



Testing normality via a distributional fixed point property in the Stein characterization

Steffen Betsch¹ · Bruno Ebner¹

Received: 19 March 2018 / Accepted: 26 January 2019 / Published online: 22 February 2019

© Sociedad de Estadística e Investigación Operativa 2019

Abstract

We propose two families of tests for the classical goodness-of-fit problem to univariate normality. The new procedures are based on L^2 -distances of the empirical zero-bias transformation to the empirical distribution or the normal distribution function. Weak convergence results are derived under the null hypothesis, under contiguous as well as under fixed alternatives. A comparative finite-sample power study shows the competitiveness to classical procedures.

Keywords Goodness-of-fit test · Normal distribution · Stein's method · Zero-bias transformation

Mathematics Subject Classification Primary 62F03; Secondary 62F05 · 60E10 · 62E10

1 Introduction

Testing normality is commonly known as the most used and discussed goodness-of-fit technique, justified by the model assumption of normality in classical models. To be specific, let X, X_1, X_2, \dots be real-valued independent and identically distributed (iid.) random variables. The problem of interest is to test the hypothesis

$$H_0 : \mathbb{P}^X \in \mathcal{N} = \{\mathcal{N}(\mu, \sigma^2) \mid (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\} \quad (1)$$

against general alternatives. So far, a great variety of goodness-of-fit tests have been proposed, and research is of ongoing interest, as witnessed by the recent papers of

✉ Bruno Ebner
Bruno.Ebner@kit.edu

Steffen Betsch
Steffen.Betsch@student.kit.edu

¹ Institute of Stochastics, Karlsruhe Institute of Technology, Englerstr. 2, 76128 Karlsruhe, Germany

Bera et al. (2016), Villaseñor-Alva and González-Estrada (2015) and comparative studies like from Romão et al. (2010), Yap and Sim (2011). Classical procedures in goodness-of-fit methodology such as the Kolmogorov–Smirnov and the Cramér–von Mises test approach the testing problem by measuring the distance of the empirical distribution function to the estimated representative of \mathcal{N} . For a theoretical approach to goodness-of-fit tests to a family of distributions, see del Barrio et al. (2000), Neuhaus (1979). Other methods are based on skewness and kurtosis, as, for instance, proposed by Pearson et al. (1977) (known to lead to inconsistent procedures), the empirical characteristic function, see Epps and Pulley (1983), the Wasserstein distance, see del Barrio et al. (2000), del Barrio et al. (1999), the sample entropy, see Vasicek (1976), the integrated empirical distribution function, see Klar (2001), or correlation and regression tests, as the famous Shapiro–Wilk test, see Shapiro and Wilk (1965), among others. For a survey of classical methods, see del Barrio et al. (2000), Sect. 3, and Henze (1994), and for the problem of testing multivariate normality, we refer to Henze (2002), Mecklin and Mundfrom (2004).

Another natural approach to assess the distance of the distribution of a real-valued random variable X to the normal distribution is to calculate the difference between $\mathbb{E}h(X)$ and $\mathbb{E}h(N)$, where $\mathbb{P}^N = \mathcal{N}(0, 1)$, over some large class of functions $h : \mathbb{R} \rightarrow \mathbb{R}$. With the class $\{x \mapsto e^{itx} \mid t \in \mathbb{R}\}$ leading to the characteristic functions, one heavily relies on the assumption of independence when proving limit theorems. In an attempt to give an alternative proof of the central limit theorem, Charles Stein considered a different class of test functions (see, e.g. Stein 1972). Stating that X has a standard normal distribution if, and only if,

$$\mathbb{E}[f'(X)] = \mathbb{E}[Xf(X)] \quad (2)$$

holds for each absolutely continuous function f for which the expectations exist, it appears reasonable to regard $\mathbb{E}[f'(X) - Xf(X)]$, for a suitable function f , as an estimate of $\mathbb{E}h(X) - \mathbb{E}h(N)$ since both terms ought to be small whenever the distribution of X is close to standard normal. In practice, solving the differential equation

$$f'(x) - xf(x) = h(x) - \mathbb{E}h(N) \quad (3)$$

for absolutely continuous functions h , evaluating at X and taking expectations, the problem reduces to appraising $\mathbb{E}[f'_h(X) - Xf_h(X)]$, with f_h being the solution of (3). A commonly used tool to handle these terms is the so-called zero-bias transformation introduced by Goldstein and Reinert (1997). Namely, if $\mathbb{E}X = 0$ and $\mathbb{V}(X) = 1$, a random variable X^* is said to have the X -zero-bias distribution if

$$\mathbb{E}[f'(X^*)] = \mathbb{E}[Xf(X)] \quad (4)$$

holds for all absolutely continuous functions f for which these expectations exist. The use of this distribution, if it exists, lends itself easily to the purpose of distributional approximation. For instance, starting with the solution of (3), the mean value theorem

gives

$$|\mathbb{E}h(X) - \mathbb{E}h(N)| = |\mathbb{E}[f'_h(X) - f'_h(X^*)]| \leq \|f''_h\|_\infty \mathbb{E}|X - X^*|.$$

Thus, the problem reduces to bounding the derivatives of the solution f_h of (3) and constructing X^* such that $\mathbb{E}|X - X^*|$ is accessible. Bounds on f_h and its derivatives are well known, and a comprehensive treatment as well as explicit constructions for X^* may be found in Chen et al. (2011) (for the bounds, see also Stein 1986). For a general introduction to Stein’s method, see Chen et al. (2011), Ross (2011). One of the main reasons Stein’s method, particularly for the normal distribution, has been studied to a remarkable extent are various central limit type results, also giving convergence rates, even in dependency settings.

It seems reasonable to ask whether Stein’s characterization (2) may be used to construct a goodness-of-fit statistic. Apparently, we can hardly evaluate a quantity for all absolutely continuous functions which makes the direct application of equation (2) rather complicated (cf. Liu et al. 2016). Instead, we propose a test based on the zero-bias distribution. To this end, we first recall the explicit formula for the density and distribution function of the zero-bias distribution.

Lemma 1 *If X is a centred, real-valued random variable with $\mathbb{V}(X) = 1$, the X -zero-bias distribution exists and is unique. Moreover, it is absolutely continuous with respect to the Lebesgue measure with density*

$$d^X(t) = \mathbb{E}[X\mathbb{1}\{X > t\}] = -\mathbb{E}[X\mathbb{1}\{X \leq t\}]$$

and distribution function

$$F^X(t) = \mathbb{E}[X(X - t)\mathbb{1}\{X \leq t\}].$$

A proof can be found in Chen et al. (2011) or in the original treatment (Goldstein and Reinert 1997). Now, interpreting (4) as a distributional transformation $\mathbb{P}^X \mapsto \mathbb{P}^{X^*}$, the standard normal distribution is characterized as the unique fixed point of this transformation [see also Goldstein and Reinert 1997, Lemma 2.1 (i)]. Writing this in terms of the formula from Lemma 1, the characterization reads as follows.

Theorem 1 *A random variable X with distribution function F and $\mathbb{E}X = 0$, $\mathbb{V}(X) = 1$ has the standard normal distribution if, and only if,*

$$F^X = F$$

which in turn holds if, and only if,

$$F^X = \Phi,$$

where Φ is the distribution function of the standard normal distribution.

Proof By Lemma 1 and the presumptions on X , the zero-bias distribution \mathbb{P}^{X^*} of \mathbb{P}^X exists, is unique and has distribution function F^X . Hence, if \mathbb{P}^X is the standard normal distribution, (2) is satisfied and the definition of the zero-bias distribution through formula (4) and its uniqueness imply $\mathbb{P}^{X^*} = \mathbb{P}^X$, that is, $F^X = F$. Conversely, if $F^X = F$, Lemma 1 implies $\mathbb{P}^{X^*} = \mathbb{P}^X$ and Stein’s characterization (2) yields that X is standard Gaussian.

For the second equivalence note that if X follows the standard normal law, $F^X = F = \Phi$ by the first part. Finally, assume that $F^X = \Phi$. Since F^X is the distribution function of X^* , \mathbb{P}^{X^*} is the standard normal distribution, so

$$\mathbb{E}[f'(X^*)] = \mathbb{E}[X^* f(X^*)]$$

holds for each absolutely continuous function f for which these expectations exist. The definition of the zero-bias distribution implies that for any such function f ,

$$\mathbb{E}[Xf(X)] = \mathbb{E}[X^* f(X^*)].$$

Noticing that these functions include any monomial (since X^* has the standard normal distribution for which moments of all orders exist) and that the normal distribution is uniquely determined through its sequence of moments (see Theorem 30.1 of Billingsley 1995), this last equation shows $\mathbb{P}^X = \mathbb{P}^{X^*} = \mathcal{N}(0, 1)$. □

This theorem paves the way for the construction of goodness-of-fit tests using a measure of deviation between an empirical version of F^X and Φ or the empirical distribution, respectively. Heuristically, the above characterization indicates that the difference between these empirical quantities ought to be small when the underlying sample comes from a normal distribution and large whenever it does not. Thus, tests based on this characterization should be able to detect deviations from the normality hypothesis (1).

In Sect. 2, we use a weighted L^2 -measure to construct two statistics for our testing problem (1). We derive the limit null distributions in Sect. 3 and study the behaviour under contiguous alternatives in Sect. 4. The consistency of these classes of tests is established in Sect. 5, and we obtain the limit distributions of the statistics under fixed alternatives. To analyse the actual performance, empirical results in form of a power study are presented in Sect. 6. Conclusions and outlines complete the article.

2 The new test statistics

Let X, X_1, X_2, \dots be real-valued iid. random variables defined on an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Further, let F be the distribution function of X and assume that $\mathbb{E}[X^2] < \infty$. To reflect the invariance of the family of normal distributions \mathcal{N} with respect to affine transformations, the proposed statistics only depend on the so-called scaled residuals, namely $Y_{n,1}, \dots, Y_{n,n}$,

$$Y_{n,j} = \frac{X_j - \bar{X}_n}{S_n},$$

where $\bar{X}_n = n^{-1} \sum_{k=1}^n X_k$ and $S_n^2 = n^{-1} \sum_{k=1}^n (X_k - \bar{X}_n)^2$ are the sample mean and variance, respectively. This way, the values of our statistics themselves and thus the tests based on them are invariant under affine transformations of the data. We note that if X has a normal distribution with some parameters μ and σ^2 , $Y_{n,1}$ is approximately standard normal since (\bar{X}_n, S_n^2) is a strongly consistent estimator of (μ, σ^2) . Due to the affine invariance, we assume, w.l.o.g., $\mathbb{E}X = 0$ and $\mathbb{V}(X) = 1$.

In view of Theorem 1 and the heuristics given thereafter, we suggest the Cramér–von Mises-type (or weighted L^2 -type) test statistics

$$G_n^{(1)} = n \int_{\mathbb{R}} \left(\frac{1}{n} \sum_{j=1}^n Y_{n,j} (Y_{n,j} - t) \mathbb{1}\{Y_{n,j} \leq t\} - \frac{1}{n} \sum_{j=1}^n \mathbb{1}\{Y_{n,j} \leq t\} \right)^2 \omega(t) dt \tag{5}$$

and

$$G_n^{(2)} = n \int_{\mathbb{R}} \left(\frac{1}{n} \sum_{j=1}^n Y_{n,j} (Y_{n,j} - t) \mathbb{1}\{Y_{n,j} \leq t\} - \Phi(t) \right)^2 \omega(t) dt. \tag{6}$$

