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 $X, X_1, X_2, \ldots$ 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

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

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
“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

\[
\begin{cases}
\varphi'(t) = -t\varphi(t), \\
\varphi(0) = 1,
\end{cases}
\]

(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

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

where \(w(\cdot)\) is a suitable bounded weight function and \(|x|^2 = \text{Re}(x)^2 + \text{Im}(x)^2\) 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). Tactily, we assume the conditions

\[w(t) = w(-t), \quad t \in \mathbb{R}, \quad \int_{-\infty}^{\infty} w(t)dt < \infty.
\]

(1.3)

Note that \(Z_n\) depends only on the scaled residuals \(Y_{1,1}, \ldots, Y_{n,n}\) 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

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

(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 $\varphi''(0) = \sigma^2$. 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

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

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

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

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 $L^2 = L^2(\mathbb{R}, \mathcal{B}, w d\mathcal{L}^1)$, where $\mathcal{B}$ is the Borel-$\sigma$-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

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

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

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

where

$$W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (Y_{n,j} + t) \cos(tY_{n,j}) + (t - Y_{n,j}) \sin(tY_{n,j}), \quad t \in \mathbb{R}.$$  \hspace{1cm} (2.1)
Motivated by a multivariate Taylor expansion we consider the processes

\[ W_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j + t) \cos(tX_j) + (t - X_j) \sin(tX_j) \]

\[ + \left( \left( tX_j - (t^2 + 1) \right) \cos(tX_j) + (tX_j + (t^2 + 1)) \sin(tX_j) \right) X_n \]

\[ + X_j \left( (tX_j - (t^2 + 1)) \cos(tX_j) + (tX_j + (t^2 + 1)) \sin(tX_j) \right) (S_n - 1) \]

and

\[ \hat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} (X_j + t) \cos(tX_j) + (t - X_j) \sin(tX_j) \]

\[ - \exp \left( -\frac{t^2}{2} \right) X_j + t \exp \left( -\frac{t^2}{2} \right) (X_j^2 - 1), \]

t \in \mathbb{R}. In what follows let \( X_1, X_2, \ldots \) be i.i.d. 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 \( \hat{W}_n \) are asymptotically equivalent. The proof is found in Appendix A.1.

**Lemma 2.1.** We have under \( H_0 \)

\[ \| W_n - W_n^* \|_{L^2} \xrightarrow{p} 0 \quad \text{and} \quad \| W_n^* - \hat{W}_n \|_{L^2} \xrightarrow{p} 0. \]

In order to derive the asymptotic null distribution of \( Z_n \), it suffices to show the weak convergence of \( \hat{W}_n \) 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 \( W_n \xrightarrow{D} W \) 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 \( \hat{W}_n \). Note that

\[ \hat{W}_n(t) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \hat{W}_{n,j}(t), \quad t \in \mathbb{R}, \]

where

\[ \hat{W}_{n,j}(t) = (X_j + t) \cos(tX_j) + (t - X_j) \sin(tX_j) \]

\[ - \exp \left( -\frac{t^2}{2} \right) X_j + t \exp \left( -\frac{t^2}{2} \right) (X_j^2 - 1), \quad t \in \mathbb{R}, \]

\( j = 1, 2, \ldots, n \) and \( \mathbb{E}\hat{W}_{n,1} = 0 \). Since \( \hat{W}_{n,1}, \hat{W}_{n,2}, \ldots \) are i.i.d. 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). Tidious calculations then show that the covariance kernel \( K_Z(s, t) = \mathbb{E} \left( \hat{W}_{n,1}(s) \hat{W}_{n,1}(t) \right) \) takes the given form.

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

**Corollary 2.3.** We have as \( n \to \infty \)

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

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 $X_{n,1}, \ldots, X_{n,n}$, $n \in \mathbb{Z}_+$, with Lebesgue density

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

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 $\int_{-\infty}^{\infty} c(t) f(t) \, dt = 0$. Notice that, since $c$ is bounded, we may assume $n$ to be large enough to ensure $f_n \geq 0$. Setting

$$\mu_n = \bigotimes_{j=1}^{n} f_{L^1} \quad \text{and} \quad \nu_n = \bigotimes_{j=1}^{n} f_{nL^1},$$

it is shown in Betsch and Ebner (2020), section 4, that by LeCam’s first Lemma $\nu_n$ is contiguous to $\mu_n$. Writing

$$\eta(x,s) = (x+s) \cos(sx) + (s-x) \sin(sx)$$

$$- \exp\left(-\frac{s^2}{2}\right) x + s \exp\left(-\frac{s^2}{2}\right) (x^2 - 1), \ x, s \in \mathbb{R},$$

