Received: 23 October 2019 | Revised: 26 April 2020 | Accepted: 28 April 2020

DOI: 10.1111/sjos.12470

ORIGINAL ARTICLE

Scandinavian Journal of Statistics

A test for Gaussianity in Hilbert spaces via the empirical characteristic functional

Norbert Henze[1] | **María Dolores Jiménez-Gamero**[2]

[1]Institute of Stochastics, Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany

[2]Department of Statistics and Operations Research, University of Seville, Seville, Spain

Correspondence
María Dolores Jiménez-Gamero, Facultad de Matemáticas, Dpto Estadística e Investigación Operativa, Avda. Reina Mercedes. s.n. 41.012-Sevilla, Spain.
Email: dolores@us.es

Funding information
Consejería de Economía, Innovación, Ciencia y Empleo, Junta de Andalucía, Grant/Award Number: P18-FR-2369; Spanish Ministry of Economy, Industry and Competitiveness, the State Agency of Investigation, the European Regional Development Fund., Grant/Award Number: MTM2017-89422-P

Abstract

Let $X_1, X_2, \dots$ be independent and identically distributed random elements taking values in a separable Hilbert space $\mathbb{H}$. With applications for functional data in mind, $\mathbb{H}$ may be regarded as a space of square-integrable functions, defined on a compact interval. We propose and study a novel test of the hypothesis H_0 that X_1 has some unspecified nondegenerate Gaussian distribution. The test statistic $T_n = T_n(X_1, \dots, X_n)$ is based on a measure of deviation between the empirical characteristic functional of $X_1, \dots, X_n$ and the characteristic functional of a suitable Gaussian random element of $\mathbb{H}$. We derive the asymptotic distribution of T_n as $n \to \infty$ under H_0 and provide a consistent bootstrap approximation thereof. Moreover, we obtain an almost sure limit of T_n and the limit distributions of T_n under fixed and contiguous alternatives to Gaussianity. Simulations show that the new test is competitive with respect to the hitherto few competitors available.

KEYWORDS

characteristic functional, functional data, Gaussianity, goodness-of-fit test, Hilbert space

1 | INTRODUCTION

The normal distribution continues to play a prominent role, since many statistical procedures for finite-dimensional data assume an underlying normal distribution. It is thus not surprising that a myriad of tests for multivariate normality have been proposed. For some more recent approaches, see for example, Arcones (2007), Doornik and Hansen (2008), Ebner (2012), Henze and Jiménez-Gamero (2019), Henze, Jiménez-Gamero, and Meintanis (2019), Henze and Visagie (2019), Kankainen, Taskinen, and Oja (2007), Pudelko (2005), Székely and Rizzo (2005), Thulin (2014), Villaseñor-Alva and Estrada (2009), and Voinov, Pya, Makarov, and Voinov (2016). A survey of affine invariant tests for multivariate normality is given in Henze (2002).

While some of these (and other) tests make use of certain properties that uniquely determine the normal law, others are based on a comparison of a nonparametric estimator of a function that characterizes a probability law with a parametric estimator of that function, obtained under the null hypothesis. A member of the latter class is the test of Epps and Pulley (1983). Although originally designed for the univariate case, the approach of Epps and Pulley was extended to test for multivariate normality by Baringhaus and Henze (1988) and Henze and Zirkler (1990). The resulting procedure, which is usually referred to as the Baringhaus-Henze-Epps-Pulley (BHEP) test, is based on a comparison of the empirical characteristic function (ECF) associated with suitably standardized data, with the characteristic function (CF) of the standard normal law in $\mathbb{R}^d$. Because of its nice properties (see Section 2), the BHEP test has been extended in several directions, such as testing for normality of the errors in linear models Jiménez-Gamero, Muñoz-García, and Pino-Mejías (2005), in nonparametric regression models, Hušková and Meintanis (2010), and in GARCH models Jiménez-Gamero (2014), just to cite a few.

The assumption of normality is important not only in the so-called *classical* context, in which the data take values in $\mathbb{R}^d$ for some fixed $d \in \mathbb{N}$, but also in other settings, such as functional data analysis. In fact, there are some inferential procedures, designed for functional data, that assume *Gaussianity* (which is a synonym for normality in that context). Examples are the test for the equality of covariance operators in chapter 5 of Horváth and Kokoszka (2012), or the test in Zhang, Liang, and Xiao (2010) for the equality of means. On the other hand, some methods are valid under quite general assumptions, but they greatly simplify when the assumption of normality is added, see, for example, Boente, Rodríguez, and Sued (2018). Thus, the problem of testing for Gaussianity is also of interest when dealing with functional data.

The literature of goodness-of-fit tests in the context of functional data, or more general, of data taking values in infinite-dimensional separable Hilbert spaces, is still rather sparse. Cuesta-Albertos, del Barrio, Fraiman, and Matrán (2007) consider a test based on random projections, while Bugni, Hall, Horowitz, and Neumann (2009) study an extension of the Cramér–von Mises test. The test in Bugni et al. (2009) assumes that the distribution in the null hypothesis depends on a finite-dimensional parameter. Górecki, Hörmann, Horváth, and Kokoszka (2018) propose Jarque–Bera type tests for Gaussianity. For a simple null hypothesis, Ditzhaus and Gaigall (2018) employ a test statistic that integrates, along all possible projections, univariate Cramér-von Mises test statistics, obtained by projecting the data. Since the probability distribution of a random element taking values in a separable Hilbert space is uniquely determined by its characteristic functional, see Laha and Rohatgi (1979), the objective of this paper is to extend the BHEP test to the functional data context.

The paper is organized as follows: Section 2 reviews the BHEP test. Section 3 highlights some key differences between the finite-dimensional case and the functional data context, and it introduces the test statistic, whose almost sure limit is derived in Section 4. The asymptotic null distribution of the test statistic is obtained in Section 5. Since this limit distribution depends on unknown quantities, we prove the consistency of a suitable bootstrap approximation. In Section 6, we derive the asymptotic distribution of the test statistic under alternatives. Section 7 presents the results of a simulation study, designed to assess the finite-sample performance of the test, and to compare it with some competitors. It also shows a real dataset application. The test statistic depends on a measure. Section 8 comments on such measure. The paper concludes with several remarks that point out some future research lines.

Throughout the manuscript, we will make use of the following standard notation: The Euclidean norm in $\mathbb{R}^d$, $d \in \mathbb{N}$, will be denoted by $\|\cdot\|$. The superscript $\top$ means transposition of column vectors and matrices. We write $\mathrm{N}_d(\mu,\Sigma)$ for the d-variate normal distribution with mean vector μ and nondegenerate covariance matrix Σ, and $\mathcal{N}_d$ stands for the class of all nondegenerate d-dimensional normal distributions. The symbol I_d denotes the unit matrix of order d, $\mathrm{i} = \sqrt{-1}$ is the imaginary unit, and $\mathbb{C}$ denotes the set of complex numbers. All random vectors and random elements will be defined on a sufficiently rich probability space $(\Omega, \mathcal{A}, \mathbb{P})$. The symbols $\mathbb{E}$ and $\mathbb{V}$ denote expectation and variance, respectively, and $\stackrel{\mathcal{D}}{=}$ and $\stackrel{\mathcal{D}}{\rightarrow}$ mean equality in distribution and convergence in distribution of random vectors and random elements, respectively. All limits are taken when $n\to\infty$, where n denotes the sample size.

2 | THE BHEP TEST IN $\mathbb{R}^d$ REVISITED

In this section, we revisit the BHEP test for finite-dimensional data. To this end, let $X_1,\ldots,X_n,\ldots$ be independent and identically distributed (iid) copies of a d-variate random column vector X. We assume that the distribution $\mathbb{P}^X$ of X is absolutely continuous with respect to Lebesgue measure. For testing the hypothesis $H_{0,d} : \mathbb{P}^X \in \mathcal{N}_d$, the rationale of the BHEP test is as follows: Write $\overline{X}_n = n^{-1}\sum_{j=1}^n X_j$ and $S_n = n^{-1}\sum_{j=1}^n (X_j - \overline{X}_n)(X_j - \overline{X}_n)^\top$ for the sample mean and the sample covariance matrix of $X_1,\ldots,X_n$, respectively, and let $Y_{n,j} = S_n^{-1/2}(X_j - \overline{X}_n)$, $j = 1,\ldots,n$, be the so-called *scaled residuals* of $X_1,\ldots,X_n$, which provide an empirical standardization of $X_1,\ldots,X_n$. Here, $S_n^{-1/2}$ denotes the unique symmetric square root of S_n^{-1}. If $n \geq d+1$, the matrix S_n is invertible with probability one, see Eaton and Perlman (1973). Since, under $H_{0,d}$ and for large n, the distribution of the scaled residuals should be close to the standard d-variate normal distribution $\mathrm{N}_d(0,\mathrm{I}_d)$, it is tempting to compare the ECF

$$\psi_n(t) = \frac{1}{n}\sum_{j=1}^n \exp(\mathrm{i}t^\top Y_{n,j}), \qquad t \in \mathbb{R}^d,$$

of $Y_{n,1},\ldots,Y_{n,n}$ with $\exp(-\|t\|^2/2)$, which is the CF of the law $\mathrm{N}_d(0,\mathrm{I}_d)$. The BHEP test rejects $H_{0,d}$ for large values of the weighted L^2-statistic

$$T_{n,d,\beta} = \int_{\mathbb{R}^d} \left|\psi_n(t) - \exp\left(-\frac{1}{2}\|t\|^2\right)\right|^2 w_{d,\beta}(t)\,\mathrm{d}t, \tag{1}$$

where $w_{d,\beta}(t) = (\beta^2 2\pi)^{-d/2}\exp(-\|t\|^2/(2\beta^2))$ is the probability density function of the d-variate normal distribution $\mathrm{N}_d(0, \beta^2\mathrm{I}_d)$, and $\beta > 0$ is a parameter. In the univariate case, this statistic has been proposed by Epps and Pulley (1983), and the extension to the case $d \geq 2$ has been studied by Baringhaus and Henze (1988) for the special case $\beta = 1$ and, for general β, by Henze and Zirkler (1990) and Henze and Wagner (1997). The acronym BHEP, after early developers of the idea, was coined by Csörgő (1989), who proved that the BHEP test is consistent against each non-normal alternative distribution (without any restriction on $\mathbb{P}^X$). The test statistic $T_{n,d,\beta}$ may be written as

$$T_{n,d,\beta} = \frac{1}{n^2}\sum_{j,k=1}^{n} \exp\left(-\frac{\beta^2}{2}\|Y_{n,j} - Y_{n,k}\|^2\right) - \frac{2}{n(1+\beta^2)^{d/2}}\sum_{j=1}^{n} \exp\left(-\frac{\beta^2}{2(1+\beta^2)}\|Y_{n,j}\|^2\right) + \frac{1}{(1+2\beta^2)^{d/2}}.$$

This representation shows that $T_{n,d,\beta}$ is a function of the scalar products $Y_{n,j}^\top Y_{n,k} = (X_j - \overline{X}_n)^\top S_n^{-1}(X_k - \overline{X}_n)$, $1 \le j,k \le n$, and is thus invariant with respect to full rank affine transformations of $X_1, \ldots, X_n$. Moreover, not even the computation of the square root $S_n^{-1/2}$ is needed. Affine invariance is a "soft necessary condition" for any genuine test for normality, since the class $\mathcal{N}_d$ is closed with respect to such transformations, see Henze (2002).

