A Random-Projection Based Test of Gaussianity for Stationary Processes[☆]

Alicia Nieto-Reyes[a,*], Juan Antonio Cuesta-Albertos[a], Fabrice Gamboa[b]

[a] *Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Spain*
[b] *Institut de Mathématiques de Toulouse - Université Paul Sabatier (Toulouse III), France*

Abstract

Gaussianity tests have being widely studied in the literature. Regarding the study of Gaussianity tests for stationary processes, these only verify the Gaussianity of a marginal at a fixed finite order, generally order one. Therefore, they do not reject stationary non-Gaussian processes with the one-dimensional Gaussian marginal. Thus, a consistent test is proposed for Gaussianity of stationary processes when a finite sample path of the process is observed. Using random projections, decision rules are applied to the whole distribution of the process and not only on its marginal distribution at a fixed order, as in previous tests. The main idea is to test the Gaussianity of the one-dimensional marginal distribution of some random linear transformations of the process. Note that testing the one-dimensional marginal distribution can be done with previous tests of Gaussianity for stationary processes. It is shown by both theoretical and empirical studies that the proposed test procedure has good properties for a wide range of alternatives.

Keywords: Normality Test, Strictly Stationary Random Process, Random Projection, Consistent Test

1. Introduction

Very often, observed data are a finite path of real temporal phenomena modeled as a second order stationary process. Adding the Gaussianity assumption, the process possesses a lot of beneficial properties as regards their statistics or prediction and, in particular, it becomes strictly stationary. This means that the law of the process is invariant if the time is shifted.

[☆] The Matlab codes, to obtain the figure and tables of the paper, appear as an annex in the electronic version.

[*] Corresponding author: Departamento de Matemáticas, Estadística y Computación, Universidad de Cantabria, Avda. Los Castros s/n, Santander E-39005, Spain, Tel.: +34 942 202289; Fax: +34 942 201402.

Email address: alicia.nieto@unican.es (Alicia Nieto-Reyes)

In this paper, the processes are assumed to be integrable and stationary stands for strictly stationary. We consider tests for Gaussianity of stationary processes. Let $\mathbf{X} := (X_t)_{t\in\mathbb{Z}}$ be a stationary process of real-valued random variables (r.v.'s). Our aim is to test:

$$H_0 : \mathbf{X} \text{ is Gaussian} \quad \text{ versus } \quad H_a : \mathbf{X} \text{ is not Gaussian.} \tag{1}$$

Conventional goodness-of-fit tests such as Kolmogorov-Smirnov or Cramér von Mises (D'Agostino and Stephens, 1986) cannot be used here, since their asymptotic distributions are not clear in the context of stationary processes and they are in general not distribution free under the null hypothesis. Other recent tests as in Liu and Maharaj (2013); Ghoudi and Rémillard (2013) do not apply here either since they are intended for particular types of dynamic data generating schemes. The problem has attracted attention over the last three decades. Researchers have proposed tests based on the analysis of the empirical characteristic function (ch.f.) (Epps, 1987), of the skewness and kurtosis (Lobato and Velasco, 2004, the so-called Jarque-Bera test), of both the empirical ch.f. and the skewness and kurtosis (Moulines and Choukri, 1996) and the bispectral density function (Rao and Gabr, 1980). An important drawback of these tests is that they only consider a finite order marginal of the process (generally order one). Obviously, this provides tests at the right level for the intended problem; but these are at the nominal power against some non-Gaussian alternatives, such as stationary non-Gaussian processes having one-dimensional Gaussian marginal.

This paper is inspired by the work on random projections of Cuesta-Albertos et al. (2007), particularly in Theorem 1.1 below. It is well known that a distribution is Gaussian if and only if all of its one-dimensional projections are Gaussian. The existence of non-Gaussian distributions with some Gaussian one-dimensional projections is also well known. However, Theorem 1.1 shows that the Gaussian projections of a non-Gaussian distribution, if any, are very scarce. To be more precise, if we take at random a one-dimensional projection of a non-Gaussian distribution, then, with probability one, this projection will be non-Gaussian.

The following theorem uses the notion of dissipative distribution (Definition 2.1). $\mathbb{H}$ denotes a separable Hilbert space.

Theorem 1.1 (Cuesta-Albertos et al., 2007)**.** *If η is a dissipative distribution on $\mathbb{H}$ and $\boldsymbol{D}$ is an $\mathbb{H}$-valued random element, then, $\boldsymbol{D}$ is Gaussian if and only if $\eta(E) > 0$, where*

$$E = \{\boldsymbol{h} \in \mathbb{H} : \textit{ the distribution of } \langle \boldsymbol{D}, \boldsymbol{h}\rangle \textit{ is Gaussian }\}.$$

The importance of this result lies in the fact that if η is dissipative then the following $0-1$ law holds

$$\eta(\{\mathbf{h} \in \mathbb{H} : \text{ the distribution of } \langle \mathbf{D}, \mathbf{h}\rangle \text{ is Gaussian}\}) \in \{0, 1\}.$$

Moreover, $\mathbf{D}$ is not Gaussian if, and only if,

$$\eta(\{\mathbf{h} \in \mathbb{H} : \text{ the distribution of } \langle \mathbf{D}, \mathbf{h}\rangle \text{ is Gaussian}\}) = 0.$$

Additionally, $\mathbf{D}$ is Gaussian if, and only if,

$$\eta(\{\mathbf{h} \in \mathbb{H} : \text{ the distribution of } \langle \mathbf{D}, \mathbf{h} \rangle \text{ is Gaussian}\}) = 1.$$

In other words, if we are interested in whether the distribution of $\mathbf{D}$ is Gaussian, then the only thing we have to do is to select at random a point $\mathbf{h} \in \mathbb{H}$ using a dissipative distribution and check if the real-valued random variable $\langle \mathbf{D}, \mathbf{h} \rangle$ is Gaussian. We will obtain the right answer with probability one.

From this result, we have that a test at the level α for the Gaussianity of a randomly chosen (one-dimensional) projection, is also a test at the same level to test the Gaussianity of the process $\mathbf{X}$. We also have that a consistent Gaussianity test applied to the projection, is, in fact, a consistent test for the Gaussianity of the full process $\mathbf{X}$. This property was used in Cuesta-Albertos et al. (2007) to construct a Gaussianity test when a random sample of trajectories is available. Let us explain these aspects in greater detail with the help of an example.

Example 1.2. Let us begin by a question: Is it worth considering that the text you have read in this paper so far has been produced by a random letter generator? I.e., is it worth to consider the assumption that the text in the previous paragraphs have been produced by a mechanism which chooses with equal probability every key in a computer keyboard, with probability a half to press the "capital" key? Obviously, the answer is no, because a text like the one in the previous paragraphs has not been written by chance, despite the fact that the probability of obtaining such a text just by chance is not zero: it is extremely small but strictly positive.

Moving back to the Gaussianity problem, let $\mathbf{X}$ be a process and let us select randomly a one-dimensional projection. According to Theorem 1.1, if the partner distribution is not Gaussian, the probability for the chosen projection to be Gaussian is zero. Let us stress this point: this probability is zero; it is not extremely small, like in the previous example, but zero. Therefore, if the partner distribution is not Gaussian, we will not obtain a Gaussian projection; and if we obtain a Gaussian projection, we will conclude that $\mathbf{X}$ is Gaussian.

Returning to the Gaussianity test of stationary processes; in this setting, we do not have a random sample of trajectories but a sequence of observations taken on a fixed trajectory. Thus, the theory developed in Cuesta-Albertos et al. (2007) can not be applied directly. The good news is that, similarly as in Example 1.2, Theorem 1.1 transforms the analysis of the Gaussianity of the process $\mathbf{X}$ in the analysis of the Gaussianity of a randomly chosen one-dimensional projection as follows.

Assume that we have the stationary process $\mathbf{X} = (X_t)_{t\in\mathbb{Z}}$. Then, randomly select a vector $\mathbf{h} := (h_t)_{t\in\mathbb{N}}$ (technical details on its construction are given in Section 3) and construct the new process $\mathbf{Y}^h = (Y_t^h)_{t\in\mathbb{Z}}$ where

$$Y_t^h := \sum_{i=0}^{\infty} h_i X_{t-i,\ t\in\mathbb{Z}}.$$

Theorem 1.1 implies that if $\mathbf{X}$ is not Gaussian, then, with probability one, the $\mathbf{h}$ we have chosen makes Y_t^h non-Gaussian. In other words, if $\mathbf{X}$ is not Gaussian, then the one-dimensional marginal of $\mathbf{Y}^h$ is not Gaussian for almost every $\mathbf{h}$.

Once $\mathbf{h}$ has been fixed, the process $\mathbf{Y}^h$ is also stationary and we can employ one of the above mentioned tests to test the Gaussianity of the marginals of $\mathbf{Y}^h$ as those tests were designed, precisely, to test the Gaussianity of a given one-dimensional marginal. With this, according to the preceding reasoning, in fact, we are testing the full Gaussianity of the $\mathbf{X}$.

The particular procedure used to test the Gaussianity of the marginal distribution of $\mathbf{Y}^h$ is left to the practitioner. Here, we use improved versions of the tests proposed in Epps (1987) and Lobato and Velasco (2004), obtaining a test consistent against every stationary alternative satisfying some regularity conditions (see Section 4.3 the explanation on how we have applied the test).

However, an important fact to be taken into account is that, under the alternative, it may happen that we we are dealing with a projection in which the non-Gaussianity is not too easy to be ascertained. To tackle this problem, we follow the proposal made in Cuesta-Albertos et al. (2007) consisting in taking more than one projection, carrying out the test on each projection and then, mixing the p-values using the False Discovery Rate, as suggested in Benjamini and Yekutieli (2001).

The paper is structured as follows. Section 2 contains the notation, some results concerning the random projection method and previous Gaussianity tests for stationary processes. The new test is proposed in Section 3, followed by details on its practical application. Section 4 reports some simulation studies and includes a section to learn to apply the proposed test without the need to read the paper in detail. An study on real data sets is carried out in Section 5. A discussion on the procedure appears in Section 6. All the proofs are deferred to the Appendix. Computations were performed using MatLab (except where otherwise stated).

2. Notation and preliminaries

We assume that all random elements (r.e.'s) are defined on the same sufficiently rich probability space $(\Omega, \sigma, I\!\!P)$. $\mathbb{H}$ denotes a separable Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. $\{v_n\}_{n=1}^{\infty}$ is a generic orthonormal basis of $\mathbb{H}$ and V_n is the n-dimensional subspace spanned by $\{v_1, \ldots, v_n\}$. For any $V \subset \mathbb{H}$, $V^\perp$ denotes its orthogonal complement. If $\mathbf{D}$ is an $\mathbb{H}$-valued r.e., $\mathbf{D}_V$ is the projection of $\mathbf{D}$ on V.

The beta distribution with parameters α_1, α_2 will be denoted $\beta(\alpha_1, \alpha_2)$; $N(\nu, \rho)$ is the one-dimensional normal distribution with mean ν and variance ρ, and $\Phi_{\nu,\rho}$ is its ch.f.

Let $\mathbf{X}$ be a stationary process and let $\mu_X := \mathbb{E}[X_0]$, $\mu_{X,k} := \mathbb{E}[(X_0 - \mu_X)^k]$ $k > 1$, and $\gamma_X(t) := \mathbb{E}[(X_0 - \mu_X)(X_t - \mu_X)]$, $t \in \mathbb{Z}$. In this section we will handle $X_1, X_2, ..., X_n$, $n \in \mathbb{N}$ a sample of equally spaced observations of $\mathbf{X}$. We set $\hat{\mu}_X := n^{-1}\sum_{i=1}^n X_i$, $\hat{\mu}_{X,k} := n^{-1}\sum_{i=1}^n (X_i - \hat{\mu}_X)^k$ $k \in \mathbb{N}$ and

$$\hat{\gamma}_X(t) := n^{-1}\sum_{i=1}^{n-|t|}(X_i - \hat{\mu}_X)(X_{i+|t|} - \hat{\mu}_X), |t| \leq n-1.$$

When it is clear, we write $\mu_{X,k}$ as μ_k, $\gamma_X(0)$ as γ_X and $\hat{\gamma}_X(0)$ as $\hat{\gamma}_X$.

