ALEA, Lat. Am. J. Probab. Math. Stat. 18, 1029–1045 (2021)
DOI: 10.30757/ALEA.v18-38

On combining the zero bias transform and the empirical characteristic function to test normality

Bruno Ebner

Institute of Stochastics, Karlsruhe Institute of Technology (KIT),
Englerstr. 2, 76133 Karlsruhe, Germany.
E-mail address: Bruno.Ebner@kit.edu
URL: http://www.math.kit.edu/stoch/∼ebner/

Abstract. We propose a new powerful family of tests of univariate normality. These tests are based on an initial value problem in the space of characteristic functions originating from the fixed point property of the normal distribution in the zero bias transform. Limit distributions of the test statistics are provided under the null hypothesis, as well as under contiguous and fixed alternatives. Using the covariance structure of the limiting Gaussian process from the null distribution, we derive explicit formulas for the first four cumulants of the limiting random element and apply the results by fitting a distribution from the Pearson system. A comparative Monte Carlo power study shows that the new tests are serious competitors to the strongest well established tests.

1. Introduction

In view of the assumption of normality in many classical models, testing for normality is commonly known as the mostly used and discussed goodness-of-fit technique. To be specific, let $X, X_1, X_2, \dots$ be real-valued independent and identically distributed (iid.) random variables defined on an underlying probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The problem of interest is to test the hypothesis

$$H_0 : \mathbb{P}^X \in \mathcal{N} = \{N(\mu, \sigma^2) \,|\, (\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)\} \tag{1.1}$$

against general alternatives. This testing problem has been considered extensively and a multitude of different test statistics is available. The classical tests are based on the empirical distribution function, like the Kolmogorov-Smirnov test (modified in Lilliefors, 1967), the Anderson-Darling test, see Anderson and Darling (1952), the empirical characteristic function, see Epps and Pulley (1983), the empirical moment generating function, see Henze and Visagie (2020), on empirical measures of skewness and kurtosis, see Bowman and Shenton (1975); Jarque and Bera (1980); Pearson et al. (1977) (known to lead to inconsistent procedures), the Wasserstein distance, see del Barrio et al. (2000), measures of entropy, see Tavakoli et al. (2019); Vasicek (1976), the integrated empirical distribution function, see Klar (2001), or correlation and regression tests, like the time-honored

Received by the editors March 20th, 2020; accepted January 26th, 2021.

2010 *Mathematics Subject Classification.* 62G10, 62E10.

Key words and phrases. Goodness-of-fit, Normal Distribution, Stein's Method, Zero Bias Transformation, Empirical Characteristic Function.

"bench-mark test" of Shapiro-Wilk, see Shapiro and Wilk (1965), among others. For a survey of classical methods see del Barrio et al. (2000), section 3, and Henze (1994), and for comparative simulation studies, see 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). For a survey on tests of multivariate normality see Henze (2002), for recent multivariate tests see Dörr et al. (2021), and for new developments on normality tests for Hilbert space valued random elements, see Henze and Jiménez-Gamero (2021); Kellner and Celisse (2019).

Our novel approach relies on an initial value problem of the ordinary differential equation

$$\begin{cases} \varphi'(t) = -t\varphi(t), \\ \varphi(0) = 1, \end{cases} \tag{1.2}$$

where the characteristic function $\varphi(t) = \exp\left(-t^2/2\right)$, $t \in \mathbb{R}$, of the standard normal distribution is the unique solution. Hence the normal distribution is characterised by the considered initial value problem. Notice that the family of normal distributions $\mathcal{N}$ is closed under transformations $X \to \sigma X + \mu$, $\mu, \sigma \in \mathbb{R}$, in the sense that $X \sim N(0,1)$ is equivalent to $\sigma X + \mu \sim N(\mu, \sigma^2)$, and hence it suffices to characterise the standard normal distribution to propose a normality test. To model the standardisation assumption, we consider the so called scaled residuals

$$Y_{n,j} = \frac{X_j - \overline{X}_n}{S_n}, \quad j = 1, \ldots, n,$$

where $\overline{X}_n = \frac{1}{n}\sum_{j=1}^n X_j$ is the mean and $S_n^2 = \frac{1}{n}\sum_{j=1}^n (X_j - \overline{X}_n)^2$ is the sample variance. Denoting the empirical characteristic function by $\varphi_n(t) = \frac{1}{n}\sum_{j=1}^n \exp(itY_{n,j})$, $t \in \mathbb{R}$, we have $\varphi_n'(t) = \frac{1}{n}\sum_{j=1}^n iY_{n,j}\exp(itY_{n,j})$, $t \in \mathbb{R}$, and by estimating both sides of (1.2) we propose the test statistic

$$Z_n = n\int_{-\infty}^{\infty} \left| \frac{1}{n}\sum_{j=1}^n \left(iY_{n,j} + t\right)\exp(itY_{n,j}) \right|^2 w(t)\mathrm{d}t,$$

where $w(\cdot)$ is a suitable bounded weight function and $|x|^2 = \mathrm{Re}(x)^2 + \mathrm{Im}(x)^2$ is the squared absolute value of a complex number $x \in \mathbb{C}$. If X originates from a normal distribution, Z_n should be close to zero, and thus rejection of H_0 in (1.1) will be for large values of Z_n (empirical and asymptotic critical values are specified in Section 5). Tacitly, we assume the conditions

$$w(t) = w(-t), \quad t \in \mathbb{R}, \quad \int_{-\infty}^{\infty} w(t)\mathrm{d}t < \infty. \tag{1.3}$$

Note that Z_n depends only on the scaled residuals $Y_{n,1}, \ldots, Y_{n,n}$ and is hence invariant under translation or rescaling of the data set $X_1, \ldots, X_n$, which indeed is a desirable property, since the family $\mathcal{N}$ is closed under affine transformations.

Remark 1.1. Note that the initial value problem (1.2) is connected to the famous Stein characterisation of the normal law, see Stein (1972), and the zero bias transform. It is well-known that the normal distribution is the fixed point of the zero bias transform, see Chen et al. (2011); Goldstein and Reinert (1997). Let X be a centred random variable with $\sigma^2 = \mathbb{E}(X^2) < \infty$. Following Shevtsova (2013), the characteristic function of the X-zero bias transformed random variable X^* is

$$\mathbb{E}\left(e^{itX^*}\right) = \begin{cases} -\dfrac{1}{\sigma^2}\dfrac{\varphi'(t)}{t}, & t \in \mathbb{R}\setminus\{0\}, \\ 1, & t = 0, \end{cases} \tag{1.4}$$

where $\varphi(\cdot)$ is the characteristic function of X, and i stands for the imaginary unit. Together with the assumption $\sigma^2 = 1$ and the fixed point argument this leads to the initial value problem (1.2). Interestingly, (1.4) represents an operator A mapping from the space of characteristic functions into itself, where $A\varphi(t) \to (\varphi'(t) - \varphi'(0))/(t\varphi''(0))$, see statement (a) of Theorem 12.2.5 in Lukacs (1970), and apply $\varphi'(0) = i\mathbb{E}(X) = 0$ and $\varphi''(0) = -\sigma^2$. This fact shows that the zero bias transform was already studied in the late 60ies of the last century, hence some years before its introduction in Goldstein and Reinert (1997) and the earliest reference therein.