3 | THE SETTING AND THE TEST STATISTIC

Assume that X is a random element of the separable Hilbert space $\mathbb{H} = L^2([0,1], \mathbb{R})$ of (equivalence classes of) square-integrable real-valued functions, defined on the compact interval [0,1], with the inner product $\langle f, g\rangle = \int_0^1 f(t)g(t)\,\mathrm{d}t$, norm $\|f\|_{\mathbb{H}} = \langle f, f\rangle^{1/2}$, $f, g \in \mathbb{H}$, and equipped with the Borel σ-algebra. Throughout the paper we assume that X is square integrable, that is, we have $\mathbb{E}\|X\|_{\mathbb{H}}^2 < \infty$. As a consequence, we have $\mathbb{E}\|X\|_{\mathbb{H}} < \infty$, and thus there is a unique mean function $\mu = \mathbb{E}(X) \in \mathbb{H}$, which satisfies $\mathbb{E}\langle X, x\rangle = \langle \mu, x\rangle$ for each $x \in \mathbb{H}$. It follows that $\mu(t) = \mathbb{E}X(t)$ for almost all $t \in [0,1]$. Let $c(s,t) = \mathbb{E}[\{X(s) - \mu(s)\}\{X(t) - \mu(t)\}]$, $s,t \in [0,1]$, stand for the covariance function of X, and write $\mathcal{C} : \mathbb{H} \mapsto \mathbb{H}$ for the covariance operator of X, defined as $\mathcal{C}f = \mathbb{E}\left(\langle X - \mu, f\rangle X\right)$, or equivalently, as $\mathcal{C}f(s) = \int c(s,t)f(t)\,\mathrm{d}t$ for each $f \in \mathbb{H}$. The operator $\mathcal{C} : \mathbb{H} \to \mathbb{H}$ is linear, compact, symmetric and positive, and it is of trace class. In what follows, we denote this class of operators by $\mathcal{L}_{\mathrm{tr}}^{+}(\mathbb{H})$.

Let $X_1, \ldots, X_n$ be iid copies of X. The mean function and the covariance function of X can be consistently estimated by means of

$$\overline{X}_n(t) = \frac{1}{n}\sum_{j=1}^{n} X_j(t), \quad c_n(s,t) = \frac{1}{n}\sum_{j=1}^{n}\{X_j(s) - \overline{X}_n(s)\}\{X_j(t) - \overline{X}_n(t)\},$$

$s,t \in [0,1]$, respectively. The sample covariance operator $\mathcal{C}_n$, say, is given by $\mathcal{C}_n f(s) = \int_0^1 c_n(s,t)f(t)\,\mathrm{d}t$, $f \in \mathbb{H}$.

The characteristic functional of X, which uniquely determines the distribution $\mathbb{P}^X$ of X, is defined as the function $\varphi : \mathbb{H} \mapsto \mathbb{C}$, with $\varphi(f) = \mathbb{E}[\exp(\mathrm{i}\langle X, f\rangle)]$. By definition, $\mathbb{P}^X$ is Gaussian if, and only if, there is a $\mu \in \mathbb{H}$ (the *expectation* of X) and $\mathcal{C} \in \mathcal{L}_{\mathrm{tr}}^{+}(\mathbb{H})$ (the *covariance operator* of X), such that

$$\varphi(f) = \varphi(f; \mu, \mathcal{C}) = \exp\left(\mathrm{i}\langle \mu, f\rangle - \frac{1}{2}\langle \mathcal{C}f, f\rangle\right), \quad f \in \mathbb{H}. \tag{2}$$

In this case, we write $X \stackrel{\mathcal{D}}{=} \mathrm{N}(\mu, C)$.

Based on the data $X_1,\dots,X_n$, we are interested in testing the hypothesis H_0 that $\mathbb{P}^X$ is nondegenerate Gaussian, that is, in a test of

$$H_0 : \varphi(\cdot) = \varphi(\cdot; \mu, C), \quad \text{for some } \mu \in \mathbb{H} \text{ and some } C \in \mathcal{L}^+_{\mathrm{tr}}(\mathbb{H}).$$

Two main problems arise when one tries to extend the BHEP test for functional data.

First, a main difference between the finite-dimensional case and the functional data one is that in the latter case we have strict inclusion

$$\mathrm{sp}_n \subsetneq \mathbb{H}, \tag{3}$$

where $\mathrm{sp}_n = \mathrm{sp}(X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n)$ denotes the set of finite linear combinations of $X_1 - \overline{X}_n, \dots, X_n - \overline{X}_n$, while in the d-dimensional case $\mathrm{sp}_n = \mathbb{R}^d$ almost surely for each $n \geq d+1$. This point has an important implication related to invariance.

Section 2 made the case for affine invariance of any genuine test for normality in $\mathbb{R}^d$. In the infinite-dimensional case we have the following: If $X \in \mathbb{H}$ is Gaussian with mean μ and covariance operator C, and $A : \mathbb{H} \to \mathbb{H}$ is a bounded linear operator, then AX is also Gaussian (with mean $A\mu$ and covariance operator ACA^*, A^* being the adjoint of A). Therefore, arguing as in the previous section, any genuine test for Gaussianity should be invariant under bounded linear operators. However, since (3) holds for each fixed n, it is not reasonable to impose that the test statistic be invariant under any bounded linear operator.

This lack of invariance entails that the null distribution of any test statistic of H_0 depends on the population parameters μ and C. Therefore, the critical points must be approximated by (for example) some resampling method. Since our test statistic (to be defined in a moment) is translation invariant, its distribution does not depend on μ.

Second, recall that the BHEP statistic in $\mathbb{R}^d$ compares the ECF of the scaled residuals $Y_{n,1},\dots,Y_{n,n}$ with the CF of the standard normal law in $\mathbb{R}^d$. There is, however, no standard normal law in $\mathbb{H}$. Nevertheless, the BHEP test statistic (1) can be rewritten in the form

$$T_{n,d,\beta} = \int_{\mathbb{R}^d} \left| \phi_n(t) - \phi(t; \overline{X}_n, S_n) \right|^2 F_\beta(\mathrm{d}t).$$

Here, $\phi_n(\cdot)$ stands for the ECF of the data $X_1,\dots,X_n$, $\phi(\cdot;\mu,\Sigma)$ is the CF of the distribution $\mathrm{N}_d(\mu,\Sigma)$, and F_β is a certain distribution function on $\mathbb{R}^d$, see Lemma 2 in Jiménez-Gamero, Alba-Fernández, Muñoz-García, and Chalco-Cano (2009) for details. In view of the above expression, we consider the test statistic

$$T_n = T_n(X_1, \dots, X_n) = \int_{\mathbb{H}} \left| \varphi_n(f) - \varphi(f; \overline{X}_n, C_n) \right|^2 Q(\mathrm{d}f), \tag{4}$$

for testing H_0. Here, Q is some suitable probability measure on (the σ-field of Borel subsets of) $\mathbb{H}$, and φ_n is the empirical characteristic functional

$$\varphi_n(f) = \frac{1}{n} \sum_{j=1}^{n} \exp(\mathrm{i}\langle f, X_j \rangle), \quad f \in \mathbb{H}, \tag{5}$$

of $X_1,\ldots,X_n$. Straightforward algebra gives

$$T_n(X_1,\ldots,X_n) = \frac{1}{n} + \frac{2}{n^2}\sum_{1\le j<k\le n}\int_{\mathbb{H}} \cos\left(\langle f, X_j - X_k\rangle\right) Q(\mathrm{d}f) - \frac{2}{n}\sum_{j=1}^{n}\int_{\mathbb{H}} \cos\left(\langle f, X_j - \overline{X}_n\rangle\right)\exp\left(-\frac{\langle C_n f, f\rangle}{2}\right) Q(\mathrm{d}f) + \int_{\mathbb{H}} \exp\left(-\langle C_n f, f\rangle\right) Q(\mathrm{d}f). \tag{6}$$

Notice that T_n depends solely on the differences $X_j - X_k$ and $X_j - \overline{X}_n$. Consequently, the distribution of T_n does not depend on the unknown expectation $\mu = \mathbb{E}(X)$ of X.

In what follows, we will restrict the probability measure Q to be symmetric with respect to the zero element 0 of $\mathbb{H}$, that is, Q is invariant with respect to the mapping $x \mapsto -x$, $x \in \mathbb{H}$. With this assumption, the addition rule $\cos(\alpha - \beta) = \cos\alpha\ \cos\beta + \sin\alpha\ \sin\beta$ and considerations of symmetry yield

$$nT_n = \int_{\mathbb{H}} V_n^2(f)\ Q(\mathrm{d}f), \tag{7}$$

where

$$V_n(f) = \frac{1}{\sqrt{n}}\sum_{j=1}^{n}\left\{\cos\langle f, X_j - \overline{X}_n\rangle + \sin\langle f, X_j - \overline{X}_n\rangle - \exp\left(-\frac{1}{2}\langle C_n f, f\rangle\right)\right\}, \quad f \in \mathbb{H}. \tag{8}$$

The statistic (4) is similar to that considered in Bugni et al. (2009), which is based on a comparison of the empirical *distribution* functional with a parametric estimator of that functional, obtained under the null hypothesis. As argued in Bugni et al. (2009), Q must be chosen so that the resulting test statistic is tractable computationally. To this end, notice that the test statistic T_n is the expected value of the function $V_n^2(f)$ with respect to the measure Q. Hence, Monte Carlo integration is an option for computation, provided that Q can be easily sampled (from a computational point of view). Thus, if $f_1,\ldots,f_M$ is a random sample from Q, for some large M, then T_n can be approximated by

$$T_{M,n} = \frac{1}{M}\sum_{m=1}^{M} V_n^2(f_m). \tag{9}$$

4 | AN ALMOST SURE LIMIT FOR T_n

This section deals with an almost sure limit of T_n under general distributional assumptions.

Theorem 1. *Let $X_1,\ldots,X_n,\ldots$ be iid copies of a random element X of $\mathbb{H}$ satisfying $\mathbb{E}\|X\|_{\mathbb{H}}^2 < \infty$. Writing φ_X for the characteristic functional of X, and letting μ and C denote the expectation and the covariance operator of X, respectively, we have*

$$T_n \to^{\text{a.s.}} \tau_Q = \int_{\mathbb{H}} |\varphi_X(f) - \varphi(f;\mu,C)|^2 Q(\mathrm{d}f). \tag{10}$$

Proof. Recall $\varphi_n(f)$ from (5) and $\varphi(f;\mu,C)$ from (2). To stress the dependence on $\omega\in\Omega$, we write

$$T_n(\omega)=\int_{\mathbb{H}}\left|\varphi_n(f,\omega)-\varphi(f;\overline{X}_n(\omega),C_n(\omega))\right|^2 Q(\mathrm{d}f), \tag{11}$$

where

$$\varphi_n(f,\omega)=\frac{1}{n}\sum_{j=1}^{n}\exp\left(\mathrm{i}\langle f,X_j(\omega)\rangle\right),$$
$$\varphi(f,\overline{X}_n(\omega),C_n(\omega))=\exp\left(\mathrm{i}\langle\overline{X}_n(\omega),f\rangle-\frac{1}{2}\langle C_n(\omega)f,f\rangle\right),$$

and $\overline{X}_n(\omega)=n^{-1}\sum_{j=1}^{n}X_j(\omega)$. Moreover, $C_n(\omega)$ is the sample covariance operator based on $X_1(\omega),\dots,X_n(\omega)$.