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 $X_{n,1}, \ldots, X_{n,n}$, we have

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

where $W$ is the limiting Gaussian process of Theorem 2.2 and

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

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 $n \to \infty$,

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

**Proof:** Let

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

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

$$\left\| n^{-1/2}(W_n - W_n^0) \right\|_{L^2} \xrightarrow{a.s.} 0. \quad (4.1)$$
Next, we consider (for the definition of \( \Psi(\cdot, \cdot) \) see (A.1) in Appendix A.1)

\[
\begin{align*}
n^{-1/2}(W_n^*(t) - W_n^0(t)) &= \frac{1}{n} \sum_{j=1}^{n} \left( (tX_j - (t^2 + 1)) \cos(tX_j) \\
&+ (tX_j + (t^2 + 1)) \sin(tX_j) \right) \mathcal{X}_n \\
&+ X_j ((tX_j - (t^2 + 1)) \cos(tX_j) \\
&+ (tX_j + (t^2 + 1)) \sin(tX_j)) (S_n - 1) \\
&= \mathcal{X}_n \frac{1}{n} \sum_{j=1}^{n} \Psi(t, X_j) + (S_n - 1) \frac{1}{n} \sum_{j=1}^{n} X_j \Psi(t, X_j).
\end{align*}
\]

By the triangle inequality, we have

\[
\begin{align*}
\left\| n^{-1/2}(W_n^* - W_n^0) \right\|_{L^2}^2 &\leq 2 \left\| \mathcal{X}_n \right\|_2^2 \left\| \frac{1}{n} \sum_{j=1}^{n} \Psi(\cdot, X_j) \right\|_{L^2}^2 \\
&+ 2|S_n - 1|^2 \left\| \frac{1}{n} \sum_{j=1}^{n} X_j \Psi(\cdot, X_j) \right\|_{L^2}^2. \tag{4.2}
\end{align*}
\]

By the strong law of large numbers in Hilbert spaces and \( (\mathcal{X}_n, S_n) \overset{a.s.}{\to} (0, 1) \), 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

\[
n^{-1/2}W_n^0(t) \overset{a.s.}{\to} \mathbb{E}[(X + t) \cos(tX) + (t - X) \sin(tX)]
\]

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

\[
\frac{Z_n}{n} = \left\| n^{-1/2}W_n^0 \right\|_{L^2}^2 \overset{a.s.}{\to} \int_{-\infty}^{\infty} \left| \mathbb{E}[(X + t) \cos(tX) + (t - X) \sin(tX)] \right|^2 w(t) dt = \Delta.
\]

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 \( Z_n \overset{a.s.}{\to} \infty \) 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 \( z(t) = \mathbb{E}[(X + t) \cos(tX) + (t - X) \sin(tX)] \) and \( W_n^*(\cdot) = n^{-1/2}W_n(t) \), then we have

\[
\sqrt{n} \left( \frac{Z_n}{n} - \Delta \right) = \sqrt{n} \left( \|W_n^*\|_{L^2}^2 - \|z\|_{L^2}^2 \right) = \sqrt{n} \langle W_n^* - z, W_n^* + z \rangle_{L^2}
\]

\[
= \sqrt{n} \langle W_n^* - z, 2z + W_n^* - z \rangle_{L^2}
\]

\[
= 2 \langle \sqrt{n}(W_n^* - z), z \rangle_{L^2} + \frac{1}{\sqrt{n}} \| \sqrt{n}(W_n^* - z) \|_{L^2}^2. \tag{4.3}
\]