Setting $w(t) = w_a(t) = \exp(-at^2)$, $a > 0$, a direct evaluation of integrals shows that Z_n takes the form

$$\begin{aligned} Z_{n,a} &= \frac{1}{n}\sqrt{\frac{\pi}{a}} \sum_{j,k=1}^{n} \left(\frac{1}{4a^2}\left(2a - (Y_{n,j} - Y_{n,k})^2\right) \right. \\ &\quad \left. - \frac{1}{2a}(Y_{n,j} - Y_{n,k})^2 + Y_{n,j}Y_{n,k} \right) \exp\left(-\frac{1}{4a}(Y_{n,j} - Y_{n,k})^2 \right), \end{aligned}$$

which represents a computational stable and easy to implement version of Z_n. By some expansion of the exponential function and noting that $\sum_{j=1}^n Y_{n,j} = 0$ and $\sum_{j=1}^n Y_{n,j}^2 = n$, we have elementwise on the probability space

$$\lim_{a\to\infty} \frac{16a^{\frac{5}{2}}}{3n\sqrt{\pi}} Z_{n,a} = \left(\frac{1}{n}\sum_{j=1}^{n} Y_{n,j}^3 \right)^2 \quad \text{and} \quad \lim_{a\to 0} \sqrt{\frac{a}{\pi}} Z_{n,a} - \frac{1}{2a} = 1.$$

It is interesting to see that the limit for $a \to \infty$ is squared sample skewness, and that this limiting behaviour coincides with the one observed in Henze and Visagie (2020), section 4.

The rest of the paper is organized as follows. In Section 2 we derive the limit distribution of $Z_{n,a}$ under the null hypothesis. Section 3 states results under a sequence of contiguous alternatives, while in Section 4 we show that the new tests are consistent against alternatives satisfying a weak moment condition. Furthermore, we obtain a central limit result for the test. In Section 5, we derive explicit formulas for the first four cumulants of the limit null distribution of $Z_{n,a}$ and fit the Pearson-system of distributions to approximate the critical values of the test statistic. We complete the paper by a competitive Monte Carlo simulation study in Section 6 and finally draw conclusions and identify some open problems for further research in Section 7. The paper is concluded by an Appendix that contains proofs and the formula of the fourth cumulant.

2. Asymptotic null distribution

A suitable setup for deriving asymptotic theory is the Hilbert space of measurable, square integrable functions $L^2 = L^2(\mathbb{R}, \mathcal{B}, w\mathrm{d}\mathcal{L}^1)$, where $\mathcal{B}$ is the Borel-σ-field of $\mathbb{R}$ and $\mathcal{L}^1$ is the Lebesgue measure on $\mathbb{R}$. Notice that the functions figuring within the integral in the definition of Z_n are $(\mathcal{A} \otimes \mathcal{B}, \mathcal{B})$-measurable random elements of L^2. We denote by

$$\|f\|_{L^2} = \left(\int_{\mathbb{R}} \left|f(t)\right|^2 \omega(t)\,\mathrm{d}t \right)^{1/2}, \qquad \langle f, g\rangle_{L^2} = \int_{\mathbb{R}} f(t)g(t)\,\omega(t)\,\mathrm{d}t$$

the usual norm and inner product in L^2. After straightforward calculations using (1.3) and symmetry arguments, we have

$$Z_n = \int_{-\infty}^{\infty} W_n^2(t) w(t)\mathrm{d}t,$$

where

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_{n,j} + t)\cos(tY_{n,j}) + (t - Y_{n,j})\sin(tY_{n,j}), \quad t \in \mathbb{R}. \tag{2.1}$$

Motivated by a multivariate Taylor expansion we consider the processes

$$\begin{aligned} W_n^*(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n (X_j+t)\cos(tX_j)+(t-X_j)\sin(tX_j) \\ &\quad+\left((tX_j-(t^2+1))\cos(tX_j)+(tX_j+(t^2+1))\sin(tX_j)\right)\overline{X}_n \\ &\quad+X_j\left((tX_j-(t^2+1))\cos(tX_j)+(tX_j+(t^2+1))\sin(tX_j)\right)(S_n-1) \end{aligned}$$

and

$$\begin{aligned} \tilde{W}_n(t) &= \frac{1}{\sqrt{n}}\sum_{j=1}^n (X_j+t)\cos(tX_j)+(t-X_j)\sin(tX_j) \\ &\quad-\exp\left(-\frac{t^2}{2}\right)X_j+t\exp\left(-\frac{t^2}{2}\right)(X_j^2-1), \end{aligned}$$

$t\in\mathbb{R}$. In what follows let $X_1, X_2, \ldots$ be iid. random variables, and in view of affine invariance of Z_n we assume w.l.o.g. $X_1\sim N(0,1)$. The following Lemma shows that the processes W_n, W_n^* and $\tilde{W}_n$ are asymptotically equivalent. The proof is found in Appendix A.1.

Lemma 2.1. *We have under H_0*

$$\|W_n-W_n^*\|_{L^2}\xrightarrow{\mathbb{P}}0 \quad \text{and} \quad \|W_n^*-\tilde{W}_n\|_{L^2}\xrightarrow{\mathbb{P}}0.$$

In order to derive the asymptotic null distribution of Z_n, it suffices to show the weak convergence of $\tilde{W}_n$ in L^2 to a centred Gaussian process.

Theorem 2.2. *Under the standing assumptions, there is a centred Gaussian random element W of L^2 with covariance kernel*

$$K_Z(s,t)=(st+1)\exp\left(-\frac{(s-t)^2}{2}\right)-(2st+1)\exp\left(-\frac{s^2+t^2}{2}\right), \quad s,t\in\mathbb{R},$$

such that with W_n defined in (2.1), we have $W_n\xrightarrow{\mathcal{D}}W$ in L^2 as $n\to\infty$.

Proof: By Lemma 2.1 it follows that the limit distribution of W_n is the same as that of $\tilde{W}_n$. Note that

$$\tilde{W}_n(t)=\frac{1}{\sqrt{n}}\sum_{j=1}^n\widetilde{W}_{n,j}(t), \quad t\in\mathbb{R},$$

where

$$\begin{aligned} \widetilde{W}_{n,j}(t) &= (X_j+t)\cos(tX_j)+(t-X_j)\sin(tX_j) \\ &\quad-\exp\left(-\frac{t^2}{2}\right)X_j+t\exp\left(-\frac{t^2}{2}\right)(X_j^2-1), \quad t\in\mathbb{R}, \end{aligned}$$

$j=1,2,\ldots,n$ and $\mathbb{E}\widetilde{W}_{n,1}=0$. Since $\widetilde{W}_{n,1},\widetilde{W}_{n,2},\ldots$ are iid. centred elements of L^2, we can directly apply the central limit theorem in Hilbert spaces, see Corollary 10.9 in Ledoux and Talagrand (1991). Tedious calculations then show that the covariance kernel $K_Z(s,t)=\mathbb{E}\left(\widetilde{W}_{n,1}(s)\widetilde{W}_{n,1}(t)\right)$ takes the given form. □

The next result follows from a direct application of the continuous mapping theorem.

Corollary 2.3. *We have as $n\to\infty$*

$$Z_n\xrightarrow{\mathcal{D}}\int_{-\infty}^{\infty}W^2(t)w(t)\,dt=\|W\|_{L^2}^2.$$

We will use this result in Section 5 to derive the first four cumulants of the limit random element. As a consequence we obtain approximate critical values by the Pearson system of distributions.

3. Contiguous alternatives

In this section we consider a triangular array of row-wise iid. random variables $X_{n,1}, \dots, X_{n,n}$, $n \in \mathbb{Z}_+$, with Lebesgue density

$$f_n(t) = f(t) \cdot \left(1 + \tfrac{c(t)}{\sqrt{n}}\right), \, t \in \mathbb{R}.$$

Here, $f(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$, $t \in \mathbb{R}$, is the density of $N(0,1)$, and $c : \mathbb{R} \to \mathbb{R}$ is a measurable, bounded function satisfying $\int_{-\infty}^{\infty} c(t)\, f(t)\, \mathrm{d}t = 0$. Notice that, since c is bounded, we may assume n to be large enough to ensure $f_n \geq 0$. Setting

$$\mu_n = \bigotimes_{j=1}^{n} f\mathcal{L}^1 \quad \text{and} \quad \nu_n = \bigotimes_{j=1}^{n} f_n\mathcal{L}^1,$$

it is shown in Betsch and Ebner (2020), section 4, that by LeCam's first Lemma ν_n is contiguous to μ_n. Writing

$$\begin{aligned} \eta(x,s) \quad &= \quad (x+s)\cos(sx) + (s-x)\sin(sx) \\ &\quad\; - \exp\left(-\frac{s^2}{2}\right) x + s\exp\left(-\frac{s^2}{2}\right)(x^2-1), \quad x, s \in \mathbb{R}, \end{aligned}$$

and following the same lines of proof as in Betsch and Ebner (2020), section 4, we can show the following result.