Let $D\subset\mathbb{H}$ be a countable dense set. By the strong law of large numbers and the fact that the intersection of a countable collection of sets of probability one has probability one, there is a measurable subset Ω_0 of Ω such that $\mathbb{P}(\Omega_0)=1$, and for each $\omega\in\Omega_0$ we have, as $n\to\infty$, $\overline{X}_n(\omega)\to\mu$, $C_n(\omega)\to C$, $n^{-1}\sum_{j=1}^{n}\|X_j(\omega)\|_{\mathbb{H}}\to\mathbb{E}\|X\|_{\mathbb{H}}$, and

$$\lim_{n\to\infty}\varphi_n(g,\omega)=\varphi_X(g)\quad\text{for each }\ g\in D. \tag{12}$$

Now, fix $\omega\in\Omega_0$ and $f,g\in\mathbb{H}$, and notice that

$$|\varphi_n(f,\omega)-\varphi_X(f)|\le|\varphi_n(f,\omega)-\varphi_n(g,\omega)|+|\varphi_n(g,\omega)-\varphi_X(g)|+|\varphi_X(g)-\varphi_X(f)|.$$

If $g\in D$, (12) and the inequality $|e^{\mathrm{i}u}-e^{\mathrm{i}v}|\le|u-v|$, valid for real numbers u and v, yield

$$\begin{aligned}\limsup_{n\to\infty}|\varphi_n(f,\omega)-\varphi_X(f)|&\le\limsup_{n\to\infty}|\varphi_n(f,\omega)-\varphi_n(g,\omega)|+|\varphi_X(g)-\varphi_X(f)|\\&\le 2\ \|f-g\|_{\mathbb{H}}\ \mathbb{E}\|X\|_{\mathbb{H}}.\end{aligned}$$

Since $2\ \|f-g\|_{\mathbb{H}}\ \mathbb{E}\|X\|_{\mathbb{H}}$ can be made arbitrarily small because D is dense, the continuity of the exponential function entails

$$\lim_{n\to\infty}|\varphi_n(f,\omega)-\varphi(f;\overline{X}_n(\omega),C_n(\omega))|^2=|\varphi_X(f)-\varphi(f;\mu,C)|^2, \tag{13}$$

for each $f\in\mathbb{H}$ if $\omega\in\Omega_0$. Since the integrand in (11) is bounded from above by 4, the result follows from dominated convergence. ∎

Notice that τ_Q is nonnegative, and that τ_Q vanishes under H_0. Since the function $t(\cdot)$ that maps f into $t(f)=|\varphi_X(f)-\varphi(f;\mu,C)|^2$ is continuous, and since H_0 does not hold if and only if $t(f)>0$ for some $f\in\mathbb{H}$, a sufficient condition for the validity of H_0 if $\tau_Q=0$ is

$$Q(B(f,\varepsilon))>0\quad\text{for each }\ f\in\mathbb{H}\ \text{ and each }\ \varepsilon>0, \tag{14}$$

where $B(f,\varepsilon)=\{g\in\mathbb{H}:\|f-g\|_{\mathbb{H}}\le\varepsilon\}$, because then $t(\cdot)$ vanishes on $\mathbb{H}$ and hence $\varphi(f)=\varphi(f;\mu,C)$, $f\in\mathbb{H}$.

Observe that (14) holds if, for example, Q is the measure associated with the random element

$$Y = \sum_{j=1}^{\infty} a_j Z_j v_j, \tag{15}$$

where $\{v_j\}_{j=1}^{\infty}$ is an orthonormal basis of $\mathbb{H}$, $\{a_j\}_{j=1}^{\infty}$ is a sequence of positive numbers satisfying $\sum_{j=1}^{\infty} a_j < \infty$, and $Z_1, Z_2, \ldots$ are iid univariate standard normal variates. A specific instance, is the measure associated with a Wiener process, which in addition can be sampled easily. Clearly, if $a_j = 0$ for sufficiently large j, then (14) may fail, and thus it is possible to have $\tau_Q = 0$ under alternatives.

Obviously, a reasonable test for Gaussianity should reject H_0 for large values of T_n. In this respect, it is indispensable to have some information on the distribution of T_n under the null hypothesis, or at least an approximation to this distribution. This will be the topic of the next section.

5 | THE LIMIT NULL DISTRIBUTION OF T_n

In this section we assume that H_0 holds, that is, $X \overset{\mathcal{D}}{=} \mathrm{N}(\mu, \mathcal{C})$ for some $\mu \in \mathbb{H}$ and some $\mathcal{C} \in \mathcal{L}_{\mathrm{tr}}^{+}(\mathbb{H})$. Since the distribution of T_n defined in (4) does not depend on μ, we will make the tacit standing assumption $\mu = 0$ in what follows. Let L_Q^2 denote the Hilbert space of (equivalence classes of) measurable functions $\Upsilon : \mathbb{H} \mapsto \mathbb{R}$ satisfying $\int_{\mathbb{H}} \Upsilon(f)^2 Q(\mathrm{d}f) < \infty$. The scalar product and the resulting norm in L_Q^2 will be denoted by $\langle \Upsilon, \Phi \rangle_Q$ and $\|\Upsilon\|_Q = \sqrt{\langle \Upsilon, \Upsilon \rangle_Q}$, respectively. Notice that L_Q^2 is separable since $\mathbb{H}$ is separable.

The main result of this section is as follows.

Theorem 2. *Let $X_1, \ldots, X_n, \ldots$ be iid copies of a Gaussian random element X of $\mathbb{H}$ with covariance operator $\mathcal{C}$. Assume that $\int_{\mathbb{H}} \|f\|_{\mathbb{H}}^4 \, Q(\mathrm{d}f) < \infty$. Then $nT_n \overset{\mathcal{D}}{\to} \|\mathcal{V}\|_Q^2$, where $\mathcal{V}$ is a centred Gaussian random element of L_Q^2 having covariance kernel*

$$\mathbb{E}[\mathcal{V}(f)\mathcal{V}(g)] = \exp\left(-\frac{1}{2}(\sigma_f^2 + \sigma_g^2)\right)\left\{\exp(\sigma_{f,g}) - 1 - \sigma_{f,g} - \frac{1}{2}\sigma_{f,g}^2\right\}, \tag{16}$$

where $\sigma_{f,g} = \langle \mathcal{C}f, g \rangle$ and $\sigma_f^2 = \sigma_{f,f}$, $f, g \in \mathbb{H}$.

Proof. From (7), we have $nT_n = \|V_n\|_Q^2$, where V_n is given in (8). The idea is to approximate the random element V_n of L_Q^2 by a random element $V_{n,0}$ such that $\|V_n - V_{n,0}\|_Q = o_{\mathbb{P}}(1)$, and $V_{n,0}$ takes the form

$$V_{n,0}(f) = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \Psi(f, X_j), \tag{17}$$

where $\Psi : \mathbb{H} \times \mathbb{H} \to \mathbb{R}$ is some measurable function satisfying $\mathbb{E}\Psi(f, X) = 0$ for each $f \in \mathbb{H}$ and

$$\mathbb{E}\|\Psi(\cdot, X_1)\|_Q^2 = \int_{\mathbb{H}} \mathbb{E}\left[\Psi^2(f, X_1)\right] \, Q(\mathrm{d}f) < \infty. \tag{18}$$

In the sequel, the notation $W_n = o_{\mathbb{P}}(1)$ always refers to a random element of L^2_Q such that $\|W_n\|_Q$ tends to zero in probability as $n \to \infty$. Starting with (8), the addition theorems for the cosine and the sine function yield

$$\cos\langle f, X_j - \overline{X}_n\rangle = \cos\langle f, X_j\rangle + \langle f, \overline{X}_n\rangle \ \sin\langle f, X_j\rangle + O\left(\langle f, \overline{X}_n\rangle^2\right), \tag{19}$$

$$\sin\langle f, X_j - \overline{X}_n\rangle = \sin\langle f, X_j\rangle - \langle f, \overline{X}_n\rangle \ \cos\langle f, X_j\rangle + O\left(\langle f, \overline{X}_n\rangle^2\right). \tag{20}$$

Moreover, we have

$$\exp\left(-\frac{1}{2}\langle C_n f, f\rangle\right) = \exp\left(-\frac{1}{2}\sigma_f^2\right)\left(1 - \frac{1}{2}\langle (C_n - C)f, f\rangle\right) + o_{\mathbb{P}}\left(\langle (C_n - C)f, f\rangle\right), \tag{21}$$

and it follows that

$$\begin{aligned} V_n(f) = {} & \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{\cos\langle f, X_j\rangle + \sin\langle f, X_j\rangle - \exp\left(-\frac{1}{2}\sigma_f^2\right)\right\} \\ & + \langle f, \overline{X}_n\rangle \ \frac{1}{\sqrt{n}} \sum_{j=1}^n \left\{\sin\langle f, X_j\rangle - \cos\langle f, X_j\rangle\right\} \qquad (22) \\ & + \frac{1}{2} \ \langle \sqrt{n}(C_n - C)f, f\rangle \exp\left(-\frac{1}{2}\sigma_f^2\right) + o_{\mathbb{P}}(1). \qquad (23) \end{aligned}$$

Now, the term figuring in (22) equals $-n^{-1/2}\sum_{j=1}^n \exp\left(-\frac{1}{2}\sigma_f^2\right)\langle f, X_j\rangle + o_{\mathbb{P}}(1)$. As for the term figuring in (23), we have

$$\langle \sqrt{n}(C_n - C)f, f\rangle = \frac{1}{\sqrt{n}} \sum_{j=1}^n \left(\langle X_j, f\rangle^2 - \sigma_f^2\right) + \sqrt{n}\langle \overline{X}_n, f\rangle^2. \tag{24}$$

Upon combining we obtain (17), where

$$\Psi(f, x) = \cos\langle f, x\rangle + \sin\langle f, x\rangle - \exp\left(-\frac{1}{2}\sigma_f^2\right)\left\{1 + \langle f, x\rangle - \frac{1}{2}\left(\langle f, x\rangle^2 - \sigma_f^2\right)\right\}, \tag{25}$$

and

$$\|V_n - V_{n,0}\|_Q = o_{\mathbb{P}}(1). \tag{26}$$

From the central limit theorem in separable Hilbert spaces, see Theorem 2.7 in Bosq (2000), there is a centered Gaussian random element $\mathcal{V}$ of L^2_Q with covariance kernel $\mathbb{E}[\mathcal{V}(f)\mathcal{V}(g)] = \mathbb{E}[\Psi(f, X)\Psi(g, X)]$, such that $V_{n,0} \xrightarrow{\mathcal{D}} \mathcal{V}$ in L^2_Q. From Sluzki's lemma, we thus have $V_n \xrightarrow{\mathcal{D}} \mathcal{V}$ in L^2_Q, and the assertion follows from the continuous mapping theorem. Using the fact that the joint distribution of $\langle f, X\rangle$ and $\langle g, X\rangle$ is the joint distribution of $\sigma_f N_1$ and $\sigma_g(\rho N_1 + \sqrt{1-\rho^2}N_2)$, where N_1, N_2 are independent standard normal random variables and $\rho = \sigma_{f,g}/(\sigma_f\sigma_g)$, one easily obtains

$$\mathbb{E}[\langle f, X\rangle \sin\langle g, X\rangle] = \sigma_{f,g} \exp\left(-\frac{1}{2}\sigma_g^2\right),$$

$$\mathbb{E}[\cos\langle f,X\rangle\langle g,X\rangle^2] = (\sigma_g^2 - \sigma_{f,g}^2)\exp\left(-\frac{1}{2}\sigma_f^2\right),$$
$$\mathbb{E}[\langle f,X\rangle^2\langle g,X\rangle^2] = \sigma_f^2\sigma_g^2 + 2\sigma_{f,g}^2,$$

and straightforward algebra shows that $\mathbb{E}[\Psi(f,X)\Psi(g,X)] = \mathbb{E}[\mathcal{V}(f)\mathcal{V}(g)]$. Notice that the condition $\int_{\mathbb{H}} \|f\|_{\mathbb{H}}^4 \, Q(\mathrm{d}f) < \infty$ implies the validity of (18). ∎

Remark 1. Notice that the kernel given in (16) is in accordance with the kernel figuring in display (2.3) of Henze and Wagner (1997), if one replaces σ_f^2 with $\|s\|^2$, σ_g^2 with $\|t\|^2$ and $\sigma_{f,g}$ with $s^\top t$.

