A random-projection based Gaussianity test for stationary process

Juan A. Cuesta-Albertos¹, F. Gamboa² and A. Nieto-Reyes¹

¹ Universidad de Cantabria,
Facultad de Ciencias, Avda. los Castros s.n., 39005 Santander, Spain
(e-mail: juan.cuesta@unican.es and alicia.nieto@unican.es)

² Université Paul Sabatier,
IMT Universit Paul Sabatier, F-31062 Toulouse cedex 9, France
(e-mail: fabrice.gamboa@math.univ-toulouse.fr)

Abstract. In this talk we present a procedure to test if a stationary process is Gaussian. The observation consists in a finite sample of a path of the process. The test is based on the fact (established in Cuesta-Albertos et al. (2007)) that, almost surely, a distribution is Gaussian iff a randomly chosen one-dimensional projection is Gaussian, thus transforming the problem of testing the infinite-dimensional Gaussianity in testing the Gaussianity of a one-dimensional distribution.

Most of known tests only check if the one-dimensional marginals of the process under consideration are Gaussian, thus being at the nominal power against those non-Gaussian alternatives with Gaussian one-dimensional marginals. However, the procedure that we present here is consistent against every alternative (under some regularity conditions).

The talk will also include some simulations and the analysis of some real data sets to compare our procedure with some other well-known tests proposed in the literature.

Keywords. Gaussianity Test, Strictly Stationary Random Process, Random Projection, Consistent Test.

1 Introduction

The problem we deal with is the following: We have a sample \( \{X_1, \ldots, X_n\} \) taken from a path of a stationary random process and we are interested into know if the distribution of this process is Gaussian or not. Many tests have been proposed to solve this problem (see, for instance, Cuesta-Albertos et al. (2009)). However, most of the proposed procedures take advantage from the fact that, since the process is stationary, then all the r.v.’s \( X_1, \ldots, X_n \) are i.d.’s. Therefore, if the process is Gaussian, this one-dimensional distribution has to be Normal. Then, most of proposed tests only test whether this one-dimensional distribution is Gaussian, or even, when this distribution enjoys certain property of the Gaussian distributions. For instance, the skewness-kurtosis tests (SK-tests) check if the data show evidence of skewness or kurtosis; and the tests based on the characteristic function (CF-tests) check if the characteristic function of the marginal has the right expression in a finite set of points. Obviously, the power of those tests against non-Gaussian distributions with Gaussian one-dimensional marginals is the level of the test and the same happens, for instance, if we employ a SK-test and we have marginals non-Gaussian, with zero skewness and kurtosis.

The objective of this talk is to present a test which does not suffer from those problems. This test will be consistent against every non-Gaussian alternative which satisfies some regularity conditions (see Cuesta-Albertos (2009)).

Spanish Ministerio de Ciencia y Tecnología partially supported first and third authors supported by the grant MTM2008-0607-C02-02 and first author by the grant PR2009-0355
The proposed test is based on Theorem 3.6 in Cuesta-Albertos et al. (2007) which, roughly speaking, states that a infinite-dimensional distribution is Gaussian iff almost every randomly chosen one-dimensional projection is Gaussian (a more precise statement appears in Proposition 1), thus transforming the problem of testing the infinite-dimensional Gaussianity in testing a one-dimensional Gaussianity. This, in turn, can be done using some of the above mentioned tests. In particular, here we have chosen the SK-test proposed in Lobato and Velasco (2004) and the well-known CF-test introduced in Epps (1987).

When combining the random projections with those tests, by the reasons we gave above, we do not obtain a procedure consistent against every possible alternative. However, as shown in Cuesta-Albertos et al. (2009), if the points involved in the CF-test are chosen at random, then we obtain a test which is consistent against every alternative under the regularity conditions stated in that paper.

2 The procedure

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a stationary process. First, we need a Hilbert space in which to include $X$ and a probability distribution on this space to choose the projections. To this, 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

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

We construct a distribution on $l^2$ using the so-called stick breaking method: let $\alpha_1, \alpha_2 > 0$ and let $\eta$ be the probability distribution which selects a random point in $l^2$ as follows:

- $h_0 \in [0, 1]$ is chosen with the beta distribution of parameters $\alpha_1$ and $\alpha_2$.
- Given $n \geq 1$, $h_n \in [0, 1 - \sum_{i=0}^{n-1} h_i]$ is chosen with the beta distribution of parameters $\alpha_1$ and $\alpha_2$, times $(1 - \sum_{i=0}^{n-1} h_i)$.

First leg of our procedure is the following result which is deduced trivially from Theorem 3.6 in Cuesta-Albertos et al. (2007).

**Proposition 1.** If $Z$ is an $l^2$-valued random element then

$$\eta \{ h \in l^2 : \text{ the distribution of } (Z, h) \text{ is Gaussian} \} \in \{0, 1\},$$

and, its value is 0 iff $Z$ is not Gaussian.

In other words, if we want to see if the distribution of $Z$ is Gaussian, then we only need to select at random a point $h \in l^2$ using $\eta$ and check if the real-valued random variable $(Z, h)$ is Gaussian. We will obtain the right answer with probability one.

In order to implement this result in practice, notice that $X$ is Gaussian iff $(X_1, \ldots, X_t)^T$ is a Gaussian vector for every $t$. As $X$ is stationary, this is equivalent to the Gaussianity of the process $X^{(t)} := (X_j)_{j \leq t}$, for any $t \in \mathbb{Z}$.
A Gaussianity test for stationary process

It is easy to see that if \(X\) is a stationary process and if the variance of \(X_t\) is finite, then, almost surely, \(X^{(t)} \in l^2\). Now, select \(h \in l^2\) using \(\eta\), independently from \(X\), and for each \(t \in \mathbb{Z}\) compute the real r.v.

\[
Y^h_t = \sum_{i=0}^{\infty} h_i X_{t-i} a_i.
\]  

(1)

Taking into account Proposition 1 and that \(X\) is stationary, we obtain the following result:

**Theorem 1.** It happens that \(X\) is Gaussian iff

\[
\eta\{h \in l^2 : \text{the distribution of } Y^h_t \text{ is Gaussian for every } t \in \mathbb{Z}\} = 1.
\]

From Theorem 1 we have that if we choose \(h \in l^2\) using \(\eta\), independently from \(X\), then a test of Gaussianity at level \(\alpha\) applied to the marginals of \(Y^h := (Y^h_t)_{t \in \mathbb{Z}}\) is also a Gaussianity test for \(X\) at the same level. Moreover, since \(Y^h\) is stationary we can test the Gaussianity of their marginals using, for instance, the SK or CF-tests we mentioned above.

However, those tests can not be applied automatically. There are some required conditions (as, for instance, ergodicity), but, it can be proved that the process \(Y^h\) inherits those conditions from the process \(X\) (see Cuesta-Albertos (2009)).

### 3 The procedure in practice

At a first view it could look suspicious that a test based on just one random projection could have enough power as to be reliable in practice. This is true. Then, we propose to choose not only one random projection but a finite number of them, \(h_1, \ldots, h_k\), then, apply a Gaussianity test to each of the processes \(Y^{h_1}, \ldots, Y^{h_k}\) that we obtain and, finally, to employ a procedure to combine multiple test to make the final decision. Our selection here is the False Discovery Rate as proposed in Benjamini and Yekutiel (2001).

In this talk we will present some simulations and analyze two real data sets to show that the proposed procedure attains similar power against many alternatives as the SK and CF-tests do when applied directly to the \(X\) initial process. Moreover, the new test clearly beats SK and CF-tests in the case of non-Gaussian distributions with Gaussian marginals.

**References**