The dissipative distributions (see Definition 2.1) were introduced in Cuesta-Albertos et al. (2007). In the finite dimensional case, the dissipative distributions and the absolutely continuous distributions with respect to the Lebesgue measure coincide. Thus, the dissipative distributions can be considered as a generalization of the absolutely continuous distributions to the infinite dimensional case in which there is no measure to play the role of the Lebesgue measure.

It should be noted that all non-degenerate Gaussian distributions are dissipative. In Section 3 we introduce a non-Gaussian dissipative distribution well suited for the problem at hand.

Definition 2.1. *Let $\boldsymbol{D}$ be an $\mathbb{H}$-valued r.e. We will say that its distribution is dissipative if there exists an orthonormal basis $\{v_n\}_{n=1}^{\infty}$ of $\mathbb{H}$, such that $I\!P(\boldsymbol{D}_{V_n^\perp} = 0) = 0$, for all $n \geq 2$, and if the conditional distribution of $\boldsymbol{D}_{V_n}$ given $\boldsymbol{D}_{V_n^\perp}$ is absolutely continuous with respect to the n-dimensional Lebesgue measure.*

2.1. The Epps test

Let

$$\Lambda_N := \{\lambda := (\lambda_1, \ldots, \lambda_N)^T \in \mathbb{R}_N^+ : \lambda_i \neq \lambda_j, i \neq j,\ i,j = 1, ..., N\},$$

and for $\lambda \in \Lambda_N$, let

$$\hat{g}(\lambda) := \frac{1}{n}\sum_{i=1}^n (\cos(\lambda_1 X_i), \sin(\lambda_1 X_i), \ldots, \cos(\lambda_N X_i), \sin(\lambda_N X_i))^T.$$

We set

$$g_{\nu,\rho}(\lambda) := (\mathrm{Re}(\Phi_{\nu,\rho}(\lambda_1)), \mathrm{Im}(\Phi_{\nu,\rho}(\lambda_1)), \ldots, \mathrm{Re}(\Phi_{\nu,\rho}(\lambda_N)), \mathrm{Im}(\Phi_{\nu,\rho}(\lambda_N)))^T.$$

The spectral density matrix of the process

$$(g(X_t,\lambda))_{t\in\mathbb{Z}} := ((\cos(\lambda_1 X_t), \sin(\lambda_1 X_t), \ldots, \cos(\lambda_N X_t), \sin(\lambda_N X_t)))_{t\in\mathbb{Z}}^T$$

at frequency 0 is denoted by $f_{\mathbf{X}}(0, (\mu_X, \gamma_X), \lambda)$ and is estimated by

$$\hat{f}(0,\lambda) = (2\pi n)^{-1}\left(\sum_{t=1}^n \hat{G}(X_{t,0},\lambda) + 2\sum_{i=1}^{\lfloor n^{2/5}\rfloor}(1 - i/\lfloor n^{2/5}\rfloor)\sum_{t=1}^{n-i}\hat{G}(X_{t,i},\lambda)\right),$$

where $\hat{G}(X_{t,i},\lambda) = (g(X_t,\lambda) - \hat{g}(\lambda))(g(X_{t+i},\lambda) - \hat{g}(\lambda))^T$ and $\lfloor\cdot\rfloor$ denotes the integer part.

Let $\Theta \subset \mathbb{R} \times \mathbb{R}^+$ be an open bounded set and let $\lambda \in \Lambda_N$. Let $G_n^+(\lambda)$ be the generalized inverse of $2\pi \hat{f}(0,\lambda)$ and

$$Q_n(\nu,\rho,\lambda) := \left(\hat{g}(\lambda) - g_{\nu,\rho}(\lambda)\right)^T G_n^+(\lambda) \left(\hat{g}(\lambda) - g_{\nu,\rho}(\lambda)\right).$$

Given $i = 1,\dots,N$, using the modulus-argument form to write complex numbers, it is obvious that there exist (ν_i,ρ_i) such that

$$\{(\nu,\rho) \in \mathbb{R} \times \mathbb{R}^+ : e^{i\nu\lambda_i} e^{-\lambda_i^2 \rho/2} = \Phi_X(\lambda_i)\} = \{(\nu_i + 2k\pi/\lambda_i, \rho_i) : k = 1,\dots\}.$$

Thus, the set

$$\Theta_0(\lambda) := \{(\nu,\rho) \in \Theta : \Phi_{\nu,\rho}(\lambda_i) = \Phi_X(\lambda_i), i = 1,\dots,N\}$$

is discrete. Notice that this set will contain at most one element unless the λ_j's are rational multiple of λ_1.

Next, we include an assumption on some regularity conditions of the involved functions on the points in $\Theta_0(\lambda)$. This assumption, taken from Epps (1987), will be employed in the results related to the Epps test.

Assumption A. For each $(\nu,\rho) \in \Theta_0(\lambda)$ it happens that $f_{\mathbf{X}}(0,(\nu,\rho),\lambda) = f_{\mathbf{X}}$ $(0,(\mu_X,\gamma_X),\lambda)$ and that

$$\left.\frac{\partial \Phi_{x,y}(\lambda_i)}{\partial(x,y)}\right|_{(x,y)=(\nu,\rho)} = \left.\frac{\partial \Phi_{x,y}(\lambda_i)}{\partial(x,y)}\right|_{(x,y)=(\mu_X,\gamma_X)}, i = 1,\dots,N.$$

The following theorem, proved in Epps (1987), shows the asymptotic distribution of the statistic involved in the Epps test under the null hypothesis.

Theorem 2.2 (Epps, 1987)**.** *Let* $\mathbf{X}$ *be a stationary Gaussian process satisfying*

$$\sum_{t\in\mathbb{Z}} |t|^\zeta |\gamma_{\mathbf{X}}(t)| < \infty, \text{ for some } \zeta > 0. \tag{2}$$

Let $\Theta \subset \mathbb{R} \times \mathbb{R}^+$ *be open and bounded and let* $\lambda \in \Lambda_N$ *such that Assumption* A *holds. Let* (μ_n,γ_n) *be the minimizer on* Θ *nearest to* $(\hat{\mu}_X,\hat{\gamma}_X)$ *of the map* $(\nu,\rho) \to Q_n(\nu,\rho,\lambda)$*. Assume moreover that* $f_{\mathbf{X}}(0,(\mu_X,\gamma_X),\lambda)$ *is positive definite. Then,* $nQ_n(\mu_n,\gamma_n,\lambda)$ *converges in distribution to* χ^2_{2N-2}*.*

Remark 2.2.1. Since the Epps test only checks whether the ch.f. of the marginal of the involved process coincides with that of a Gaussian distribution on a finite number of points fixed in advance (see Theorem 2.2), this test is non-consistent against alternatives with Gaussian marginals or, even, against distributions with non-Gaussian marginals whose ch.f.s take the appropriate values on the selected points.

Below, in Theorem 3.6, we show that this last problem is alleviated if the set employed in the Epps test is selected at random, thus making this test consistent against every alternative with non-Gaussian one-dimensional marginals.

2.2. The Lobato and Velasco test

Theorem 2.3 shows the behavior of the Lobato and Velasco test. We use the following functions.

$$\begin{aligned} F_k &:= \sum_{t=-\infty}^{\infty} \gamma_X(t)^k, \\ \tilde{F}_k &:= 2\sum_{t=1}^{n-1} \hat{\gamma}_X(t)(\hat{\gamma}_X(t)+\hat{\gamma}_X(n-t))^{k-1} + \hat{\gamma}_X^k \text{ and} \\ \tilde{G}_X &:= n\hat{\mu}_{X,3}^2/(6\tilde{F}_3) + n(\hat{\mu}_{X,4} - 3\hat{\mu}_{X,2}^2)^2/(24\tilde{F}_4). \end{aligned}$$

Theorem 2.3 (Lobato and Velasco, 2004)**.** *Let* $\mathbf{X}$ *be an ergodic stationary process. If* $\mathbf{X}$ *is Gaussian and satisfies* $\sum_{t=0}^{\infty} |\gamma_{\mathbf{X}}(t)| < \infty$*, then* $\tilde{G}_X \longrightarrow \chi_2^2$ *in distribution.*

$\tilde{G}_X$ *diverges to infinity whenever* $\mu_{X,3} \neq 0$ *or* $\mu_{X,4} \neq 3\mu_{X,2}^2$*, if* $\mathbb{E}[X_t^{16}] < \infty$ *and*

- $\sum_{t_1=\infty}^{\infty} \cdots \sum_{t_{q-1}=\infty}^{\infty} |k_q(t_1, ..., t_{q-1})| < \infty$*, for q=2,...,16, where* $k_q(t_1, ..., t_{q-1})$ *denotes the qth-order cumulant of* $X_1, X_{1+t_1}, ..., X_{1+t_{q-1}}$*,*
- $\sum_{t=1}^{\infty} [\mathbb{E}|(\mathbb{E}(X_0-\mu)^k|\mathcal{F}_{-t}) - \mu_k|^2]^{1/2} < \infty$*, for* $k = 3, 4$*, where* $\mathcal{F}_{-t}$ *denotes the* σ*-field generated by* X_j*,* $j \leq -t$*, and*
- $\mathbb{E}[(X_0-\mu)^k - \mu_k]^2 + 2\sum_{t_1=\infty}^{\infty} \mathbb{E}([(X_0-\mu)^k - \mu_k][(X_t-\mu)^k - \mu_k]) > 0$*,* $k = 3, 4$*.*

Remark 2.3.1. As shown in Theorem 2.3, the Lobato and Velasco test is not consistent, since this test only checks whether the kurtosis and the skewness of the marginal coincide with those of a Gaussian distribution.

3. Main results

In (1) the null hypothesis holds if, and only if, $(X_1, \ldots, X_t)^T$ is a Gaussian vector for all $t \in \mathbb{N}$. Due to the stationarity of $\mathbf{X}$, this is equivalent to $(X_t)_{t\leq 0}$ being Gaussian and so, to the Gaussianity of the process $X^{(t)} := (X_j)_{j\leq t}$ for any $t \in \mathbb{Z}$. Given $t \in \mathbb{Z}$, we want to use Theorem 1.1 to check whether $X^{(t)}$ is Gaussian. Hence, we have to include $X^{(t)}$ in an appropriate Hilbert space, $\mathbb{H}$, and select at random a point $\mathbf{h} \in \mathbb{H}$ using a dissipative distribution. Once this is done, we will have that, with probability one, $X^{(t)}$ is Gaussian if, and only if, the real-valued r.v. $\langle X^{(t)}, \mathbf{h}\rangle$ is Gaussian.

Concerning $\mathbb{H}$, let us consider the space of sequences

$$l^2 = \left\{ (x_n)_{n\in\mathbb{N}} : \sum_{n\in\mathbb{N}} x_n^2 a_n < \infty \right\},$$

with $a_0 := 1$ and $a_n := n^{-2}, n \geq 1$, endowed with the scalar product

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n \in \mathbb{N}} x_n y_n a_n, \text{ where } \mathbf{x} = (x_n)_{n \in \mathbb{N}} \text{ and } \mathbf{y} = (y_n)_{n \in \mathbb{N}}.$$

If the variance of X_t is finite, then $E[\sum_{n \in \mathbb{N}} X_{t-n}^2 a_n]$ is also finite. This implies that, almost surely, $X^{(t)} \in l^2$.

Now we need a dissipative distribution on l^2 to be employed to select the direction in which we will project the data. To do this, we use the so-called Dirichlet distribution (Pitman, 2006) and build it using the stick-breaking method: Let $\alpha_1, \alpha_2 > 0$ and consider the following distribution:

- $l_0 \in [0, 1]$ is drawn with the $\beta(\alpha_1, \alpha_2)$ distribution and,
- for $n \geq 1$, $l_n \in [0, 1 - \sum_{i=0}^{n-1} l_i]$ is drawn multiplying an independent $\beta(\alpha_1, \alpha_2)$ r.v. by $1 - \sum_{i=0}^{n-1} l_i$.

Define $H_n = (l_n/a_n)^{1/2}$ for $n \geq 0$ and set $\mathbf{H} = (H_0, H_1, ...)^T$. It can be easily checked that the distribution of $\mathbf{H}$ is dissipative. The only point remaining is to show that the elements generated from this distribution belong to l^2. This is discussed in what follows.