Here, $n^{-1} \sum_{j=1}^n Y_{n,j} (Y_{n,j} - t) \mathbb{1}\{Y_{n,j} \leq t\}$ is an empirical version of the zero-bias distribution function and $n^{-1} \sum_{j=1}^n \mathbb{1}\{Y_{n,j} \leq t\}$ is the empirical distribution function of $Y_{n,1}, \dots, Y_{n,n}$. By $\omega : \mathbb{R} \rightarrow \mathbb{R}$, we denote a positive, continuous weight function satisfying

$$\int_{\mathbb{R}} t^6 \omega(t) dt < \infty \tag{7}$$

and

$$n \int_{\mathbb{R}} \left| \omega \left(\frac{s - \bar{X}_n}{S_n} \right) - \omega(s) \right|^3 (\omega(s))^{-2} ds = o_{\mathbb{P}}(1), \tag{8}$$

where $o_{\mathbb{P}}(1)$ denotes convergence to 0 in probability as $n \rightarrow \infty$. A test based on $G_n^{(1)}$ or $G_n^{(2)}$ rejects H_0 for large values of the statistic. For the implementation of our tests, we need to specify the weight function ω . To that end, we use the density function of a centred normal distribution

$$\omega_a(t) = \frac{1}{\sqrt{2\pi a}} e^{-\frac{t^2}{2a}},$$

where the variance is chosen to be some tuning parameter $a > 0$. We prove in Lemma 2 of ‘‘Appendix A’’ that ω_a satisfies the above conditions. Note that this type of weight has also been employed by Henze and Zirkler (1990). For this explicit function, our statistics have the expressions

$$\begin{aligned}
 G_{n,a}^{(1)} = & \frac{2}{n} \sum_{1 \leq j < k \leq n} \left\{ \left(1 - \Phi \left(\frac{Y_{(k)}}{\sqrt{a}} \right) \right) \left((Y_{(j)}^2 - 1)(Y_{(k)}^2 - 1) + aY_{(j)}Y_{(k)} \right) \right. \\
 & \left. + \frac{a}{\sqrt{2\pi a}} \exp \left(-\frac{Y_{(k)}^2}{2a} \right) \left(-Y_{(j)}^2 Y_{(k)} + Y_{(k)} + Y_{(j)} \right) \right\} \\
 & + \frac{1}{n} \sum_{j=1}^n \left\{ \left(1 - \Phi \left(\frac{Y_j}{\sqrt{a}} \right) \right) \left(Y_j^4 + (a - 2)Y_j^2 + 1 \right) \right. \\
 & \left. + \frac{a}{\sqrt{2\pi a}} \exp \left(-\frac{Y_j^2}{2a} \right) \left(2Y_j - Y_j^3 \right) \right\} \tag{9}
 \end{aligned}$$

and

$$\begin{aligned}
 G_{n,a}^{(2)} = & \frac{2}{n} \sum_{1 \leq j < k \leq n} \left\{ Y_{(j)}Y_{(k)} \left[(Y_{(j)}Y_{(k)} + a) \left(1 - \Phi \left(\frac{Y_{(k)}}{\sqrt{a}} \right) \right) - aY_{(j)} \omega_a(Y_{(k)}) \right] \right\} \\
 & + \sum_{j=1}^n \left\{ \frac{Y_j^2}{n} \left[(Y_j^2 + a) \left(1 - \Phi \left(\frac{Y_j}{\sqrt{a}} \right) \right) - aY_j \omega_a(Y_j) \right] \right. \\
 & - 2Y_j \left[Y_j \int_{Y_j}^{\infty} \Phi(t) \omega_a(t) dt - a\Phi(Y_j) \omega_a(Y_j) \right. \\
 & \left. \left. - \frac{a}{\sqrt{2\pi(1+a)}} \left(1 - \Phi \left(\sqrt{\frac{1+a}{a}} Y_j \right) \right) \right] \right\} \\
 & + n \int_{\mathbb{R}} \Phi(t)^2 \omega_a(t) dt, \tag{10}
 \end{aligned}$$

where Y_1, \dots, Y_n is shorthand for the normalized sample $Y_{n,1}, \dots, Y_{n,n}$ and $Y_{(1)} \leq \dots \leq Y_{(n)}$ is the ordered sample. Those expressions make the statistics amenable to computations and, with critical values like those given in Sect. 6, the tests can be implemented immediately for any fixed $a > 0$.

The tuning parameter a determines the decay of the weight function. For tests based on the Laplace or the Fourier transform, the properties that those transformations reflect on the underlying distribution often give a good heuristic for which values of the tuning parameter lead to a high power of the test (see Baringhaus et al. 2000 for examples and explanations). Since the zero-bias transformation is known to preserve many properties of the original distribution, we expect that, at least for our first statistic, the tuning parameter will have little influence on the power for most alternative distributions. Indeed, we will observe that both tests are very stable in this regard. Nevertheless, for some (symmetric) alternative distributions the choice of the tuning parameter is crucial; therefore, we additionally implement our test with an adaptive, data-dependent choice as proposed by Allison and Santana (2015). Particularly interesting is the case $a \searrow 0$. Here, Baringhaus et al. (2000) have shown that, after suitable rescaling, this limit can be obtained explicitly for many test statistics by using an Abelian theorem for the Laplace transform. (Note that due to different parametrization they let $a \rightarrow \infty$.)

For our statistics, we have

$$2n \lim_{a \searrow 0} G_{n,a}^{(1)} = \left(\sum_{j=1}^n (Y_j^2 - 1) \mathbb{1}\{Y_j \leq 0\} \right)^2 + \left(\sum_{j=1}^n (Y_j^2 - 1) \mathbb{1}\{Y_j < 0\} \right)^2 \tag{11}$$

and

$$2n \lim_{a \searrow 0} G_{n,a}^{(2)} = \left(\frac{n}{2} - \sum_{j=1}^n Y_j^2 \mathbb{1}\{Y_j \leq 0\} \right)^2 + \left(\frac{n}{2} - \sum_{j=1}^n Y_j^2 \mathbb{1}\{Y_j < 0\} \right)^2, \tag{12}$$

that is, in the limit $a \searrow 0$, $G_{n,a}^{(1)}$ and $G_{n,a}^{(2)}$ reject the normality hypothesis for large values of the respective limits in (11) and (12). A proof of those limit relations is given in ‘‘Appendix C’’. If the underlying distribution of X is continuous, the indicator functions in the above limits are equal almost surely, and the terms can be simplified. A related question is the limit for $a \rightarrow \infty$. Starting from (9), direct but tedious calculations, mostly involving L’Hospital’s rule, give

$$\begin{aligned} \sqrt{2\pi} n \lim_{a \rightarrow \infty} \sqrt{a} G_{n,a}^{(1)} &= \sum_{1 \leq j < k \leq n} \left\{ 2Y_{(j)}^2 Y_{(k)} - Y_{(j)} Y_{(k)}^2 + \frac{1}{3} Y_{(j)} Y_{(k)}^4 - Y_{(j)}^2 Y_{(k)}^3 \right\} \\ &\quad + \sum_{j=1}^n \left\{ -\frac{1}{3} Y_{(j)}^5 + j Y_{(j)}^3 - 2(j-1) Y_{(j)} \right\} \end{aligned}$$

We omit the calculations as they provide no further insight. It remains open if a similar limit exists for the second statistic $G_{n,a}^{(2)}$.

Having discussed the framework for the implementation of our tests, it remains to introduce the setting for our theoretical studies. Namely, to develop the asymptotic theory, we let \mathcal{B} be the Borel- σ -field of \mathbb{R} and \mathcal{L}^1 the Lebesgue measure on \mathbb{R} , and consider the Hilbert space

$$\mathcal{H} = L^2(\mathbb{R}, \mathcal{B}, \omega \, d\mathcal{L}^1)$$

of measurable, square-integrable functions $f : \mathbb{R} \rightarrow \mathbb{R}$. Notice that the functions figuring within the integral in the definition of $G_n^{(1)}$ and $G_n^{(2)}$ are $(\mathcal{A} \otimes \mathcal{B}, \mathcal{B})$ -measurable and random elements of \mathcal{H} . We denote by

$$\|f\|_{\mathcal{H}} = \left(\int_{\mathbb{R}} |f(t)|^2 \omega(t) \, dt \right)^{1/2}, \quad \langle f, g \rangle_{\mathcal{H}} = \int_{\mathbb{R}} f(t)g(t) \omega(t) \, dt$$

the usual norm as well as the usual inner product in \mathcal{H} . Furthermore, we write $\mathcal{U}_n(s) \approx \mathcal{V}_n(s)$ whenever

$$\|\mathcal{U}_n - \mathcal{V}_n\|_{\mathcal{H}} = o_{\mathbb{P}}(1).$$

Here, \mathcal{U}_n and \mathcal{V}_n are random elements of our Hilbert space. For the approximations associated with this notation, Lemma 3 stated in ‘‘Appendix A’’ will be essential. We have also deferred some asymptotic expansions to ‘‘Appendix B’’ so it is easier to grasp the main ideas of the proofs. In the following, we denote convergence in distribution by $\xrightarrow{\mathcal{D}}$ and write $O_{\mathbb{P}}(1)$ for boundedness in probability.

3 The limit null distributions

Our first results for the statistics concern the study of their behaviour under the hypothesis. In particular, we derive the limit distributions for $n \rightarrow \infty$ when the normality hypothesis (1) holds. Therefore, we assume in this section that X, X_1, X_2, \dots are iid. random variables with $\mathbb{P}^X = \mathcal{N}(0, 1)$. By φ , we denote the density function of the standard normal law.

Theorem 2 *There exists a centred Gaussian element $\mathcal{W}^{(2)}$ of \mathcal{H} with covariance kernel*

$$\begin{aligned} \mathcal{K}^{(2)}(s, t) = & (2(s + t) - (st + 3)(s \wedge t) + (s + t)(s \wedge t)^2 - (s \wedge t)^3)\varphi(s \wedge t) \\ & + (st + 3)(\Phi(s \wedge t) - \Phi(s)\Phi(t)) + (t - 2s)\varphi(t)\Phi(s) \\ & + (s - 2t)\varphi(s)\Phi(t) - \frac{st}{2}\varphi(s)\varphi(t) - 4\varphi(s)\varphi(t), \quad s, t \in \mathbb{R}, \end{aligned}$$

where $s \wedge t = \min\{s, t\}$, such that

$$G_n^{(2)} \xrightarrow{\mathcal{D}} \left\| \mathcal{W}^{(2)} \right\|_{\mathcal{H}}^2, \quad \text{as } n \rightarrow \infty.$$

Proof Note that a simple change of variable in the integral gives

$$G_n^{(2)} = \frac{1}{S_n} \int_{\mathbb{R}} |\sqrt{n} U_n(s)|^2 \omega \left(\frac{s - \bar{X}_n}{S_n} \right) ds, \tag{13}$$

where

$$U_n(s) = \widehat{F}_n^X(s) - \Phi \left(\frac{s - \bar{X}_n}{S_n} \right)$$

and

$$\widehat{F}_n^X(s) = \frac{1}{n} \sum_{j=1}^n \frac{X_j - \bar{X}_n}{S_n^2} (X_j - s) \mathbb{1}\{X_j \leq s\}, \quad s \in \mathbb{R}.$$

The idea of the proof is to show that U_n converges weakly to the Gaussian element of \mathcal{H} stated in the theorem and to use Lemma 3 from ‘‘Appendix A’’ to replace the shifted

weight function in the integral above by $\omega(s)$. Indeed, with (26) from Lemmata 4 and 5 we have

$$\begin{aligned} \sqrt{n} U_n(s) \approx & \frac{1}{S_n^2} \sqrt{n} \left\{ \frac{1}{n} \sum_{j=1}^n X_j(X_j - s) \mathbb{1}\{X_j \leq s\} - \bar{X}_n \mathbb{E}[(X - s) \mathbb{1}\{X \leq s\}] \right. \\ & \left. - \frac{1}{n} \sum_{j=1}^n X_j^2 \Phi(s) - S_n \varphi(s) ((1 - S_n) \cdot s - \bar{X}_n) \right\}. \end{aligned}$$

Since

$$\sqrt{n} (1 - S_n) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \frac{1}{2} (1 - X_j^2) + o_{\mathbb{P}}(1),$$

we obtain

$$\sqrt{n} U_n(s) \approx \frac{1}{S_n^2} \frac{1}{\sqrt{n}} \sum_{j=1}^n W_j(s), \tag{14}$$

where

$$\begin{aligned} W_j(s) = & X_j(X_j - s) \mathbb{1}\{X_j \leq s\} + X_j(\varphi(s) + s\Phi(s)) \\ & - X_j^2 \Phi(s) - \left(\frac{1}{2} (1 - X_j^2) \cdot s - X_j\right) \varphi(s). \end{aligned}$$

Notice that W_1, \dots, W_n are iid. random elements of \mathcal{H} with $\mathbb{E}W_1 = 0$ (as $F^X = \Phi$ under H_0 , cf. Theorem 1) and $\mathbb{E} \|W_1\|_{\mathcal{H}}^2 < \infty$. The central limit theorem for separable Hilbert spaces, see Corollary 10.9 in Ledoux and Talagrand (2011), provides the existence of a centred Gaussian element $\mathcal{W}^{(2)} \in \mathcal{H}$ with

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n W_j(\cdot) \xrightarrow{\mathcal{D}} \mathcal{W}^{(2)}(\cdot).$$

By (14), $\|\sqrt{n} U_n\|_{\mathcal{H}} = O_{\mathbb{P}}(1)$ and Lemma 4 implies $\sup_{s \in \mathbb{R}} |U_n(s)| \leq 2$ \mathbb{P} -almost surely (a.s.) for each $n \in \mathbb{N}$. Thus, with Lemma 3, (13) reads as

$$G_n^{(2)} = \|\sqrt{n} U_n\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

The continuous mapping theorem and Slutsky’s lemma imply

$$G_n^{(2)} \xrightarrow{\mathcal{D}} \left\| \mathcal{W}^{(2)} \right\|_{\mathcal{H}}^2.$$

Since the function $\mathcal{K}^{(2)}$ defined in the statement of the theorem satisfies $\mathcal{K}^{(2)}(s, t) = \mathbb{E}[W_1(s)W_1(t)]$, it is the covariance kernel of $\mathcal{W}^{(2)}$ and we are done. \square

For $G_n^{(1)}$, the limit distribution under the hypothesis can be obtained in a similar manner. Starting with

$$G_n^{(1)} = \frac{n}{S_n} \int_{\mathbb{R}} \left(\widehat{F}_n^X(s) - \frac{1}{n} \sum_{j=1}^n \mathbb{1}\{X_j \leq s\} \right)^2 \omega \left(\frac{s - \bar{X}_n}{S_n} \right) ds, \quad (15)$$

the reasoning closely parallels that of the proof of Theorem 2.