Theorem 3.1. *Under the triangular array $X_{n,1}, \dots, X_{n,n}$, we have*

$$Z_n \xrightarrow{\mathcal{D}} \|W + \zeta\|_{L^2}^2,$$

where W is the limiting Gaussian process of Theorem 2.2 and

$$\zeta(s) = \int_{-\infty}^{\infty} \eta(x,s)c(x)f(x)\,dx, \quad s \in \mathbb{R}.$$

4. Consistency and behaviour under fixed alternatives

In this section we assume that the underlying distribution is a fixed alternative to H_0 and that the distribution is absolutely continuous, as well as in view of affine invariance of the test statistic, we assume $\mathbb{E}(X) = 0$ and $\mathbb{E}(X^2) = 1$.

Theorem 4.1. *Under the standing assumptions, we have as $n \to \infty$,*

$$\frac{Z_n}{n} \xrightarrow{\text{a.s.}} \int_{-\infty}^{\infty} \left|\mathbb{E}\left((iX+t)\exp(itX)\right)\right|^2 w(t)\,dt = \Delta.$$

Proof: Let

$$W_n^0(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j + t)\cos(tX_j) + (t - X_j)\sin(tX_j), \quad t \in \mathbb{R}.$$

In this setting, we have $(\overline{X}_n, S_n) \xrightarrow{\text{a.s.}} (0,1)$ and hence we can apply a similar reasoning as in the proof of Lemma 2.1 to show that

$$\left\| n^{-1/2}(W_n - W_n^*) \right\|_{L^2}^2 \xrightarrow{\text{a.s.}} 0. \tag{4.1}$$

Next, we consider (for the definition of $\Psi(\cdot,\cdot)$ see (A.1) in Appendix A.1)

$$\begin{aligned}
n^{-1/2}(W_n^*(t)-W_n^0(t)) &= \frac{1}{n}\sum_{j=1}^n \big((tX_j-(t^2+1))\cos(tX_j)\\
&\quad +(tX_j+(t^2+1))\sin(tX_j)\big)\,\overline{X}_n\\
&\quad +X_j\big((tX_j-(t^2+1))\cos(tX_j)\\
&\quad +(tX_j+(t^2+1))\sin(tX_j)\big)\,(S_n-1)\\
&= \overline{X}_n\frac{1}{n}\sum_{j=1}^n\Psi(t,X_j)+(S_n-1)\frac{1}{n}\sum_{j=1}^n X_j\Psi(t,X_j).
\end{aligned}$$

By the triangle inequality, we have

$$\begin{aligned}
\left\|n^{-1/2}(W_n^*-W_n^0)\right\|_{L^2}^2 &\le 2\left|\overline{X}_n\right|^2\left\|\frac{1}{n}\sum_{j=1}^n\Psi(\cdot,X_j)\right\|_{L^2}^2\\
&\quad +2|S_n-1|^2\left\|\frac{1}{n}\sum_{j=1}^n X_j\Psi(\cdot,X_j)\right\|_{L^2}^2. \qquad (4.2)
\end{aligned}$$

By the strong law of large numbers in Hilbert spaces and $(\overline{X}_n,S_n)\xrightarrow{\text{a.s.}}(0,1)$, the right hand side of (4.2) converges to zero almost surely. Note that the expectations exist due to the existence of the first two derivatives of the characteristic function of X, which is implied by $\mathbb{E}(X^2)=1<\infty$. Again, by the strong law of large numbers in Hilbert spaces, we have

$$n^{-1/2}W_n^0(t)\xrightarrow{\text{a.s.}}\mathbb{E}\left[(X+t)\cos(tX)+(t-X)\sin(tX)\right]$$

in L^2. In view of (4.1), (4.2), and the symmetry of the weight function $w(\cdot)$, some calculations give

$$\frac{Z_n}{n}=\left\|n^{-1/2}W_n\right\|_{L^2}^2\xrightarrow{\text{a.s.}}\int_{-\infty}^{\infty}\left|\mathbb{E}\left[(X+t)\cos(tX)+(t-X)\sin(tX)\right]\right|^2 w(t)\mathrm{d}t=\Delta.$$

□

Notice that, if $g(t)=\mathbb{E}[(\exp(itX)]$ denotes the characteristic function of X, we have $\Delta=0$ if and only if $g=\varphi$, which is shown by the unique solution of the initial value problem (1.2). This implies that $Z_n\xrightarrow{\text{a.s.}}\infty$ for any alternative with existing second moment. Thus we conclude that the test based on Z_n is consistent against each such alternative.

To derive further asymptotic results, we follow the methodology in Baringhaus et al. (2017). Put $z(t)=\mathbb{E}\left[(X+t)\cos(tX)+(t-X)\sin(tX)\right]$ and $W_n^\bullet(\cdot)=n^{-1/2}W_n(t)$, we then have

$$\begin{aligned}
\sqrt{n}\left(\frac{Z_n}{n}-\Delta\right) &= \sqrt{n}\left(\|W_n^\bullet\|_{L^2}^2-\|z\|_{L^2}^2\right)=\sqrt{n}\langle W_n^\bullet-z,W_n^\bullet+z\rangle_{L^2}\\
&= \sqrt{n}\langle W_n^\bullet-z,2z+W_n^\bullet-z\rangle_{L^2}\\
&= 2\langle\sqrt{n}(W_n^\bullet-z),z\rangle_{L^2}+\frac{1}{\sqrt{n}}\|\sqrt{n}\,(W_n^\bullet-z)\,\|_{L^2}^2. \qquad (4.3)
\end{aligned}$$

The following structural Lemma is needed in the subsequent derivations and is proved in Appendix A.2.

Lemma 4.2. *For $\mathbb{E}(X^4)<\infty$ we have*

$$\sqrt{n}(W_n^\bullet-z)\xrightarrow{\mathcal{D}}\mathcal{W},$$

in L^2, where $\mathcal{W}$ is a centred Gaussian process in L^2 with covariance kernel

$$\begin{aligned} K_W(s,t) &= \mathbb{E}\left[(st+X^2)\cos((s-t)X)+(st-X^2)X\sin((s+t))\right] \\ &+\mathbb{E}\left[((s-t)\sin((s-t)X)+(s+t)X\cos((s+t)X))\right] \\ &+\frac{1}{2}\left[\mathbb{E}\left[(X+s)(b(t)(X^2-1)+2a(t)X)\cos(sX)\right]\right. \\ &+\mathbb{E}\left[(b(s)(X^2-1)+2a(s)X)(t+X)\cos(tX)\right] \\ &+\mathbb{E}\left[(b(t)(X^2-1)+2a(t)X)(s-X)\sin(sX)\right] \\ &\left.+\mathbb{E}\left[(b(s)X^2+2a(s)X-b(s))(t-X)\sin(tX)\right]\right] \\ &+\frac{1}{4}\left[b(t)\mathbb{E}(X^4)+2a(t)\mathbb{E}(X^3)-b(t)\right]b(s) \\ &+\frac{1}{2}a(s)(b(t)\mathbb{E}(X^3)+2a(t))-z(s)z(t), \quad s,t\in\mathbb{R}, \end{aligned}$$

where

$$a(t)=\mathbb{E}(\Psi(t,X)) \quad \textit{and} \quad b(t)=\mathbb{E}(X\Psi(t,X)).$$

Lemma 4.2 shows that $\sqrt{n}\,(W_n^\bullet - z)$ is a tight sequence in L^2, thus we see by Slutzky's Lemma and $\frac{1}{\sqrt{n}}\|\sqrt{n}\,(W_n^\bullet - z)\|_{L^2}^2 \xrightarrow{\mathbb{P}} 0$ that the limit distribution of $\sqrt{n}\left(\frac{Z_n}{n}-\Delta\right)$ in (4.3) only depends on $2\langle\sqrt{n}(W_n^\bullet - z), z\rangle_{L^2}$. A direct application of Theorem 1 in Baringhaus et al. (2017) and Lemma 4.2 yields the following result.