Proposition 3.1. *Let $\mathbf{H} = (H_n)_{n \geq 0}$ be a stochastic process constructed as described above. Then, $\|\mathbf{H}\| = 1$, a.s.*

Using this distribution, we obtain the random projections as follows. Let $\mathbf{h} = (h_i)_{i \in \mathbb{N}}$ be a fixed realization of the random element $\mathbf{H}$. We assume that $\mathbf{H}$ is independent of the process $\mathbf{X}$. Let us consider the process $\mathbf{Y}^h = (Y_t^h)_{t \in \mathbb{Z}}$ given by the projections of $(X^{(t)})_{t \in \mathbb{Z}}$ on the one-dimensional subspace generated by $\mathbf{h}$, i.e.

$$Y_t^h = \sum_{i=0}^{\infty} h_i X_{t-i} a_i, t \in \mathbb{Z}. \tag{3}$$

Henceforth, when no ambiguity arises, the superscript ^{h} is omitted to simplify notation.

We will denote $\gamma_{Y|\mathbf{h}}(t) := \mathbb{E}[(Y_0 - \mu_{Y|\mathbf{h}})(Y_t - \mu_{Y|\mathbf{h}})|\mathbf{h}]$, where $\mu_{Y|\mathbf{h}} := \mathbb{E}[Y_0|\mathbf{h}]$.

The following proposition shows that the projected process inherits the properties of the original one.

Proposition 3.2. *Let $(X_t)_{t \in \mathbb{Z}}$ be an ergodic and stationary process such that $\sum_{t=0}^{\infty} t^{\zeta}$ $|\gamma_X(t)| < \infty$, with $\zeta \geq 0$. Then, conditionally on $\mathbf{h}$, the process $(Y_t)_{t \in \mathbb{Z}}$ defined in (3) is ergodic and stationary. Additionally, $\mathbb{E}[|Y_0||\mathbf{h}]$ and $\sum_{t=0}^{\infty} t^{\zeta}$ $|\gamma_{Y|\mathbf{h}}(t)|$ are finite.*

Therefore, the process $\mathbf{Y}$ is stationary. Thus, it is possible to assess the Gaussianity of the one-dimensional marginal distribution of $\mathbf{Y}$ with the tests we mentioned in the introduction. In particular, it is possible to use the procedures proposed in Epps (1987) or Lobato and Velasco (2004). We denote these procedures, respectively, by E-test and LV-test. Thus, the only remaining task is to find a suitable set of hypotheses which allow these procedures to be applied to the process $\mathbf{Y}$.

Theorem 3.6 shows that if the points involved in the E-test are selected at random, then, the consistency of this test improves. For the sake of simplicity, in this result we have not made explicit the dependence of Q_n on $\mathbf{h}$. To establish the result, we need some preliminary results, that include a corollary which shows that the E-test behaves properly when applied to the process $\mathbf{Y}$.

We denote by $k_{lmno}(q, r, q+r; \lambda)$ the fourth-order cumulant of $Z_{0,l}$, $Z_{q,m}$, $Z_{r,n}$, and $Z_{q+r,o}$, where, for instance, $Z_{q,m}$ is the m-th component of the vector $g(Y_q, \lambda) - g_{\mu_Y,\gamma_Y}(\lambda)$.

Lemma 3.3. *Let $\lambda \in \Lambda_N$ and let $\boldsymbol{Y}$ be a stationary process such that*

$$\sup_{-\infty<q<\infty} \sum_{r=-\infty}^{\infty} |k_{lmno}(q, r, q+r; \lambda)| < \infty, \text{ for each } l, m, n, o \in \{1, ..., N\}. \quad (4)$$

Then, $\hat{f}(0, \lambda) \to f_{\mathbf{Y}}(0, (\mu_Y, \gamma_Y), \lambda)$ almost surely.

Lemma 3.4. *If $\lambda = (\lambda_1, \ldots, \lambda_N)^T \in \Lambda_N$ $(N > 1)$ is randomly drawn in such a way that λ_1 and λ_2 are i.i.d. and have a density, then, Assumption* A *is fulfilled a.s.*

The following corollary follows trivially from Theorem 2.2 and the previous lemma.

Corollary 3.5. *Let $(Y_t)_{t\in\mathbb{Z}}$ be a stationary Gaussian process satisfying (2) and λ as in Lemma 3.4. Let (μ_n, γ_n) be the minimizer on Θ nearest to $(\hat{\mu}, \hat{\gamma})$ of the map $(\nu, \rho) \to Q_n(\nu, \rho, \lambda)$. If $f_{\mathbf{Y}}(0, (\mu_Y, \gamma_Y), \lambda)$ is positive definite, then $nQ_n(\mu_n, \gamma_n, \lambda)$ converges in distribution to χ^2_{2N-2}.*

The following result provides the conditions that allow the E-test to be applied to the projected process. Here, we modify the E-test to select the values of λ at random. This improves the consistency of the initial procedure which is now able to detect (with a sufficiently large sample) every non-Gaussian alternative which satisfies the assumptions.

Theorem 3.6. *Let $\boldsymbol{X}$ be an stationary process satisfying (2). Draw λ as in Lemma 3.4 and $\boldsymbol{h}$ independently of λ using $P_{\mathbf{H}}$. Assume that, conditionally on $\mathbf{h}$, $\mathbf{Y}$ defined in (3) satisfies (4). Assume further that the modulus of the ch.f. of its one-dimensional marginal is analytic and that $f_{\mathbf{Y}|\mathbf{h}}(0, (\mu_{Y|\mathbf{h}}, \gamma_{Y|\mathbf{h}}), \lambda)$ exists and is positive definite for almost every $\mathbf{h}$.*

Let $Q_n(\cdot, \cdot, \lambda)$ be the quadratic form defined in Section 2.1 applied to $\boldsymbol{Y}$ and (μ_n, γ_n) be the minimizer on Θ nearest to $(\hat{\mu}_{Y|\mathbf{h}}, \hat{\gamma}_{Y|\mathbf{h}})$ of $Q_n(\cdot, \cdot, \lambda)$. Let moreover

$$B := \{(\lambda, h) : nQ_n(\mu_n, \gamma_n, \lambda) \to_d \text{ a non-degenerate distribution}\}.$$

Then, $\boldsymbol{X}$ is Gaussian if, and only if, $(P_{\lambda} \otimes P_{\mathbf{H}})[B] > 0$.

Remark 3.6.1.

1. The assumption that $\mathbf{X}$ is ergodic is required only to prove the inverse part of Theorem 3.6. Indeed, any stationary Gaussian process satisfying (2) is ergodic (Ibragimov and Rozanov, 1978).
2. A slightly more involved proof would allow Theorem 3.6 to be proved under the assumption that the ch.f. is analytic.

The following corollary shows that the consistency of the Epps test improves if the involved points are chosen at random. Then, we include Corollary 3.8, where we state a kind of zero-one law to reinforce the statements of Theorem 3.6 and Corollary 3.7.

Corollary 3.7. *Let $\boldsymbol{X}$ be an ergodic stationary process. Assume that the modulus of the ch.f. of its one-dimensional marginal is analytic. Assume further that (2) holds. Take λ as in Lemma 3.4 and $Q_n(\cdot,\cdot,\lambda)$ as in Section 2.1. Let (μ_n,γ_n) be the minimizer on Θ nearest to $(\hat{\mu}_X,\hat{\gamma}_X)$ of $Q_n(\cdot,\cdot,\lambda)$. Let*

$$C := \{\lambda : nQ_n(\mu_n,\gamma_n,\lambda) \to_d \text{ a non-degenerate distribution}\}.$$

If we assume that $f_{\mathbf{X}}(0,(\mu_X,\gamma_X),\lambda)$ exists and is positive definite, then, $\boldsymbol{X}$ is Gaussian if, and only if, $P_{\lambda}(C) > 0$.

Corollary 3.8. *Under the assumptions of Theorem 3.6, $(P_{\lambda} \otimes P_{\mathbf{H}})[B] \in \{0,1\}$ and $\boldsymbol{X}$ is Gaussian if, and only if, $(P_{\lambda} \otimes P_{\mathbf{H}})[B] = 1$.*

Analogously, under the assumptions of Corollary 3.7, $P_{\lambda}(C) \in \{0,1\}$ and $\boldsymbol{X}$ is Gaussian if, and only if, $P_{\lambda}(C) = 1$.

Remark 3.8.1. From Theorem 2.2, we have that Theorem 3.6 and Corollaries 3.7 and 3.8 remain true if we substitute in the definition of sets B and C "non-degenerate distribution" by "chi-squared distribution with $2(N-1)$ degrees of freedom"; this allows a test to be constructed based on the asymptotic distribution of $nQ_n(\mu_n,\gamma_n,\lambda)$.

We end this section with a result which shows the applicability of the LV-test to the projected process under different assumptions than the ones used in Lobato and Velasco (2004). To this end, we replace the statistics $\tilde{G}_Y$ by

$$G_Y = n\hat{\mu}_3^2/(6|\hat{F}_3|) + n(\hat{\mu}_4 - 3\hat{\mu}_2^2)^2/(24|\hat{F}_4|),$$

with

$$\hat{F}_k = 2\sum_{t=1}^{\tau_n} \hat{\gamma}(t)(\hat{\gamma}(t) + \hat{\gamma}(\tau_n + 1 - t))^{k-1} + \hat{\gamma}^k, \tau_n < cn^{\beta_0}, 0 < \beta_0 < .5 \text{ and } c > 0. \tag{5}$$

Thus, the differences between G_Y and $\tilde{G}_Y$ are the absolute values in the denominator and the number of terms involved in $\hat{F}_k$.

Theorem 3.9. *Let $\boldsymbol{X}$ be an ergodic and stationary process which satisfies $\sum_{t=0}^{\infty} |\gamma_X(t)| < \infty$. Then,*

1. *If $\boldsymbol{X}$ is a Gaussian process, then $G_Y \longrightarrow_d \chi_2^2$.*

2. *Assume that* $X_t - \mu_X = \sum_{i=1}^{\infty} k(i)\epsilon_{t-i}$ *with* $\sum_{i=1}^{\infty} |k(i)| < \infty, \sum_{i=1}^{\infty} ik(i) < \infty$*, and* (ϵ_t) *are i.i.d. r.v.'s with* $\mathbb{E}[\epsilon_n] = 0$*, and* $\mathbb{E}[X_0^4] < \infty$*. Then, conditionally on* $\mathbf{h}$*,* G_Y *diverges a.s. to infinity whenever* $\mu_3 \neq 0$ *or* $\mu_4 \neq 3\mu_2^2$*.*

Applying Theorem 3.9 directly to the process $\mathbf{X}$, we obtain the following corollary.

Corollary 3.10. *Under the assumptions of Theorem 3.9, we have that if* $\boldsymbol{X}$ *is a Gaussian process, then* $G_X \longrightarrow_d \chi_2^2$*. Moreover, if the assumptions in point 2 of this theorem hold, then, conditionally on* $\mathbf{h}$*,* G_X *diverges a.s. to infinity whenever* $\mu_{X,3} \neq 0$ *or* $\mu_{X,4} \neq 3\mu_{X,2}^2$*.*

3.1. Context of our test procedure

The Gaussian goodness of fit tests have a long and rich history. On one hand, in the framework of i.i.d. data, this story began in the early days of statistics. First, in the one dimensional case and when the parameters are known, the classical generic procedures based on a distance between the empirical distribution function and the Gaussian target distribution are a simple old procedure. They have been widely used for a long time. In this context, the tests based on Kolmogorov or Cramér von Mises statistics (D'Agostino and Stephens, 1986) are among the most popular. Notice that when the parameters of the Gaussian distribution are unknown, everything gets more complicated. Indeed, the previous generic procedures may still be applied in a plug-in approach. Nevertheless, care should be taken in the way the unknown parameters are estimated. We refer to Cabaña (1996) for these kind of plug-in procedures. Smart procedures specially devoted to the one dimensional Gaussian case appear in Shapiro and Wilk (1965); D'Agostino (1971). The procedures therein are based on some self-normalized L-statistics and rely on some statistical approximation of the Wasserstein distance W_2 between the sample distribution and the one dimensional Gaussian distribution family (Barrio et al., 1999, for example). The multidimensional case is rather more complicated and requires new tools. In Csorgo (1986), the author uses a version of the empirical characteristic function to build general procedures. As a matter of fact, under the null hypothesis (Gaussianity), this random process is asymptotically distribution-free. This allows, for example, by sampling and whitening, to build a quadratic form statistic having an asymptotic χ^2 distribution (under the null hypothesis). Notice that, here the dimension of the space is finite and greater than 1 but fixed. This pioneering work has subsequently been skillfully extended in many directions for example to test the symmetry or isotropy of a multidimensional distribution and using the cumulant generating function (Ghosh and Ruymgaart, 1992; Fang and Liang, 1998; Liang and Ng, 2009, and references therein). The infinite dimension case is tackled using random projections in Cuesta-Albertos et al. (2007).