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

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

\[
\sqrt{n}(W_n^* - z) \overset{\mathcal{D}}{\to} \mathcal{W},
\]
in $L^2$, where $W$ is a centred Gaussian process in $L^2$ with covariance kernel

$$K_W(s, t) = \mathbb{E} [(st + X^2) \cos((s - t)X) + (st - X^2)X \sin((s + t))]$$

where

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

Lemma 4.2 shows that $\sqrt{n} (W_n^* - z)$ is a tight sequence in $L^2$, thus we see by Slutsky’s Lemma and $\frac{1}{\sqrt{n}} \|\sqrt{n} (W_n^* - z)\|_{L^2}^2 \xrightarrow{p} 0$ that the limit distribution of $\sqrt{n} \left( \frac{Z_n}{n} - \Delta \right)$ in (4.3) only depends on $2\langle \sqrt{n}(W_n^* - z), z \rangle_{L^2}$. 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

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

where

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

\[
K f(s) = \int_{-\infty}^{\infty} K_Z(s, t) f(t) w_a(t) \, dt,
\]

on \( L^2 \). Notice that \( K \) 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 \( \lambda_j \), 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

\[
\kappa_m(a) = 2^{m-1}(m-1)! \int_{-\infty}^{\infty} h_m(t, t) w_a(t) \, dt,
\]

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

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

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

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

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

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

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.

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

Table 5.1. Values of mean, variance, skewness and kurtosis of \( Z_{\infty, a} \) rounded to 4 decimals
Table 6.2. Empirical quantiles of $Z_{n,a}$ for $n = 20, 50, 100$ (100 000 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 $t_\nu$-distribution with $\nu \in \{3, 5, 10\}$ degrees of freedom, as well as the uniform distribution $U(-\sqrt{3}, \sqrt{3})$, and asymmetric distributions, such as the $\chi^2$-distribution with $\nu \in \{5, 15\}$ degrees of freedom, the beta distributions $B(1, 4)$ and $B(2, 5)$, and the gamma distributions $\Gamma(1, 5)$ and $\Gamma(5, 1)$, both parametrized by their shape and rate parameter, the Gumbel distribution $\text{Gum}(1, 2)$ with location parameter 1 and scale parameter 2, the Weibull distribution $W(1, 0.5)$ with scale parameter 1 and shape parameter 0.5, and the lognormal distribution $\text{LN}(0, 1)$. As representatives of bimodal distributions, we simulate the mixture of normal distributions $\text{NMix}(p, \mu, \sigma^2)$, where the random variables are generated by $(1 - p)N(0, 1) + p N(\mu, \sigma^2)$, $p \in (0, 1)$, $\mu \in \mathbb{R}$, $\sigma > 0$. 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

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

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

$$BE = \frac{2}{n} \sum_{1 \leq j < k \leq n} \left\{ 1 - \Phi\left(\frac{Y_{n,j}}{\sqrt{a}}\right) \left(\frac{Y_{n,j}^2}{\sigma} - 1\right)(Y_{n,k}^2 - 1) + a Y_{n,j} Y_{n,k} \right\}$$
$$+ \frac{a}{\sqrt{2\pi a}} \exp\left(-\frac{Y_{n,j}^2}{2a}\right) \left(-Y_{n,j}^2 Y_{n,k} + Y_{n,k} + Y_{n,j}\right)$$
$$+ \frac{1}{n} \sum_{j=1}^{n} \left( 1 - \Phi\left(\frac{Y_{n,j}}{\sqrt{a}}\right) \right) \left(Y_{n,j}^4 + (a - 2) Y_{n,j}^2 + 1\right)$$
$$+ \frac{a}{\sqrt{2\pi a}} \exp\left(-\frac{Y_{n,j}^2}{2a}\right) \left(2 Y_{n,j}^2 - Y_{n,j}^2\right),$$
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

\[
\text{BHEP} = \frac{1}{n} \sum_{j,k=1}^{n} \exp \left( -\frac{\beta^2}{2} (Y_{n,j} - Y_{n,k})^2 \right) \left[ \frac{2}{\sqrt{1 + \beta^2}} \sum_{j=1}^{n} \exp \left( -\frac{\beta^2}{2(1 + \beta^2)} Y_{n,j}^2 \right) + \frac{n}{\sqrt{1 + 2\beta^2}} \right].
\]

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 \texttt{R} package \texttt{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