Theorem 4.3. *Under the standing assumptions and $\mathbb{E}(X^4)<\infty$, we have*

$$\sqrt{n}\left(\frac{Z_n}{n}-\Delta\right) \xrightarrow{\mathcal{D}} N(0,\tau^2),$$

where

$$\tau^2 = 4\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K_W(s,t)z(s)z(t)w(s)w(t)\,dsdt.$$

In principle, one can calculate for a fixed alternative the explicit version of K_W and z and finally τ^2. For most of the alternatives this will be too complicated, thus we suggest to estimate τ^2 by a consistent estimator $\widehat{\tau}_n^2$. Corollary 1 in Baringhaus et al. (2017) then states that

$$\frac{\sqrt{n}}{\widehat{\tau}_n}\left(\frac{Z_n}{n}-\Delta\right) \xrightarrow{\mathcal{D}} N(0,1),$$

which opens grounds to applications as suggested in Section 3 in Baringhaus et al. (2017), i.e., computation of an asymptotic confidence interval for Δ, approximation of the power function or neighborhood-of-model validation. For specific examples of the needed methodology, see Section 4 of Baringhaus et al. (2017), or Betsch and Ebner (2020); Dörr et al. (2021) respectively.

5. Approximation of the limit null distribution

In this section we follow the methodology in Henze (1990) to approximate the critical values of the asymptotic level α test based on Z_n by exploiting the covariance structure of the limiting centred Gaussian process of Theorem 2.2. Let $Z_{\infty,a}=\|W\|_{L^2}$ be the random variable with the limit null distribution of Z_n in Corollary 2.3. Hence W is the random element in Theorem 2.2 and the weight function $w_a(t)=\exp\left(-at^2\right)$, $t\in\mathbb{R}$, of Section 1 is used. By the results of Corollary 2.3 it is well-known that the limiting null distribution of $Z_{n,a}$ is given by the infinite series

$$Z_{\infty,a}=\sum_{j=1}^{\infty}\lambda_j(a)Y_j^2.$$

a	$\mathbb{E}(Z_{\infty,a})$	$\mathrm{Var}(Z_{\infty,a})$	$\sqrt{\beta_1}(a)$	$\beta_2(a)$
0.1	30.4036	304.1938	1.4542	6.4513
0.25	7.7811	31.2928	1.7549	7.8821
0.5	2.6013	4.7153	1.9576	8.9907
0.75	1.3056	1.3821	2.0799	9.7885
1	0.7787	0.5430	2.1780	10.4822
3	0.0861	0.0094	2.5812	13.3852
5	0.0277	0.0011	2.7053	14.2265
10	0.0055	0.0001	2.7885	14.7597

TABLE 5.1. Values of mean, variance, skewness and kurtosis of $Z_{\infty,a}$ rounded to 4 decimals

Here, $Y_1, Y_2, \ldots$ being independent $N(0,1)$ distributed random variables and $(\lambda_j(a))_{j\geq 1}$ is the decreasing sequence of the positive eigenvalues of the integral operator

$$\mathcal{K}f(s) = \int_{-\infty}^{\infty} K_Z(s,t)f(t)w_a(t)\,\mathrm{d}t,$$

on L^2. Notice that $\mathcal{K}$ depends solely on the covariance kernel K_Z of Theorem 2.2 and the weight function $w_a(\cdot)$. It seems hopeless to obtain closed-form expressions for the eigenvalues λ_j, hence we derive the first four moments of $Z_{\infty,a}$ in order to fit the Pearson system of distributions, see Johnson et al. (1994), chapter 12, section 4.1. The m-th cumulant $\kappa_m(a)$ of $Z_{\infty,a}$ is

$$\kappa_m(a) = 2^{m-1}(m-1)! \int_{-\infty}^{\infty} h_m(t,t)w_a(t)\,\mathrm{d}t,$$

where $h_1(s,t) = K_Z(s,t)$ and

$$h_m(s,t) = \int_{-\infty}^{\infty} h_{m-1}(s,u)K_Z(u,t)w_a(u)\,\mathrm{d}u, \quad m \geq 2.$$

The formulae for the first four cumulants are (the computations were performed by using the computer algebra system Maple 2019, Maplesoft (2019))

$$\kappa_1(a) = -1/2\,\frac{\sqrt{\pi}\left(2\,a^{5/2} - 2\sqrt{a+1}a^2 + 4\,a^{3/2} - 3\sqrt{a+1}a - \sqrt{a+1}\right)}{(a+1)^{3/2}\,a^{3/2}},$$

where the formulae for $\kappa_j(a)$ for $j = 2,3,4$ can be found in the arXiv version, see Ebner (2020). From the first four cumulants we can approximate the distribution of $Z_{\infty,a}$ by a member of the Pearson system of distributions, since

$$\mathbb{E}(Z_{\infty,a}) = \kappa_1(a) \quad \text{and} \quad \mathrm{Var}(Z_{\infty,a}) = \kappa_2(a),$$

as well as the parameters of skewness and kurtosis of $Z_{\infty,a}$ are given by

$$\sqrt{\beta_1}(a) = \frac{\kappa_3(a)}{(\kappa_2(a))^{3/2}} \quad \text{and} \quad \beta_2(a) = 3 + \frac{\kappa_4(a)}{(\kappa_2(a))^2}.$$

These values can directly be used in packages that implement the Pearson system, for concrete values see Table 5.1. In the statistical computing language `R`, see R Core Team (2019), we will use the package `PearsonDS`, see Becker and Klößner (2017), to approximate critical values, see Table 6.2, and p-values of the corresponding tests.