On the other hand, in the case of time series data, only the goodness of fit test for Gaussianity of a fixed finite marginal has been studied (Epps, 1987;

Lobato and Velasco, 2004; Moulines and Choukri, 1996, for example). Notice that the regularity assumption has recently been weakened in Ghosh (2013) in order to work with long range dependence processes. Our work appears to be at the crossroad between all these works; that is, those on time series, on i.i.d. data in infinite dimension or on finite dimension. As a matter of fact, the random projection trick allows only a one dimensional marginal to be considered. Nevertheless, as the sample is not i.i.d., tools similar to those used in the multidimensional case appear. For example, we have to whiten empirically the sample in order to obtain an asymptotic χ^2 distribution (under the null hypothesis). Furthermore, the procedure based on the empirical characteristic function is a method related to the one developed in Csorgo (1986) although the sampled frequencies are, in our work, chosen randomly.

3.2. The test in practice

Let $X_0, \ldots, X_n$ be the available measurements. To compute $\mathbf{h}$, let $\delta > 0$ be a fixed number (equal to 10^{-15} in the simulations that we carry out in Sections 4 and 5), and take $\mathbf{h} = (h_0, \ldots, h_m)^T$ with

$$m = 1 + \min\{\min\{t : \|(h_0, \ldots, h_t)^T\| \geq 1 - \delta\}, n - 1\},$$

where $h_0, \ldots, h_{m-1}$ are drawn as follows and h_m is such that $\|\mathbf{h}\| = 1$. Then, define

$$Y_t = \sum_{i=0}^{\min(m,t)} h_i X_{t-i} a_i, t = 0, \ldots, n.$$

To draw $h_0, \ldots, h_{m-1}$, let's fix $\alpha_1, \alpha_2 > 0$. Then, we choose $(\beta_n)_{n\in\mathbb{N}}$ independent and identically distributed with beta distribution of parameters α_1 and α_2. Further, we consider the probability distribution which selects a random point in l^2 according to the following iterative procedure:

- $l_0 = \beta_0 \in [0, 1]$.
- Given $n \geq 1$, $l_n \in [0, 1 - \sum_{i=0}^{n-1} l_i]$ equal to $\beta_n(1 - \sum_{i=0}^{n-1} l_i)$.

Let us define $H_n = (l_n/a_n)^{1/2}$ for $n \in \mathbb{N}$ with $a_0 = 1$ and $a_n = n^{-2}, n \geq 1$, and take $\mathbf{H} = (H_n)_{n\in\mathbb{N}}$. Thus, $\mathbf{h} = (h_i)_{i\in\mathbb{N}}$ is a fixed realization of the random element $\mathbf{H}$.

Selecting α_1, α_2, leads to the following problem: if m is large, the r.v.'s Y_t are linear combinations of many r.v.'s from the first sample and, by the Central Limit Theorem (CLT), its common distribution becomes closer to a normal law, causing loss of power when the marginal of $\mathbf{X}$ is not Gaussian. To have a small m, we take $\alpha_2 = 1$ and $\alpha_1 \gg 1$ (we take $\alpha_1 = 100$ in Sections 4 and 5). However, in this case the samples $Y_0, \ldots, Y_n$ and $X_0, \ldots, X_n$ are quite similar. Thus, the test will not be able to detect properly non-Gaussian alternatives with the Gaussian marginal. To overcome this drawback, the projections should mix several r.v.'s from the initial sample. To do this, we take $\alpha_2 > \alpha_1$, with α_2 not too big to avoid the effect of the CLT. Values such as $\alpha_1 = 2$ and $\alpha_2 = 7$ seem appropriate, and this is our selection in Sections 4 and 5.

To avoid selecting α_1 and α_2 based on an assumed alternative, we use two tests, one with each pair of parameters, and apply the false discovery rate for dependent tests (Benjamini and Yekutieli, 2001, FDR) to mix the p-values. Taking into account that the null hypotheses are the same in both cases, it happens that the FDR coincides with the level of the whole procedure.

Furthermore, in the simulations, we discovered that the relative power performance of both the E-test and the LV-test changes across the different alternatives. This led us to conclude that, rather than using either the improved version of the E-test or that of the LV-test, we should use both tests and use, again, the FDR to mix all the p-values.

Regarding the selection of the random points to be used in the improved E-test, it happens that in the simulations in Epps (1987) and Lobato and Velasco (2004), the authors take $\xi_j/\sqrt{\hat{\gamma}}$, with $\xi_j = j,\ j = 1, 2$, where $\hat{\gamma}$ denotes the sample variance of the process. Here, we take ξ_j distributed as the absolute value of a $N(0, j),\ j = 1, 2$.

Finally, even if $\mathbf{X}$ is not Gaussian, the chosen projections may, just by chance, be close enough to Gaussianity as to lead to non-rejection. To alleviate this, Cuesta-Albertos et al. (2007) suggests increasing the number of random projections. Following this idea, our full proposal is to choose $k > 0$ and select independent random vectors $\mathbf{h}^{(i,j)},\ i = 1, \ldots, k,\ j = 1, \ldots, 4$, where $\mathbf{h}^{(i,1)}$ and $\mathbf{h}^{(i,2)}$ are drawn with the $\beta(100, 1)$ distribution and $\mathbf{h}^{(i,3)}$ and $\mathbf{h}^{(i,4)}$ with the $\beta(2, 7)$ distribution, $i = 1, \ldots, k$. Then, for $i = 1, \ldots, k$:

1. Draw $\mathbf{h}^{(i,1)}$ with the $\beta(100, 1)$ distribution and apply the E-test to the projections to obtain the p-value $p^{(i,1)}$.
2. Draw $\mathbf{h}^{(i,2)}$ (independently of $\mathbf{h}^{(i,1)}$) with the $\beta(100, 1)$ distribution and apply the L-V-test to the projections to obtain the p-value $p^{(i,2)}$.
3. Draw $\mathbf{h}^{(i,3)}$ (independently of $\mathbf{h}^{(i,1)}$ and $\mathbf{h}^{(i,2)}$) with the $\beta(2, 7)$ distribution and apply the E-test to the projections to obtain the p-value $p^{(i,3)}$.
4. Draw $\mathbf{h}^{(i,4)}$ (independently of $\mathbf{h}^{(i,1)}$, $\mathbf{h}^{(i,2)}$ and $\mathbf{h}^{(i,3)}$) with the $\beta(2, 7)$ distribution and apply the L-V-test to the projections to obtain the p-value $p^{(i,4)}$.

Finally, combine the $p^{(i,j)}$'s with the FDR to obtain a p-value for the global procedure.

The parameter k remains free and, to underline this, we call this procedure the k *random projection* test (${}_k$RP-test). In Cuesta-Albertos and Nieto-Reyes (2008), it is suggested that around 250 random projections are enough; in keeping with this, here we take $k = 64$.

4. Simulation results

First, we study the behavior of the ${}_k$RP-test against the distributions employed in Lobato and Velasco (2004). Then, we study a non-Gaussian process with the Gaussian marginal (Section 4.1) and pay some attention to the effect of taking different numbers of projections (Section 4.2). We end the section by giving a concise explanation of the test to those who want to use it while skipping the details of the paper (Section 4.3).

In Lobato and Velasco (2004), the following AR(1) processes are used:

$$X_t = qX_{t-1} + \varepsilon_t,\ t \in \mathbb{Z},\ \text{for } q \in \{0, \pm.5, .6, .7, .8, \pm.9\}, \tag{6}$$

where ε_t are i.i.d. r.v.'s with distribution $D_\varepsilon : N(0,1)$, standard log-normal, Student t with 10 degrees of freedom, chi-squared with 1 and 10 degrees of freedom, uniform on $[0,1]$ and $\beta(2,1)$.

We simulate the process taking, $X_1 = \varepsilon_1,$ and $X_t = qX_{t-1}+\varepsilon_t,\ t = 2, ..., M$, where the ε_t's are i.i.d. with distribution D_ε. To alleviate the non-stationarity of the process if $q \neq 0$ (notice that, for instance, $\mathrm{Var}[X_t] = \mathrm{Var}[\varepsilon_1](1-q^{2t})/(1-q^2)$ which is non-constant), we discarded the first *past* $= 1000$ observations. As in Lobato and Velasco (2004), we take sample sizes $n = 100, 500, 1000$, and so, $M = n + \textit{past}$. Additionaly, here we also take $n = 50$.

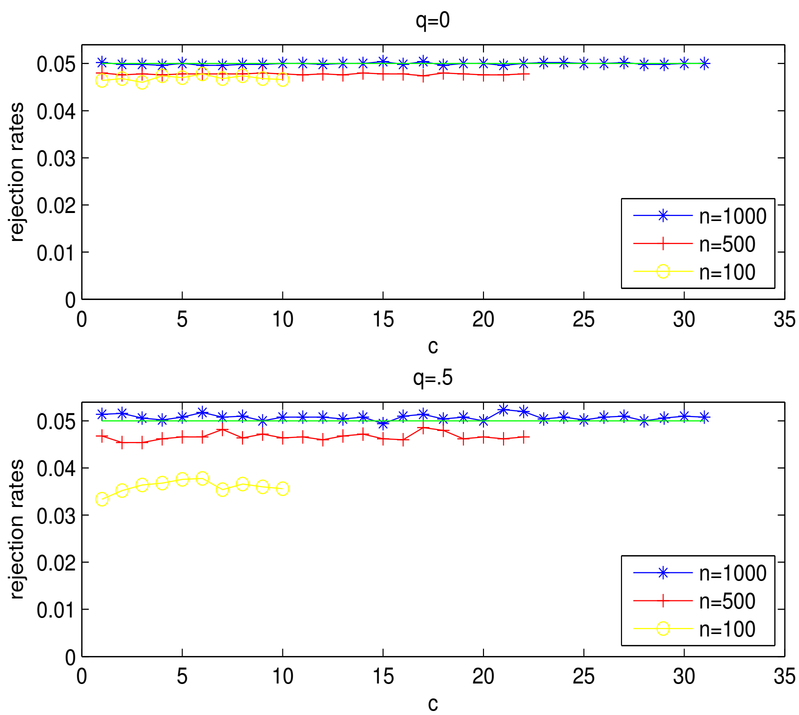
We performed 5,000 simulations in each situation, computing the p-values using the asymptotic distributions. A slow convergence to the asymptotic distribution might be the reason why the rejection rates at level .05 under the null hypothesis are sometimes far from the nominal level (mostly when $n = 50, 100$) and even decrease under some alternatives with n (mostly for high values of $|q|$), (see Tables 2, 3, 4 and 5).

Table 1: Rejection rates for 5,000 simulations for different *past*, with the E-test $n = 100$, D_ε a $\beta(2,1)$ and $q = .7$.

past	0	1	2	10
rejections	.0750	.1378	.1998	.2210

There are some differences between our rates and those in Lobato and Velasco (2004) which might be due to the fact that their *past* is not large enough. For instance, if $n = 100$, $q = .7$ and D_ε is $\beta(2,1)$, with the E-test, we obtain a rejection rate of .2214 while they obtain .080, noticeably worse. In Table 1, we see that .080 is appropriate for *past* =0 and that the rejection rates increase with *past*, approaching the value obtained here. The same happens with the LV-test, but here we obtain lower rejection rates than in Lobato and Velasco (2004). Another difference with Lobato and Velasco (2004) is in the computation of $\hat{F}_k$ (see (5)). Indeed, in our case, it is necessary to fix β_0 and c. As β_0 may be as close as desired to .5, we take $\beta_0 = .5$ in the simulations. We have studied the sensitivity to c by running the LV-test under the null hypothesis for all q's and $\tau_n < n$, $n = 100, 500, 1000$. It seems that c has little influence on the rejection rates (except, perhaps, in the case $q = -.9$ where the rejection rate first appears to be constant and then sharply decreases); thus, we choose $c = 1$. Cases $q = 0$ and $q = .5$ are shown in Figure 1, for $n = 100, 500, 1000$.