Since the asymptotic null distribution of nT_n depends on the unknown covariance operator $\mathcal{C}$, it cannot be used to approximate the actual null distribution of nT_n. To this end, we consider a parametric bootstrap estimator, defined as follows: Given $X_1,\dots,X_n$, let $X_1^*,\dots,X_n^*$ be iid copies of $X^* \stackrel{\mathcal{D}}{=} \mathrm{N}(0,\mathcal{C}_n)$. Let T_n^* be the bootstrap version of T_n, which is obtained by replacing $X_1,\dots,X_n$ with $X_1^*,\dots,X_n^*$ in the expression of T_n given in (6). Let $\mathbb{P}_*$ denote the conditional distribution, given $X_1,\dots,X_n$, and let $\mathbb{P}_0$ denote the null distribution. The bootstrap estimates $\mathbb{P}_0(nT_n \le t)$ by means of $\mathbb{P}_*(nT_n^* \le t)$. The next result gives the limit law of the bootstrap distribution of nT_n.

Theorem 3. *Let $X_1,\dots,X_n,\dots$ be iid copies of a random element X of $\mathbb{H}$ with covariance operator $\mathcal{C}$. Assume that $\int_{\mathbb{H}} \|f\|_{\mathbb{H}}^4 \, Q(\mathrm{d}f) < \infty$. Then $nT_n^* \stackrel{\mathcal{D}}{\to} \|\mathcal{V}\|_Q^2$ $\mathbb{P}^X$-almost surely, where $\mathcal{V}$ is the centered Gaussian random element given in the statement of Theorem 2.*

Proof. Let V_n^* be the bootstrap version of V_n in (8), which is obtained by replacing $X_1,\dots,X_n$, $\overline{X}_n$ and $\mathcal{C}_n$ with $X_1^*,\dots,X_n^*$, $\overline{X}_n^*$ and $\mathcal{C}_n^*$, respectively, where $\overline{X}_n^*$ is the sample mean and $\mathcal{C}_n^*$ denotes the sample covariance operator associated with the bootstrap sample $X_1^*,\dots,X_n^*$. Then $nT_n^* = \|V_n^*\|_Q^2$. Proceeding as in the proof of Theorem 2, one obtains $\|V_n^* - V_{n,0}^*\|_Q = o_{\mathbb{P}_*}(1)$ $\mathbb{P}^X$-almost surely, where

$$V_{n,0}^*(f) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \hat{\Psi}_n(f,X_j^*),$$

$$\hat{\Psi}_n(f,x) = \cos\langle f,x\rangle + \sin\langle f,x\rangle - \exp\left(-\frac{1}{2}\hat{\sigma}_{n,f}^2\right)\left\{1 + \langle f,x\rangle - \frac{1}{2}\left(\langle f,x\rangle^2 - \hat{\sigma}_{n,f}^2\right)\right\},$$

and $\hat{\sigma}_{n,f}^2 = \langle \mathcal{C}_n f, f\rangle$. It thus only remains to show that $V_{n,0}^* \stackrel{\mathcal{D}}{\to} \mathcal{V}$ in L_Q^2 $\mathbb{P}^X$-a.s. With this aim, we apply Theorem 1.1 in Kundu, Majumdar, and Mukherjee (2000). To verify the conditions (i)–(iii) of that theorem, let $\mathcal{C}_{n,V}^*$ and $c_{n,V}^*$ be the covariance operator and the covariance kernel of $V_{n,0}^*$, respectively. Likewise, let $\mathcal{C}_{\mathcal{V}}$ and $c_{\mathcal{V}}$ denote the covariance operator and the covariance kernel of $\mathcal{V}$, respectively. Notice that $c_{n,V}^*$ has the same expression as $c_{\mathcal{V}}$ in (16), with σ_f^2, σ_g^2 and $\sigma_{f,g}$ replaced by $\hat{\sigma}_{n,f}^2$, $\hat{\sigma}_{n,g}^2$ and $\hat{\sigma}_{n,f,g} = \langle \mathcal{C}_n f, g\rangle$, respectively. Notice also that $c_{n,V}^*(f,g) \to^{\text{a.s.}} c_{\mathcal{V}}(f,g)$, for each $f,g \in \mathbb{H}$, and that $c_V^*(f,g)$ is a bounded function, that is, for some finite constant M we have $|c_V^*(f,g)| \le M$ for each f,g. Let $\{e_k\}_{k\ge1}$ be an orthonormal basis of L_Q^2. By dominated convergence, it follows that

$$\lim_{n\to\infty}\langle \mathcal{C}_{n,V}^* e_k, e_\ell\rangle_Q = \lim_{n\to\infty}\int_{\mathbb{H}} c_{n,V}^*(f,g)e_k(f)e_\ell(g)\, Q(\mathrm{d}f)Q(\mathrm{d}g)$$
$$= \int_{\mathbb{H}} c_{\mathcal{V}}(f,g)e_k(f)e_\ell(g)\, Q(\mathrm{d}f)Q(\mathrm{d}g) = \langle \mathcal{C}_{\mathcal{V}} e_k, e_\ell\rangle_Q \quad \mathbb{P}^X - \text{a.s.}$$

Setting $a_{k,\ell} = \langle C_{\mathcal{V}} e_k, e_\ell \rangle_Q$ in the notation of Theorem 1.1 of Kundu et al. (2000), this proves that condition (i) holds. Let $\mathbb{E}_*$ denote the conditional expectation, given $X_1, \ldots, X_n$. To verify condition (ii) of Theorem 1.1 of Kundu et al. (2000), we use Beppo Levi's theorem, Parseval's relation and dominated convergence and obtain

$$\lim_{n\to\infty} \sum_{k\geq 1} \langle C^*_{n,V} e_k, e_k \rangle_Q = \lim_{n\to\infty} \sum_{k\geq 1} \mathbb{E}_* \langle V^*_{n,0}, e_k \rangle_Q^2 = \lim_{n\to\infty} \mathbb{E}_* \| V^*_{n,0} \|_Q^2$$
$$= \int_{\mathbb{H}} \lim_{n\to\infty} c^*_{n,V}(f,f)\, Q(\mathrm{d}f) = \int_{\mathbb{H}} c_{\mathcal{V}}(f,f)\, Q(\mathrm{d}f) = \mathbb{E}\|\mathcal{V}\|_Q^2 < \infty.$$

$\mathbb{P}^X$-almost surely. Finally, we must prove that $L_n(\varepsilon,\Theta) \to 0$ for each $\varepsilon > 0$ and each $\Theta \in L_Q^2$, where

$$L_n(\varepsilon,\Theta) = \sum_{j=1}^n \mathbb{E}_* \left[\langle \frac{1}{\sqrt{n}} \hat{\Psi}_n(\cdot, X_j^*), \Theta \rangle_Q^2 \mathbf{1}\left\{ \left| \langle \frac{1}{\sqrt{n}} \hat{\Psi}_n(\cdot, X_j^*), \Theta \rangle_Q \right| > \varepsilon \right\} \right]$$
$$= \mathbb{E}_* \left[\langle \hat{\Psi}_n(\cdot, X_1^*), \Theta \rangle_Q^2 \mathbf{1}\left\{ \left| \langle \hat{\Psi}_n(\cdot, X_1^*), \Theta \rangle_Q \right| > \varepsilon\sqrt{n} \right\} \right],$$

and $\mathbf{1}\{\cdot\}$ stands for the indicator function. In the sequel, let $\Theta \neq 0$ without loss of generality. Using the inequality $t\exp(-t/2) \leq 2/e$, $t \geq 0$, it follows that $|\hat{\Psi}_n(f,x)| \leq 4 + \|f\|_{\mathbb{H}}\|x\|_{\mathbb{H}} + \frac{1}{2}\|f\|_{\mathbb{H}}^2\|x\|_{\mathbb{H}}^2$ and thus $\hat{\Psi}_n^2(f,x) \leq \sum_{j=0}^4 a_j \|f\|_{\mathbb{H}}^j \|x\|_{\mathbb{H}}^j$ for each $f, x \in \mathbb{H}$, where $a_0, \ldots, a_4$ are positive constants. By the Cauchy–Schwarz inequality, we have $\langle \hat{\Psi}_n(\cdot, X_1^*), \Theta \rangle_Q^2 \leq \|\hat{\Psi}_n(\cdot, X_1^*)\|_Q^2 \|\Theta\|_Q^2$. Moreover, $\left|\langle \hat{\Psi}_n(\cdot, X_1^*), \Theta \rangle_Q\right| > \varepsilon\sqrt{n}$ implies $\|\hat{\Psi}_n(\cdot, X_1^*)\|_Q^2 > \varepsilon^2 n/\|\Theta\|_Q^2$ and thus $\sum_{j=0}^4 a_j \|X_1^*\|_{\mathbb{H}}^j \int_{\mathbb{H}} \|f\|_{\mathbb{H}}^j Q(\mathrm{d}f) > \varepsilon^2 n/\|\Theta\|_Q^2$. As a consequence, we have

$$\mathbf{1}\left\{ \left| \langle \hat{\Psi}_n(\cdot, X_1^*), \Theta \rangle_Q \right| > \varepsilon\sqrt{n} \right\} \leq \sum_{k=0}^4 \mathbf{1}\left\{ \|X_1^*\|_{\mathbb{H}}^k > \frac{n\varepsilon^2}{5a_k b_k \|\Theta\|_Q^2} \right\},$$

where $b_k = \int_{\mathbb{H}} \|f\|_{\mathbb{H}}^k Q(\mathrm{d}f)$, $k \in \{0, \ldots, 4\}$, and thus $L_n(\varepsilon,\Theta) \to 0$ will follow if

$$\lim_{n\to\infty} \mathbb{E}_* \left[\|X_1^*\|_{\mathbb{H}}^j \mathbf{1}\left\{ \|X_1^*\|_{\mathbb{H}}^k > cn \right\} \right] = 0, \quad j,k \in \{0, \ldots, 4\}, \tag{27}$$

where c is a positive constant. Now, (27) holds trivially if $k = 0$, and if $k > 0$ we have to show that $\mathbb{E}_* \left[\|X_1^*\|_{\mathbb{H}}^j \mathbf{1}\left\{ \|X_1^*\|_{\mathbb{H}} > (cn)^{1/k} \right\} \right]$ tends to zero $\mathbb{P}^X$-almost surely for each $j \in \{0, \ldots, 4\}$. The latter convergence follows since $X_1^* \stackrel{\mathcal{D}}{=} \mathrm{N}(0, C_n)$ and $C_n \to C$ $\mathbb{P}^X$-almost surely. Hence, condition (iii) of Theorem 1.1 of Kundu et al. (2000) holds and thus $V_{n,0} \xrightarrow{\mathcal{D}} \mathcal{V}$ in L_Q^2 $\mathbb{P}^X$-almost surely. Now, reasoning as in the proof of Theorem 2, the result follows. ∎

Theorem 3 (which holds regardless of whether H_0 is true or not) states that the conditional distribution of nT_n^* given $X_1, \ldots, X_n$ and the distribution of nT_n when the sample is drawn from a Gaussian population with covariance operator C, are close to each other for large n. In particular, under the null hypothesis H_0, the conditional distribution of nT_n^* given the data is close to the null distribution of nT_n. More precisely, letting $nT_{n,\mathrm{obs}} = nT_n(X_1, \ldots, X_n)$ denote the observed value of the test statistic and, for given $\alpha \in (0,1)$, writing $t^*_{n,\alpha}$ for the upper α-percentile of the bootstrap distribution of nT_n, the test function

$$\Psi_n^* = \begin{cases} 1, & \text{if} \quad nT_{n,\text{obs}} \geq t_{n,\alpha}^*, \\ 0, & \text{otherwise,} \end{cases}$$

or, equivalently, the test that rejects H_0 when $\mathbb{P}_*\{nT_n^* \geq nT_{n,\text{obs}}\} \leq \alpha$, is asymptotically correct in the sense that when H_0 is true, we have $\lim_{n\to\infty}\mathbb{P}(\Psi_n^* = 1) = \alpha$.