Corollary 1 *There exists a centred Gaussian element $\mathcal{W}^{(1)}$ of \mathcal{H} with covariance kernel*

$$\begin{aligned} \mathcal{K}^{(1)}(s, t) = & ((s+t) - (st+1)(s \wedge t) + (s+t)(s \wedge t)^2 - (s \wedge t)^3) \varphi(s \wedge t) \\ & + (st+2)(\Phi(s \wedge t) - \Phi(s)\Phi(t)) - s\Phi(s)\varphi(t) - t\Phi(t)\varphi(s) \\ & - \varphi(s)\varphi(t), \quad s, t \in \mathbb{R}, \end{aligned}$$

such that

$$G_n^{(1)} \xrightarrow{\mathcal{D}} \left\| \mathcal{W}^{(1)} \right\|_{\mathcal{H}}^2, \quad \text{as } n \rightarrow \infty.$$

Remark The distribution of $\left\| \mathcal{W}^{(k)} \right\|_{\mathcal{H}}^2$, $k = 1, 2$, that is, the limit distribution of $G_n^{(k)}$ under the hypothesis, is that of $\sum_{j=1}^{\infty} \lambda_j^{(k)} N_j^2$. Here, N_1, N_2, \dots are independent standard Gaussian random variables and $\lambda_1^{(k)}, \lambda_2^{(k)}, \dots$ are the nonzero eigenvalues of the operator

$$\mathcal{H} \rightarrow \mathcal{H}, \quad f \mapsto \int_{\mathbb{R}} \mathcal{K}^{(k)}(\cdot, t) f(t) \omega(t) dt,$$

$k = 1, 2$. Considering the complexity of $\mathcal{K}^{(k)}$, it does not seem possible to determine $\lambda_j^{(k)}$ explicitly. Thus, in practice, critical values are obtained by simulation rather than by using asymptotic results. An alternative approach to gain theoretically justified (approximate) critical values is to calculate the first four moments of the limit null distribution and fit a representative of the Pearson- or Johnson-family of distributions to those moments (see Henze 1990 for an example). Since we do not face any complications in computing the critical values, we will only pursue the empirical approach.

4 Contiguous alternatives

Adjusting the argumentation of Henze and Wagner (1997), we will derive non-degenerate limit distributions for our statistics under contiguous alternatives converging to the normal distribution at rate $n^{-1/2}$. To this end, we introduce a triangular array

of row-wise iid. random variables $X_{n,1}, \dots, X_{n,n}, n \in \mathbb{N}$, with Lebesgue density

$$p_n(x) = \varphi(x) \cdot \left(1 + \frac{1}{\sqrt{n}} c(x)\right), \quad x \in \mathbb{R}.$$

Here, $c : \mathbb{R} \rightarrow \mathbb{R}$ is a measurable, bounded function satisfying

$$\int_{\mathbb{R}} c(x) \varphi(x) dx = 0.$$

Notice that, by the boundedness of c , we may assume n to be large enough to ensure $p_n \geq 0$. We set

$$\mu_n = \bigotimes_{j=1}^n (\varphi \mathcal{L}^1), \quad \nu_n = \bigotimes_{j=1}^n (p_n \mathcal{L}^1)$$

which are measures on $(\mathbb{R}^n, \mathcal{B}^n)$, where \mathcal{B}^n is the Borel- σ -field of \mathbb{R}^n . Apparently, ν_n is absolutely continuous with respect to μ_n and we can look upon the Radon–Nikodym derivative $L_n = \frac{d\nu_n}{d\mu_n}$. By a Taylor expansion,

$$\begin{aligned} \log(L_n(X_{n,1}, \dots, X_{n,n})) &= \sum_{j=1}^n \log\left(1 + \frac{1}{\sqrt{n}} c(X_{n,j})\right) \\ &= \sum_{j=1}^n \left(\frac{1}{\sqrt{n}} c(X_{n,j}) - \frac{1}{2n} c(X_{n,j})^2\right) + o_{\mathbb{P}}(1) \end{aligned}$$

whenever $(X_{n,1}, \dots, X_{n,n})$ has distribution μ_n . (Note that in this case the triangular array essentially reduces to a sequence of iid. random variables with density φ .) Therefore, viewing L_n as a random element $(\mathbb{R}^n, \mathcal{B}^n, \mu_n) \rightarrow (\mathbb{R}, \mathcal{B})$, the central limit theorem and the law of large numbers give

$$\log(L_n) \xrightarrow{\mathcal{D}_{\mu_n}} \mathcal{N}\left(-\frac{\tau^2}{2}, \tau^2\right),$$

where

$$\tau^2 = \int_{\mathbb{R}} c(x)^2 \varphi(x) dx$$

and $\xrightarrow{\mathcal{D}_{\mu_n}}$ denotes convergence in distribution under μ_n . By LeCam’s first Lemma (see, for instance, Hájek et al. 1999, p. 253, Corollary 1), ν_n is contiguous to μ_n . Interpreting U_n from the proof of Theorem 2 as $U_n : \mathbb{R}^n \rightarrow \mathcal{H}$, we have shown that $G_n^{(2)} = \|\sqrt{n} U_n\|_{\mathcal{H}}^2 + o_{\mu_n}(1)$ and (14) reads as

$$\|\sqrt{n} U_n - W_n^*\|_{\mathcal{H}}^2 = o_{\mu_n}(1),$$

where $W_n^*(s) = \frac{1}{\sqrt{n}} \sum_{j=1}^n W_{n,j}(s)$ and

$$W_{n,j}(s) = X_{n,j}(X_{n,j} - s)\mathbb{1}\{X_{n,j} \leq s\} + (X_{n,j} \cdot s - X_{n,j}^2)\Phi(s) + \left(2X_{n,j} - \frac{1}{2}(1 - X_{n,j}^2) \cdot s\right)\varphi(s).$$

Thus, by contiguity, $\|\sqrt{n} U_n - W_n^*\|_{\mathcal{H}}^2 = o_{v_n}(1)$ and, in particular,

$$G_n^{(2)} = \|W_n^*\|_{\mathcal{H}}^2 + o_{v_n}(1). \tag{16}$$

Defining

$$\eta(x, s) = x(x - s)\mathbb{1}\{x \leq s\} + (x \cdot s - x^2)\Phi(s) + \left(2x - \frac{1}{2}(1 - x^2) \cdot s\right)\varphi(s),$$

we have, under μ_n ,

$$\text{Cov}(W_{n,1}(s), c(X_{n,1})) = \int_{\mathbb{R}} \eta(x, s) c(x) \varphi(x) dx = \zeta(s),$$

with $\zeta \in \mathcal{H}$. Consequently, for any $k \in \mathbb{N}$, $v \in \mathbb{R}^k$ and $s_1, \dots, s_k \in \mathbb{R}$, the multivariate central limit theorem, the law of large numbers and Slutsky’s lemma imply

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{ \begin{pmatrix} v_1 W_{n,j}(s_1) + \dots + v_k W_{n,j}(s_k) \\ c(X_{n,j}) - \frac{1}{2\sqrt{n}} c(X_{n,j})^2 \end{pmatrix} - \begin{pmatrix} 0 \\ -\frac{\tau^2}{2\sqrt{n}} \end{pmatrix} \right\} \xrightarrow{\mathcal{D}_{\mu_n}} \mathcal{N}_2 \left(\mathbf{0}, \begin{pmatrix} v^\top \Sigma v & v^\top \zeta_k \\ \zeta_k^\top v & \tau^2 \end{pmatrix} \right).$$

Here, $\Sigma = (\mathcal{K}^{(2)}(s_i, s_j))_{1 \leq i, j \leq k}$, with $\mathcal{K}^{(2)}$ the covariance kernel of $\mathcal{W}^{(2)}$ from Theorem 2, and $\zeta_k = (\zeta(s_1), \dots, \zeta(s_k))^\top$. Therefore,

$$\begin{pmatrix} v_1 W_n^*(s_1) + \dots + v_k W_n^*(s_k) \\ \log(L_n) \end{pmatrix} \xrightarrow{\mathcal{D}_{\mu_n}} \mathcal{N}_2 \left(\begin{pmatrix} 0 \\ -\frac{\tau^2}{2} \end{pmatrix}, \begin{pmatrix} v^\top \Sigma v & v^\top \zeta_k \\ \zeta_k^\top v & \tau^2 \end{pmatrix} \right)$$

and LeCam’s third Lemma (see Hájek et al. 1999, p. 259, Lemma 2) implies

$$W_n^* \xrightarrow{\mathcal{D}_{v_n, fidi}} \mathcal{W}^{(2)} + \zeta, \tag{17}$$

where $\xrightarrow{\mathcal{D}_{v_n, fidi}}$ denotes convergence of the finite-dimensional distributions (under v_n). In the proof of Theorem 2, we have shown that $W_n^* \xrightarrow{\mathcal{D}_{\mu_n}} \mathcal{W}^{(2)}$ which entails

the tightness of $\{W_n^* \mid n \in \mathbb{N}\}$ under μ_n . As $\{W_n^* \mid n \in \mathbb{N}\}$ remains tight under ν_n by contiguity, (17) yields

$$W_n^* \xrightarrow{\mathcal{D}_{\nu_n}} \mathcal{W}^{(2)} + \zeta.$$

Combining this with (16), we have shown the following theorem.

Theorem 3 *Under the triangular array $X_{n,1}, \dots, X_{n,n}$ with $\mathbb{P}^{X_{n,1}} = p_n \mathcal{L}^1$, we have*

$$G_n^{(2)} \xrightarrow{\mathcal{D}} \left\| \mathcal{W}^{(2)} + \zeta \right\|_{\mathcal{H}}^2.$$

Since Corollary 1 is obtained with the same line of proof used in Theorem 2, we can likewise conclude

$$G_n^{(1)} \xrightarrow{\mathcal{D}} \left\| \mathcal{W}^{(1)} + \tilde{\zeta} \right\|_{\mathcal{H}}^2,$$

where $\tilde{\zeta}(s) = \int \tilde{\eta}(s, x) c(x) \varphi(x) dx$ and

$$\tilde{\eta}(s, x) = (x(x - s) - 1)\mathbb{1}\{x \leq s\} + (1 + x \cdot s - x^2) \Phi(s) + x\varphi(s).$$

From these statements, we discern that tests based on any of our statistics are able to detect contiguous alternatives which converge, at rate $n^{-1/2}$, to the class of normal distributions. For further insights on contiguity, we refer to Roussas (1972) and Sen (1981).

5 Consistency and limit distributions under fixed alternatives

The major goal of this section is to establish that our test procedures can detect any fixed alternative satisfying a weak moment condition. For some of those distributions, we can even extend the consistency results and derive weak limits of our statistics.

We return to the nonparametric setting of Sect. 2 and let X, X_1, X_2, \dots be iid. random variables with distribution function F and $\mathbb{E}[X^2] < \infty$. Further, we assume $\mathbb{E}X = 0$ and $\mathbb{V}(X) = 1$.

Theorem 4 *As $n \rightarrow \infty$, we have*

$$\frac{G_n^{(1)}}{n} \longrightarrow \int_{\mathbb{R}} \left(F^X(s) - F(s) \right)^2 \omega(s) ds = \left\| F^X - F \right\|_{\mathcal{H}}^2 = \Delta^{(1)}$$

and

$$\frac{G_n^{(2)}}{n} \longrightarrow \int_{\mathbb{R}} \left(F^X(s) - \Phi(s) \right)^2 \omega(s) ds = \left\| F^X - \Phi \right\|_{\mathcal{H}}^2 = \Delta^{(2)},$$

where each convergence is in probability.