Figure 1: Rejection rates under the null hypothesis for an AR(1) process with $q = 0$ (top graph), and $q = .5$ (bottom graph), using the LV-test, for different values of c and sample sizes.



Now, we mention the tests used in Tables 2, 3, 4 and 5 and comment on the results.

E-test We choose $(\xi_1, \xi_2) = (1, 2)$ as in the simulations of Epps (1987) and Lobato and Velasco (2004).

Under the null hypothesis, the rejection rates are above the level of the test, except for $q = 0$ with $n = 1000$, and increase with $|q|$ except for $|q| = .5$ with $n = 50$. Under the alternatives, this test behaves poorly when D_ε is t_{10} and its power decreases when $|q|$ increases (with very low powers when $|q| = .9$) and also when the sample size increases if $|q| = .9$ and D_ε is t_{10}, χ^2_{10}, $U(0,1)$ or $\beta(2,1)$ (even with $q = .8$ when $D_\varepsilon = t_{10}$).

LV-test We report the results obtained using G_X instead of $\tilde{G}_X$, but this is not too important because the rejection rates are similar in both cases.

The rejections under the null hypothesis are above the level of the test only in 5 cases out of 32 and, in contrast with the E-test, they generally decrease when $|q|$ increases. This test has very low powers when $|q|$ is large (sometimes

Table 2: Rejection rates at level .05 of a process defined by (6). Sample size $n = 50$.

q	Test	N(0,1)	log N	t_{10}	χ^2_1	χ^2_{10}	$U(0,1)$	$\beta(2,1)$
	E	.1652	.0956	.1740	.1210	.1676	.1868	.1886
-.9	LV	.0288	.0750	.0290	.0532	.0350	.0236	.0280
	ELV	.1444	.1346	.1288	.1362	.1252	.1468	.1478
	$_k$RP	.1706	.5920	.1888	.5404	.2650	.2598	.31662
	E	.0818	.2386	.0584	.2964	.1318	.3682	.3092
-.5	LV	.0568	.9312	.1342	.9120	.2600	.0146	.0884
	ELV	.0940	.9084	.1256	.8760	.2676	.2852	.2602
	$_k$RP	.0946	.9528	.1546	.9532	.3288	.4158	.5126
	E	.0854	.6980	.0620	.8330	.2942	.8778	.7940
0	LV	.0420	.9964	.1778	.9986	.4414	.0006	.1256
	ELV	.0792	.9980	.1678	.9992	.5008	.7386	.6872
	$_k$RP	.0888	.9934	.1490	.9968	.4676	.6686	.7326
	E	.0826	.4464	.0532	.5694	.1364	.3896	.3476
.5	LV	.0312	.9080	.0996	.8958	.2036	.0004	.0434
	ELV	.0730	.9220	.1142	.9182	.2370	.2978	.2690
	$_k$RP	.0860	.8134	.1128	.7746	.4230	.2572	.6490
	E	.0920	.2460	.0656	.3410	.1284	.2502	.2472
.6	LV	.0262	.8084	.0720	.7546	.1366	.0018	.0286
	ELV	.0750	.7934	.0936	.7502	.1684	.1880	.1868
	$_k$RP	.0898	.6630	.1036	.5694	.4788	.2542	.6852
	E	.0924	.1326	.0812	.1826	.1116	.1772	.1650
.7	LV	.0220	.6116	.0472	.4946	.0786	.0030	.0182
	ELV	.0768	.5818	.0834	.4746	.1220	.1226	.1202
	$_k$RP	.0896	.4786	.1076	.3416	.5548	.3250	.7506
	E	.1186	.1060	.1088	.1108	.1234	.1506	.1480
.8	LV	.0164	.3142	.0256	.2108	.0346	.0054	.0134
	ELV	.0830	.3090	.0826	.2186	.0972	.0974	.089
	$_k$RP	.1088	.2784	.1034	.1866	.6808	.4588	.8274
	E	.1876	.1372	.1778	.1328	.1662	.1812	.1806
.9	LV	.0124	.0696	.0140	.0490	.0152	.0038	.0072
	ELV	.1150	.1192	.1090	.0904	.1040	.1088	.0938
	$_k$RP	.1460	.3190	.1522	.2162	.8734	.7484	.9370

even lower than the E-test) and suffers from a lack of power when D_ϵ is $U(0,1)$ or $\beta(2,1)$.

ELV-test We take ξ_i as described in Section 3.2, to take advantage of Corollary 3.7.

This test combines the E and LV-tests using the FDR, thus, obtaining rejection rates between those of the E and LV-tests although closer to the highest one. However, the fact that ξ_i is chosen at random improves the performance of

Table 3: Rejection rates at level .05 of a process defined by (6). Sample size $n = 100$.

q	Test	N(0,1)	log N	t_{10}	χ^2_1	χ^2_{10}	$U(0,1)$	$\beta(2,1)$
	E	.1264	.0508	.1104	.0656	.1124	.1390	.1354
-.9	LV	.0292	.1414	.0310	.0840	.0332	.0290	.0266
	ELV	.0942	.1422	.0908	.1072	.0920	.1020	.1010
	${}_k$RP	.1380	.8070	.1742	.7576	.3076	.2620	.3902
	E	.0724	.6780	.0556	.8514	.2058	.5408	.4914
-.5	LV	.0504	.9986	.1692	.9986	.4602	.0102	.1696
	ELV	.0774	.9976	.1582	.9972	.4552	.4454	.4154
	${}_k$RP	.0752	.9998	.1980	1	.5824	.6404	.7460
	E	.0632	.9616	.0830	.9964	.5372	.9918	.9704
0	LV	.0458	1	.2820	1	.7898	.5404	.7520
	ELV	.0732	1	.2402	1	.8074	.8596	.8706
	${}_k$RP	.0772	1	.2288	1	.7640	.8496	.9054
	E	.0682	.8594	.0608	.9582	.2610	.5618	.5562
.5	LV	.0384	.9990	.1696	.9982	.4118	.0010	.1102
	ELV	.0642	.9990	.1444	.9988	.4700	.4680	.4882
	${}_k$RP	.0750	.9908	.1132	.9880	.5226	.3256	.7500
	E	.0710	.6118	.0582	.8106	.2006	.3462	.3650
.6	LV	.0358	.9884	.1162	.9772	.2858	.0012	.0592
	ELV	.0640	.9882	.1144	.9832	.3218	.2800	.3086
	${}_k$RP	.0802	.9536	.1030	.9262	.5164	.2580	.7744
	E	.0838	.3250	.0626	.4640	.1492	.2032	.2214
.7	LV	.0260	.9076	.0814	.8196	.1610	.0036	.0334
	ELV	.0714	.9042	.0866	.8448	.1998	.1634	.1802
	${}_k$RP	.0784	.8022	.0926	.7010	.5754	.2902	.8060
	E	.1034	.1552	.0810	.2004	.1324	.1620	.1596
.8	LV	.0206	.6146	.0466	.4406	.0708	.0046	.0166
	ELV	.0726	.6118	.0796	.4488	.1122	.1154	.1136
	${}_k$RP	.0896	.4928	.0932	.3264	.6766	.3950	.8782
	E	.1752	.1264	.1618	.1368	.1612	.1870	.1680
.9	LV	.0106	.1558	.0094	.0714	.0150	.0054	.0086
	ELV	.1074	.1844	.0968	.1190	.0980	.1182	.1072
	${}_k$RP	.1168	.1982	.1174	.1338	.8702	.6788	.9662

the E-test and this means that, sometimes, the ELV-test performs better than the E and LV-tests.

${}_k$**RP-test** We apply this test following the steps described in Subsection 3.2. However, here we take $k = 1$ to stress the usefulness of the test in its worst circumstances.

Under the alternative, the ${}_1$RP-test has the highest rejection rates when $q < 0$. The most striking behavior occurs for $q = .9$ and $D_\varepsilon = \chi^2_{10}$ or $\beta(2,1)$,

Table 4: Rejection rates at level .05 of a process defined by (6). Sample size $n = 500$.

q	Test	N(0,1)	log N	t_{10}	χ^2_1	χ^2_{10}	$U(0,1)$	$\beta(2,1)$
	E	.0744	.3720	.0584	.2162	.0712	.0918	.0850
-.9	LV	.0708	.8838	.0840	.6202	.1142	.0462	.0754
	ELV	.0780	.8604	.0924	.5400	.1116	.0866	.0952
	$_k$RP	.0810	.9990	.2260	.9928	.6924	.4630	.6918
	E	.0594	1	.1334	1	.7730	.9924	.9922
-.5	LV	.0472	1	.4580	1	.9960	.9656	.9976
	ELV	.0476	1	.3784	1	.9912	.9514	.9914
	$_k$RP	.0490	1	.5090	1	.9998	.9946	1
	E	.0566	1	.3292	1	.9982	1	1
0	LV	.0480	1	.7428	1	1	1	1
	ELV	.0510	1	.6756	1	1	1	1
	$_k$RP	.0554	1	.6188	1	1	1	1
	E	.0654	1	.1476	1	.8808	.9918	.9960
.5	LV	.0454	1	.4340	1	.9972	.9704	.9988
	ELV	.0516	1	.3816	1	.9924	.9504	.9962
	$_k$RP	.0618	1	.2656	1	.9610	.7440	.9634
	E	.0566	.9998	.1026	1	.7084	.8286	.9090
.6	LV	.0470	1	.3336	1	.9582	.4678	.8858
	ELV	.0570	1	.2692	1	.9388	.6944	.8870
	$_k$RP	.0610	1	.1794	1	.8604	.4730	.9006
	E	.0708	.9996	.0786	1	.4704	.4042	.5810
.7	LV	.0474	1	.1970	1	.7592	.0644	.4040
	ELV	.0598	1	.1670	1	.7332	.3640	.5768
	$_k$RP	.0702	1	.1282	1	.6986	.2616	.8786
	E	.0776	.9780	.0710	.9638	.2500	.1948	.2564
.8	LV	.0744	.9998	.0976	.9980	.3908	.1524	.2628
	ELV	.0702	.9998	.1102	.9978	.3972	.1848	.2960
	$_k$RP	.0710	.9986	.0910	.9908	.6834	.2484	.9208
	E	.1156	.5708	.0944	.4674	.1526	.1430	.1560
.9	LV	.0232	.8356	.0370	.5404	.0764	.0138	.0336
	ELV	.0802	.8708	.0838	.6378	.1490	.1092	.1390
	$_k$RP	.0860	.7996	.0770	.5510	.8430	.4818	.9772

as the rejection rates are larger than .8 while the second most successful test remains below .25. The remaining rejection rates are among those obtained with the E, G and ELV tests but closer to the highest, sometimes, even outperforming the other tests.