An immediate consequence of Theorems 1 and 3 is that, if Q satisfies (14), then the test Ψ_n^* is consistent, that is, it is able to detect any fixed alternative in the sense that $\lim_{n\to\infty}\mathbb{P}(\Psi_n^* = 1) = 1$ whenever X is not Gaussian.

In practice, the bootstrap distribution of nT_n cannot be calculated exactly. It can be approximated, however, as follows:

1. Generate a bootstrap sample $X_1^*, \dots, X_n^*$, where $X_1^*, \dots, X_n^*$ are iid from $\mathrm{N}(0, C_n)$.
2. Calculate the sample mean $\overline{X}_n^*$ and the sample covariance operator C_n^* of $X_1^*, \dots, X_n^*$, and compute $nT_n^* = nT_n(X_1^*, \dots, X_n^*)$ as given in (6), with $\overline{X}_n$ replaced by $\overline{X}_n^*$ and C_n replaced by C_n^*, respectively.
3. Repeat steps 1–2 B times (say), thus obtaining $nT_n^{*1}, \dots, nT_n^{*B}$. Approximate the upper α-percentile of the null distribution of nT_n by the upper α-percentile of the empirical distribution of $nT_n^{*1}, \dots, nT_n^{*B}$.

6 | THE LIMIT DISTRIBUTION OF T_n UNDER ALTERNATIVES

In this section, we derive the limit distribution of T_n both under fixed and contiguous alternatives to Gaussianity. Notice that, by Theorem 1, we have $T_n \to^{\text{a.s.}} \tau_Q$, where τ_Q is given in (10). We first show that, under slightly more restrictive conditions on the underlying distribution, $\sqrt{n}(T_n - \tau_Q)$ has a centred limit normal distribution. To this end, we first present an alternative representation of τ_Q. Recall the standing assumption that Q is symmetric. We first notice that τ_Q does not depend on the expectation $\mu = \mathbb{E}(X)$ of the underlying distribution, since $\varphi_X(f) = \varphi_{X-\mu}(f)\exp(\mathrm{i}\langle f, \mu\rangle)$ and thus

$$|\varphi_X(t) - \varphi(f, \mu, C)|^2 = \left|\varphi_{X-\mu}(f) - \exp\left(-\frac{1}{2}\sigma_f^2\right)\right|^2,$$

where $X - \mu$ is centred. Since the covariance operator is invariant with respect to translations, the result follows.

Proposition 1. *We have*

$$\tau_Q = \|z\|_Q^2 = \int_{\mathbb{H}} z^2(f)\, Q(\mathrm{d}f),$$

where

$$z(f) = \mathbb{E}[\cos\langle f, X\rangle] + \mathbb{E}[\sin\langle f, X\rangle] - \exp\left(-\frac{1}{2}\sigma_f^2\right), \quad f \in \mathbb{H}. \tag{28}$$

Proof. In view of the discussion above, we assume w.l.o.g. $\mu = 0$. Using Fubini's theorem and the symmetry of Q, (10) entails $\tau_Q = \tau_Q^{(1)} - \tau_Q^{(2)} + \exp(-\sigma_f^2)$, where

$$\begin{aligned}
\tau_Q^{(1)} &= \int_{\mathbb{H}} \left(\int_\Omega e^{\mathrm{i}\langle f, X(\omega)\rangle} \mathbb{P}(\mathrm{d}\omega) \cdot \int_\Omega e^{-\mathrm{i}\langle f, X(\omega')\rangle} \mathbb{P}(\mathrm{d}\omega') \right) Q(\mathrm{d}f) \\
&= \int_\Omega \int_\Omega \left(\int_{\mathbb{H}} e^{\mathrm{i}\langle f, X(\omega) - X(\omega')\rangle} Q(\mathrm{d}f) \right) \mathbb{P}(\mathrm{d}\omega)\mathbb{P}(\mathrm{d}\omega') \\
&= \int_\Omega \int_\Omega \left(\int_{\mathbb{H}} \cos\langle f, X(\omega) - X(\omega')\rangle Q(\mathrm{d}f) \right) \mathbb{P}(\mathrm{d}\omega)\mathbb{P}(\mathrm{d}\omega'),
\end{aligned} \tag{29}$$

and

$$\begin{aligned}
\tau_Q^{(2)} &= \int_{\mathbb{H}} \left(\mathbb{E}\left[e^{-\mathrm{i}\langle f, X\rangle}\right] e^{-\sigma_f^2/2} + \mathbb{E}\left[e^{\mathrm{i}\langle f, X\rangle}\right] e^{-\sigma_f^2/2} \right) Q(\mathrm{d}f) \\
&= 2\int_{\mathbb{H}} \mathbb{E}[\cos\langle f, X\rangle] e^{-\sigma_f^2/2}\, Q(\mathrm{d}f).
\end{aligned}$$

Using $\cos(\alpha - \beta) = \cos\alpha\ \cos\beta + \sin\alpha\ \sin\beta$ and again Fubini's theorem, the expression given in (29) equals

$$\int_{\mathbb{H}} \left((\mathbb{E}[\cos\langle f, X\rangle])^2 + (\mathbb{E}[\sin\langle f, X\rangle])^2 \right) Q(\mathrm{d}f).$$

Since the symmetry of Q implies $\int_{\mathbb{H}} \mathbb{E}[\sin\langle f, X\rangle] e^{-\sigma_f^2/2} Q(\mathrm{d}f) = 0$, the assertion follows. ∎

We now present the rationale why the limit distribution of $\sqrt{n}(T_n - \tau_Q)$ under fixed alternatives to Gaussianity is a normal distribution. The main idea is borrowed from Baringhaus, Ebner, and Henze (2017), who consider weighted L^2-statistics in a more specialized setting. Putting $V_n'(\cdot) = V_n(\cdot)/\sqrt{n}$, where $V_n(\cdot)$ is given in (8), display (7) and Proposition 1 yield

$$\begin{aligned}
\sqrt{n}\left(T_n - \tau_Q\right) &= \sqrt{n}\left(\|V_n'\|_Q^2 - \|z\|_Q^2 \right) = \sqrt{n}\langle V_n' - z, V_n' + z\rangle_Q \\
&= \sqrt{n}\langle V_n' - z, 2z + V_n' - z\rangle_Q \\
&= 2\langle \sqrt{n}(V_n' - z), z\rangle_Q + \frac{1}{\sqrt{n}} \| \sqrt{n}(V_n' - z)\|_Q^2.
\end{aligned}$$

Hence, if we can prove the convergence in distribution of $\sqrt{n}(V_n' - z)$ in L_Q^2 to a centred Gaussian element $\mathcal{V}'$ of L_Q^2, then the continuous mapping theorem and Slutsky's lemma yield $\sqrt{n}(T_n - \tau_Q) \xrightarrow{\mathcal{D}} 2\langle \mathcal{V}', z\rangle_Q$, where the distribution of $2\langle \mathcal{V}', z\rangle_Q$ is centered normal with variance $4\mathbb{E}[\langle \mathcal{V}', z\rangle_Q^2]$.

Theorem 4. *Assume that $\int_{\mathbb{H}} \|f\|_{\mathbb{H}}^4\, Q(\mathrm{d}f) < \infty$, and let $X_1, \dots, X_n, \dots$ be iid copies of a random element X satisfying $\mathbb{E}\|X\|_{\mathbb{H}}^4 < \infty$. Let $V_n'(\cdot) = V_n(\cdot)/\sqrt{n}$, where $V_n(\cdot)$ is given in (8). With $z(\cdot)$ defined in (28), there is a centred Gaussian random element $\mathcal{V}'$ of L_Q^2 with covariance kernel*

$$K'(f, g) = \mathbb{E}[\xi(f, X)\xi(g, X)], \quad f, g \in \mathbb{H},$$

where

$$\xi(f,x) = \cos\langle f,x\rangle - \mathbb{E}\cos\langle f,X\rangle + \sin\langle f,x\rangle - \mathbb{E}\sin\langle f,X\rangle$$
$$+ \langle f,x\rangle\ \mathbb{E}[\sin\langle f,X\rangle - \cos\langle f,X\rangle] + \frac{1}{2}e^{-\sigma_f^2/2}\left(\langle f,x\rangle^2 - \sigma_f^2\right), \quad x,f \in \mathbb{H},$$

such that $\sqrt{n}\left(V_n' - z\right) \xrightarrow{\mathcal{D}} \mathcal{V}'$.

Proof. Notice that $\sqrt{n}\left(V_n'(f) - z(f)\right) = n^{-1/2}\sum_{j=1}^n \left(R_{n,j}(f) + S_{n,j}(f) - T_{n,j}(f)\right)$, where

$$R_{n,j}(f) = \cos\langle f, X_j - \overline{X}_n\rangle - \mathbb{E}[\cos\langle f,X\rangle],$$
$$S_{n,j}(f) = \sin\langle f, X_j - \overline{X}_n\rangle - \mathbb{E}[\sin\langle f,X\rangle],$$
$$T_{n,j}(f) = \exp\left(-\frac{1}{2}\langle C_n f, f\rangle\right) - \exp\left(-\frac{1}{2}\langle Cf,f\rangle\right).$$

Using (19), (20), (21), and (24), straightforward calculations yield

$$\sqrt{n}\left(V_n'(f) - z(f)\right) = \frac{1}{\sqrt{n}}\sum_{j=1}^n \xi(f,X_j) + o_{\mathbb{P}}(1).$$

Since $\mathbb{E}[\xi(f,X)] = 0$, $f \in \mathbb{H}$, and since $\mathbb{E}\|\xi(\cdot,X)\|_Q^2 < \infty$ due to the conditions $\mathbb{E}\|X\|_{\mathbb{H}}^4 < \infty$ and $\int_{\mathbb{H}}\|f\|_{\mathbb{H}}^4 Q(\mathrm{d}f) < \infty$, the central limit theorem in Hilbert spaces and Slutsky's lemma yield the assertion. ∎

Corollary 1. *Under the conditions of Theorem 4 we have*

$$\sqrt{n}\left(T_n - \tau_Q\right) \xrightarrow{\mathcal{D}} \mathrm{N}(0,\sigma^2),$$

where

$$\sigma^2 = 4\int_{\mathbb{H}}\int_{\mathbb{H}} K'(f,g)z(f)z(g)\ Q(\mathrm{d}f)Q(\mathrm{d}g).$$

Proof. In view of the reasoning preceding Theorem 4, the proof follows from Fubini's theorem, since

$$\begin{aligned}\sigma^2 &= 4\ \mathbb{E}[\langle \mathcal{V}', z\rangle_Q^2] \\ &= 4\ \mathbb{E}\left[\left(\int_{\mathbb{H}}\mathcal{V}'(f)z(f)\ Q(\mathrm{d}f)\right)\left(\int_{\mathbb{H}}\mathcal{V}'(g)z(g)\ Q(\mathrm{d}g)\right)\right] \\ &= 4\int_{\mathbb{H}}\int_{\mathbb{H}}\mathbb{E}[\mathcal{V}'(f)\mathcal{V}'(g)]\ z(f)z(g)\ Q(\mathrm{d}f)Q(\mathrm{d}g).\end{aligned}$$

∎

We now show that the test for Gaussianity based on T_n is able to detect contiguous alternatives that approach H_0 at the rate $n^{-1/2}$. To this end, let $P = \mathrm{N}(0,C)$ a distribution from H_0. Suppose that, for each n, $X_{n,1},\dots,X_{n,n}$ are iid random elements of $\mathbb{H}$ with distribution P_n, where

$$\frac{\mathrm{d}P_n}{\mathrm{d}P} = 1 + \frac{g}{\sqrt{n}},$$

and $g : \mathbb{H} \to \mathbb{R}$ is a measurable bounded function satisfying $\int_{\mathbb{H}} g \,\mathrm{d}P = 0$.