Proof We denote by \widehat{F}_n the empirical distribution function of X_1, \dots, X_n . By the classical Glivenko–Cantelli theorem and (25) from Lemma 4,

$$\begin{aligned} & \left| \left\| \widehat{F}_n^X - \widehat{F}_n \right\|_{\mathcal{H}}^2 - \Delta^{(1)} \right| \\ & \leq 4 \left(\sup_{s \in \mathbb{R}} \left| \widehat{F}_n^X(s) - F^X(s) \right| + \sup_{s \in \mathbb{R}} \left| \widehat{F}_n(s) - F(s) \right| \right) \int_{\mathbb{R}} \omega(s) \, ds \\ & \longrightarrow 0 \end{aligned}$$

\mathbb{P} -a.s., as $n \rightarrow \infty$. Rewriting $G_n^{(1)}$ as in (15) and applying the second part of Lemma 3, we have

$$\frac{G_n^{(1)}}{n} = \left\| \widehat{F}_n^X - \widehat{F}_n \right\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

The proof of the second claim is almost identical. □

We recall that a level- α -test based on $G_n^{(k)}$, $k = 1, 2$, rejects the hypothesis if $G_n^{(k)} > c_n^{(k)}$, where $c_n^{(k)}$ is the $(1 - \alpha)$ -quantile of the distribution of $G_n^{(k)}$ under H_0 . Theorem 2 (and Corollary 1) ensures that $\sup_{n \in \mathbb{N}} c_n^{(k)} < \infty$. Now, by Theorem 1, the limits $\Delta^{(1)}$ and $\Delta^{(2)}$ figuring in Theorem 4 are positive if X has a non-normal distribution. Consequently,

$$\mathbb{P} \left(G_n^{(k)} > c_n^{(k)} \right) \geq \mathbb{P} \left(n^{-1} G_n^{(k)} > n^{-1} \sup_{n \in \mathbb{N}} c_n^{(k)} \right) \longrightarrow 1,$$

as $n \rightarrow \infty$, and a level- α -test based on $G_n^{(1)}$ or $G_n^{(2)}$ is consistent against each alternative with existing second moment.

The following theorem concerns the limit distributions of our statistics under fixed alternatives.

Theorem 5 *Let X, X_1, X_2, \dots be iid., non-normal random variables with distribution function F and $\mathbb{E}X^4 < \infty$. Assume that X has a continuously differentiable density function p satisfying $\sup_{s \in \mathbb{R}} |p(s)| \leq K_1 < \infty$ and $\sup_{s \in \mathbb{R}} |p'(s)| \leq K_2 < \infty$. W.l.o.g., $\mathbb{E}X = 0$ and $\mathbb{V}(X) = 1$. Then, as $n \rightarrow \infty$,*

$$\sqrt{n} \left(\frac{G_n^{(1)}}{n} - \Delta^{(1)} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \tau_{(1)}^2 \right), \tag{18}$$

where

$$\tau_{(1)}^2 = 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \mathfrak{C}^{(1)}(s, t) \left(F^X(s) - F(s) \right) \left(F^X(t) - F(t) \right) \omega(s) \omega(t) \, ds \, dt$$

with

$$c^{(1)}(s, t) = \mathbb{E}[C^{(1)}(s) C^{(1)}(t)], \quad s, t \in \mathbb{R},$$

and

$$C^{(1)}(s) = (X(X - s) - 1)\mathbb{1}\{X \leq s\} - X^2 F^X(s) + (1 + X \cdot s)F(s) - \left(\frac{1}{2}(1 - X^2) \cdot s - X\right) p(s) + \left(2X - \frac{1}{2}(1 - X^2) \cdot s\right) d^X(s), \quad s \in \mathbb{R}.$$

Proof Setting

$$V_n(s) = \widehat{F}_n^X(s) - \frac{1}{n} \sum_{j=1}^n \mathbb{1}\{X_j \leq s\} - F^X\left(\frac{s - \bar{X}_n}{S_n}\right) + F\left(\frac{s - \bar{X}_n}{S_n}\right), \quad s \in \mathbb{R},$$

a change of variable in both integrals and an integral decomposition, as used by Chapman (1958), gives

$$\begin{aligned} \sqrt{n} \left(\frac{G_n^{(1)}}{n} - \Delta^{(1)} \right) &= \frac{1}{S_n} \left\{ 2 \int_{\mathbb{R}} \sqrt{n} V_n(s) \cdot [F^X(\tilde{s}) - F(\tilde{s})] \omega(\tilde{s}) ds \right. \\ &\quad \left. + \frac{1}{\sqrt{n}} \int_{\mathbb{R}} |\sqrt{n} V_n(s)|^2 \omega(\tilde{s}) ds \right\}, \end{aligned}$$

where $\tilde{s} = (s - \bar{X}_n)/S_n$. Under the assumption $\|\sqrt{n} V_n\|_{\mathcal{H}} = O_{\mathbb{P}}(1)$, Hölder’s inequality, Lebesgue’s theorem and Slutsky’s lemma give

$$\begin{aligned} &\left| \int_{\mathbb{R}} \sqrt{n} V_n(s) \cdot [F^X(\tilde{s}) - F(\tilde{s})] \omega(s) ds - \langle \sqrt{n} V_n, F^X - F \rangle_{\mathcal{H}} \right| \\ &\leq \|\sqrt{n} V_n\|_{\mathcal{H}} \left(\int_{\mathbb{R}} |F^X(\tilde{s}) - F^X(s) + F(s) - F(\tilde{s})|^2 \omega(s) ds \right)^{1/2} \\ &= o_{\mathbb{P}}(1). \end{aligned}$$

Here, we used that both F^X and F are continuous distribution functions. Using that $\sup_{s \in \mathbb{R}} |V_n(s)| \leq 4$ \mathbb{P} -a.s. for each $n \in \mathbb{N}$ (note that \widehat{F}_n^X is a distribution function as well by Lemma 4), Lemma 3 from “Appendix A” yields

$$\sqrt{n} \left(\frac{G_n^{(1)}}{n} - \Delta^{(1)} \right) = 2 \langle \sqrt{n} V_n, F^X - F \rangle_{\mathcal{H}} + \frac{1}{\sqrt{n}} \|\sqrt{n} V_n\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1). \quad (19)$$

To verify the above assumption, we show that $\sqrt{n} V_n$ converges in distribution in \mathcal{H} . In this regard, (26) and Lemma 5 imply

$$\begin{aligned} \sqrt{n} V_n(s) \approx & \frac{\sqrt{n}}{S_n^2} \left\{ \frac{1}{n} \sum_{j=1}^n X_j(X_j - s) \mathbb{1}\{X_j \leq s\} - \bar{X}_n \mathbb{E}[(X - s) \mathbb{1}\{X \leq s\}] \right. \\ & - S_n^2 \left(\frac{1}{n} \sum_{j=1}^n \mathbb{1}\{X_j \leq s\} - F(s) \right) - \frac{1}{n} \sum_{j=1}^n X_j^2 F^X(s) \\ & \left. - S_n ((1 - S_n) \cdot s - \bar{X}_n) (d^X(s) - p(s)) \right\}. \end{aligned}$$

By the classical Glivenko–Cantelli theorem and $\sqrt{n}(S_n^2 - 1) = O_{\mathbb{P}}(1)$,

$$\sqrt{n} S_n^2 \left(\frac{1}{n} \sum_{j=1}^n \mathbb{1}\{X_j \leq s\} - F(s) \right) \approx \frac{1}{\sqrt{n}} \sum_{j=1}^n (\mathbb{1}\{X_j \leq s\} - F(s)).$$

Together with the expansion $\sqrt{n}(1 - S_n) = n^{-1/2} \sum_{j=1}^n \frac{1}{2}(1 - X_j^2) + o_{\mathbb{P}}(1)$, we have

$$\sqrt{n} V_n(s) \approx \frac{1}{S_n^2} \frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j(s), \tag{20}$$

where

$$\begin{aligned} Z_j(s) = & X_j(X_j - s) \mathbb{1}\{X_j \leq s\} + X_j (d^X(s) + sF(s)) - X_j^2 F^X(s) \\ & - \mathbb{1}\{X_j \leq s\} + F(s) - \left(\frac{1}{2}(1 - X_j^2) \cdot s - X_j \right) (d^X(s) - p(s)). \end{aligned}$$

Since Z_1, \dots, Z_n are iid. random elements of \mathcal{H} with $\mathbb{E}Z_1 = 0$ as well as $\mathbb{E} \|Z_1\|_{\mathcal{H}}^2 < \infty$, the central limit theorem for separable Hilbert spaces implies

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n Z_j(\cdot) \xrightarrow{\mathcal{D}} \mathcal{Z}(\cdot),$$

where $\mathcal{Z} \in \mathcal{H}$ is a centred Gaussian element. In particular, $\|\sqrt{n} V_n\|_{\mathcal{H}}$ is bounded in probability by (20), and (19) holds. The continuous mapping theorem and Slutsky’s Lemma imply

$$\sqrt{n} \left(\frac{G_n^{(1)}}{n} - \Delta^{(1)} \right) \xrightarrow{\mathcal{D}} 2 \langle \mathcal{Z}, F^X - F \rangle_{\mathcal{H}}.$$

Denoting the covariance kernel of \mathcal{Z} by

$$\mathfrak{C}^{(1)}(s, t) = \mathbb{E}[\mathcal{Z}(s) \mathcal{Z}(t)],$$

the limiting random variable $2\langle \mathcal{Z}, F^X - F \rangle_{\mathcal{H}}$ has the normal distribution $\mathcal{N}(0, \tau_{(1)}^2)$, where

$$\begin{aligned} \tau_{(1)}^2 &= 4 \mathbb{E} \left[\langle \mathcal{Z}, F^X - F \rangle_{\mathcal{H}}^2 \right] \\ &= 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \mathfrak{C}^{(1)}(s, t) \left(F^X(s) - F(s) \right) \left(F^X(t) - F(t) \right) \omega(s) \omega(t) \, ds \, dt. \end{aligned}$$

□

Applying the reasoning of Theorem 5 to $G_n^{(2)}$, Φ will drop out when considering the decomposition of the integrals. Proceeding with the remaining terms exactly as before, we obtain an analogous statement for the second statistic under slightly weaker conditions.

Corollary 2 *Let X, X_1, X_2, \dots be iid., non-normal random variables with distribution function F , Lebesgue density p and $\mathbb{E}X^4 < \infty$. Further, assume $\sup_{s \in \mathbb{R}} |p(s)| < \infty$ and $\mathbb{E}X = 0, \mathbb{V}(X) = 1$. Then, as $n \rightarrow \infty$,*

$$\sqrt{n} \left(\frac{G_n^{(2)}}{n} - \Delta^{(2)} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \tau_{(2)}^2 \right), \tag{21}$$

where

$$\tau_{(2)}^2 = 4 \int_{\mathbb{R}} \int_{\mathbb{R}} \mathfrak{C}^{(2)}(s, t) \left(F^X(s) - \Phi(s) \right) \left(F^X(t) - \Phi(t) \right) \omega(s) \omega(t) \, ds \, dt$$

with

$$\mathfrak{C}^{(2)}(s, t) = \mathbb{E}[C^{(2)}(s) C^{(2)}(t)], \quad s, t \in \mathbb{R},$$

and

$$\begin{aligned} C^{(2)}(s) &= X(X - s) \mathbb{1}\{X \leq s\} - X^2 F^X(s) + X \cdot s F(s) \\ &\quad + \left(2X - \frac{1}{2}(1 - X^2) \cdot s \right) d^X(s), \quad s \in \mathbb{R}. \end{aligned}$$

Remark Note that for Theorem 5, we redeployed a line of proof put forward by Baringhaus et al. (2017). The asymptotic normality also qualifies our statistics for the applications they propose (see also Baringhaus and Henze 2017).

First, we fix $\alpha \in (0, 1)$ and denote by $q_\alpha = \Phi^{-1}(1 - \alpha/2)$ the $(1 - \alpha/2)$ -quantile of the standard normal distribution. Letting

$$\widehat{\tau}_{(k),n}^2 = \widehat{\tau}_{(k),n}^2(X_1, \dots, X_n)$$

be a (weakly) consistent estimator of $\tau_{(k)}^2, k = 1, 2$, figuring in Theorem 5 and Corollary 2, respectively, (18) and (21) immediately indicate that

$$I_n = \left[\frac{G_n^{(k)}}{n} - \frac{q_\alpha \widehat{\tau}_{(k),n}}{\sqrt{n}}, \frac{G_n^{(k)}}{n} + \frac{q_\alpha \widehat{\tau}_{(k),n}}{\sqrt{n}} \right] \tag{22}$$

is an asymptotic confidence interval for $\Delta^{(k)} = \Delta^{(k)}(F)$ at level $1 - \alpha$. Here, F satisfies the assumptions of Theorem 5 (or Corollary 2). As was briefly explained in the introduction, one objective of Stein’s method for the normal distribution is to assess how close a given distribution is to being normal. Thus, seeing $\Delta^{(1)}(F)$ and $\Delta^{(2)}(F)$ as ‘measures’ of how far F differs from the standard normal distribution, we also developed a procedure for empirical assessments of this kind.