The reason for the behavior of the $_1$RP-test when $q = .9$ is that when computing the random projections, the structure of the process changes and it happens that the new structure of the process makes it easier to detect the

Table 5: Rejection rates at level .05 of a process defined by (6). Sample size $n = 1000$.

q	Test	N(0,1)	log N	t_{10}	χ^2_1	χ^2_{10}	$U(0,1)$	$\beta(2,1)$
	E	.0648	.7836	.0578	.4572	.0826	.0888	.0942
-.9	LV	.0902	.9934	.1206	.8932	.2448	.0760	.1358
	ELV	.0880	.9856	.1002	.8560	.2190	.1004	.1450
	$_k$RP	.0940	1	.3344	.9998	.8686	.5876	.8056
	E	.0530	1	.2574	1	.9764	1	1
-.5	LV	.0436	1	.6778	1	1	1	1
	ELV	.0450	1	.6040	1	1	1	1
	$_k$RP	.0378	1	.7498	1	1	1	1
	E	.0490	1	.5946	1	1	1	1
0	LV	.0546	1	.9364	1	1	1	1
	ELV	.0486	1	.9162	1	1	1	1
	$_k$RP	.0422	1	.8734	1	1	1	1
	E	.0550	1	.2534	1	.9966	1	1
.5	LV	.0482	1	.6788	1	1	1	1
	ELV	.0424	1	.6016	1	1	1	1
	$_k$RP	.0484	1	.4348	1	.9994	.9738	.9996
	E	.0566	1	.1718	1	.9580	.9800	.9974
.6	LV	.0472	1	.5112	1	.9996	.9724	.9996
	ELV	.0464	1	.4234	1	.9996	.9550	.9986
	$_k$RP	.0584	1	.2812	1	.9902	.7110	.9804
	E	.0594	1	.1162	1	.7720	.6338	.8632
.7	LV	.0418	1	.3104	1	.9744	.3642	.8830
	ELV	.0558	1	.2380	1	.9672	.5642	.8724
	$_k$RP	.0598	1	.1754	1	.8888	.3554	.9036
	E	.0690	.9998	.0720	1	.4342	.2288	.4108
.8	LV	.0500	1	.1638	1	.6804	.0432	.3284
	ELV	.0670	1	.1294	1	.6708	.2216	.4450
	$_k$RP	.0654	1	.0996	1	.7144	.1920	.9076
	E	.0902	.9152	.0880	.7690	.1836	.1170	.1686
.9	LV	.0346	.9944	.0636	.9136	.1574	.0174	.0574
	ELV	.0690	.9926	.0798	.9206	.2178	.1040	.1596
	$_k$RP	.0736	.9844	.0678	.8580	.8328	.3946	.9774

non-Gaussianity.

4.1. An alternative with the Gaussian marginal

To check the power of the $_k$RP test against an alternative with the Gaussian marginal, we use Example 2.3 of Cuesta-Albertos and Matrán (1991). For its construction, let $p > 2$ be a prime number, and Y_0, U and $\{Z_{m\cdot p}, m \in \mathbb{N}\}$ be i.i.d. r.v.'s uniformly distributed on $S := \{0, 1, \ldots, p-1\}$. Let $Z_{m\cdot p+j}$ be the sum modulus p of $Z_{m \cdot p}$ and jY_0, $j \in S, m \in \mathbb{N}$. In Cuesta-Albertos and Matrán (1991), it is proved that the sequence $W_n = Z_{n+U}$ is stationary and composed of pairwise independent r.v.'s, which are not mutually independent as $W_{n-U}, \ldots, W_{n-U+p-2}$ determine $W_{n-U+p-1}$ because, for every $m \in \mathbb{N}$,

$$\begin{array}{ll} p(p-1)/2 = \sum_{i=0}^{p-1} W_{mp-U+i} & \text{if } Y_0 \neq 0 \text{ and} \\ W_{mp-U} = \ldots = W_{mp-U+p-1} & \text{if } Y_0 = 0. \end{array} \tag{7}$$

Given $k \in S$, let q_j be the j/p-th quantile of the $N(0,1)$. If $W_n = j$, let W_n^* be a $N(0,1)$ conditioned to (q_j, q_{j+1}), independent of all the other r.v.'s. Since W_n is uniformly distributed on S, W_n^* is $N(0,1)$, and $\{W_n^*\}$ is a stationary sequence of pairwise independent Gaussian r.v.'s. However, if $n > p-1$ and we know the values $W_{n-U}^*, \ldots, W_{n-U+p-2}^*$, we can recover the values $W_{n-U}, \ldots, W_{n-U+p-2}$ and, by (7), deduce the value $W_{n-U+p-1}$, knowing the interval in which $W_{n-U+p-1}^*$ lies. Thus, the process is not Gaussian as $\{W_n^*\}$ are not mutually i.r.v.'s.

We have simulated the previous process 5000 times for different values of p and sample sizes $n = 100, 500, 1000$ and applied the ${}_1$RP test at level $\alpha = .05$, with the results shown in Table 6.

Table 6: Rejection rates of the process $\mathbf{W}^*$ tested with ${}_k$RP at level $\alpha = .05$.

Sample sizes	$p=3$	$p=5$	$p=7$	$p=11$	$p=13$	$p=17$
$n=$ 100	.1268	.1676	.1516	.1602	.1380	.1146
$n=$ 500	.3654	.4938	.5154	.5822	.5590	.5588
$n=1000$	.6386	.6814	.7250	.7802	.7608	.7700

For the sake of comparison, Table 7 shows the rejection rates of the E, LV and ELV tests for $p = 5$. Those tests are intended to detect non-Gaussian marginals and since this process has the Gaussian marginal, these tests should give powers at the nominal level. Surprisingly, we see that the obtained rates decrease with n and are well below the intended level, except for the ELV test with $n = 100$.

The reason for this is that the process is not conditionally stationary given Y_0: for instance, we have that given $Y_0 = 0$, $I\!P[W_{mp-U} = W_{mp-U+1}] = 1 \neq I\!P[W_{mp-U-1} = W_{mp-U}]$. In fact, what happens is that, in some sense, the observed trajectories of the process $\mathbb{W}^*$ are closer to Gaussianity than those produced by a Gaussian process, because when we generate observations of a Gaussian process, *approximately* a proportion of $1/p$ of observations are in the interval (q_k, q_{k+1}), with $k \in \{0, \ldots, p-1\}$. However, when $Y_0 \neq 0$ (which happens with probability $1-p^{-1}$) the process $\mathbb{W}^*$ generates *exactly* a proportion of $1/p$ of observations in each interval (q_k, q_{k+1}). Thus, it has a "more Gaussian" behavior than expected. Consequently, the rejection rates are lower than .05 (Table 7) and this fact becomes more apparent when n increases.

Table 7: Rejection rates of $\mathbf{W}^*$ with $p = 5$ using E, LV, EVL and ${}_k$RP, $\alpha = .05$.

Sample size	E-test	LV-test	EVL-test	${}_k$RP-test. $k = 1$	$k = 2$	$k = 8$	$k = 64$
$n =$ 100	.0338	.0372	.0520	.1676	.1906	.2288	.2674
$n =$ 500	.0266	.0336	.0336	.4938	.5772	.6988	.8064
$n = 1000$	.0186	.0326	.0206	.6814	.7688	.8498	.8628

4.2. Taking more random projections

Although the rejection rates in Table 6 are above the nominal levels, they are not high, especially when the sample size is 100. To improve them, we can increase the number of random projections. That is, we take a higher k. Table 7 shows how an increase in the number of random projections noticeably improves the rejection rates.

4.3. Proposed test for practitioners

This subsection is mainly devoted to readers whose aim is to understand how to apply the proposed test, and possibly use the provided code, while skipping the details of the paper. Thus, here we detail comprehensively how the test should be applied from start to finish on a given dataset.

Simulation of the dataset

Let us simulate a dataset from an AR(1) process $X_t = .5X_{t-1} + \varepsilon_t,\ t \in \mathbb{Z}$, where ε_t are i.i.d. r.v.'s with distribution, say, chi-squared with 10 degrees of freedom. We simulate the process taking $X_1 = \varepsilon_1$, and $X_t = .5X_{t-1} + \varepsilon_t,\ t = 2, ..., M$. To get a process of size n, for instance $n = 100$, $M = n + 1000$ in order to alleviate the non-stationarity of the process. Thus, we get a realization of the process we can denote by $x = (x_1, \ldots, x_n)$.

What follows is done k times, with for example $k = 64$.

The vectors in which to project

For each $i \in \{1, \ldots, k\}$ the dataset, x, is projected in four different vectors $h^{(i,j)}$, $j = 1, \ldots, 4$, that are drawn using the following procedure. For each (i, j), let $\beta_0, \ldots, \beta_{n-1}$ be fixed realizations of independent and identically distributed beta distributions with parameters $(100, 1)$ for $j = 1, 2$ and $(2, 7)$ for $j = 3, 4$. Thus,

$$h = (\sqrt{l_0}, \sqrt{l_1}, 2\sqrt{l_2}, \ldots, m\sqrt{l_m})^T,$$

with $m = 1 + \min\{\min\{t : \|h\| \geq 1 - 10^{-15}\}, n - 1\}$, $l_0 = \beta_0$ and $l_t = \beta_t \cdot (1 - \sum_{s=0}^{t-1} l_s)$ for $t = 1, \ldots m$. Note the abuse of notation w.r.t. Subsection 3.2.

The projected process

For each $i \in \{1, \ldots, k\}$ there are four projections of x, they are $y^{(i,j)} = (y_1^{(i,j)}, \ldots, y_n^{(i,j)})$, $j = 1, \ldots, 4$, where

$$y_t^{(i,j)} = \sum_{s=0}^{\min(m,t)} h_s^{(i,j)} x_{t-s} a_s, \text{ with } t = 1, \ldots, n,$$

with $a_0 := 1$ and $a_t = t^{-2}, t \geq 1$. Then, let us apply:

1. The E-test to $y^{(i,1)}$ to obtain the p-value $p^{(i,1)}$.
2. The L-V-test to $y^{(i,2)}$ to obtain the p-value $p^{(i,2)}$.
3. The E-test to $y^{(i,3)}$ to obtain the p-value $p^{(i,3)}$.
4. The L-V-test to $y^{(i,4)}$ to obtain the p-value $p^{(i,4)}$.

Note that for the E-test and the L-V-test we mean the improved versions of Epps (1987) and Lobato and Velasco (2004) developed in this paper.

The E-test

Using the notation in Section 2, the statistic $nQ_n(\mu_n, \gamma_n, (\xi_1/\sqrt{\hat{\gamma}}, \xi_2/\sqrt{\hat{\gamma}}))$ follows a chi-squared with 2 degrees of freedom, where, with slight abuse of notation, ξ_j is the realization of the absolute value of a $N(0, j)$, $j = 1, 2$.

The L-V-test

Taking into account the notation in Section 2, the statistics is $n\hat{\mu}_3^2/(6|\hat{F}_3|) + n(\hat{\mu}_4 - 3\hat{\mu}_2^2)^2/(24|\hat{F}_4|)$, with $\hat{F}_k = 2\sum_{t=1}^{\lfloor\sqrt{n}\rfloor} \hat{\gamma}(t)(\hat{\gamma}(t) + \hat{\gamma}(\lfloor\sqrt{n}\rfloor + 1 - t))^{k-1} + \hat{\gamma}^k$. This statistics follows a chi-squared with 2 degrees of freedom. Thus, the computation of this test is quite straightforward.

The p-value

Finally, we have $p^{(i,j)}$, $i = 1, \ldots, k$, $j = 1, \ldots, 4$ that we combine using the FDR to obtain a p-value for the global procedure, p_0. For that we first order the p-values increasingly $p_{(1)} \leq \ldots \leq p_{(4k)}$. Thus,

$$p_0 = 4k \sum_{i=1}^{4k} i^{-1} \min_{i=1,\ldots,4k} p_{(i)}/i.$$

5. An empirical study

We analyze two real data sets: the Canadian lynx and the Wolfer sunspot, which were previously found to be non-Gaussian (Rao and Gabr, 1980; Epps, 1987).

The Canadian lynx data consists in the annual record of the number of lynxes trapped in the Mackenzie River district of the North-West Canada for the period from 1821 to 1934 while the Wolfer sunspot data consists in the annual record of the sunspot activity in the period from 1700 to 1960. We have applied the ${}_1$RP-test to both, with the p-values displayed in Table 8 together with those obtained in previous studies. The similarity between all of them is quite clear.

Table 8: p-values for the $_k$RP-test, E-test and RG-test Rao and Gabr (1980) tests.

Data set	$_1$RP-test	E-test	RG-test
lynx	1.029×10^{-4}	1.402×10^{-5}	1.084×10^{-4}
sunspot	1.314×10^{-6}	7.356×10^{-6}	2.818×10^{-4}

6. Discussion

In this paper, we introduce the $_k$RP-test, intended to check the Gaussianity of stationary processes. Given a sample, this test is based on a three-step procedure. First, a vector $\mathbf{h}$ must be drawn at random in a suitable Hilbert space. Then, the sample is sequentially projected on the one-dimensional space spanned by $\mathbf{h}$. Finally, we take advantage of the fact that, with probability one, if the marginal of the projected process is Gaussian, then the initial process is Gaussian. Therefore, all that is needed is to use a test that can check the Gaussianity of the marginal of a stationary process, such us, for instance the E-test or LV-test or a combination of these.