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

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 \( \chi^2_{15} \), the \( \Gamma(5, 1) \) and the Gum(1, 2) distribution. Very interesting is the behaviour of the HV-test for the uniform \( U(-\sqrt{3}, \sqrt{3}) \), where it fails to detect the alternative for any value of \( \gamma \) 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
<table>
<thead>
<tr>
<th>Alt.</th>
<th>$n \backslash a$</th>
<th>$Z_{n,a}$</th>
<th>$HV_{\gamma}$</th>
<th>SW</th>
<th>BCMR</th>
<th>BHEP</th>
<th>AD</th>
<th>JB</th>
<th>BE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>0.1</td>
<td>0.25</td>
<td>0.5</td>
<td>0.75</td>
<td>1</td>
<td>3</td>
<td>5</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.5</td>
<td>5</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N(0, 1)</td>
<td>20</td>
<td>5</td>
<td>5</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>25</td>
<td>28</td>
<td>27</td>
<td>24</td>
<td>23</td>
<td>18</td>
<td>17</td>
<td>17</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>29</td>
<td>30</td>
<td>7</td>
<td>25</td>
</tr>
<tr>
<td>NMix(0.3, 1, 0.25)</td>
<td>50</td>
<td>61</td>
<td>65</td>
<td>62</td>
<td>59</td>
<td>57</td>
<td>54</td>
<td>41</td>
<td>38</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>92</td>
<td>93</td>
<td>91</td>
<td>89</td>
<td>86</td>
<td>76</td>
<td>72</td>
<td>68</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>35</td>
<td>42</td>
<td>42</td>
<td>41</td>
<td>40</td>
<td>33</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>98</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>96</td>
<td>87</td>
<td>77</td>
<td>75</td>
</tr>
<tr>
<td>NMix(0.5, 1, 4)</td>
<td>50</td>
<td>78</td>
<td>85</td>
<td>83</td>
<td>80</td>
<td>77</td>
<td>59</td>
<td>53</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>98</td>
<td>97</td>
<td>86</td>
<td>77</td>
<td>66</td>
</tr>
<tr>
<td>$t_3$</td>
<td>20</td>
<td>19</td>
<td>27</td>
<td>33</td>
<td>35</td>
<td>36</td>
<td>36</td>
<td>36</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>37</td>
<td>52</td>
<td>61</td>
<td>64</td>
<td>65</td>
<td>62</td>
<td>61</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>63</td>
<td>79</td>
<td>85</td>
<td>87</td>
<td>87</td>
<td>84</td>
<td>82</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>18</td>
<td>19</td>
<td>20</td>
<td>20</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>$t_5$</td>
<td>20</td>
<td>14</td>
<td>22</td>
<td>30</td>
<td>34</td>
<td>35</td>
<td>36</td>
<td>35</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>23</td>
<td>38</td>
<td>48</td>
<td>52</td>
<td>54</td>
<td>53</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>23</td>
<td>38</td>
<td>48</td>
<td>52</td>
<td>54</td>
<td>53</td>
<td>50</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>$t_{10}$</td>
<td>20</td>
<td>5</td>
<td>6</td>
<td>8</td>
<td>11</td>
<td>14</td>
<td>16</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>9</td>
<td>12</td>
<td>19</td>
<td>20</td>
<td>23</td>
<td>22</td>
<td>21</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>8</td>
<td>12</td>
<td>16</td>
<td>19</td>
<td>20</td>
<td>23</td>
<td>22</td>
<td>22</td>
</tr>
<tr>
<td>$U(-\sqrt{3}, \sqrt{3})$</td>
<td>20</td>
<td>18</td>
<td>21</td>
<td>15</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>47</td>
<td>59</td>
<td>59</td>
<td>50</td>
<td>36</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>87</td>
<td>94</td>
<td>95</td>
<td>94</td>
<td>91</td>
<td>81</td>
<td>84</td>
<td>84</td>
</tr>
<tr>
<td>$\chi_5^2$</td>
<td>20</td>
<td>18</td>
<td>28</td>
<td>36</td>
<td>40</td>
<td>41</td>
<td>40</td>
<td>39</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>44</td>
<td>66</td>
<td>78</td>
<td>82</td>
<td>84</td>
<td>85</td>
<td>84</td>
<td>83</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>80</td>
<td>95</td>
<td>98</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>7</td>
<td>11</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>13</td>
<td>23</td>
<td>33</td>
<td>39</td>
<td>42</td>
<td>45</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>$\chi_{15}^2$</td>
<td>20</td>
<td>32</td>
<td>41</td>
<td>47</td>
<td>48</td>
<td>48</td>
<td>43</td>
<td>41</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>78</td>
<td>88</td>
<td>92</td>
<td>93</td>
<td>94</td>
<td>91</td>
<td>89</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>98</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>$B(1, 4)$</td>
<td>20</td>
<td>9</td>
<td>12</td>
<td>14</td>
<td>15</td>
<td>14</td>
<td>13</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>18</td>
<td>31</td>
<td>39</td>
<td>42</td>
<td>43</td>
<td>39</td>
<td>37</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>38</td>
<td>63</td>
<td>75</td>
<td>79</td>
<td>81</td>
<td>79</td>
<td>76</td>
<td>73</td>
</tr>
<tr>
<td>$\Gamma(1, 5)$</td>
<td>20</td>
<td>55</td>
<td>66</td>
<td>73</td>
<td>74</td>
<td>75</td>
<td>71</td>
<td>70</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>96</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>$W(1, 0.5)$</td>
<td>20</td>
<td>96</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>96</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
<tr>
<td>$\text{Gum}(1, 2)$</td>
<td>20</td>
<td>12</td>
<td>19</td>
<td>26</td>
<td>29</td>
<td>31</td>
<td>31</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>24</td>
<td>44</td>
<td>58</td>
<td>65</td>
<td>68</td>
<td>71</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td>$\text{LN}(0, 1)$</td>
<td>20</td>
<td>76</td>
<td>84</td>
<td>88</td>
<td>90</td>
<td>90</td>
<td>88</td>
<td>87</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
<td>99</td>
</tr>
</tbody>
</table>