Second, we emphasize that our statistics can be employed for inverse testing problems. Namely, if $\Delta_0 > 0$ is a given distance of tolerance, tests that reject H_{Δ_0} if

$$\frac{G_n^{(k)}}{n} \leq \Delta_0 - \frac{\widehat{\tau}_{(k),n}}{\sqrt{n}} \Phi^{-1}(1 - \alpha)$$

are asymptotic level- α -tests for the problem

$$H_{\Delta_0} : \Delta^{(k)}(F) \geq \Delta_0 \text{ against } K_{\Delta_0} : \Delta^{(k)}(F) < \Delta_0.$$

These tests are consistent against each alternative and aim at validating a whole non-parametric neighbourhood of the hypothesized, underlying normality. Unfortunately, the direct approach to obtain estimators for $\tau_{(k)}^2$ does not lead to feasible results.

Finally, we suppose $\{c_n^{(k)}\} \subset (0, \infty)$ is the sequence of critical values for a level- α -test based on $G_n^{(k)}, k = 1, 2$. For an alternative distribution F satisfying the relevant prerequisites of Theorem 5 or Corollary 2, we can approximate the power of the test against this alternative by

$$\begin{aligned} \mathbb{P}_F \left(G_n^{(k)} > c_n^{(k)} \right) &= \mathbb{P}_F \left(\frac{\sqrt{n}}{\tau_{(k)}} \left\{ \frac{G_n^{(k)}}{n} - \Delta^{(k)} \right\} > \frac{\sqrt{n}}{\tau_{(k)}} \left\{ \frac{c_n^{(k)}}{n} - \Delta^{(k)} \right\} \right) \\ &\approx 1 - \Phi \left(\frac{\sqrt{n}}{\tau_{(k)}} \left\{ \frac{c_n^{(k)}}{n} - \Delta^{(k)} \right\} \right). \end{aligned} \tag{23}$$

Note that this last application does (in theory) not need an estimator of $\tau_{(k)}^2$. Instead, $\tau_{(k)}^2$ and $\Delta^{(k)}$ have to be calculated for the particular fixed alternative.

6 Empirical results

In this section, we investigate the behaviour of our two statistics given through the explicit formulas (9) and (10). It is organized as follows: First, we compare the two

Table 1 Empirical 0.95 quantiles for $G_{n,a}^{(1)}$ under H_0 (100,000 replications)

$n \setminus a$	0.1	0.25	0.5	1	1.5	2	3
20	1.027	0.891	0.754	0.609	0.525	0.469	0.396
50	1.066	0.928	0.789	0.642	0.555	0.497	0.421
100	1.062	0.927	0.791	0.643	0.558	0.499	0.423

Table 2 Empirical 0.95 quantiles for $G_{n,a}^{(2)}$ under H_0 (100,000 replications)

$n \setminus a$	0.1	0.25	0.5	1	1.5	2	3
20	0.391	0.380	0.357	0.315	0.284	0.261	0.227
50	0.420	0.408	0.384	0.342	0.309	0.284	0.247
100	0.424	0.413	0.389	0.347	0.315	0.290	0.253

tests over a range of possible tuning parameters and alternative distributions. Based on these results, we choose our final procedure and describe its implementation. Then, a brief summary of the competing tests for an empirical power study is given. We display the performance of our test in comparison with the established tests in a finite-sample power study. Finally, we add results for the applications from the last section (as described in the Remark) for three alternative distributions. The simulations are performed using the statistical computing environment R, see R Core Team (2017). Notice that there are several comparative simulation studies for testing normality in the literature, as witnessed by Baringhaus et al. (1989), Farrell and Rogers-Stewart (2006), Landry and Lepage (1992), Pearson et al. (1977), Romão et al. (2010), Shapiro et al. (1968), Yap and Sim (2011) and others.

Since we consider two new families of tests both depending on the choice of the tuning parameter a , we will calculate the finite-sample power for a range of different parameters. In each simulation, we consider the sample sizes $n = 20$, $n = 50$ and $n = 100$, and fix the nominal level of significance α to 0.05. To implement the tests for any of the (fixed) values $a \in \{0.1, 0.25, 0.5, 1, 1.5, 2, 3\}$, we calculate the critical values by a Monte Carlo simulation with 100,000 repetitions. These critical values for $G_{n,a}^{(k)}$, $k = 1, 2$, can be found in Tables 1 and 2 and are taken from there throughout the simulations.

We choose the alternative distributions to fit the extensive power study of normality tests by Romão et al. (2010), in order to ease the comparison to other tests. Namely, we choose as symmetric distributions the Student t_ν -distribution with $\nu \in \{3, 5, 10\}$ degrees of freedom, as well as the uniform distribution $\mathcal{U}(-\sqrt{3}, \sqrt{3})$. The asymmetric distributions are the χ_ν^2 -distribution with $\nu \in \{5, 15\}$ degrees of freedom, the Beta distributions $B(1, 4)$ and $B(2, 5)$, the Gamma distributions $\Gamma(1, 5)$ and $\Gamma(5, 1)$ parametrized by their shape and rate parameter, the Gumbel distribution $\text{Gum}(1, 2)$ with location parameter 1 and scale parameter 2, the lognormal distribution $\text{LN}(0, 1)$ as well as the Weibull distribution $W(1, 0.5)$ with scale parameter 1 and shape parameter 0.5. As representatives of bimodal distributions, we take the mixture of normal distributions $\text{Mix}\mathcal{N}(p, \mu, \sigma^2)$, where the random variables are generated by

$$(1 - p)\mathcal{N}(0, 1) + p\mathcal{N}(\mu, \sigma^2), \quad p \in (0, 1), \mu \in \mathbb{R}, \sigma > 0.$$

Each entry in Table 3 referring to the finite-sample power of the tests is based on 10,000 replications. From the results in this table, we infer that for asymmetric alternative distributions our tests perform almost identical and are extremely stable over the range of tuning parameters. For the symmetric and bimodal alternatives however, the choice of a can have considerable influence on the power of the test. Moreover, with particular focus on the uniform distribution and the normal mixtures, the test based on $G_{n,a}^{(1)}$ shows a significantly better performance than $G_{n,a}^{(2)}$.

Taking this superiority of $G_{n,a}^{(1)}$, and the fact that the choice of tuning parameter influences the power, into account, we propose as a final procedure a test based on $G_{n,a}^{(1)}$ calculated by (9) with a data-dependent choice of the tuning parameter a . To implement the latter, we use the algorithm from Allison and Santana (2015), which has already been applied in the recent simulation study by Allison et al. (2017) for tests of exponentiality. Given the standardized sample Y_1, \dots, Y_n as in Sect. 2, our test is carried out as follows:

- (a) Fix a grid of possible tuning parameters $a \in \{a_1, \dots, a_\ell\}$ (here: $a \in \{0.1, 0.25, 0.5, 1, 1.5, 2, 3\}$).
- (b) Sample from Y_1, \dots, Y_n with replacement and, for the obtained bootstrap sample, calculate $G_{n,a_i}^{(1)}, i = 1, \dots, \ell$, via (9).
- (c) Repeat step (b) B times (here: $B = 400$) and denote the resulting values of the statistic by $G_{1,a_i}^*, \dots, G_{B,a_i}^*, i = 1, \dots, \ell$.
- (d) Calculate the bootstrap powers by $\widehat{P}_{a_i} = B^{-1} \sum_{b=1}^B \mathbb{1}\{G_{b,a_i}^* > c_{n,a_i}(\alpha)\}, i = 1, \dots, \ell$, where $c_{n,a_i}(\alpha)$ is the critical value for a level- α -test based on $G_{n,a_i}^{(1)}$ (for the Monte Carlo approximations, see Table 1).
- (e) Choose as the tuning parameter $\widehat{a} = \arg \max\{\widehat{P}_a | a \in \{a_1, \dots, a_\ell\}\}$ and apply the test based on $G_{n,\widehat{a}}^{(1)}$ to Y_1, \dots, Y_n .

We consider the following competitors to this test. As classical and well-known tests, we include the Shapiro–Wilk test (SW), see Shapiro and Wilk (1965), the Shapiro–Francia test (SF), see Shapiro and Francia (1972), and the Anderson–Darling test (AD), see Anderson and Darling (1952). For the implementation of these tests in R, we refer to the package `nortest` by Gross and Ligges (2015). Tests based on the empirical characteristic function are represented by the Baringhaus–Henze–Epps–Pulley test (BHEP), see Baringhaus and Henze (1988), Epps and Pulley (1983). The BHEP test with tuning parameter $\beta > 0$ is based on

$$\begin{aligned} \text{BHEP} &= \frac{1}{n} \sum_{j,k=1}^n \exp\left(-\frac{\beta^2}{2} (Y_j - Y_k)^2\right) \\ &\quad - \frac{2}{\sqrt{1+\beta^2}} \sum_{j=1}^n \exp\left(-\frac{\beta^2}{2(1+\beta^2)} Y_j^2\right) + \frac{n}{\sqrt{1+2\beta^2}}, \end{aligned}$$

where Y_1, \dots, Y_n is the standardized sample. We fix $\beta = 1$ and take the critical values from Henze (1990) but also restate them in Table 4.

Table 3 Empirical rejection rates for $G_{n,a}^{(k)}$, $k = 1, 2$ ($\alpha = 0.05$, 10,000 replications)

Alt.	$G_{n,a}^{(1)}$							$G_{n,a}^{(2)}$							
	$n \setminus a$	0.1	0.25	0.5	1	1.5	2	3	0.1	0.25	0.5	1	1.5	2	3
$\mathcal{N}(0, 1)$	20	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Mix $\mathcal{N}(0.3, 1, 0.25)$	20	25	25	24	24	23	23	23	22	22	21	20	20	19	19
	50	56	57	57	56	56	55	55	50	50	49	48	47	46	46
	100	86	87	87	87	86	86	86	80	80	79	78	77	77	76
Mix $\mathcal{N}(0.5, 1, 4)$	20	34	35	35	36	37	37	37	32	32	33	33	33	33	34
	50	52	56	60	63	64	64	65	42	43	44	46	47	48	49
	100	75	84	89	91	92	92	92	55	57	60	66	68	70	71
t_3	20	30	31	33	34	35	35	35	32	32	33	34	34	35	35
	50	41	46	50	54	56	57	58	44	45	47	50	52	53	54
	100	54	63	70	76	78	79	81	53	56	61	67	70	72	74
t_5	20	16	17	18	19	19	19	20	18	18	19	19	19	19	20
	50	22	25	27	29	31	31	32	26	26	27	29	30	31	31
	100	27	31	36	41	43	44	46	30	31	34	37	39	41	42
t_{10}	20	9	9	10	10	10	10	10	10	10	10	10	11	11	11
	50	11	11	12	13	14	15	15	13	13	13	14	15	15	16
	100	11	12	14	16	17	17	18	14	14	15	16	17	17	18
$\mathcal{U}(-\sqrt{3}, \sqrt{3})$	20	4	4	3	3	3	3	3	2	2	2	2	1	1	1
	50	3	4	5	7	7	8	8	2	2	2	2	1	1	1
	100	5	10	19	32	38	41	45	2	2	3	3	3	3	3
χ_5^2	20	44	45	45	45	44	44	44	45	45	44	44	43	43	43
	50	84	86	87	87	87	87	87	87	87	87	87	87	87	87
	100	99	99	99	99	99	99	99	99	99	99	99	99	99	99
χ_{15}^2	20	18	18	19	19	19	19	19	19	19	19	19	19	19	19
	50	44	45	46	46	46	46	46	47	47	47	47	46	46	46
	100	74	76	77	78	78	78	78	78	78	79	78	78	78	78
$B(1, 4)$	20	49	51	51	51	51	50	50	49	49	49	48	47	47	46
	50	90	93	94	94	94	94	94	92	92	92	92	92	92	92
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$B(2, 5)$	20	15	16	15	15	15	15	15	15	15	15	15	14	14	14
	50	40	42	43	44	44	43	43	42	42	42	42	41	41	41
	100	73	77	79	80	81	81	81	76	77	78	78	78	78	78
$\Gamma(1, 5)$	20	76	78	78	78	78	78	78	77	77	76	76	75	75	74
	50	99	100	100	100	100	100	100	99	99	100	100	100	100	99
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
$\Gamma(5, 1)$	20	25	25	25	25	25	25	25	26	26	26	25	25	25	25
	50	58	60	61	61	61	61	61	62	62	62	62	62	61	61
	100	88	90	90	91	91	91	91	91	91	91	91	91	91	91

Table 3 continued

Alt.	$n \setminus a$	$G_{n,a}^{(1)}$						$G_{n,a}^{(2)}$							
		0.1	0.25	0.5	1	1.5	2	3	0.1	0.25	0.5	1	1.5	2	3
W(1, 0.5)	20	76	78	78	78	78	78	78	77	77	77	76	75	75	75
	50	99	100	100	100	100	100	100	99	99	100	100	99	99	99
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100
Gum(1, 2)	20	32	33	33	33	33	33	33	34	34	34	33	33	33	33
	50	70	71	72	72	72	72	72	73	73	73	72	72	72	72
	100	94	95	95	95	96	96	96	96	96	96	96	96	96	96
LN(0, 1)	20	90	91	91	91	91	91	91	90	90	90	90	90	89	89
	50	100	100	100	100	100	100	100	100	100	100	100	100	100	100
	100	100	100	100	100	100	100	100	100	100	100	100	100	100	100

Furthermore, we include the quantile correlation test of del Barrio–Cuesta–Albertos–Mátran–Rodríguez–Rodríguez (BCMR) based on the L^2 -Wasserstein distance, see del Barrio et al. (1999) and Sect. 3.3 of del Barrio et al. (2000). The BCMR statistic is given by

$$BCMR = n \left(1 - \frac{1}{S_n^2} \left(\sum_{k=1}^n X_{(k)} \int_{\frac{k-1}{n}}^{\frac{k}{n}} \Phi^{-1}(t) dt \right)^2 \right) - \int_{\frac{1}{n+1}}^{\frac{n}{n+1}} \frac{t(1-t)}{(\varphi(\Phi^{-1}(t)))^2} dt,$$

where $X_{(k)}$ is the k -th order statistic of X_1, \dots, X_n , S_n^2 is the sample variance and Φ^{-1} is the quantile function of the standard normal distribution. Simulated critical values be found in the work of Krauczi (2009), or in Table 4.