From a theoretical point of view, just one random projection is enough to carry out the test. However, a way to improve the power is to consider a larger number of projections. These projections are computed using two different beta distributions. The obtained p-values are mixed together using the FDR, which, in spite of being slightly conservative, gives reasonable results.

The comparison of the $_k$RP-test with the E and LV-tests (as well as with a combination of these) in situations where the marginal is not Gaussian is acceptable and there are even cases in which the $_k$RP-test is clearly better. Moreover, the $_k$RP-test is able to detect alternatives with the Gaussian marginal, while the other tests are not designed to perform this task.

Acknowledgements

We warmly thank both the associate editor and the anonymous referees for their constructive comments and suggestions that have allowed to improve the paper. We are grateful to thank Jean-Noel Kien for some last minute help. J.A. Cuesta- Albertos and A. Nieto-Reyes have been partially supported by the Spanish Ministery of Science and Innovation, grant MTM2008-0607-C02-02 and MTM2011-28657-C02-02.

Appendix

Proof of Proposition 3.1. Let $\alpha = \alpha_1/(\alpha_1 + \alpha_2)$ be the mean of the $\beta(\alpha_1, \alpha_2)$ distribution. Let us begin by proving that

$$\mathbb{E}[l_n] = \alpha(1-\alpha)^n, \text{ for every } n \in \mathbb{N}^*. \tag{8}$$

This, obviously holds for $n = 0$. Let us assume that it is satisfied for $n \in \mathbb{N}^*$ and let us show that it holds for $n+1$. By construction, if B is a r.v. with distribution $\beta(\alpha_1, \alpha_2)$, then

$$\mathbb{E}[l_{n+1}] = \mathbb{E}[B](1 - \sum_{i=0}^{n} \mathbb{E}[l_i]) = \alpha(1 - \sum_{i=0}^{n} \alpha(1-\alpha)^i) = \alpha(1-\alpha)^{n+1},$$

where the last equality comes from the application of the formula giving the sum of n numbers in a geometric progression.

We have that $\|\mathbf{H}\| = \sum_{i=0}^{\infty} H_i^2 a_i = \sum_{i=0}^{\infty} l_i \leq 1$ by construction; and the proposition follows from here because (*8*) gives that $\mathbb{E}\left[\|\mathbf{H}\|\right] = \sum_{i=0}^{\infty} \alpha(1-\alpha)^i = 1$. $\square$

Lemma A.1. *Let $\boldsymbol{X}$ be an ergodic and stationary process such that $\sum_{t=0}^{\infty} |\gamma_X(t)| < \infty$. If we select $\boldsymbol{H}$ as in Section 3. Then,*

1. $\sum_{i=0}^{\infty} H_i a_i < \infty$ *almost surely.*
2. $L := \sum_{i,j=0}^{\infty} H_i H_j a_i a_j |X_{-i} - \mu_X||X_{t-j} - \mu_X|$ *is integrable conditionally to* $\mathbf{H}$ *a.s.*

Proof. Item *1* is straightforward since the Cauchy-Schwartz inequality gives

$$\sum_{i=0}^{\infty} H_i a_i \leq \left(\sum_{i=0}^{\infty} l_i\right)^{1/2} \left(1 + \sum_{i=1}^{\infty} 1/i^2\right)^{1/2} = \left(1 + \sum_{i=1}^{\infty} 1/i^2\right)^{1/2} < \infty \text{ a. s.},$$

where the last equality comes from Proposition 3.1.

To prove *2*, let $\mathbf{h} = (h_0, h_1, \ldots)$ be a fixed realization of $\mathbf{H}$. We have

$$\begin{aligned}
\mathbb{E}[L|\mathbf{h}] &= \sum_{i,j=0}^{\infty} h_i h_j a_i a_j \mathbb{E}[|X_{-i} - \mu_X||X_{t-j} - \mu_X|] \\
&\leq \sum_{i,j=0}^{\infty} h_i h_j a_i a_j (\mathbb{E}[(X_{-i} - \mu_X)^2])^{1/2} (\mathbb{E}[(X_{t-j} - \mu_X)^2])^{1/2} \\
&= \gamma_X \left(\sum_{i}^{\infty} h_i a_i\right)^2
\end{aligned}$$

and so, L is conditionally integrable due to *1* and that $\gamma_X \leq \sum_{t=0}^{\infty} |\gamma_X(t)| < \infty$. $\square$

Proof of Proposition 3.2. $(X_t)_{t\in\mathbb{Z}}$ is a stationary ergodic process. Thus, conditionally on $\mathbf{h}$, $(Y_t)_{t\in\mathbb{Z}}$ is also a stationary ergodic process (Doob, 1953, p. 458). From the definition of the process $\mathbf{Y}$ we have

$$\mathbb{E}[|Y_0| \ |\mathbf{h}] \leq \mathbb{E}\left[\sum_{i=0}^{\infty} h_i a_i |X_{-i}| \middle| \mathbf{h}\right] = \mathbb{E}[|X_0|] \sum_{i=0}^{\infty} h_i a_i < \infty, \quad \text{a.s.}$$

because of *1* in Lemma A.1. From *2* in Lemma A.1, we may deduce that

$$\gamma_{Y|\mathbf{h}}(t) = \mathbb{E}[\sum_{i,j=0}^{\infty} h_i h_j a_i a_j (X_{-i} - \mu_X)(X_{t-j} - \mu_X) \ |\mathbf{h}]$$

exists. Thus, using the dominated convergence theorem, we obtain that

$$\gamma_{Y|\mathbf{h}}(t) = \sum_{i,j=0}^{\infty} h_i h_j a_i a_j \gamma_X(t-j+i),$$

and

$$\sum_{t=0}^{\infty} t^{\zeta} |\gamma_{Y|\mathbf{h}}(t)| \leq \sum_{i,j=0}^{\infty} h_i h_j a_i a_j \sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t-j+i)|.$$

Obviously,

$$\sum_{i,j=0}^{\infty} h_i h_j a_i a_j \sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t-j+i)| =: T_1 + T_2 + T_3,$$

where

$$\begin{aligned}
T_1 &= \sum_{j=0}^{\infty} h_j a_j \sum_{i=j}^{\infty} h_i a_i \sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t-j+i)|, \\
T_2 &= \sum_{j=0}^{\infty} h_j a_j \sum_{i=0}^{j-1} h_i a_i \sum_{t=2j+1}^{\infty} t^{\zeta} |\gamma_X(t-j+i)| \text{ and} \\
T_3 &= \sum_{j=0}^{\infty} h_j a_j \sum_{i=0}^{j-1} h_i a_i \sum_{t=0}^{2j} t^{\zeta} |\gamma_X(t-j+i)|.
\end{aligned}$$

If $i \geq j$, as $t \in \mathbb{N}^*$ and $\zeta \geq 0$, we have $t^{\zeta} \leq (t-j+i)^{\zeta}$. Thus,

$$T_1 \leq \sum_{j=0}^{\infty} h_j a_j \sum_{i=j}^{\infty} h_i a_i \sum_{t=0}^{\infty} (t-j+i)^{\zeta} |\gamma_X(t-j+i)| \leq \sum_{j=0}^{\infty} h_j a_j \sum_{i=j}^{\infty} h_i a_i \sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t)|,$$

because $t-j+i \geq t$. Then, $\sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t)| < \infty$ and so, *1* in Lemma A.1 implies $T_1 < \infty$. Concerning T_2, as $j > i$ and $t-j+i > 0$, we can apply the $c_{\zeta}-$inequality (Loève, 1977, p. 157) to $t = (t-j+i) + (j-i)$ to obtain that there exists $c_{\zeta} > 0$ such that $t^{\zeta} \leq c_{\zeta}(t-j+i)^{\zeta} + c_{\zeta}(j-i)^{\zeta} \leq 2c_{\zeta}(t-j+i)^{\zeta}$. Thus,

$$\begin{aligned}
T_2 &\leq 2c_{\zeta} \sum_{j=0}^{\infty} h_j a_j \sum_{i=0}^{j-1} h_i a_i \sum_{t=2j+1}^{\infty} (t-j+i)^{\zeta} |\gamma_X(t-j+i)| \\
&\leq 2c_{\zeta} \sum_{j=0}^{\infty} h_j a_j \sum_{i=0}^{j-1} h_i a_i \sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t)|.
\end{aligned}$$

Using the same tricks as for T_1, we obtain that $T_2 < \infty$. For T_3, the fact that $\sum_{t=0}^{\infty} t^{\zeta} |\gamma_X(t)| < \infty$ implies that there exists $R > 0$ such that $|\gamma_X(t)| \le R$ for all $t \in \mathbb{Z}$. Therefore,

$$T_3 \le R(\sum_{i=0}^{\infty} h_i a_i) \sum_{j=0}^{\infty} h_j a_j (2j)^{\zeta} (2j+1) =: R(\sum_{i=0}^{\infty} h_i a_i) T_3^*.$$

By *1* in Lemma A.1, in order to show that $T_3 < \infty$, we only need to prove that $T_3^* < \infty$. Furthermore, applying Jensen's inequality and *8* in Proposition 3.1, we have that

$$\mathbb{E}[T_3^*] \le \sum_{j=0}^{\infty} a_j^{1/2} (2j)^{\zeta} (2j+1) \alpha^{1/2} (1-\alpha)^{j/2}.$$

This series converges because $\alpha \in (0,1)$; so that, T_3^* is almost surely finite and the proof ends. □

Proof of Lemma 3.3. It is straightforward from the proof of Lemma 4.1 in Epps (1987) but substituting (2) and Gebelein's inequality for Gaussian processes by (4). □

Proof of Lemma 3.4. Proceeding as in Epps (1987) we have that $\Theta_0(\lambda)$ is contained or equal to

$$\{(\nu, \gamma_Y) : \nu\lambda_1 = \mu_Y \lambda_1 + 2\pi k \text{ and } \nu\lambda_2 = \mu_Y \lambda_2 + 2\pi k^*, \text{ with } k, k^* \in \mathbb{Z}\}.$$

In order to get that the cardinal of $\Theta_0(\lambda)$ is larger than one, we need λ_2 to be equal to a rational number times λ_1. However, this happens with probability zero and so, with probability one $\Theta_0(\lambda) \subseteq \{(\mu_Y, \gamma_Y)\}$. Thus, the lemma follows directly. □

In the next lemma, we use analytic ch.f.'s. A precise definition and some properties appear in Laha and Rohatgi (1979).

Lemma A.2. *Let P be a Borel probability measure defined on $\mathbb{R}$. Assume that P is absolutely continuous with respect to the Lebesgue measure. Let Y be a r.v. such that the modulus of its ch.f. $|\Phi_Y|$ is analytic. Then, Y is Gaussian if, and only if,*

$$\exists m \in \mathbb{R},\ \exists s \in \mathbb{R}^+ \ s.t.\ P(\{y \in \mathbb{R} : |\Phi_Y(y)| = |\Phi_{m,s}(y)|\}) > 0. \tag{9}$$

Proof. The necessary part is obvious. Let us prove the sufficiency. As Y satisfies (9), and P is absolutely continuous, we have that $C := \{y \in \mathbb{R} : |\Phi_Y(y)| = |\Phi_{m,s}(y)|\}$ is infinite and not denumerable and so, it has at least one accumulation point.

Furthermore, the function $y \to |\Phi_Y(y) - \Phi_{m,s}(y)|$ is analytic, and vanishes on C. Thus, this function has one non-isolated zero but the only analytical function with at least one non-isolated zero is the null function (Rudin, 1966). The proof ends because a distribution is Gaussian if and only if the modulus of the ch.f. coincides with the modulus of a Gaussian ch.f. □

Proof of Theorem 3.6. If $\mathbf{X}$ is Gaussian, then $\mathbf{Y}$ is Gaussian and Proposition 3.2 gives that $\mathbf{Y}$ satisfies the assumptions of Corollary 3.5. This fulfills the necessary part.