Test	$a\backslash q$	0.01	0.05	0.1	0.9	0.95	0.99
	0.1	6.86169	10.51119	13.04177	53.02253	62.95695	84.99265
	0.25	1.04896	1.85667	2.48811	15.17277	18.74151	26.51209
	0.5	0.21479	0.42809	0.61130	5.43755	6.81869	9.96010
$Z_{20,a}$	0.75	0.07838	0.16585	0.24721	2.75949	3.53893	5.41328
	1	0.03616	0.08152	0.12623	1.63633	2.14516	3.40884
	3	0.00136	0.00467	0.00866	0.17785	0.25076	0.44301
	5	0.00030	0.00120	0.00225	0.05674	0.08146	0.14904
	10	0.00004	0.00017	0.00032	0.01116	0.01631	0.03051
	0.1	6.44914	9.96704	12.49225	53.05553	63.40399	86.79261
	0.25	0.99974	1.79193	2.40350	15.05561	18.79428	26.90751
	0.5	0.20190	0.42203	0.60621	5.40910	6.89826	10.24374
$Z_{50,a}$	0.75	0.07554	0.16688	0.25100	2.79329	3.59782	5.50480
	1	0.03526	0.08298	0.12909	1.69165	2.20391	3.46662
	3	0.00144	0.00464	0.00847	0.19527	0.27076	0.46805
	5	0.00030	0.00116	0.00221	0.06399	0.09031	0.16034
	10	0.00004	0.00017	0.00033	0.01292	0.01859	0.03359
	0.1	6.35269	9.91258	12.40696	53.26156	63.85351	87.88282
	0.25	0.99012	1.75319	2.36455	15.13173	18.81316	27.23384
	0.5	0.20292	0.41909	0.60298	5.43381	6.88596	10.38987
$Z_{100,a}$	0.75	0.07478	0.16796	0.25341	2.80970	3.63695	5.56701
	1	0.03583	0.08448	0.13175	1.70456	2.23136	3.47844
	3	0.00149	0.00478	0.00876	0.20069	0.27509	0.45774
	5	0.00031	0.00118	0.00228	0.06636	0.09233	0.15819
	10	0.00004	0.00017	0.00034	0.01354	0.01923	0.03358
	0.1	6.89295	9.89596	12.27245	53.39952	63.92766	87.89731
	0.25	1.29920	1.83683	2.35713	15.10009	18.73029	27.15089
	0.5	0.32955	0.45838	0.60581	5.41750	6.89193	10.35260
$Z_{\infty,a}$	0.75	0.13650	0.19046	0.25773	2.81741	3.63395	5.57065
	1	0.07292	0.10059	0.13743	1.71902	2.23934	3.48445
	3	0.00826	0.00931	0.01142	0.20558	0.27903	0.45910
	5	0.00254	0.00274	0.00322	0.06843	0.09441	0.15828
	10	0.00038	0.00041	0.00048	0.01414	0.01980	0.03371

Table 6.2. Empirical quantiles of $Z_{n,a}$ for $n = 20, 50, 100$ (100000 replications) and approximation of the quantiles of $Z_{\infty,a}$ by a Pearson family

6. Simulations

This section presents results of a comparative finite sample power simulation study. The study is designed to match and complement the counterparts in Dörr et al. (2021), Section 7, and in Betsch and Ebner (2020), Section 6, since we take exactly the same setting with regard to sample size, nominal level of significance and selected alternative distributions. In this way, we facilitate the comparison with existing procedures, even with some procedures not covered here. We consider sample sizes $n \in \{20, 50, 100\}$ and fix the nominal level of significance throughout all simulations to 0.05. All simulations are performed using the statistical computing environment `R`, see R Core Team (2019). We simulated empirical critical values under H_0 for $Z_{n,a}$ with 100 000 replications, see Table 6.2. The row segment entitled '$Z_{\infty,a}$' gives approximations by the method described in Section 5. Each entry in Table 6.3 was simulated with 10 000 replications, and an asterisk '*' denotes a perfect rejection rate of 100%.

The following alternatives are considered: symmetric distributions, like the Student t_ν-distribution with $\nu \in \{3, 5, 10\}$ degrees of freedom, as well as the uniform distribution $\mathrm{U}(-\sqrt{3}, \sqrt{3})$, and asymmetric distributions, such as the χ^2_ν-distribution with $\nu \in \{5, 15\}$ degrees of freedom, the beta distributions $\mathrm{B}(1, 4)$ and $\mathrm{B}(2, 5)$, and the gamma distributions $\Gamma(1, 5)$ and $\Gamma(5, 1)$, both parametrized by their shape and rate parameter, the Gumbel distribution $\mathrm{Gum}(1, 2)$ with location parameter 1 and scale parameter 2, the Weibull distribution $\mathrm{W}(1, 0.5)$ with scale parameter 1 and shape parameter 0.5, and the lognormal distribution $\mathrm{LN}(0, 1)$. As representatives of bimodal distributions, we simulate the mixture of normal distributions $\mathrm{NMix}(p, \mu, \sigma^2)$, where the random variables are generated by $(1-p)\,\mathrm{N}(0, 1) + p\,\mathrm{N}(\mu, \sigma^2)$, $p \in (0, 1)$, $\mu \in \mathbb{R}$, $\sigma > 0$. Note that these alternatives can also be found in the simulation studies presented in Betsch and Ebner (2020); Dörr et al. (2021); Dörr et al. (2021); Romão et al. (2010). We chose these alternatives in order to ease the comparison with many other existing tests.

The considered competing test statistics are the following:

- the Anderson-Darling test, see Anderson and Darling (1952),
- the Shapiro-Wilk test, see Shapiro and Wilk (1965),
- the Jarque-Bera test, see Jarque and Bera (1980),
- the Henze-Visagie test, see Henze and Visagie (2020),
- the Betsch-Ebner test, see Betsch and Ebner (2020),
- the BHEP test, see Henze and Wagner (1997),
- the BCMR test, see del Barrio et al. (1999).

Note that these tests are very strong competitors as witnessed by extensive simulation studies, see Romão et al. (2010).

We used the implementation of the Anderson-Darling (AD) test in the package `nortest` from Gross and Ligges (2015) and the implementation of the Shapiro-Wilk (SW) test from the `stats` package. The Jarque-Bera (JB) test was implemented in the package `tseries`, see Trapletti and Hornik (2019). The Henze-Visagie (HV) test uses a weighted L^2-type statistic based on a characterization of the moment generating function that similarly as the newly proposed test employs a first-order differential equation. The univariate statistic is defined by

$$\mathrm{HV}_\gamma = \sqrt{\frac{\pi}{\gamma}}\frac{1}{n}\sum_{j,k=1}^{n} \exp\left(\frac{(Y^+_{n,j,k})^2}{4\gamma}\right)\left(Y_{n,j}Y_{n,k} + (Y^+_{n,j,k})^2\left(\frac{1}{4\gamma^2} - \frac{1}{2\gamma}\right) + \frac{1}{2\gamma}\right),$$

where $Y^+_{n,j,k} = Y_{n,j} + Y_{n,k}$ and $\gamma > 2$. In what follows, we consider three different tuning parameters $\gamma \in \{2.5, 5, 10\}$. Simulated critical values can be found in the arXiv version of Henze and Visagie (2020). The Betsch-Ebner (BE) test is based on a L^2-distance between the empirical zero-bias transformation and the empirical distribution function. By the same fixed point argument, this distance is minimal under normality. The statistic is given by

$$\begin{aligned}
\mathrm{BE} \;=\; & \frac{2}{n}\sum_{1\le j<k\le n}\Bigg\{\left(1-\Phi\left(\tfrac{Y_{(k)}}{\sqrt{a}}\right)\right)\left((Y^2_{(j)}-1)(Y^2_{(k)}-1) + aY_{(j)}Y_{(k)}\right)\\
& + \frac{a}{\sqrt{2\pi a}}\exp\left(-\tfrac{Y^2_{(k)}}{2a}\right)\left(-Y^2_{(j)}Y_{(k)} + Y_{(k)} + Y_{(j)}\right)\Bigg\}\\
& + \frac{1}{n}\sum_{j=1}^{n}\Bigg\{\left(1-\Phi\left(\tfrac{Y_{n,j}}{\sqrt{a}}\right)\right)\left(Y^4_{n,j} + (a-2)Y^2_{n,j} + 1\right)\\
& + \frac{a}{\sqrt{2\pi a}}\exp\left(-\tfrac{Y^2_{n,j}}{2a}\right)\left(2Y_{n,j} - Y^3_{n,j}\right)\Bigg\},
\end{aligned}$$

where $Y_{(1)} \leq \ldots \leq Y_{(n)}$ are the order statistics of the scaled residuals $Y_{n,1}, \ldots, Y_{n,n}$, and $\Phi(\cdot)$ stands for the distribution function of the standard normal law. The parameter $a > 0$ and the corresponding critical values were chosen by the algorithm presented in Betsch and Ebner (2020).