The Henze–Jiménez-Gamero test (HJG), see Henze and Jiménez-Gamero (2018), uses a weighted L^2 -distance between the empirical moment-generating function of the standardized sample and the moment-generating function of the standard normal distribution. The test is based on

$$HJG_\beta = \frac{1}{n\sqrt{\beta}} \sum_{j,k=1}^n \exp\left(\frac{(Y_j + Y_k)^2}{4\beta}\right) - \frac{2}{\sqrt{\beta-1/2}} \sum_{j=1}^n \exp\left(\frac{Y_j^2}{4\beta-2}\right) + \frac{n}{\sqrt{\beta-1}}$$

with $\beta > 2$. We consider the tuning parameters $\beta \in \{2.5, 5, 10\}$. Since Henze and Jiménez-Gamero (2018) did not simulate critical values in the univariate case, the empirical critical values can be found in Table 4. This test was proposed recently, so it is not yet included in any other power study. All of the simulated critical values displayed in Table 4 have been confirmed in a simulation with 100,000 replications (compare to Henze 1990; Krauczi 2009).

In Table 5, we display the results of the competitive simulation study, where our test based on the steps (a)–(e) (with bootstrap size $B = 400$ and values for a as

Table 4 Empirical 0.95 quantiles for BCMR, BHEP and HJG $_{\beta}$ under H_0 (100,000 replications)

$n \setminus$ Test	BCMR	BHEP	HJG $_{2.5}$	HJG $_5$	HJG $_{10}$
20	0.31	0.368	0.1503	0.2568	0.3258
50	0.30	0.374	6.349e-3	8.568e-3	9.556e-3
100	0.29	0.376	4.001e-4	4.988e-4	5.369e-4

Table 5 Empirical rejection rates for competing procedures ($\alpha = 0.05$, 10,000 replications)

Alt.	n	BE \hat{a}	SW	BCMR	BHEP	AD	SF	HJG $_{2.5}$	HJG $_5$	HJG $_{10}$
$\mathcal{N}(0, 1)$	20	5	5	5	5	5	5	5	5	5
	50	5	5	5	5	5	5	5	5	5
	100	5	5	5	5	5	5	5	5	5
Mix $\mathcal{N}(0.3, 1, 0.25)$	20	23	28	28	27	30	25	11	13	14
	50	54	60	60	62	68	57	16	26	32
	100	85	89	89	90	94	88	28	49	58
Mix $\mathcal{N}(0.5, 1, 4)$	20	36	40	43	42	46	48	34	33	33
	50	64	78	80	80	86	83	49	49	46
	100	92	97	98	98	99	98	69	68	61
t_3	20	34	35	37	34	33	40	38	37	36
	50	57	64	65	61	60	69	64	62	59
	100	81	88	89	86	85	91	86	84	78
t_5	20	20	19	20	18	17	22	22	22	21
	50	31	35	37	32	31	41	40	38	36
	100	46	56	58	50	48	63	59	55	50
t_{10}	20	11	10	11	9	9	12	12	12	12
	50	14	16	17	13	12	20	20	19	18
	100	19	22	24	16	15	28	28	26	23
$\mathcal{U}(-\sqrt{3}, \sqrt{3})$	20	3	21	17	13	17	8	0	0	0
	50	8	75	70	55	58	47	0	0	0
	100	45	100	99	95	95	97	0	0	0
χ^2_5	20	44	44	44	42	38	42	33	36	39
	50	87	89	88	84	80	85	65	76	80
	100	100	100	100	99	99	100	91	98	99
χ^2_{15}	20	19	18	18	17	16	18	16	17	18
	50	45	42	42	39	33	40	32	38	41
	100	77	75	74	68	61	71	54	68	73
$B(1, 4)$	20	50	59	58	52	51	53	28	34	38
	50	94	98	98	94	95	97	57	76	83
	100	100	100	100	100	100	100	89	99	100

Table 5 continued

Alt.	n	$BE_{\hat{a}}$	SW	BCMR	BHEP	AD	SF	HJG _{2,5}	HJG ₅	HJG ₁₀
$B(2, 5)$	20	15	16	16	16	14	14	9	11	12
	50	43	50	47	45	39	40	16	25	30
	100	81	90	89	80	76	82	29	54	64
$\Gamma(1, 5)$	20	77	83	83	77	77	80	57	63	67
	50	100	100	100	100	100	100	91	97	98
	100	100	100	100	100	100	100	100	100	100
$\Gamma(5, 1)$	20	25	24	24	23	20	24	20	22	23
	50	61	59	59	55	49	56	42	51	55
	100	91	90	90	85	81	88	69	83	87
$W(1, 0.5)$	20	100	84	83	78	77	80	58	64	67
	50	100	100	100	100	100	100	91	97	98
	100	100	100	100	100	100	100	100	100	100
Gum(1, 2)	20	34	31	32	31	28	32	28	30	31
	50	72	69	69	66	60	67	55	64	67
	100	96	94	94	91	89	93	83	92	94
LN(0, 1)	20	91	93	93	91	90	91	78	83	85
	50	100	100	100	100	100	100	99	100	100
	100	100	100	100	100	100	100	100	100	100

before) is denoted by $BE_{\hat{a}}$. Each entry is based on 10,000 Monte Carlo replications, and the best-performing test for each distribution and sample size is highlighted for easy reference.

Starting with the symmetric distributions, we see that the SF and SW tests perform best for these models. Interestingly, the HJG_{β} test has the highest power against Students t_{10} -distribution but completely fails to detect the uniform alternative. The finite-sample power of our new test for the t_{ν} -distributions is comparable to the BHEP test, but the uniform distribution seems to be a weak spot. Bimodal distributions are best detected by the AD test. The performance of the SW, BCMR, BHEP and SF tests is comparable, while the new $BE_{\hat{a}}$ procedure has a slightly weaker power and the HJG_{β} test is clearly inferior for those distributions. Considering the asymmetric alternatives, our new procedure shows its potential by dominating all other procedures for the χ^2 -, the Gamma as well as the Gumbel distributions. All procedures do a good job in rejecting the Weibull and the lognormal alternatives.

To conclude the simulation study, we investigate the confidence interval (22) and the power approximation in (23) for the fixed tuning parameter $a = 1$. As an example, we examine three alternatives also partially considered by Baringhaus et al. (2017), namely the uniform distribution $\mathcal{U}(-\sqrt{3}, \sqrt{3})$ and the Laplace distribution $L(0, 1/\sqrt{2})$ with density $p(x) = \exp(-\sqrt{2}|x|)$, $x \in \mathbb{R}$, as well as the Logistic distribution $Lo(0, \sqrt{3}/\pi)$ with density $p(x) =$

Table 6 Empirical power and approximation (23) against three alternatives

n	$\mathcal{U}(-\sqrt{3}, \sqrt{3})$				$L(0, 1/\sqrt{2})$				$L_o(0, \sqrt{3}/\pi)$			
	$G_{n,1}^{(1)}$		$G_{n,1}^{(2)}$		$G_{n,1}^{(1)}$		$G_{n,1}^{(2)}$		$G_{n,1}^{(1)}$		$G_{n,1}^{(2)}$	
	MC	Apr	MC	Apr	MC	Apr	MC	Apr	MC	Apr	MC	Apr
20	3	0	2	0	25	6	25	0	12	0	13	0
50	7	0	2	0	40	27	34	10	16	0	16	0
100	32	10	3	0	62	53	44	57	21	0	20	0

$\pi \exp(-\pi x/\sqrt{3})/\sqrt{3} \left(1 + \exp(-\pi x/\sqrt{3})\right)^2$, $x \in \mathbb{R}$. Notice that these alternatives are standardized. The values of $\Delta^{(k)}$ are 0.004167 ($k = 1$) and 0.000985 ($k = 2$) for $\mathcal{U}(-\sqrt{3}, \sqrt{3})$, 0.005225 ($k = 1$) and 0.001367 ($k = 2$) for $L(0, 1/\sqrt{2})$, and 0.000819 ($k = 1$) and 0.000241 ($k = 2$) for $L_o(0, \sqrt{3}/\pi)$.

Note that the uniform and the Laplace distribution do not satisfy the differentiability condition of Theorem 5, but since they do satisfy all conditions of Corollary 2, we include simulations for $G_{n,1}^{(1)}$ in those cases as well. The results in Tables 6 and 7 however, when compared to the Logistic distribution or the second statistic $G_{n,1}^{(2)}$ where all requirements are formally fulfilled, indicate that Theorem 5 should also cover the uniform and, in particular, the Laplace distribution, and thus the result might hold under weaker conditions.

Since the limit variance $\tau_{(k)}^2$, $k = 1, 2$, seems inaccessible by computation in each case, we decided to estimate $\tau_{(k)}^2$, $k = 1, 2$, by means of simulation. For a sample size of $n = 1000$ with 10,000 repetitions, the estimated values are 0.000302 ($k = 1$) and 0.000016 ($k = 2$) for $\mathcal{U}(-\sqrt{3}, \sqrt{3})$, 0.002452 ($k = 1$) and 0.000565 ($k = 2$) for $L(0, 1/\sqrt{2})$, as well as 0.000430 ($k = 1$) and 0.000125 ($k = 2$) for $L_o(0, \sqrt{3}/\pi)$. Table 6 displays the empirical power (in %) of $G_{n,1}^{(1)}$ and $G_{n,1}^{(2)}$ against the three alternatives. The nominal level is $1 - \alpha = 0.95$ which explains the considerably smaller power compared to Table 2 from Baringhaus et al. (2017). The columns denoted by 'Apr' show the corresponding approximations given by (23), while 'MC' stands for the empirical rejection rates for 10,000 repetitions. Table 6 shows that the approximate power function (23) often appears as a lower bound to the power of the test statistics, confirming the observations by Baringhaus et al. (2017).

Because no consistent estimator $\hat{\tau}_{(k),n}^2(X_1, \dots, X_n)$ of $\tau_{(k)}^2$, $k = 1, 2$, is known, a nonparametric bootstrap procedure with $B = 500$ bootstrap samples (drawn with repetition) is implemented to calculate the empirical coverage probabilities shown in Table 7. The confidence level is set to $1 - \alpha = 0.9$, and each value is based on 10,000 repetitions. Obviously the empirical coverage probabilities are higher than the confidence level, indicating that the approximate lower and upper bounds of the confidence intervals are too conservative. This might be an effect of the bootstrap variance estimation procedure, but regarding the results of the power approximation,

Table 7 Empirical coverage probabilities of I_n from (22) for $\Delta^{(k)}$, $k = 1, 2$, (at nominal level 0.9, with 10,000 replications)

n	$\mathcal{U}(-\sqrt{3}, \sqrt{3})$		$L(0, 1/\sqrt{2})$		$Lo(0, \sqrt{3}/\pi)$	
	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(1)}$	$\Delta^{(2)}$	$\Delta^{(1)}$	$\Delta^{(2)}$
20	99	100	92	91	95	96
50	99	100	96	96	98	98
100	97	100	98	98	99	99

we rather think that an appropriate error-correction term in (22) and (23) will lead to better results.