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 $\tau^2$ 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.

Acknowledgements

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 $t, x \in \mathbb{R}$

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

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 $\mu$ and $(\sigma - 1)$. With that notation and $(\mathbf{X}_n, S_n) \xrightarrow{p} (0, 1)$ (assuring that the stochastic remainder is $o_p(1)$) we have

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

and hence $\|W_n - W_n^*\|_{L^2} = o_p(1)$ 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

$$
\Psi(t, x) = (tx - (t^2 + 1)) \cos(tx) + (tx + (t^2 + 1)) \sin(tx), \quad (A.1)
$$
Hence, we have for $t \in \mathbb{R}$

$$
\mathbb{E}(\Psi(t, X_1)) = -\exp(-t^2/2) = -\varphi(t),
\mathbb{E}(X_1\Psi(t, X_1)) = 2t\exp(-t^2/2) = -2\varphi'(t).
$$

Now, it is easy to see that

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

Since

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

we have

$$
\| W_n - \tilde{W}_n \|_{L^2}^2 = \int_{-\infty}^{\infty} \left| \frac{1}{n} \sum_{j=1}^{n} \Psi(t, X_1)X_n + X_j \Psi(t, X_j)(S_n - 1) + \varphi(t)X_n \\
+ \varphi'(t)(X_j^2 - 1) \right|^2 w(t) dt \\
\leq \int_{-\infty}^{\infty} \left| \left( \frac{1}{n} \sum_{j=1}^{n} \Psi(t, X_1) + \varphi(t) \right) \frac{1}{\sqrt{n}} \sum_{l=1}^{n} X_l \\
+ \frac{1}{n} \sum_{j=1}^{n} X_j \Psi(t, X_j) + \varphi'(t) \right| \frac{1}{\sqrt{n}} \sum_{l=1}^{n} (X_l^2 - 1)^2 \right|^2 w(t) dt + o_p(1)
$$

$$
\leq 2 \left\{ \left\| \frac{1}{n} \sum_{j=1}^{n} \Psi(\cdot, X_j) + \varphi(\cdot) \right\|_{L^2}^2 \left( \frac{1}{\sqrt{n}} \sum_{l=1}^{n} X_l \right)^2 \\
+ \left\| \frac{1}{n} \sum_{j=1}^{n} X_j \Psi(\cdot, X_j) + 2\varphi'(\cdot) \right\|_{L^2}^2 \left( \frac{1}{\sqrt{n}} \sum_{l=1}^{n} (X_l^2 - 1) \right)^2 \right\} + o_p(1)
$$

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

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

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

$$
\sqrt{n} \left( \frac{\tilde{W}_n}{\sqrt{n}} - z \right) \overset{D}{\to} \mathcal{W},
$$

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


Tenreiro, C. On the automatic selection of the tuning parameter appearing in certain families of