Tests based on the empirical characteristic function are represented by the Baringhaus-Henze-Epps-Pulley (BHEP) test, see Baringhaus and Henze (1988); Epps and Pulley (1983). The univariate BHEP test with tuning parameter $\beta > 0$ uses the test statistic

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

We fix $\beta = 1$ and took the critical values from Henze (1990). Note that the BHEP test and the HV test are included in the `R` package `mnt`, see Butsch and Ebner (2020). 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. (2000), section 3.3, and del Barrio et al. (1999). The BCMR statistic is given by

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

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

The results presented in Table 6.3 show that the power of $Z_{n,a}$ depends on the choice of the tuning parameter $a > 0$. In most cases one is able to find a value of a in which the tests are nearly as good or better than the competitors. Note that for higher values of a $Z_{n,a}$ performs best for the χ^2_{15}, the $\Gamma(5,1)$ and the $\mathrm{Gum}(1,2)$ distribution. Very interesting is the behaviour of the HV-test for the uniform $\mathrm{U}(-\sqrt{3},\sqrt{3})$, where it fails to detect the alternative for any value of γ for any n. Another interesting comparison can be made for this uniform distribution between $Z_{n,a}$ and the BE-test if one also takes Table 3 of Betsch and Ebner (2020) into consideration, since it seems that even though both procedures are based on the zero bias transform, $Z_{n,a}$ seems to attain higher power for some values of a, while the BE-test seems to be much less sensitive to the actual choice of a. The AD-test performs best for the normal mixture distributions, while the overall the SW-test has a strong power for most asymmetric distributions.

Depending of the nature of the alternatives, we would suggest to use $a = 0.25$ for symmetric alternatives and $a = 3$ for asymmetric alternatives for performing the test. If nothing is known about the nature of the alternative, we suggest to use $a = 1$, as it seems to have a good overall power performance. Naturally, it would be interesting to implement a data driven choice of the tuning parameter as suggested by Allison and Santana (2015) and corrected in Tenreiro (2019), but we leave this problem open for further research.

7. Conclusions and outlook

We have proposed a new family of tests for normality based on an initial value problem of an ordinary differential equation, connected to the fixed point property of the zero bias transformation and its corresponding characteristic function. These tests are universally consistent under weak moment conditions and show a remarkable power performance in comparison to the strongest time-honored tests of normality. Weak convergence results under the null hypothesis, under contiguous

		$Z_{n,a}$								HV_γ								
Alt.	$n \backslash a$	0.1	0.25	0.5	0.75	1	3	5	10	2.5	5	10	SW	BCMR	BHEP	AD	JB	BE
$\mathrm{N}(0,1)$	20	5	5	4	4	4	5	5	5	5	5	5	5	5	5	5	2	5
	50	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5
	100	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	4	5
$\mathrm{NMix}(0.3,1,0.25)$	20	25	28	27	24	23	18	17	17	11	13	15	29	29	29	30	7	25
	50	61	65	62	59	57	45	41	38	15	24	31	60	60	62	68	19	56
	100	92	93	91	89	86	76	72	68	23	45	56	88	88	90	94	50	86
$\mathrm{NMix}(0.5,1,4)$	20	35	42	42	41	40	33	32	32	32	32	31	39	41	41	46	25	34
	50	78	85	83	80	77	59	53	48	49	50	46	77	78	79	86	59	52
	100	98	99	99	98	97	86	77	66	68	68	63	98	98	98	99	87	75
t_3	20	19	27	33	35	36	36	36	36	39	38	37	35	37	34	34	32	30
	50	37	52	61	64	65	62	61	58	66	64	60	64	66	62	61	67	41
	100	63	79	85	87	87	84	82	78	85	84	80	87	88	86	85	89	54
t_5	20	8	12	16	18	19	20	20	21	22	22	22	19	20	18	17	17	16
	50	14	22	30	34	35	36	35	34	40	39	36	36	38	31	30	39	22
	100	23	38	48	52	54	53	50	47	59	57	51	56	58	50	48	63	27
t_{10}	20	6	6	8	9	10	11	11	11	12	12	11	10	10	9	9	8	9
	50	6	8	11	13	14	16	16	16	20	19	18	15	17	12	11	18	11
	100	8	12	16	19	20	23	22	21	29	27	24	24	25	18	16	29	11
$\mathrm{U}(-\sqrt{3},\sqrt{3})$	20	18	21	15	8	4	1	1	1	0	0	0	20	17	13	17	0	4
	50	47	59	59	50	36	2	1	0	0	0	0	75	69	54	57	0	3
	100	87	94	95	94	91	8	1	0	0	0	0	*	99	94	95	57	5
χ^2_5	20	18	28	36	40	41	40	39	40	32	35	38	43	43	42	38	24	44
	50	44	66	78	82	84	85	84	83	62	74	79	89	88	83	80	68	84
	100	80	95	98	99	99	99	99	99	89	97	98	*	*	99	99	97	99
χ^2_{15}	20	7	11	15	16	17	18	18	18	16	17	18	18	18	17	15	11	18
	50	13	23	33	39	42	45	45	45	31	37	42	43	43	40	35	31	44
	100	21	42	59	67	71	76	77	76	50	65	72	74	74	68	61	60	74
$\mathrm{B}(1,4)$	20	32	41	47	48	48	43	41	41	27	34	38	59	58	52	51	20	49
	50	78	88	92	93	94	91	89	87	51	73	81	99	98	94	95	67	90
	100	99	*	*	*	*	*	*	*	84	98	99	*	*	*	*	99	*
$\mathrm{B}(2,5)$	20	9	12	14	15	14	13	13	13	9	11	12	17	17	17	15	5	15
	50	18	31	39	42	43	39	37	36	14	22	29	50	48	44	39	15	40
	100	38	63	75	79	81	79	76	73	23	51	64	90	89	80	76	51	73
$\Gamma(1,5)$	20	55	66	73	74	75	71	70	69	54	62	66	83	82	77	77	47	76
	50	96	99	99	99	*	99	99	99	90	96	98	*	*	*	*	96	99
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\Gamma(5,1)$	20	9	14	19	22	23	23	24	24	20	22	23	23	24	23	20	14	25
	50	18	34	47	53	56	59	59	59	40	49	54	59	58	54	48	42	58
	100	35	63	79	84	87	90	90	89	65	81	86	90	90	85	81	78	88
$\mathrm{W}(1,0.5)$	20	56	68	74	75	76	72	71	70	56	63	67	84	83	78	78	49	76
	50	96	99	99	*	*	99	99	99	90	97	98	*	*	*	*	96	99
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
$\mathrm{Gum}(1,2)$	20	12	19	26	29	31	31	32	32	27	29	31	31	31	31	27	20	32
	50	24	44	58	65	68	71	70	70	53	62	67	68	69	65	60	55	70
	100	47	76	87	91	93	95	95	95	80	90	93	94	94	92	89	89	94
$\mathrm{LN}(0,1)$	20	76	84	88	90	90	88	87	87	77	82	85	93	93	91	90	72	90
	50	99	*	*	*	*	*	*	*	99	*	*	*	*	*	*	*	*
	100	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*

TABLE 6.3. Empirical rejection rates of $Z_{n,a}$ and competitors ($\alpha = 0.05$, 10 000 replications).

and fixed alternatives were derived, which open ground to further insights on the behaviour of the tests.

Finally, we point out some open problems concerning the test statistic. A first step to further investigation, would be to derive a consistent estimator of the limiting variance τ^2 in Theorem 4.3. The approximation of the eigenvalues connected to the limiting random element $Z_{\infty,a}$ would give some further insight to approximate Bahadur efficiency statements and the structure of the initial value problem gives hope to extend the procedure to the multivariate case. Note that the approach in Remark 1.1 that leads to the initial value problem (1.2) of the differential equation can be modified to characterise other parametric families of distributions, see, e.g., Section 2 of Gaunt (2019) for Student's t distribution or Gaunt (2017) for the family of generalized hyperbolic distributions. This approach leads to feasible motivations of new test statistics even if the formula of the characteristic function is unknown or intractable.