7 Conclusions and outlines

Starting with Charles Stein's insight that a random variable X has a standard normal distribution if, and only if,

$$\mathbb{E}[f'(X)] = \mathbb{E}[Xf(X)]$$

holds for any absolutely continuous function, we developed two classes of goodness-of-fit statistics for testing the normality hypothesis. We utilized the zero-bias transformation to bypass the problem of calculating an empirical property for all absolutely continuous functions. An advantage of the underlying zero-bias identity over many other types of transformation applied in goodness-of-fit testing, like the characteristic function or the Laplace transform, is that the distribution inserted into the mapping is not associated with a purely analytic quantity but is mapped to another distribution and, thereby, stays accessible to a stochastically intuitive examination (cf. Lemmata 1 and 4). The conducted power study suggests that our tests are serious competitors to established tests and even set new markers in terms of the highest power achieved for many asymmetrical alternatives. Both procedures are consistent against any alternative distribution satisfying a weak moment condition.

We want to emphasize that some problems remain open for further research. One issue concerns our choice of weight function. The integrals figuring in the second sum of $G_{n,a}^{(2)}$ in (10), though they are accessible to stable numerical integration, are a slight drawback in terms of calculation time as compared to $G_{n,a}^{(1)}$. It is conceivable to replace the term $\omega(t)dt$ in (5) and (6) by $dF(t)$ and to estimate F by the empirical distribution function. However, this type of test is not included in the framework for our theoretical results. Another question is whether there is some limiting statistic, as $a \rightarrow \infty$, when considering $G_{n,a}^{(2)}$ from Sect. 2. Finally, since we have not succeeded in calculating consistent estimators for $\tau_{(1)}^2$ and $\tau_{(2)}^2$ (see the Remark in Sect. 5), it remains to derive appropriate estimators and, in view of the results in Tables 6 and 7, to find better power approximations as well as suitable confidence intervals.

Acknowledgements The authors thank Norbert Henze for useful comments and also express their gratitude to three anonymous referees for careful reading and suggestions that helped improve the article.

A Preliminary results concerning the weight functions

We first prove that the density function of a centred normal distribution is an admissible weight function. Then, we give a general result for the asymptotic behaviour of integral terms involving weight functions of the type we consider. In the whole section, we adopt the setting and notation from Sect. 2.

Lemma 2 *The functions $\omega_a(s) = (2\pi a)^{-1/2} \exp(-s^2/(2a))$, $s \in \mathbb{R}$, $a > 0$, satisfy the weight function conditions stated in Sect. 2.*

Proof The only non-trivial statement is that ω_a satisfies (8). Let $0 < \varepsilon < 1/8$ be arbitrary. In the case $|S_n^{-1} - 1| \leq \varepsilon$ and $|\bar{X}_n|/S_n \leq \varepsilon$, a Taylor expansion gives

$$\omega_a \left(\frac{s - \bar{X}_n}{S_n} \right) - \omega_a(s) = \omega'_a(\xi_n(s)) \left(\frac{s - \bar{X}_n}{S_n} - s \right), \tag{24}$$

where $|\xi_n(s) - s| \leq |(s - \bar{X}_n)/S_n - s| \leq (|s| + 1)/8$. Consequently,

$$\begin{aligned} (\xi_n(s))^2 - s^2 &\geq \min \left\{ \left| s - \frac{|s| + 1}{8} \right|, \left| s + \frac{|s| + 1}{8} \right| \right\}^2 - s^2 \\ &= -\frac{15}{64}s^2 - \frac{7}{32}|s| + \frac{1}{64} \end{aligned}$$

from which we conclude

$$\begin{aligned} \frac{|\omega'_a(\xi_n(s))|^3}{(\omega_a(s))^2} &= \frac{|\xi_n(s)|^3}{a^3 \sqrt{2\pi a}} \exp \left(-\frac{3}{2a} \left((\xi_n(s))^2 - s^2 \right) - \frac{1}{2a}s^2 \right) \\ &\leq \frac{1}{a^3 \sqrt{2\pi a}} (2|s| + 1)^3 \exp \left(-\frac{s^2}{8a} + \frac{|s|}{a} \right). \end{aligned}$$

Combining this with (24),

$$\begin{aligned} &n \int_{\mathbb{R}} \left| \omega_a \left(\frac{s - \bar{X}_n}{S_n} \right) - \omega_a(s) \right|^3 (\omega_a(s))^{-2} ds \\ &\leq \varepsilon \int_{\mathbb{R}} n \left| \left(\frac{1}{S_n} - 1 \right) s - \frac{\bar{X}_n}{S_n} \right|^2 \frac{(2|s| + 1)^4}{a^3 \sqrt{2\pi a}} \exp \left(-\frac{s^2}{8a} + \frac{|s|}{a} \right) ds. \end{aligned}$$

As ε was arbitrary, the claim follows from the boundedness in probability of $\sqrt{n}(S_n^{-1} - 1)$ and $\sqrt{n}(\bar{X}_n/S_n)$. □

Lemma 3 Let \mathcal{U}_n be a random element of \mathcal{H} , $n \in \mathbb{N}$, such that $\|\sqrt{n}\mathcal{U}_n\|_{\mathcal{H}} = O_{\mathbb{P}}(1)$. Then,

$$\int_{\mathbb{R}} |\sqrt{n}\mathcal{U}_n(s)| \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) - \omega(s) \right| ds = o_{\mathbb{P}}(1).$$

If in addition $\sup_{s \in \mathbb{R}} |\mathcal{U}_n(s)| \leq C$ \mathbb{P} -a.s. for each $n \in \mathbb{N}$ and some $C > 0$,

$$\int_{\mathbb{R}} |\sqrt{n}\mathcal{U}_n(s)|^2 \omega\left(\frac{s - \bar{X}_n}{S_n}\right) ds = \|\sqrt{n}\mathcal{U}_n\|_{\mathcal{H}}^2 + o_{\mathbb{P}}(1).$$

Proof By Hölder's inequality ($p = q = 2$) and Slutsky's lemma

$$\begin{aligned} & \int_{\mathbb{R}} |\sqrt{n}\mathcal{U}_n(s)| \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) - \omega(s) \right| ds \\ & \leq \|\sqrt{n}\mathcal{U}_n\|_{\mathcal{H}} \left(\int_{\mathbb{R}} \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) / \omega(s) - 1 \right|^2 \omega(s) ds \right)^{1/2} \\ & = o_{\mathbb{P}}(1), \end{aligned}$$

where we used the assumption on \mathcal{U}_n and the fact that (8) implies

$$\begin{aligned} & \int_{\mathbb{R}} \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) / \omega(s) - 1 \right|^2 \omega(s) ds \\ & \leq \left(\int_{\mathbb{R}} \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) / \omega(s) - 1 \right|^3 \omega(s) ds \right)^{2/3} \left(\int_{\mathbb{R}} \omega(s) ds \right)^{1/3} \\ & = o_{\mathbb{P}}(1). \end{aligned}$$

The second claim also follows from Hölder's inequality ($p = 3/2$, $q = 3$) and (8) since

$$\begin{aligned} & \left| \int_{\mathbb{R}} |\sqrt{n}\mathcal{U}_n(s)|^2 \omega\left(\frac{s - \bar{X}_n}{S_n}\right) ds - \|\sqrt{n}\mathcal{U}_n\|_{\mathcal{H}}^2 \right| \\ & \leq n \int_{\mathbb{R}} |\mathcal{U}_n(s)|^2 (\omega(s))^{2/3} \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) / \omega(s) - 1 \right| (\omega(s))^{1/3} ds \\ & \leq n^{2/3} \left(\int_{\mathbb{R}} |\mathcal{U}_n(s)|^3 \omega(s) ds \right)^{2/3} n^{1/3} \left(\int_{\mathbb{R}} \left| \omega\left(\frac{s - \bar{X}_n}{S_n}\right) / \omega(s) - 1 \right|^3 \omega(s) ds \right)^{1/3} \end{aligned}$$

$$\begin{aligned} &\leq C^{2/3} \|\sqrt{n} \mathcal{U}_n\|_{\mathcal{H}}^{4/3} \left(n \int_{\mathbb{R}} \left| \omega \left(\frac{s - \bar{X}_n}{S_n} \right) - \omega(s) \right|^3 (\omega(s))^{-2} ds \right)^{1/3} \\ &= o_{\mathbb{P}}(1). \end{aligned}$$

□

B Asymptotic expansions

We adopt the setting from Sect. 2, that is, we let X, X_1, X_2, \dots be iid. random variables with distribution function F and $\mathbb{E}[X^2] < \infty$ as well as $\mathbb{E}X = 0, \mathbb{V}(X) = 1$. The following lemma collects basic facts about a quantity closely related to the empirical zero-bias distribution function.

Lemma 4 *The function*

$$\widehat{F}_n^X(s) = \frac{1}{n} \sum_{j=1}^n \frac{X_j - \bar{X}_n}{S_n^2} (X_j - s) \mathbb{1}\{X_j \leq s\}, \quad s \in \mathbb{R},$$

is a continuous distribution function for each $n \in \mathbb{N}$ (and on a set of measure one). Furthermore,

$$\sup_{s \in \mathbb{R}} \left| \widehat{F}_n^X(s) - F^X(s) \right| \longrightarrow 0 \tag{25}$$

\mathbb{P} -a.s., as $n \rightarrow \infty$, and

$$\sqrt{n} \widehat{F}_n^X(s) \approx \frac{\sqrt{n}}{S_n^2} \left\{ \frac{1}{n} \sum_{j=1}^n X_j (X_j - s) \mathbb{1}\{X_j \leq s\} - \bar{X}_n \mathbb{E}[(X - s) \mathbb{1}\{X \leq s\}] \right\}. \tag{26}$$

Proof We fix $n \in \mathbb{N}$ and notice that

$$\widehat{d}_n^X(s) = \frac{1}{n} \sum_{j=1}^n \frac{X_j - \bar{X}_n}{S_n^2} \mathbb{1}\{X_j > s\} = -\frac{1}{n} \sum_{j=1}^n \frac{X_j - \bar{X}_n}{S_n^2} \mathbb{1}\{X_j \leq s\} (\geq 0).$$

Using the first representation when integrating over (\bar{X}_n, ∞) and the second for $(-\infty, \bar{X}_n]$, we obtain

$$\int_{\mathbb{R}} \widehat{d}_n^X(t) dt = \frac{1}{S_n^2} \left(\frac{1}{n} \sum_{j=1}^n (X_j - \bar{X}_n)^2 \right) = 1.$$

Now, we conclude from

$$\int_{-\infty}^s \widehat{d}_n^X(t) dt = -\frac{1}{n} \sum_{j=1}^n \frac{X_j - \bar{X}_n}{S_n^2} \int_{-\infty}^s \mathbb{1}\{X_j \leq t\} dt = \widehat{F}_n^X(s)$$

that \widehat{F}_n^X is a continuous distribution function. By the strong law of large numbers and the almost sure convergence $(\bar{X}_n, S_n^2) \rightarrow (0, 1)$, we have

$$\begin{aligned} \widehat{F}_n^X(s) &= \frac{1}{S_n^2} \cdot \frac{1}{n} \sum_{j=1}^n X_j (X_j - s) \mathbb{1}\{X_j \leq s\} \\ &\quad - \frac{\bar{X}_n}{S_n^2} \cdot \frac{1}{n} \sum_{j=1}^n (X_j - s) \mathbb{1}\{X_j \leq s\} \\ &\longrightarrow F^X(s) \end{aligned}$$

\mathbb{P} -a.s., as $n \rightarrow \infty$, for any fixed $s \in \mathbb{R}$. The proof of the classical Glivenko–Cantelli theorem applies to \widehat{F}_n^X which yields (25). For the last claim, we set

$$A_n(s) = \frac{1}{n} \sum_{j=1}^n (X_j - s) \mathbb{1}\{X_j \leq s\} - \mathbb{E}[(X - s) \mathbb{1}\{X \leq s\}], \quad s \in \mathbb{R}.$$

Straightforward calculations using Tonelli's theorem and the integrability condition (7) give

$$\mathbb{E} \left[\int_{\mathbb{R}} A_n(s)^2 \omega(s) ds \right] \longrightarrow 0, \quad \text{as } n \rightarrow \infty,$$

so $\|A_n\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$. Together with $\sqrt{n} \bar{X}_n = O_{\mathbb{P}}(1)$ and Slutsky's lemma, this implies (26). \square

We proceed by proving further asymptotic expansions of the same type as (26).

Lemma 5 *Assume, in addition to the above prerequisites, that X has a continuously differentiable density function p with*

$$\sup_{s \in \mathbb{R}} |p(s)| \leq K_1 < \infty \quad \text{and} \quad \sup_{s \in \mathbb{R}} |p'(s)| \leq K_2 < \infty.$$

We have

$$\sqrt{n} F \left(\frac{s - \bar{X}_n}{S_n} \right) \approx \sqrt{n} \left\{ F(s) + p(s) \left(\left(\frac{1}{S_n} - 1 \right) s - \frac{\bar{X}_n}{S_n} \right) \right\}$$

and, with F^X as in Lemma 1,

$$\sqrt{n} F^X \left(\frac{s - \bar{X}_n}{S_n} \right) \approx \sqrt{n} \left\{ F^X(s) + d^X(s) \left(\left(\frac{1}{S_n} - 1 \right) s - \frac{\bar{X}_n}{S_n} \right) \right\}.$$

Moreover,

$$\sqrt{n} S_n^2 F^X(s) \approx \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j^2 F^X(s)$$

which reads as $\sqrt{n} S_n^2 \Phi(s) \approx n^{-1/2} \sum_{j=1}^n X_j^2 \Phi(s)$ when $\mathbb{P}^X = \mathcal{N}(0, 1)$ (cf. Theorem 1).