Notice that the boundedness of g implies $\mathrm{d}P_n/\mathrm{d}P \geq 0$ if n is large enough.

Theorem 5. *Under these assumptions, we have*

$$nT_n \xrightarrow{\mathcal{D}} \|\mathcal{V} + c\|_Q^2,$$

where $\mathcal{V}$ is the Gaussian random element figuring in Theorem 2, and

$$c(f) = \int_{\mathbb{H}} \Psi(f,x)g(x)\, P(\mathrm{d}x), \quad f \in \mathbb{H}, \tag{30}$$

with $\Psi(f,x)$ given in (25).

Proof. We write $Q^{(n)}$ and $P^{(n)}$ for the n-fold product measures of P_n and P, respectively, and we put $L_n := \mathrm{d}Q^{(n)}/\mathrm{d}P^{(n)}$. Since the function g is bounded, a Taylor expansion of the function $t \mapsto \log(1+t)$ yields

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

By the Lindeberg–Feller central limit theorem and the law of large numbers, it follows that $\log L_n \xrightarrow{\mathcal{D}} \mathrm{N}(-\tau^2/2, \tau^2)$ as $n \to \infty$ under $P^{(n)}$, where $\tau^2 = \int_{\mathbb{H}} g^2 \,\mathrm{d}P$. Hence, by Le Cam's first lemma, see p. 297 of Li and Babu (2019), the sequence $Q^{(n)}$ is contiguous to $P^{(n)}$. Now, straightforward calculations yield $\lim_{n\to\infty}\mathrm{Cov}(V_{n,0}(f), \log L_n) = c(f)$, where $c(f)$ is given in (30), and $V_{n,0}(f)$ is defined in (17). Moreover, for fixed $k \geq 1$ and $f_1,\dots,f_k \in \mathbb{H}$, the joint limiting distribution of $V_{n,0}(f_1),\dots,V_{n,0}(f_k)$ and $\log L_n$ under $P^{(n)}$ is the $(k+1)$-variate normal distribution with expectation vector $(0,\dots,0,-\tau^2/2)^\top$ and covariance matrix

$$\begin{pmatrix} \Sigma & \mathbf{c} \\ \mathbf{c}^\top & \tau^2 \end{pmatrix}.$$

Here, $\mathbf{c} = (c(f_1),\dots,c(f_k))^\top$ and Σ has entries $\mathbb{E}[\mathcal{V}(f_i)\mathcal{V}(f_j)]$, $1 \leq i,j \leq k$, given in (16). From LeCam's third lemma, see p. 300 of Li and Babu (2019), we thus obtain that, under $Q^{(n)}$, the finite-dimensional distributions of $V_{n,0}$ converge to the finite-dimensional distributions of the shifted Gaussian random element $\mathcal{V} + c$. Since tightness of $V_{n,0}$ under $P^{(n)}$ and the contiguity of $Q^{(n)}$ to $P^{(n)}$ entail tightness of $V_{n,0}$ under $Q^{(n)}$, we have $V_{n,0} \xrightarrow{\mathcal{D}} \mathcal{V} + c$ under $Q^{(n)}$. In view of (26) (with $\mathbb{P}$ now being $P^{(n)}$) and the fact that also $\|V_n - V_{n,0}\|_Q = o_{Q^{(n)}}(1)$ (because of contiguity), the assertion follows from the continuous mapping theorem. ∎

TABLE 1 Description of alternatives

Alternative	Half-normal	Standard normal
Alt1	A_0, $C_1,\dots,C_5$, $S_1,\dots,S_5$	
Alt2	A_0, C_1,C_2,C_3, S_1,S_2,S_3	C_4,C_5, S_4,S_5
Alt3	A_0, C_1, S_1	$C_2,\dots,C_5$, $S_2,\dots,S_5$

Remark 2. All our results have been stated under the tacit assumption that realizations of $X_1,\dots,X_n$, that is, complete trajectories of functions, are observable. In practice, these functions are observed at a finite grid of points, and the curves $X_1,\dots,X_n$ are recovered by using nonparametric techniques, such as local linear regression. The statistics are then calculated from $\hat{X}_1,\dots,\hat{X}_n$, which stand for the resulting curve estimators. Under suitable assumptions, all previous results remain valid when the test statistic is calculated from $\hat{X}_1,\dots,\hat{X}_n$, see, for example, Jiang, Hušková, Meintanis, and Zhu (2019), in particular the comments made after the proof of their theorem 2.

7 | NUMERICAL RESULTS

In this section, we present the results of a simulation study that has been conducted in order to study the finite-sample performance of the test for Gaussianity based on T_n, and to compare the power of this novel test with respect to competing procedures. All computations have been carried out using programs written in the `R` language, R Core Team (2017), with the help of the package `fda.usc`, Febrero-Bande and Oviedo-de-la-Fuente (2012).

We first studied the performance of the bootstrap approximation to the null distribution of T_n. With this aim, the following experiment was repeated 1,000 times: Independently of each other, we generated $n = 50$ realizations of a standard Wiener process on [0,1] (denoted by W in Table 2), and we calculated $T_{M,n}$ in (9), where $M = 1{,}000$ and Q is the Wiener measure on $\mathbb{H}$. The associated p-value was then obtained by generating 200 bootstrap samples. This setting has been repeated for an Ornstein–Uhlenbeck process (denoted by OU in Table 2), and simulations have also been run for both scenarios with the sample size $n = 100$. The results for this choice of Q are labeled $T_n(W)$ in the tables. Suggested by an anonymous reviewer, we also considered the measure Q associated with $Y = \sum_{j=1}^{13} N_j SP_j$. Here, $N_1,\dots,N_{13}$ are iid univariate standard normal variables, and $SP_1,\dots,SP_{13}$ are the 13 cubic B-splines on [0,1], with interior points 0.1,0.2,…,0.9. See figure 3.5 in Ramsay and Silverman (2005) for a graphical representation of these B-splines. The results for this choice of Q are labeled $T_n(BS)$ in the tables.

To study the power, we generated samples from

$$Z(t) = A_0 + \sqrt{2}\sum_{j=1}^{5} C_j \cos(2\pi jt) + \sqrt{2}\sum_{j=1}^{5} S_j \sin(2\pi jt), \tag{31}$$

where A_0, $C_1,\dots,C_5$ and $S_1,\dots,S_5$ are independent random variables, the distributions of which are shown in Table 1.

We also considered the alternatives Alt1′, Alt2′, and Alt3′. These are the same as the alternatives given in Table 1, with the exception that the half normal distribution is throughout replaced with an equal mixture of a half normal distribution and a standard normal

TABLE 2 Empirical levels and powers for nominal levels $\alpha = 0.05, 0.10$

	$T_n(W)$		$T_n(BS)$		*JB*		*RP*							
							3		**5**		**10**		**40**	
α	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10	0.05	0.10
$n = 50$, empirical level														
W	0.054	0.106	0.046	0.090	0.041	0.064	0.067	0.114	0.059	0.117	0.067	0.127	0.074	0.130
OU	0.044	0.109	0.048	0.090	0.037	0.055	0.063	0.107	0.069	0.107	0.069	0.124	0.070	0.115
$n = 50$, empirical power														
Alt1	0.605	0.742	0.517	0.674	0.384	0.452	0.372	0.493	0.388	0.532	0.432	0.562	0.469	0.600
Alt2	0.502	0.667	0.435	0.626	0.154	0.203	0.355	0.459	0.365	0.492	0.374	0.513	0.416	0.557
Alt3	0.518	0.692	0.223	0.334	0.067	0.095	0.315	0.432	0.307	0.419	0.317	0.452	0.365	0.479
Alt1′	0.161	0.248	0.517	0.700	0.426	0.498	0.174	0.266	0.172	0.291	0.249	0.363	0.269	0.382
Alt2′	0.163	0.250	0.462	0.617	0.169	0.217	0.145	0.240	0.155	0.240	0.149	0.245	0.200	0.303
Alt3′	0.170	0.272	0.205	0.327	0.083	0.115	0.152	0.234	0.135	0.223	0.134	0.224	0.162	0.245
Alt1″	0.426	0.538	0.565	0.686	0.612	0.662	0.266	0.376	0.295	0.405	0.303	0.421	0.307	0.420
Alt2″	0.471	0.565	0.531	0.665	0.568	0.634	0.283	0.395	0.326	0.429	0.336	0.453	0.329	0.455
Alt3″	0.468	0.570	0.517	0.649	0.472	0.516	0.319	0.416	0.348	0.475	0.350	0.479	0.349	0.490
$n = 100$, empirical level														
W	0.048	0.103	0.049	0.096	0.053	0.072	0.059	0.105	0.054	0.104	0.054	0.102	0.049	0.094
OU	0.050	0.095	0.048	0.096	0.051	0.077	0.058	0.097	0.063	0.114	0.057	0.107	0.047	0.095
$n = 100$, empirical power														
Alt1	0.932	0.967	0.934	0.981	0.766	0.835	0.675	0.781	0.725	0.829	0.759	0.872	0.811	0.888
Alt2	0.910	0.956	0.876	0.954	0.468	0.538	0.638	0.734	0.658	0.776	0.695	0.828	0.747	0.851
Alt3	0.888	0.938	0.459	0.615	0.102	0.128	0.568	0.678	0.569	0.701	0.603	0.719	0.714	0.812
Alt1′	0.210	0.325	0.889	0.948	0.765	0.824	0.256	0.384	0.274	0.431	0.434	0.576	0.493	0.648
Alt2′	0.244	0.369	0.848	0.917	0.306	0.391	0.249	0.351	0.233	0.363	0.266	0.382	0.391	0.522
Alt3′	0.284	0.413	0.423	0.573	0.150	0.200	0.259	0.380	0.256	0.362	0.239	0.359	0.292	0.414
Alt1″	0.705	0.787	0.792	0.865	0.885	0.909	0.461	0.568	0.485	0.619	0.504	0.638	0.507	0.640
Alt2″	0.702	0.797	0.759	0.852	0.883	0.915	0.457	0.600	0.498	0.656	0.522	0.655	0.521	0.637
Alt3″	0.690	0.798	0.741	0.830	0.829	0.850	0.455	0.575	0.501	0.627	0.533	0.664	0.543	0.666

distribution. Likewise, the alternatives denoted by Alt1″, Alt2″, and Alt3″ originate from throughout replacing the half normal distribution with a Laplace distribution (two-sided exponential distribution).

