderate dimension and arbitrary sample sizes. In the cases of simple goodness-of-fit hypothesis and composite hypothesis of normality, the constructive approximations of the null distributions of the test statistics ζ_i, $i=1,2$, were established using the theory of large excursions of Gaussian random fields. The obtained approximations are sufficiently precise even for small sample sizes and moderate significance levels and are advisable to apply for calculation of the cut-off points for the critical regions of the examined tests.

The presented Monte Carlo power study illustrates that the relative performance of the considered ecf tests is competitive both to the classical general criteria based on the empirical distribution function, for example, the Anderson–Darling, and some popular specific normality tests, for example, the BHEP, Jarque–Bera, and Mardia tests, against a wide range of alternative distributions. The considered tests were especially powerful in the following cases:

- Hypothesis of normality: arbitrary mixtures of normal distributions with different scale parameters (H1) $(1-\epsilon)N(0,I_d)+\epsilon N(0,\sigma R)$, $\epsilon<1/2$, where the contaminating distribution has significantly larger variance, that is, $\sigma \gg 1$, and dependent marginal components, that is, $R \neq I_d$;
- Gaussian mixture model: detection of the additional distribution cluster with significantly different variance relative to the null distribution.

Taking into account all the examined alternatives, the best results among the investigated ecf criteria on average were shown by the tests ζ_1 and ζ_2, where the first one is outperforming the second only for clearly asymmetric alternatives.

References

1. T. Anderson and D. Darling, Asymptotic theory of certain goodness-of-fit criteria based on stochastic processes, *Ann. Math. Stat.*, **23**(2):193–212, 1952.
2. T. Anderson and D. Darling, A test of goodness of fit, *J. Am. Stat. Assoc.*, **49**(268):765–769, 1954.
3. T. Anderson and D. Darling, The Kolmogorov–Smirnov, Cramer–von Mises tests, *Ann. Math. Stat.*, **28**(3):823–838, 1957.
4. A. Bakshaev and R. Rudzkis, Goodness of fit tests based on kernel density estimators, *Informatica*, **24**(3):447–460, 2013.
5. A. Bakshaev and R. Rudzkis, Multivariate goodness of fit tests based on kernel density estimators, *Nonlinear Anal. Model. Control*, **20**(4):585–602, 2015.
6. P.J. Bickel and M. Rosenblatt, On some global measures of the deviations of density function estimates, *Ann. Stat.*, **1**(6):1071–1095, 1973.
7. S. Csörgő, Testing for normality in arbitrary dimension, *Ann. Stat.*, **14**(2):708–723, 1986.
8. Y. Fan, Tests for a multivariate distribution by the empirical characteristic function, *J. Multivariate Anal.*, **62**:36–63, 1997.
9. A. Feuerverger and R.A. Mureika, The empirical characteristic function and its application, *Ann. Stat.*, **5**(1):88–97, 1977.
10. N. Henze, Invariant tests for multivariate normality: A critical review, *Stat. Pap.*, **43**(4):467–506, 2002.
11. N. Henze and L. Baringhaus, A consistent test for multivariate normality based on the empirical characteristic function, *Metrika*, **35**:339–348, 1998.
12. N. Henze and T. Wagner, A new approach to the BHEP tests for multivariate normality, *J. Multivariate Anal.*, **62**:1–27, 1997.
13. N. Henze and B. Zirklers, A class of invariant and consistent tests for multivariate normality, *J. Multivariate Anal.*, **19**:3595–3617, 1990.
14. M. Huskova and S.G. Meintanis, Testing procedures based on the empirical characteristic functions I: Goodness-of-fit, testing for symmetry and independence, *Tatra Mt. Math. Publ.*, **39**:225–233, 2008.
15. M. Huskova and S.G. Meintanis, Testing procedures based on the empirical characteristic functions II: k-sample problem, change point problem, *Tatra Mt. Math. Publ.*, **39**:235–243, 2008.
16. C.M. Jarque and A.K. Bera, Efficient tests for normality, homoscedasticity and serial independence of regression residuals, *Econ. Lett.*, **6**:255–259, 1980.
17. C.M. Jarque and A.K. Bera, A test for normality of observations and regression residuals, *Int. Stat. Rev.*, **55**:163–172, 1987.
18. M.D. Jimenez-Gamero, V. Alba-Fernández, J. Munoz-Garcia, and Y. Chalco-Cano, Goodness-of-fit tests based on empirical characteristic functions, *Comput. Stat. Data Anal.*, **53**(12):3957–3971, 2009.
19. A. Justel, D. Pena, and R. Zamar, A multivariate Kolmogorov–Smirnov test of goodness of fit, *Stat. Probab. Lett.*, **35**:251–259, 1997.
20. A.N. Kolmogorov, Sulla determinazione empririca di una legge di distribuzione, *G. Ist. Ital. Attuari*, **4**:1–11, 1933.
21. I.A. Koutrouvelis, A goodness-of-fit test of simple hypothesis based on the empirical characteristic function, *Biometrika*, **67**:238–240, 1980.
22. I.A. Koutrouvelis and J. Kellermeier, A goodness-of-fit test based on the empirical characteristic function when parameters must be estimated, *J. R. Stat. Soc., Ser. B*, **43**:173–176, 1981.

23. K.V. Mardia, Measures of multivariate skewness and kurtosis with applications, *Biometrika*, **57**:519–530, 1970.
24. G. Martynov, *Omega-Square Criteria*, Nauka, Moscow, 1978.
25. S.G. Meintanis, J. Ngatchou-Wandj, and E. Taufer, Goodness-of-fit tests for multivariate stable distributions based on the empirical characteristic function, *J. Multivariate Anal.*, **140**:171–192, 2015.
26. J. Pudelko, On a new affine invariant and consistent test for multivariate normality, *Probab. Math. Stat.*, **25**:43–54, 2005.
27. L.B. Pulley and T.W. Epps, A test for normality based on the empirical characteristic function, *Biometrika*, **70**:723–726, 1983.
28. M. Rosenblatt, On the maximal deviation of k-dimensional density estimates, *Ann. Probab.*, **4**(6):1009–1015, 1976.
29. R. Rudzkis, On the distribution of supremum-type functionals of nonparametric estimates of probability and spectral densities, *Theory Probab. Appl.*, **37**(2):236–249, 1992.
30. R. Rudzkis, Probabilities of large excursions of empirical processes and fields, *Sov. Math., Dokl.*, **45**(1):226–228, 1992.
31. R. Rudzkis and A. Bakshaev, Probabilities of high excursions of Gaussian fields, *Lith. Math. J.*, **52**(2):196–213, 2012.
32. S.S. Shapiro and M.B. Wilk, An analysis of variance test for normality, *Biometrika*, **52**:591–611, 1965.
33. N.V. Smirnov, Approximate laws of distribution of random variables from empirical data, *Usp. Mat. Nauk*, **10**:179–206, 1944.
34. C. Tenreiro, An affine invariant multiple test procedure for assessing multivariate normality, *Comput. Stat. Data Anal.*, **55**(12):1980–1992, 2011.
35. M. Towhidi and M. Salmanpour, A new goodness-of-fit test based on the empirical characteristic function, *Commun. Stat., Theory Methods*, **36**:2777–2785, 2007.
36. W.K. Wong and C.H. Sim, Goodness-of-fit test based on empirical characterisitc function, *J. Stat. Comput. Simulation*, **65**:243–269, 2007.
37. G. Yang, *The Energy Goodness-of-Fit Test for Univariate Stable Distributions*, PhD dissertation, Graduate College of Bowling Green State University, Bowling Green, OH, 2012.