Proof By Taylor’s theorem,

$$\sqrt{n} F \left(\frac{s - \bar{X}_n}{S_n} \right) = \sqrt{n} \left\{ F(s) + p(s) \left(\frac{s - \bar{X}_n}{S_n} - s \right) \right\} + R_n(s),$$

where

$$R_n(s) = \sqrt{n} \frac{p'(\xi_n(s))}{2} \left(\frac{s - \bar{X}_n}{S_n} - s \right)^2$$

and $|\xi_n(s) - s| \leq |(s - \bar{X}_n)/S_n - s|$. Condition (7) assures that $R_n \in \mathcal{H}$ \mathbb{P} -a.s. and with $\sqrt{n}(S_n^{-1} - 1) = O_{\mathbb{P}}(1)$, $\sqrt{n} \bar{X}_n = O_{\mathbb{P}}(1)$ we conclude

$$\|R_n\|_{\mathcal{H}}^2 \leq \frac{K_2^2}{4} \int_{\mathbb{R}} n \left| \left(\frac{1}{S_n} - 1 \right) s - \frac{\bar{X}_n}{S_n} \right|^4 \omega(s) ds = o_{\mathbb{P}}(1).$$

Now, let $0 < \varepsilon < 1$ be arbitrary. In the case $|S_n^{-1} - 1| \leq \varepsilon$ and $|\bar{X}_n|/S_n \leq \varepsilon$, we have

$$\sqrt{n} F^X \left(\frac{s - \bar{X}_n}{S_n} \right) = \sqrt{n} \left\{ F^X(s) + d^X(s) \left(\frac{s - \bar{X}_n}{S_n} - s \right) \right\} + \tilde{R}_n(s),$$

where

$$\tilde{R}_n(s) = -\frac{\sqrt{n}}{2} \tilde{\xi}_n(s) p(\tilde{\xi}_n(s)) \left(\frac{s - \bar{X}_n}{S_n} - s \right)^2$$

and $|\tilde{\xi}_n(s) - s| \leq |(s - \bar{X}_n)/S_n - s| \leq |s| + 1$. Using $(\tilde{\xi}_n(s))^2 \leq (2|s| + 1)^2$, we get

$$\begin{aligned} \|\tilde{R}_n\|_{\mathcal{H}}^2 &\leq \frac{K_1^2}{4} \int_{\mathbb{R}} n \left| \frac{s - \bar{X}_n}{S_n} - s \right|^4 (2|s| + 1)^2 \omega(s) \, ds \\ &\leq \frac{\varepsilon^2 K_1^2}{4} \int_{\mathbb{R}} n \left| \left(\frac{1}{S_n} - 1 \right) s - \frac{\bar{X}_n}{S_n} \right|^2 (2|s| + 1)^4 \omega(s) \, ds. \end{aligned}$$

Since $\sqrt{n}(S_n^{-1} - 1)$ and $\sqrt{n}(\bar{X}_n/S_n)$ are bounded in probability and ε was arbitrary, $\|\tilde{R}_n\|_{\mathcal{H}}^2 = o_{\mathbb{P}}(1)$. The last claim of the lemma follows from

$$\left\| \sqrt{n} S_n^2 F^X - \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j^2 F^X \right\|_{\mathcal{H}} = \sqrt{n} \bar{X}_n^2 \|F^X\|_{\mathcal{H}} = o_{\mathbb{P}}(1).$$

□

C Proof of the limit relations in (11) and (12)

We will give the proof of (11), using the notation from Sect. 2. The limit in (12) is obtained by the same argument. Set

$$g(s) = s^{-1/2} \left(\frac{1}{n} \sum_{j=1}^n (Y_j(Y_j - \sqrt{2s}) - 1) \mathbb{1}\{Y_j \leq \sqrt{2s}\} \right)^2, \quad s > 0,$$

as well as

$$\tilde{g}(s) = s^{-1/2} \left(\frac{1}{n} \sum_{j=1}^n (Y_j(Y_j + \sqrt{2s}) - 1) \mathbb{1}\{Y_j \leq -\sqrt{2s}\} \right)^2, \quad s > 0.$$

Splitting the integral in the definition of $G_{n,a}^{(1)}$ (see (5)) into integrals over $(-\infty, 0]$ and $(0, \infty)$, simple changes of variable yield

$$\begin{aligned} \lim_{a \searrow 0} G_{n,a}^{(1)} &= \lim_{a \searrow 0} \frac{n}{2\sqrt{\pi}} \left(a^{-1/2} \int_0^\infty g(s) e^{-s/a} \, ds + a^{-1/2} \int_0^\infty \tilde{g}(s) e^{-s/a} \, ds \right) \\ &= \lim_{a \rightarrow \infty} \frac{n}{2\sqrt{\pi}} \left(a^{1/2} \int_0^\infty g(s) e^{-as} \, ds + a^{1/2} \int_0^\infty \tilde{g}(s) e^{-as} \, ds \right). \end{aligned}$$

Since the integrals on the right-hand side of the above equation are Laplace transforms, and since we have

$$\lim_{s \searrow 0} \Gamma(1/2) s^{1/2} g(s) = \sqrt{\pi} \left(\frac{1}{n} \sum_{j=1}^n (Y_j^2 - 1) \mathbb{1}\{Y_j \leq 0\} \right)^2$$

and

$$\lim_{s \searrow 0} \Gamma(1/2) s^{1/2} \tilde{g}(s) = \sqrt{\pi} \left(\frac{1}{n} \sum_{j=1}^n (Y_j^2 - 1) \mathbb{1}\{Y_j < 0\} \right)^2,$$

an Abelian theorem for the Laplace transform, as stated on p. 182 in the book by Widder (1959) (see also Baringhaus et al. 2000), implies the claim. Here, $\Gamma(1/2) = \sqrt{\pi}$ denotes the Gamma function evaluated at $1/2$.

References

- Allison JS, Santana L (2015) On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *J Stat Comput Simul* 85(16):3276–3288
- Allison JS, Santana L, Smit N, Visagie IJH (2017) An ‘apples to apples’ comparison of various tests for exponentiality. *Comput Stat* 32(4):1241–1283
- Anderson TW, Darling DA (1952) Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann Math Stat* 23:193–212
- Baringhaus L, Henze N (1988) A consistent test for multivariate normality based on the empirical characteristic function. *Metrika* 35(1):339–348
- Baringhaus L, Henze N (2017) Cramér–von Mises distance: probabilistic interpretation, confidence intervals, and neighbourhood-of-model validation. *J Nonparametr Stat* 29(2):167–188
- Baringhaus L, Danschke R, Henze N (1989) Recent and classical tests for normality—a comparative study. *Commun Stat Simul Comput* 18(1):363–379
- Baringhaus L, Gürtler N, Henze N (2000) Weighted integral test statistics and components of smooth tests of fit. *Aust N Z J Stat* 42(2):179–192
- Baringhaus L, Ebner B, Henze N (2017) The limit distribution of weighted L^2 -goodness-of-fit statistics under fixed alternatives, with applications. *Ann Inst Stat Math* 69(5):969–995
- Bera AK, Galvao AF, Wang L, Xiao Z (2016) A new characterization of the normal distribution and test for normality. *Econom Theory* 32(5):1216–1252
- Billingsley P (1995) *Probability and measure*, 3rd edn. Wiley, Hoboken
- Chapman DG (1958) A comparative study of several one-sided goodness-of-fit tests. *Ann Math Stat* 29(3):655–674
- Chen LHY, Goldstein L, Shao QM (2011) *Normal approximation by Steins method*. Probability and its applications. Springer, Berlin
- del Barrio E, Cuesta-Albertos JA, Matrán C, Rodríguez-Rodríguez JM (1999) Tests of goodness of fit based on the L_2 -Wasserstein distance. *Ann Stat* 27(4):1230–1239
- del Barrio E, Cuesta-Albertos JA, Matrán C, Csörgö S, Cuadras CM, de Wet T, Giné E, Lockhart R, Munk A, Stute W (2000) Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *TEST* 9(1):1–96
- Epps TW, Pulley LB (1983) A test for normality based on the empirical characteristic function. *Biometrika* 70(3):723–726
- Farrell PJ, Rogers-Stewart K (2006) Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *J Stat Comput Simul* 76(9):803–816
- Goldstein L, Reinert G (1997) Stein’s method and the zero bias transformation with application to simple random sampling. *Ann Appl Probab* 7(4):935–952

- Gross J, Ligges U (2015) nortest: tests for normality. R package version 1.0-4
- Hájek J, Šidák Z, Sen PK (1999) Theory of rank tests. Academic Press, Cambridge, Probability and Mathematical Statistics
- Henze N (1990) An approximation to the limit distribution of the Epps–Pulley test statistic for normality. *Metrika* 37:7–18
- Henze N (1994) Tests of normality (in German). *Allgemeines statistisches Archiv. J German Stat Soc* 78(3):293–317
- Henze N (2002) Invariant tests for multivariate normality: a critical review. *Stat Pap* 43(4):467–506
- Henze N, Wagner T (1997) A new approach to the BHEP tests for multivariate normality. *J Multivariate Anal* 62:1–23
- Henze N, Zirkler B (1990) A class of invariant consistent tests for multivariate normality. *Commun Stat Theory Methods* 19(10):3595–3617
- Henze N, Jiménez-Gamero MD (2018) A new class of tests for multinormality with iid. and Garch data based on the empirical moment generating function. *TEST*. <https://doi.org/10.1007/s11749-018-0589-z>
- Klar B (2001) Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. *Ann Inst Stat Math* 53(2):338–353
- Krauczi Ě (2009) A study of the quantile correlation test for normality. *TEST* 18(1):156–165
- Landry L, Lepage Y (1992) Empirical behavior of some tests for normality. *Commun Stat Simul Comput* 21(4):971–999
- Ledoux M, Talagrand M (2011) Probability in Banach spaces. Isoperimetry and processes. Springer, Berlin
- Liu Q, Lee JD, Jordan M (2016) A kernelized Stein discrepancy for goodness-of-fit tests. In: Proceedings of the 33rd international conference on international conference on machine learning, vol 48, pp 276–284
- Mecklin CJ, Mundfrom DJ (2004) An appraisal and bibliography of tests for multivariate normality. *Int Stat Rev* 72(1):123–138
- Neuhaus G (1979) Asymptotic theory of goodness of fit tests when parameters are present: a survey. *Stat J Theor Appl Stat* 10(3):479–494
- Pearson ES, D'Agostino RB, Bowman KO (1977) Tests for departure from normality: comparison of powers. *Biometrika* 64(2):231–246
- Core Team R (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna
- Romão X, Delgado R, Costa A (2010) An empirical power comparison of univariate goodness-of-fit tests for normality. *J Stat Comput Simul* 80(5):545–591
- Ross N (2011) Fundamentals of Steins method. *Probab Surv* 8:210–293
- Roussas GG (1972) Contiguity of probability measures: some applications in statistics. Cambridge University Press, Cambridge
- Sen PK (1981) Sequential nonparametrics : invariance principles and statistical inference. Wiley, New York
- Shapiro SS, Francia RS (1972) An approximate analysis of variance test for normality. *J Am Stat Assoc* 67(337):215–216
- Shapiro SS, Wilk MB (1965) An analysis of variance test for normality (complete samples). *Biometrika* 52(3/4):591–611
- Shapiro SS, Wilk MB, Chen HJ (1968) A comparative study of various tests for normality. *J Am Stat Assoc* 63(324):1343–1372
- Stein C (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. In: Proceedings of the sixth Berkeley symposium on mathematical statistics and probability, volume 2: probability theory, pp 583–602
- Stein C (1986) Approximate computation of expectations, vol 7. Lecture Notes-Monograph Series. Institute of Mathematical Statistics, Hayward
- Vasicek O (1976) A test for normality based on sample entropy. *J R Stat Soc Ser B (Methodol)* 38(1):54–59
- Villaseñor-Alva JA, González-Estrada E (2015) A correlation test for normality based on the Lévy characterization. *Commun Stat Simul Comput* 44(5):1225–1238
- Widder DV (1959) The Laplace transform, 5th printing. Princeton University Press, Princeton
- Yap BW, Sim CH (2011) Comparisons of various types of normality tests. *J Stat Comput Simul* 81(12):2141–2155