As competitors to the novel test for Gaussianity based on T_n, we considered the Jarque–Bera type test in Górecki et al. (2018) for iid data (denoted by *JB* in the tables), and the random projection test of Cuesta-Albertos et al. (2007) (denoted by *RP*) with 3, 5, 10, and 40 projections. Table 2 reports both the observed empirical level and the empirical power for the nominal levels of significance $\alpha = .05$ and $\alpha = .10$.

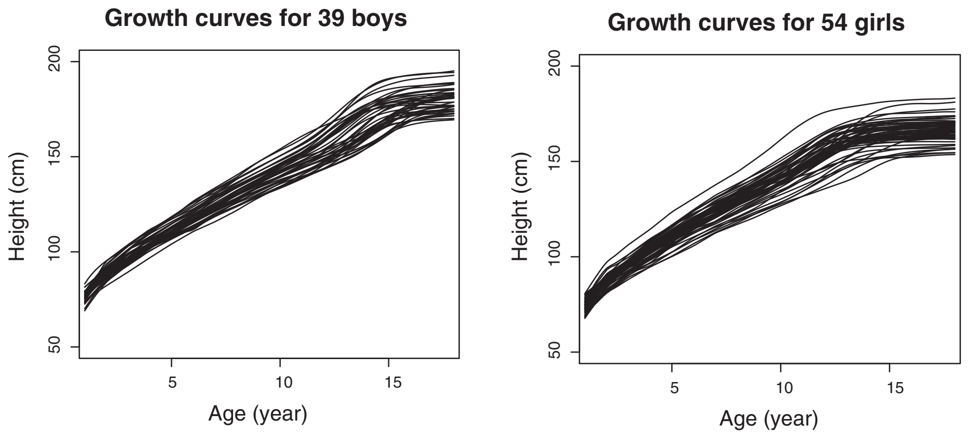

FIGURE 1 The Berkeley growth data

TABLE 3 *p*-values for testing Gaussianity for the Berkeley Growth Data

				RP			
	$T_n(W)$	$T_n(BS)$	*JB*	**3**	**5**	**10**	**40**
Boys	0.323	0.078	0.301	0.126	0.167	0.235	0.153
Girls	0.144	0.064	0.433	0.107	0.138	0.153	0.248

From Table 2, we see that the empirical levels of the tests $T_n(W)$ and $T_n(BS)$ are quite close to the nominal values, even for the moderate sample size $n = 50$. As for the power, it is not surprising that there is no test having highest power against all alternatives considered. For the alternatives Alt1, Alt2, and Alt3, the test $T_n(W)$ outperforms its competitors; for the alternatives Alt1′, Alt2′, and Alt3′, the test $T_n(BS)$ exhibits the highest power; while for the alternatives Alt1″ and Alt2″, the test *JB* presents the larger power for $n = 100$. Notice that all alternatives considered belong to the same basic model (31), in which the distribution of some coefficients is switched. The power of the T_n-tests and of the random projection test change softly as the coefficients are switched. In most cases, however, the power of the *JB* test drops as the alternative becomes closer to H_0, that is, as the number of coefficients with normal distribution increases.

We close this section with a real data set application. As explained in Section 1, some inferential procedures, designed for functional data, assume Gaussianity. An example is the test in Zhang et al. (2010) for the equality of the mean of two functional populations. Zhang et al. (2010) applied their test to the Berkeley Growth Data set. This dataset contains the heights of 39 boys and 54 girls recorded at 31 not equally spaced ages from Year 1 to Year 18. The dataset is available from the `R` package `fda`. The method in that paper is designed for Gaussian random functions, but the assumption of Gaussianity had not been checked for either sample. We applied all tests in Table 2 to each of the two datasets. First of all, proceeding as in Zhang et al. (2010), the growth curves have been reconstructed by using local polynomial smoothing. Each of the individual curves has been smoothed separately, using the same bandwidth $h = 0.3674$. Figure 1 shows the smoothed growth curves. Table 3 reports the *p*-values obtained. All tests but $T_n(BS)$ agree in not rejecting the assumption of Gaussianity for both populations, the boys and the girls.

8 | SOME COMMENTS ON THE CHOICE OF Q

This section gathers some recommendations for the choice of Q spread along the paper, and it discusses some special choices of Q not considered so far.

Since the calculation of the integral in the definition of T_n involves generation of random samples from Y having distribution Q and the p-values are calculated using a bootstrap approximation, which requires the recalculation of the test statistic for a large number of bootstrap samples, it is indispensable that Q can be sampled easily. A further criterion for the choice of Q is consistency of the resulting test. In this regard, we gave recommendations at the end of Section 4.

If the alternative were known, one could in principle choose Q to maximize the power, but this situation is unrealistic. Therefore, in practice, an optimal choice of Q (in the sense of yielding the highest power) in unfeasible.

All results have been derived under the assumption that the measure Q is fixed. We now briefly study the case that Q is allowed to vary with n, which is mainly motivated by the recommendation for Q as the measure associated with the random element Y figuring in (15). In practice, random samples from Y are obtained by truncating the infinite sum up to M_n terms, for some finite M_n. Thus we are effectively sampling from a measure Q_n that approximates Q.

Next we show that, under some weak assumptions, the hitherto obtained results still hold if the measure Q is replaced with a sequence Q_n, $n \geq 1$, of nonrandom measures that converges weakly to some measure Q_0, for short: $Q_n \to^w Q_0$. In order to emphasize the dependence of the test statistic on the sample size and on Q_n, we use the notation

$$T_n(Q_n) = T_n(X_1, \dots, X_n; Q_n) = \int_{\mathbb{H}} \left|\varphi_n(f) - \varphi(f; \overline{X}_n, C_n)\right|^2 Q_n(\mathrm{d}f),$$

We confine ourselves to state Q_n-analogues of Theorems 1 and 2. Mutatis mutandis, one can obtain Q_n-analogues of further results obtained so far.

Theorem 6. *Under the assumptions of Theorem 1 and $Q_n \to^w Q_0$, we have*

$$T_n(Q_n) \to^{\text{a.s.}} \tau_{Q_0} = \int_{\mathbb{H}} |\varphi_X(f) - \varphi(f; \mu, C)|^2 Q_0(\mathrm{d}f).$$

Proof. Let $\Omega_0 \subset \Omega$ be as in the proof of Theorem 1 and assume without loss of generality that $n^{-1} \sum_{j=1}^n \|X_j(\omega)\|_{\mathbb{H}}^2 \to \mathbb{E}\|X\|_{\mathbb{H}}^2$ for each $\omega \in \Omega_0$. Putting

$$h_n(f, \omega) = |\varphi_n(f, \omega) - \varphi(f; \overline{X}_n(\omega), C_n(\omega))|^2, \quad h(f, \omega) = |\varphi_X(f) - \varphi(f; \mu, C)|^2,$$

$\omega \in \Omega$, $f \in \mathbb{H}$, (13) shows that $h_n(f,\omega) \to h(f,\omega)$ for each $\omega \in \Omega_0$ and each $f \in \mathbb{H}$. Now, fix $\varepsilon > 0$, $\omega \in \Omega_0$ and $f \in \mathbb{H}$. Since the integrand in the expression of $T_n(Q_n)$ is bounded from above by 4, the result would follow from display (2.12) in Bosq (2000) if we can find a $\delta = \delta(f,\varepsilon,\omega) > 0$ such that, for each $g \in \mathbb{H}$ satisfying $\|f - g\|_{\mathbb{H}} \leq \delta$, we have $|h_n(g,\omega) - h(f,\omega)| < \varepsilon$ for sufficiently large n. Now, using the inequalities $\|z|^2 - |w|^2| \leq 4|z - w|$ ($z, w \in \mathbb{C}$, $|z| \leq 1, |w| \leq 1$) and $|e^{iu} - e^{iv}| \leq |u - v|$ ($u, v \in \mathbb{R}$), the Cauchy–Schwarz-inequality and the fact that $\langle C_n(\omega)h, h\rangle = n^{-1} \sum_{j=1}^n \langle X_j(\omega), h\rangle^2 - \langle \overline{X}_n(\omega), h\rangle^2$, $h \in \mathbb{H}$, straightforward calculations yield

$$|h_n(f, \omega) - h_n(g, \omega)| \leq 4\|f - g\|_{\mathbb{H}} \left(2M_n^{(1)}(\omega) + \|f + g\|_{\mathbb{H}} \left[M_n^{(2)}(\omega) + \left(M_n^{(1)}(\omega)\right)^2\right]\right),$$

where $M_n^{(\ell)}(\omega) = n^{-1}\sum_{j=1}^n \|X_j(\omega)\|_{\mathbb{H}}^{\ell}$, $\ell = 1,2$. Since $\lim_{n\to\infty} M_n^{(\ell)}(\omega) = \mathbb{E}\|X\|^{\ell}$, $\ell = 1,2$, the right-hand side of the above inequality converges to 0 if $\|f - g\| \to 0$. The assertion now follows from (13) and the triangle inequality. ∎

The next result gives the asymptotic null distribution of $T_n(Q_n)$.

Theorem 7. *Let $X_1,\dots,X_n,\dots$ be iid copies of a Gaussian random element X of $\mathbb{H}$ with covariance operator C. Assume that $\int_{\mathbb{H}} \|f\|_{\mathbb{H}}^4\, Q_n(\mathrm{d}f) \le M$, $n \ge 1$, for some positive constant M, and that $Q_n \to^w Q_0$. Then $nT_n(Q_n) \xrightarrow{\mathcal{D}} \|\mathcal{V}\|_{Q_0}^2$, where $\mathcal{V}$ is defined in Theorem 2.*

Proof. Recall that $nT_n(Q_n) = \|V_n\|_{Q_n}^2$, where V_n is given in (8). Proceeding as in the proof of Theorem 2, it follows that $\|V_n - V_{n,0}\|_{Q_n} = o_{\mathbb{P}}(1)$, where $V_{n,0}$ is defined in (17). Now, taking into account that $\|V_{n,0}\|_{Q_n}^2 = \|V_{n,0}\|_{Q_0}^2 + R_n$, where $R_n = \int_{\mathbb{H}} V_{n,0}^2(f)(Q_n - Q_0)(\mathrm{d}f)$, it suffices to prove $R_n = o_{\mathbb{P}}(1)$ in view of the proof of Theorem 2. To this end, let $\Delta_n := Q_n - Q_0$, and put $\Psi_h^2 := \Psi(h,X)^2$, $h \in \mathbb{H}$. Considerations of symmetry involving the fourfold sum $V_{n,0}^2(f)V_{n,0}^2(g)$ and the fact that $\mathbb{E}\Psi(h,X) = 0$, $h \in \mathbb{H}$, yield

$$\mathbb{E}(R_n^2) \le 3\left(\int_{\mathbb{H}} \mathbb{E}(\Psi_f^2)\ \Delta_n(\mathrm{d}f)\right)^2 + \frac{1}{n}\int_{\mathbb{H}}\int_{\mathbb{H}} \mathbb{E}(\Psi_f^2\Psi_g^2)\ \Delta_n(\mathrm{d}f)\ \Delta_n(\mathrm{d}g). \tag{32}$$

From the assumptions made, it follows that $\int_{\mathbb{H}}\int_{\mathbb{H}} \mathbb{E}(\Psi_f^2\Psi_g^2)\ \Delta_n(\mathrm{d}f)\Delta_n(\mathrm{d}g) \le K$ for some positive constant K, and thus the second term on the right-hand side of (32) converges to 0. Since $\mathbb{E}(\Psi_f^2) = 1 - \exp(-\sigma_f^2)(1 + \sigma_f^2 + 0.5 - \sigma_f^4)$ is a continuous bounded function of f, and since $Q_n \to^w Q_0$, it follows that the first term on the right-hand side of (32) also tends to 0, and the proof is finished, because $\mathbb{E}(R_n^2) \to 0$ implies $R_n \to^{\mathbb{P}} 0$. ∎