Acknoledgements

We thank Norbert Henze for numerous suggestions that led to an improvement of the paper, and also thank an anonymous referee for valuable suggestions.

Appendix A. Proofs

A.1. *Proof of Lemma 2.1.*

Proof: Let for $t, x \in \mathbb{R}$

$$\begin{aligned}\Psi_{t,x}(\mu,\sigma) &= \frac{x-\mu}{\sigma}\left(\cos\left(t\frac{x-\mu}{\sigma}\right) - \sin\left(t\frac{x-\mu}{\sigma}\right)\right)\\ &\quad +t\left(\cos\left(t\frac{x-\mu}{\sigma}\right) + \sin\left(t\frac{x-\mu}{\sigma}\right)\right),\end{aligned}$$

and notice that a first order multivariate Taylor approximation around $(\mu_0, \sigma_0) = (0, 1)$ gives

$$\Psi_{t,x}(\mu,\sigma) = \Psi_{t,x}(0,1) + \frac{\partial \Psi_{t,x}(0,1)}{\partial \mu}\mu + \frac{\partial \Psi_{t,x}(0,1)}{\partial \sigma}(\sigma - 1) + R,$$

where R is a remainder term involving higher powers of μ and $(\sigma-1)$. With that notation and $(\overline{X}_n, S_n) \xrightarrow{\mathbb{P}} (0,1)$ (assuring that the stochastic remainder is $o_{\mathbb{P}}(1)$) we have

$$W_n(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\Psi_{t,X_j}(\overline{X}_n, S_n) = W_n^*(t) + o_{\mathbb{P}}(1), \quad t \in \mathbb{R},$$

and hence $\|W_n - W_n^*\|_{L^2} = o_{\mathbb{P}}(1)$ by application of the triangle inequality, Slutzky's Lemma and the central limit theorem in L^2 implying the boundedness in probability of mixed remainder terms. Next, note that for $X_1 \sim N(0,1)$ and $t \in \mathbb{R}$, we have by symmetry

$$\mathbb{E}(\sin(tX_1)) = \mathbb{E}(X_1\cos(tX_1)) = \mathbb{E}(X_1^2\sin(tX_1)) = 0$$

and using the standard normal characteristic function $\varphi(t) = \mathbb{E}(\cos(tX_1)) = \exp(-t^2/2)$ and its derivatives, we have

$$\mathbb{E}(X_1\sin(tX_1)) = t\exp(-t^2/2), \quad \text{and} \quad \mathbb{E}(X_1^2\cos(tX_1)) = (1-t^2)\exp(-t^2/2).$$

Let

$$\Psi(t,x) = (tx - (t^2+1))\cos(tx) + (tx + (t^2+1))\sin(tx), \tag{A.1}$$

Hence, we have for $t \in \mathbb{R}$

$$\begin{aligned}\mathbb{E}\left(\Psi(t,X_1)\right) &= -\exp(-t^2/2) = -\varphi(t),\\ \mathbb{E}\left(X_1\Psi(t,X_1)\right)\right) &= 2t\exp(-t^2/2) = -2\varphi'(t).\end{aligned}$$

Now, it is easy to see that

$$\big\|\frac{1}{n}\sum_{j=1}^n \Psi(\cdot,X_j)+\varphi(\cdot)\big\|_{L^2} = o_{\mathbb{P}}(1) \quad \text{and} \quad \big\|\frac{1}{n}\sum_{j=1}^n X_j\Psi(\cdot,X_j)+2\varphi'(\cdot)\big\|_{L^2} = o_{\mathbb{P}}(1).$$

Since

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

we have

$$\begin{aligned}\|W_n^*-\tilde{W}_n\|_{L^2}^2 &= \int_{-\infty}^{\infty}\big|\frac{1}{\sqrt{n}}\sum_{j=1}^n \Psi(t,X_1)\overline{X}_n + X_j\Psi(t,X_j)(S_n-1)+\varphi(t)\overline{X}_n\\ &\quad +\varphi'(t)(X_j^2-1)\big|^2 w(t)\mathrm{d}t\\ &\le \int_{-\infty}^{\infty}\Big|\big(\frac{1}{n}\sum_{j=1}^n \Psi(t,X_1)+\varphi(t)\big)\frac{1}{\sqrt{n}}\sum_{l=1}^n X_l\\ &\quad +\big(\frac{1}{n}\sum_{j=1}^n X_j\Psi(t,X_j)+\varphi'(t)\big)\frac{1}{\sqrt{n}}\sum_{l=1}^n (X_l^2-1)\Big|^2 w(t)\mathrm{d}t + o_{\mathbb{P}}(1)\\ &\le 2\Big\{\big\|\frac{1}{n}\sum_{j=1}^n \Psi(\cdot,X_j)+\varphi(\cdot)\big\|_{L^2}^2\big(\frac{1}{\sqrt{n}}\sum_{l=1}^n X_l\big)^2\\ &\quad +\big\|\frac{1}{n}\sum_{j=1}^n X_j\Psi(\cdot,X_j)+2\varphi'(\cdot)\big\|_{L^2}^2\big(\frac{1}{\sqrt{n}}\sum_{l=1}^n (X_l^2-1)\big)^2\Big\} + o_{\mathbb{P}}(1)\end{aligned}$$

and since $n^{-1/2}\sum_{l=1}^n X_l$ and $n^{-1/2}\sum_{l=1}^n (X_l^2-1)$ are tight sequences, the result follows by Slutsky's Lemma. □

A.2. *Proof of Lemma 4.2.*

Proof: Set

$$\tilde{W}_n^{\bullet}(t) = \frac{1}{\sqrt{n}}\sum_{j=1}^n (X_j+t)\cos(tX_j)+(t-X_j)\sin(tX_j)+\mathbb{E}(\Psi(t,X))X_j+\frac{1}{2}\mathbb{E}(X\Psi(t,X))(X_j^2-1)$$

by the same arguments as in the proof of Lemma 2.1, we have $\|W_n-\tilde{W}_n^{\bullet}\|_{L^2} \xrightarrow{\mathbb{P}} 0$. Now, by the central limit theorem in Hilbert spaces, we have

$$\sqrt{n}\left(\frac{\tilde{W}_n^{\bullet}}{\sqrt{n}}-z\right) \xrightarrow{\mathcal{D}} \mathcal{W},$$

where $\mathcal{W} \in L^2$ is a centered Gaussian process with covariance kernel $K_W(s,t) = \mathrm{Cov}(\tilde{W}_1^{\bullet}(s),\tilde{W}_1^{\bullet}(t))$. The stated formula is derived by straightforward calculation. □

References

Allison, J. S. and Santana, L. On a data-dependent choice of the tuning parameter appearing in certain goodness-of-fit tests. *J. Stat. Comput. Simul.*, **85** (16), 3276–3288 (2015). MR3384048.

Anderson, T. W. and Darling, D. A. Asymptotic theory of certain "goodness of fit" criteria based on stochastic processes. *Ann. Math. Statistics*, **23**, 193–212 (1952). MR50238.

Baringhaus, L., Danschke, R., and Henze, N. Recent and classical tests for normality—a comparative study. *Comm. Stat. Simulat.*, **18** (1), 363–379 (1989). MR1000045.

Baringhaus, L., Ebner, B., and Henze, N. The limit distribution of weighted L^2-goodness-of-fit statistics under fixed alternatives, with applications. *Ann. Inst. Statist. Math.*, **69** (5), 969–995 (2017). MR3689158.

Baringhaus, L. and Henze, N. A consistent test for multivariate normality based on the empirical characteristic function. *Metrika*, **35** (6), 339–348 (1988). MR980849.