Let us prove the sufficient part. As $(P_\lambda \otimes P_{\mathbf{H}})[B] > 0$, there exist $\mathbf{h}$ and λ with $\lambda_1 \neq 0$ and $\lambda_2 \neq 0$ such that $nQ_n(\mu_n, \gamma_n, \lambda)$ converges in law to a non-degenerate distribution. We assume without loss of generality that $\Phi_{Y_0}(\lambda_1) \neq 0$ and $\Phi_{Y_0}(\lambda_2) \neq 0$. As Φ_{Y_0} is an analytic ch.f., it has only isolated zeros. Thus $Q_n(\mu_n, \gamma_n, \lambda)$ converges in probability to zero. By Lemma 3.3, $\hat{f}(0, \lambda)$ converges to $f_{\mathbf{Y}|\mathbf{h}}(0, (\mu_{Y|\mathbf{h}}, \gamma_{Y|\mathbf{h}}), \lambda)$. Thus, $\lim_n G_n^+$ is positive definite as it is the inverse of $2\pi f_{\mathbf{Y}|\mathbf{h}}(0, (\mu_{Y|\mathbf{h}}, \gamma_{Y|\mathbf{h}}), \lambda)$. This and the definition of $Q_n(\cdot, \cdot, \cdot)$ (see Section 2.1) gives

$$\hat{g}(\lambda) - g_{\mu_n, \gamma_n}(\lambda) \to_{\text{c.p.}} 0. \tag{10}$$

Since $\mathbf{X}$ is an ergodic stationary process, we have that $(g(Y_t, \lambda))_{t \in \mathbb{Z}}$ is also an ergodic stationary process (Doob, 1953, p. 458). Thus, as $\mathbb{E}|\cos(\lambda_i Y_0)| < \infty$ and $\mathbb{E}|\sin(\lambda_i Y_0)| < \infty$ for all $i = 1, ..., N$, we may conclude by Theorem 2 in Hannan (1970, chap IV) that $\hat{g}(\lambda) \to_{\text{c.p.}} \mathbb{E}[g(Y_0, \lambda)]$. This and (10) gives that $\Phi_{\mu_n, \gamma_n}(\lambda_i)$ converges in probability to $\Phi_{Y_0}(\lambda_i)$, $i = 1, ..., N$.

Now we show that this implies that the sequence $\{\gamma_n\}_{n \in \mathbb{N}}$ converges. We have that

$$\lim_{n \to \infty} |\Phi_{\mu_n, \gamma_n}(\lambda_1)| = \lim_{n \to \infty} \exp(-\lambda_1^2 \gamma_n / 2) = |\Phi_{Y_0}(\lambda_1)|,$$

in probability. Now, since $\lambda_1 \neq 0$ and $\Phi_{Y_0}(\lambda_1) \neq 0$, this implies that there exists $s \in \mathbb{R}$ such that $s = \lim_{n \to \infty} \gamma_n$ in probability.

Analogously,

$$|\Phi_{Y_0}(\lambda_2)| = \lim_{n \to \infty} \exp(-\lambda_2^2 \gamma_n / 2) = \exp(-\lambda_2^2 s / 2).$$

As λ_2 was drawn independently of λ_1 with an absolute continuous distribution and as $|\Phi_{Y_0}|$ is analytic, Lemma A.2 gives that Y_0 is Gaussian. Then, by Theorem 1.1, the process $\mathbf{X}$ is Gaussian. □

Proof of Theorem 3.9. Using Proposition 3.2 for $\zeta = 0$ we get that $(Y_t)_{t \in \mathbb{Z}}$ is an ergodic and stationary process with $\sum_{t=0}^{\infty} |\gamma(t)| < \infty$. If $(X_t)_{t \in \mathbb{Z}}$ is Gaussian, the process $(Y_t)_{t \in \mathbb{Z}}$ is also Gaussian. Thus, the assumptions of the first part of Theorem 2.3 hold for the process $(Y_t)_{t \in \mathbb{Z}}$ and so $\tilde{G}_Y \longrightarrow_d \chi_2^2$.

As $\mathbf{Y}$ is Gaussian, we have that $F_k > 0$ for $k = 3, 4$ (Gasser, 1975, p. 568). Repeating the proof of Lemma 1 in Lobato and Velasco (2004), we have that $\lim_{n \to \infty} \hat{F}_k = F_k$ and so, we may conclude that $\lim_{n \to \infty} G_Y = \lim_{n \to \infty} \tilde{G}_Y$. This shows *1*.

Let us now prove statement *2*. First, let us show that $\mathbb{E}[|Y|^k | \mathbf{h}] < \infty$, almost surely, for $k = 1, ..., 4$. By Hölder's inequality, we have

$$|Y_0| \leq \left(\sum_{i=0}^{\infty} a_i \right)^{1/2} \left(\sum_{i=0}^{\infty} h_i^2 a_i X_{-i}^2 \right)^{1/2}$$

and, as by Proposition 3.1 $\sum_{i=0}^{\infty} h_i^2 a_i = 1$, almost surely, we can apply Jensen's inequality. We obtain that

$$Y_0^4 \leq \left(\sum_{i=0}^{\infty} a_i\right)^2 \left(\sum_{i=0}^{\infty} h_i^2 a_i X_{-i}^4\right), \text{ almost surely.}$$

Thus, $\mathbb{E}[|Y_0|^k|\mathbf{h}] < \infty$, almost surely, for $k = 1, ..., 4$. By Doob (1953, p. 458), we have that $(Y_t^k)_{t\in\mathbb{Z}}$ is stationary and ergodic, for all $k = 1, ..., 4$. Thus, Theorem 2 in Hannan (1970, chap IV) implies

$$\lim_{n\to\infty} \hat{\mu}_k = \mu_k, \text{ for almost every } \mathbf{h} \text{ and } k = 2, 3, 4. \tag{11}$$

Further, let us prove that $\lim_{n\to\infty} |\hat{F}_k| < \infty$ for almost every $\mathbf{h}$ and $k = 3, 4$. As

$$\hat{F}_k = \hat{\gamma}_Y^k + 2\sum_{t=1}^{\tau_n}\sum_{j=0}^{k-1}\binom{k-1}{j}\hat{\gamma}_Y(t)^{k-j}\hat{\gamma}_Y(\tau_n+1-t)^j$$

and $|a^{k-j}b^j| \leq |a|^k + |b|^k$, with $k \in \mathbb{N}$, $j \in \mathbb{N}$ and $j < k$, we have

$$|\hat{F}_k| \leq |\hat{\gamma}_Y|^k + 2^k \sum_{t=1}^{\tau_n}(|\hat{\gamma}_Y(t)|^k + |\hat{\gamma}_Y(\tau_n+1-t)|^k)$$

and so, $|\hat{F}_k| \leq 2^{k+1}(\sum_{t=0}^{\tau_n}|\hat{\gamma}_Y(t)|)^k$. Let us prove that $\lim_{n\to\infty}\sum_{t=0}^{\tau_n}|\hat{\gamma}_Y(t)| < \infty$. Note that as $\mathbb{E}[X_0^4] < \infty$, we also have $\infty > \mathbb{E}[(X_0 - \mu_X)^4]$ which is equal to

$$\sum_{j_1,\ldots,j_4=1}^{\infty}\prod_{r=1}^{4}k(j_r)E\left[\prod_{r=1}^{4}\epsilon_{n-j_r}\right] = \mathbb{E}[\epsilon_1^4]\sum_{j=1}^{\infty}k(j)^4 + \mathbb{E}[\epsilon_1^2]^2\sum_{i,j=1,i\neq j}^{\infty}k(i)^2k(j)^2.$$

Indeed, (ϵ_n) are i.i.d. r.v.'s with $\mathbb{E}[\epsilon_1] = 0$. Thus $\mathbb{E}[\epsilon_1^4] < \infty$. By Kavalieris (2008), we obtain

$$|\sum_{t=0}^{\tau_n}(|\hat{\gamma}_X(t)| - |\gamma_X(t)|)| \leq (\tau_n+1)o(n^{2/\alpha-1}) = o(1).$$

Therefore, $\lim_{n\to\infty}\sum_{t=0}^{\tau_n}|\hat{\gamma}_X(t)| < \infty$. Proceeding similarly as in the proof of Proposition 3.2, we get $\lim_{n\to\infty}\sum_{t=0}^{\tau_n}|\hat{\gamma}_Y(t)| < \infty$ and so, $\lim_{n\to\infty}|\hat{F}_k| < \infty$ for $k = 3, 4$ a.s. Using (11) we conclude that *2* holds. □

References

del Barrio, E., Cuesta-Albertos, J.A., Matrán, C., Rodríguez-Rodríguez, J., 1999. Tests of goodness of fit based on the L2-Wasserstein distance. Annals of Statistics 27, 1230–1239.

Benjamini, Y., Yekutieli, D., 2001. The control of the false discovery rate in multiple testing under dependency. Annals of Statistics 29(4), 1165–1188.

Cabaña, A., 1996. Transformations of the empirical measure and Kolmogorov-Smirnov tests. Annals of Statistics. 24(5), 2020–2035.

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 and Data Analysis 51, 4814–4831.

Cuesta-Albertos, J.A., Matrán, C., 1991. On the asymptotic behavior of sums of pairwise independent random variables. Statistics and Probability Letters 11(3), 201–210.

Cuesta-Albertos, J.A., Nieto-Reyes, A., 2008. The Random Tukey Depth. Computational Statistics and Data Analysis 52(11), 4979–4988.

Csorgo, S., 1986. Testing for normality in arbitrary dimension. Annals of Statistics 14(2), 708–723.

D'Agostino, R.B., 1971. An omnibus test of normality for moderate and large size samples. Biometrika 58(2), 341–348.

D'Agostino, R.B., Pearson, E.S., 1973. Tests for departure from normality. Empirical results for the distributions of b2 and b1. Biometrika 60(3), 613–622.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-fit techniques. Marcel Dekker.

Doob, J.L., 1953. Stochastic Processes. John Wiley & Sons.

Epps, T. W., 1987. Testing that a stationary time series is Gaussian. Annals of Statistics 15(4), 1683–1698.

Fang, K.T., Li, R.Z., Liang, J.J., 1998. A multivariate version of Ghosh's $T3$-plot to detect non-multinormality. Computational Statistics and Data Analysis. 28(4) 371–386.

Gasser, T., 1975. Goodness-of-fit tests for correlated data. Biometrika 62(3), 563–57.

Ghosh, S., 2013. Normality testing for a long-memory sequence using the empirical moment generating function. Journal of Statistical Planning and Inference 143(5), 944–954.

Ghosh, S., Ruymgaart, F.H., 1992. Applications of empirical characteristic functions in some multivariate problems. Canadian Journal of Statistics 20(4), 429–440.

Ghoudi, K. and Rémillard, B., 2013. Comparison of specification tests for GARCH models. Computational Statistics and Data Analysis. http://dx.doi.org/10.1016/j.csda.2013.03.009

Hannan, E.J., 1970. Multiple Time Series. John Wiley & Sons.

Ibragimov, I.A., Rozanov, Y. A., 1978. Gaussian Random Processes. Springer.

Kavalieris, L., 2008. Uniform convergence of autocovariances. Statistics and Probability Letters 78(6), 830–838.

Laha, R.G., Rohatgi, V.K., 1979. Probability Theory. John Wiley & Sons.

Liang, J.J., Ng, K.W., 2009. A multivariate normal plot to detect nonnormality. Journal of Computational and Graphical Statistics 18(1), 52–72.

Liu, S. and Maharaj, E.A., 2013. A hypothesis test using bias-adjusted AR estimators for classifying time series in small samples. Computational Statistics and Data Analysis 60(0), 32–49.

Lobato, I.N., Velasco, C., 2004. A simple test of normality for time series. Econometric Theory 20(4), 671–689.

Loève, M., 1977. Probability Theory I. Springer.

Moulines, E., Choukri, K., 1996. Time-domain procedures for testing that a stationary time-series is Gaussian. IEEE Transactions on Signal Processing 44(8), 2010–2025.

Pitman, J., 2006. Combinatorial Stochastic Processes. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour. Springer.

Rao, T.S., Gabr, M.M., 1980. A test for linearity of stationary time series. Journal of Time Series Analysis 1, 145–158.

Rudin, W., 1966. Real and Complex Analysis. Mc Graw-Hill.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples). Biometrika 52, 591–611.