In the above approach, we have *truncated* Q, but a common approach in functional data analysis consists in *truncating* the data as follows. If $\mathbb{E}\|X\|_{\mathbb{H}}^2 < \infty$, then X admits a Karhunen–Loève expansion

$$X = \mu + \sum_{j\ge 1} \sqrt{\lambda_j}\ Z_j\ v_j. \tag{33}$$

Here, λ_j and v_j, $j \ge 1$, are the eigenvalues and the corresponding eigenfunctions—called principal components—associated with the covariance operator of X, and $Z_1, Z_2, \dots$ are centered uncorrelated random variables, which are even iid standard normal under the null hypothesis of Gaussianity. In practice, the λ_j's and v_j's are unknown, and they are estimated from the data. Moreover, only a finite number d_n (say) of projections are used in applications, yielding the sample analogue

$$X_j \approx \overline{X}_n + \sum_{k=1}^{d_n} \sqrt{\hat{\lambda}_k}\ \hat{\xi}_{jk}\ \hat{v}_k, \quad 1 \le j \le n,$$

of (33). Here, the $\hat{\lambda}_k$'s and the $\hat{v}_k$'s are the eigenvalues and the corresponding eigenfunctions—called sample principal components—associated with the sample covariance operator, and $\hat{\xi}_{jk} = \langle X_j, \hat{v}_k\rangle / \sqrt{\hat{\lambda}_k}$, $1 \le j \le n$, $1 \le k \le d_n$. There are several proposals for the practical determination of d_n, see Chapter 3 of Horváth and Kokoszka (2012). One of the most popular methods chooses d_n so that $\sum_{k=1}^{d_n} \hat{\lambda}_k / \sum_{k=1}^{n} \hat{\lambda}_k \geq 0.85$. Under the null hypothesis of Gaussianity, the random vectors $\hat{\xi}_1 = (\hat{\xi}_{11}, \dots, \hat{\xi}_{1d_n})^\top, \dots, \hat{\xi}_n = (\hat{\xi}_{n1}, \dots, \hat{\xi}_{nd_n})^\top$ are *approximately* iid from a d_n-variate normal law with mean zero and unit covariance matrix. These quantities play a role similar to that of the scaled residuals in the finite-dimensional case. So it is tempting to compare the ECF of $\hat{\xi}_1, \dots, \hat{\xi}_n$ with the CF of the law $\mathrm{N}_d(0,\mathrm{I}_d)$. This heuristic derivation can be formally obtained as follows. Let $\hat{Q}_n$ be the probability measure associated with the process

$$W_n = \sum_{k=1}^{d_n} \beta \hat{\lambda}_k^{-1/2} \, \hat{v}_k \, N_k,$$

where $\beta > 0$ is a positive constant and $N_1, \dots, N_{d_n}$ are iid standard normal random variables. Then, routine calculations show that

$$\int_{\mathbb{H}} \left| \varphi_n(f) - \varphi(f; \overline{X}_n, C_n) \right|^2 \hat{Q}_n(\mathrm{d}f), \tag{34}$$

coincides with $T_{n,d_n,\beta}$ defined in (1), with $Y_1, \dots, Y_n$ replaced with $\hat{\xi}_1, \dots, \hat{\xi}_n$. Notice that the results of Theorems 6 and 7 do not apply to $\hat{Q}_n$ because $\hat{Q}_n$ is a random measure. The study of the test based on (34) deserves further study.

9 | CONCLUDING REMARKS AND FURTHER RESEARCH

We have introduced and studied a novel genuine test for Gaussianity in separable Hilbert spaces that is applicable for functional data. Some preliminary simulation results show that the procedure compares favorably with the hitherto few existing competitors. It would be interesting to modify and generalize the approach with respect to testing for Gaussianity in situations in which the mean and/or the covariance operator have a certain parametric structure. For example, one could test for a Wiener process on [0,1], where $\mu = 0$ and $c(t,s) = \vartheta \min\{t,s\}$, for some $\vartheta > 0$. It would also be tempting to test for Gaussianity of multivariate functional data that take values in $L^2([0,1]^d)$.

It would be interesting to have a consistent estimator of the variance of the limit normal distribution under alternatives to obtain an asymptotic confidence interval for τ_Q.

A test for Gaussianity could be based on applying the BHEP test to the principal scores. We have seen that the resulting test statistic has the same expression as that studied in this paper, but the measure involved in the definition of the test statistic is of random nature, which requires subsequent study.

ACKNOWLEDGEMENTS

The authors thank the Associate Editor and two anonymous referees for their constructive comments and suggestions which helped to improve the presentation. M.D. Jiménez-Gamero has been partially supported by grants MTM2017-89422-P (Spanish Ministry of Economy, Industry and Competitiveness, the State Agency of Investigation, the European Regional Development Fund) and P18-FR-2369 (Junta de Andalucía).

ORCID

María Dolores Jiménez-Gamero https://orcid.org/0000-0002-8823-3292

REFERENCES

Arcones, M. A. (2007). Two tests for multivariate normality based on the characteristic function. *Mathematical Methods of Statistics*, *16*, 177–201.

Baringhaus, L., Ebner, B., & Henze, N. (2017). The limit distribution of weighted L^2-statistics under alternatives, with applications. *Annals of the Institute of Statistical Mathematics*, *69*, 969–995.

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

Boente, G., Rodríguez, D., & Sued, M. (2018). Testing equality between several population covariance operators. *Annals of the Institute of Statistical Mathematics*, *70*, 919–950.

Bosq, D. (2000). *Linear processes in function spaces*. New York. NY: Springer.

Bugni, F. A., Hall, P., Horowitz, J. L., & Neumann, G. R. (2009). Goodness-of-fit tests for functional data. *The Econometrics Journal*, *12*, S1–S18.

Csörgő, S. (1989). Consistency of some tests for multivariate normality. *Metrika*, *36*, 107–116.

Cuesta-Albertos, J. A., del Barrio, E., Fraiman, R., & Matrán, C. (2007). The random projection method in goodness of fit for functional data. *Computational Statistics & Data Analysis*, *51*, 4814–4831.

Ditzhaus, M., & Gaigall, D. (2018). A consistent goodness-of-fit test for huge dimensional data and functional data. *Journal of Nonparametric Statistics*, *30*, 834–859.

Doornik, J. A., & Hansen, H. (2008). An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics*, *70*, 927–939.

Eaton, M. L., & Perlman, M. D. (1973). The non-singularity of generalized sample covariance matrices. *The Annals of Statistics*, *1*, 710–717.

Ebner, B. (2012). Asymptotic theory for the test of multivariate normality by cox and small. *Journal of Multivariate Analysis*, *111*, 368–379.

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

Febrero-Bande, M., & Oviedo-de-la-Fuente, M. (2012). Statistical computing in functional data analysis: The R package fda.usc. *Journal of Statistical Software*, *51*(4), 1–28.

Górecki, T., Hörmann, S., Horváth, L., & Kokoszka, P. (2018). Testing normality of functional time series. *Journal of Time Series Analysis*, *39*, 471–487.

Henze, N. (2002). Invariant tests for multivariate normality: A critical review. *Statistical Papers*, *43*, 467–506.

Henze, N., & Jiménez-Gamero, M. D. (2019). A class of tests for multinormality with IID and GARCH data based on the empirical moment generating function. *Test*, *28*, 499–521.

Henze, N., Jiménez-Gamero, M. D., & Meintanis, S. G. (2019). Characterizations of multinormality and corresponding tests of fit, including for GARCH models. *Econometric Theory*, *35*, 510–546.

Henze, N., & Visagie, J. (2019). Testing for normality in any dimension based on a partial differential equation involving the moment generating function. *Annals of the Institute of Statistical Mathematics*. https://doi.org/10.1007/s10463-019-00720-8.

Henze, N., & Wagner, T. (1997). A new approach to the BHEP tests for multivariate normality. *Journal of Multivariate Analysis*, *62*, 1–23.

Henze, N., & Zirkler, B. (1990). A class of affine invariant consistent tests for multivariate normality. *Communications in Statistics-Theory and Methods*, *19*, 3595–3617.

Horváth, L., & Kokoszka, P. (2012). *Inference for functional data with applications*. New York, NY: Springer.

Hušková, M., & Meintanis, S. G. (2010). Test for the error distribution in nonparametric possiby heteroscedatic regression models. *Test*, *19*, 92–112.

Jiang, Q., Hušková, M., Meintanis, S. G., & Zhu, L. (2019). Asymptotics, finite-sample comparisons and applications for two-sample tests with functional data. *Journal of Multivariate Analysis*, *170*, 202–220.

Jiménez-Gamero, M. D. (2014). On the empirical characteristic function process of the residuals in GARCH models and applications. *Test*, *23*, 409–423.

Jiménez-Gamero, M. D., Alba-Fernández, M. V., Muñoz-García, J., & Chalco-Cano, Y. (2009). Goodness-of-fit tests based on empirical characteristic functions. *Computational Statistics & Data Analysis*, *53*, 3957–3971.

Jiménez-Gamero, M. D., Muñoz-García, J., & Pino-Mejías, R. (2005). Testing goodness of fit for the distribution of errors in multivariate linear models. *Journal of Multivariate Analysis*, *95*, 301–322.

Kankainen, K., Taskinen, S., & Oja, H. (2007). Tests of multinormality based on location vectors and scatter matrices. *Statistical Methods and Applications*, *47*, 1923–1934.

Kundu, S., Majumdar, S., & Mukherjee, K. (2000). Central limit theorems revisited. *Statistics & Probability Letters*, *47*, 265–275.

Laha, R. G., & Rohatgi, V. K. (1979). *Probability theory*. New York, NY: Wiley.

Li, B., & Babu, G. J. (2019). *A graduate course on statistical inference*. New York, NY: Springer.

Pudelko, J. (2005). On a new affine invariant and consistent test for multivariate normality. *Theory of Probability and Mathematical Statistics*, *25*, 43–54.

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing Retrieved from. https://www.R-project.org/

Ramsay, J., & Silverman, B. W. (2005). *Functional data analysis*. New York, NY: Springer.

Székely, G. J., & Rizzo, M. L. (2005). A new test for multivariate normality. *Journal of Multivariate Analysis*, *93*, 58–80.

Thulin, M. (2014). Tests for multivariate normality based on canonical correlations. *Statistical Methods & Applications*, *23*, 189–208.

Villaseñor-Alva, J. A., & Estrada, E. G. (2009). A generalization of Shapiro–Wilk's test for multivariate normality. *Communications in Statistics—Theory and Methods*, *38*, 1870–1883.

Voinov, V., Pya, N., Makarov, R., & Voinov, Y. (2016). New invariant and consistent chi-squared type goodness-of-fit tests for multivariate normality and a related comparative study. *Communications in Statistics—Theory and Methods*, *45*, 3249–3263.

Zhang, J. T., Liang, X., & Xiao, S. (2010). On the two-sample Behrens-Fisher problem for functional data. *Journal of Statistical Theory and Practice*, *4*, 571–587.

How to cite this article: Henze N, Jiménez-Gamero MD. A test for Gaussianity in Hilbert spaces via the empirical characteristic functional. *Scand J Statist*. 2021;48:406–428. https://doi.org/10.1111/sjos.12470