Becker, M. and Klößner, S. *PearsonDS: Pearson Distribution System* (2017). URL `https://CRAN.R-project.org/package=PearsonDS`. R package version 1.1.

Betsch, S. and Ebner, B. Testing normality via a distributional fixed point property in the Stein characterization. *TEST*, **29** (1), 105–138 (2020). MR4063385.

Bowman, K. O. and Shenton, L. R. Omnibus test contours for departures from normality based on $\sqrt{b_1}$ and b_2. *Biometrika*, **62** (2), 243–250 (1975). MR381079.

Butsch, L. and Ebner, B. *mnt: Affine Invariant Tests of Multivariate Normality* (2020). URL `https://CRAN.R-project.org/package=mnt`. R package version 1.3.

Chen, L. H. Y., Goldstein, L., and Shao, Q.-M. *Normal approximation by Stein's method.* Probability and its Applications (New York). Springer, Heidelberg (2011). ISBN 978-3-642-15006-7. MR2732624.

del Barrio, E., Cuesta-Albertos, J. A., and Matrán, C. Contributions of empirical and quantile processes to the asymptotic theory of goodness-of-fit tests. *Test*, **9** (1), 1–96 (2000). MR1790430.

del Barrio, E., Cuesta-Albertos, J. A., Matrán, C., and Rodríguez-Rodríguez, J. M. Tests of goodness of fit based on the L_2-Wasserstein distance. *Ann. Statist.*, **27** (4), 1230–1239 (1999). MR1740113.

Dörr, P., Ebner, B., and Henze, N. A new test of multivariate normality by a double estimation in a characterizing PDE. *Metrika*, **84** (3), 401–427 (2021). MR4233599.

Dörr, P., Ebner, B., and Henze, N. Testing multivariate normality by zeros of the harmonic oscillator in characteristic function spaces. *Scand. J. Stat.*, **48** (2), 456–501 (2021). DOI: 10.1111/sjos.12477.

Ebner, B. On combining the zero bias transform and the empirical characteristic function to test normality. *ArXiv Mathematics e-prints* (2020). arXiv: 2002.12085.

Epps, T. W. and Pulley, L. B. A test for normality based on the empirical characteristic function. *Biometrika*, **70** (3), 723–726 (1983). MR725389.

Farrell, P. J. and Rogers-Stewart, K. Comprehensive study of tests for normality and symmetry: extending the Spiegelhalter test. *J. Stat. Comput. Simul.*, **76** (9), 803–816 (2006). MR2253788.

Gaunt, R. E. A Stein characterisation of the generalized hyperbolic distribution. *ESAIM Probab. Stat.*, **21**, 303–316 (2017). MR3743916.

Gaunt, R. E. A simple proof of the characteristic function of Student's t-distribution. *Comm. Stat. Theory* (2019). DOI: 10.1080/03610926.2019.1702695.

Goldstein, L. and Reinert, G. Stein's method and the zero bias transformation with application to simple random sampling. *Ann. Appl. Probab.*, **7** (4), 935–952 (1997). MR1484792.

Gross, J. and Ligges, U. *nortest: Tests for Normality* (2015). URL `https://CRAN.R-project.org/package=nortest`. R package version 1.0-4.

Henze, N. An approximation to the limit distribution of the Epps-Pulley test statistic for normality. *Metrika*, **37** (1), 7–18 (1990). MR1046994.

Henze, N. Tests of normality (in German). *Allg. Stat. Archiv, J. German Stat. Soc.*, **78** (3), 293–317 (1994).

Henze, N. Invariant tests for multivariate normality: a critical review. *Statist. Papers*, **43** (4), 467–506 (2002). MR1932769.

Henze, N. and Jiménez-Gamero, M. D. A test for Gaussianity in Hilbert spaces via the empirical characteristic functional. *Scand. J. Stat.*, **48** (2), 406–428 (2021). DOI: 10.1111/sjos.12470.

Henze, N. and Visagie, J. Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Ann. Inst. Statist. Math.*, **72** (5), 1109–1136 (2020). MR4137748.

Henze, N. and Wagner, T. A new approach to the BHEP tests for multivariate normality. *J. Multivariate Anal.*, **62** (1), 1–23 (1997). MR1467870.

Jarque, C. M. and Bera, A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Econom. Lett.*, **6** (3), 255–259 (1980). MR615323.

Johnson, N. L., Kotz, S., and Balakrishnan, N. *Continuous univariate distributions. Vol. 1.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, second edition (1994). ISBN 0-471-58495-9. MR1299979.

Kellner, J. and Celisse, A. A one-sample test for normality with kernel methods. *Bernoulli*, **25** (3), 1816–1837 (2019). MR3961232.

Klar, B. Goodness-of-fit tests for the exponential and the normal distribution based on the integrated distribution function. *Ann. Inst. Statist. Math.*, **53** (2), 338–353 (2001). MR1841140.

Krauczi, E. A study of the quantile correlation test for normality. *TEST*, **18** (1), 156–165 (2009). MR2495969.

Landry, L. and Lepage, Y. Empirical behavior of some tests for normality. *Comm. Stat. Simulat.*, **21** (4), 971–999 (1992). DOI: 10.1080/03610919208813062.

Ledoux, M. and Talagrand, M. *Probability in Banach spaces. Isoperimetry and processes*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)]*. Springer-Verlag, Berlin (1991). ISBN 3-540-52013-9. MR1102015.

Lilliefors, H. W. On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *J. Amer. Stat. Assoc.*, **62** (318), 399–402 (1967).

Lukacs, E. *Characteristic functions.* Hafner Publishing Co., New York (1970). MR0346874.

Maplesoft. *Maple 2019.* Maplesoft, Waterloo ON, Canada (2019).

Pearson, E. S., D'Agostino, R. B., and Bowman, K. O. Tests for Departure from Normality: Comparison of Powers. *Biometrika*, **64** (2), 231–246 (1977).

R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria (2019). URL `https://www.R-project.org/`.

Romão, X., Delgado, R., and Costa, A. An empirical power comparison of univariate goodness-of-fit tests for normality. *J. Stat. Comput. Simul.*, **80** (5-6), 545–591 (2010). MR2758146.

Shapiro, S. S. and Wilk, M. B. An analysis of variance test for normality: Complete samples. *Biometrika*, **52**, 591–611 (1965). MR205384.

Shapiro, S. S., Wilk, M. B., and Chen, H. J. A comparative study of various tests for normality. *J. Amer. Statist. Assoc*, **63**, 1343–1372 (1968). MR0237069.

Shevtsova, I. G. A square bias transformation of a probability distribution with applications. *Dokl. Akad. Nauk*, **451** (1), 14–16 (2013). MR3155154.

Stein, C. 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 (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability theory*, pp. 583–602 (1972). MR0402873.

Tavakoli, M., Arghami, N., and Abbasnejad, M. A goodness of fit test for normality based on Balakrishnan-Sanghvi information. *J. Iran. Stat. Soc. (JIRSS)*, **18** (1), 177–190 (2019). MR3961538.

Tenreiro, C. On the automatic selection of the tuning parameter appearing in certain families of goodness-of-fit tests. *J. Stat. Comput. Simul.*, **89** (10), 1780–1797 (2019). MR3945298.

Trapletti, A. and Hornik, K. *tseries: Time Series Analysis and Computational Finance* (2019). URL `https://CRAN.R-project.org/package=tseries`. R package version 0.10-47.

Vasicek, O. A test for normality based on sample entropy. *J. Roy. Statist. Soc. Ser. B*, **38** (1), 54–59 (1976). MR420958.

Yap, B. W. and Sim, C. H. Comparisons of various types of normality tests. *J. Stat. Comput. Simul.*, **81** (12), 2141–2155 (2011). MR2864195.