Journal of Multivariate Analysis 139 (2015) 147–165

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Data driven smooth test of comparison for dependent sequences

P. Doukhan [a], D. Pommeret [b], L. Reboul [b,*]

[a] *University Cergy-Pontoise, Mathematics, office E 5.28, UCP site Saint-Martin, 2 Bd. Adolphe Chauvin, 95000 Cergy-Pontoise, France*
[b] *Aix-Marseille University, Campus de Luminy, 13288 Marseille cedex 9, France*

ARTICLE INFO

Article history:
Received 20 November 2013
Available online 10 March 2015

AMS subject classifications:
62G10
62E20
62M07
62M10
62P05

Keywords:
Smooth test
Schwarz's rule
Strictly stationary process
m-dependence
α-mixing
θ-dependence
Long memory
Gaussian processes

ABSTRACT

In this paper we propose a smooth test of comparison for the marginal distributions of strictly stationary dependent bivariate sequences. We first state a general test procedure and several cases of dependence are then investigated. The test is applied to both simulated data and real datasets.

1. Introduction

The comparison of two or more time series is a problem of great interest in many practical situations. In addition to the assumption of identically drawn observations, independence properties both within and between series are generally assumed. These assumptions are too restrictive in practice. For instance, most of economic or financial series exhibit dependence and a drift in time. However, it is often realistic to consider time series where time has only transitory effects, which leads to replace the independent and identically drawn assumption by the strict stationarity of the sequences. Moreover, even when a drift in time is observed, simple series transformations such as linear detrending or first differencing allow to boil down to stationarity.

In this paper, we wish to compare the marginal distributions of two possibly cross-dependent time series X and Y such that $Z = (X, Y)$ is a strictly stationary bivariate process with given within dependence structure. Our comparison problem relies in testing the nonparametric hypothesis

$$H_0 : f_X = f_Y, \tag{1}$$

* Corresponding author.
E-mail addresses: doukhan@u-cergy.fr (P. Doukhan), denys.pommeret@univ-amu.fr (D. Pommeret), laurence.reboul@univ-amu.fr (L. Reboul).

http://dx.doi.org/10.1016/j.jmva.2015.02.017

where f_X and f_Y denote the unknown marginal densities at any given time of X and Y, respectively. The omnibus alternative consists in the difference of the two marginal distributions.

This two-sample nonparametric testing problem has been widely studied when there is no dependence within and between series (see [24,17] and references therein). Most of the usual tests rely on ranks or empirical CDF (see for instance [24]). For possibly cross-dependent data, a test strategy inspired from Neyman's smooth test [29] was recently proposed by [17]. In this paper, the authors considered the series expansions of f_X and f_Y along an appropriated family of orthogonal functions. Testing (1) thus reduces to the parametric testing problem that the coefficients of the same order in both expansions are equal. In [25], a data-driven method was proposed to choose the optimal dimension of the orthogonal family. It consists of a modified Schwarz's Bayesian information criterion. Finally, they obtained a test statistic and derived its asymptotic distribution.

In the context of within-sample dependence, the development of practicable tests becomes usually more difficult. In the one sample case, [21] and recently [27] proposed goodness-of-fit tests for the marginal distribution of observations arising from an α-mixing discrete time stochastic process. [27]'s test is a data-driven version of Neyman's smooth test adapted to dependent data.

In this paper, we propose a test strategy for (1) generalizing [27,17]'s results to several dependence structures within observations, covering short and long range dependence. In so doing, we provide a wide panorama of dependence cases which can be adapted to various concrete situations. We first develop a general theory where the test statistic is defined as a normalized sum of the squared differences between the estimated coefficients in the expansions of f_X and f_Y in an appropriated basis. The number of components to keep in the sum is selected by the way of an information criterion inspired from [17,27]. For each particular dependence structure, the normalization of the test statistic is chosen in such a way that a central limit theorem obtains so that its limiting distribution under the null is derived. We also study the consistency of the test under contiguous alternatives. For the practical implementation of the test, some modifications in the definition of our test statistic and of our selection rule are considered and discussed. Our method is illustrated on simulations and real data arising from short memory financial index time series: the Dow Jones Composite Average, the NASDAQ Composite and the NYSE International 100 Index.

The rest of the paper is organized as follows. In Section 2 we develop the methodology of the test and introduce a data-driven statistic. The main theoretical results are stated in Section 3. In Section 4 we apply our general methodology to several situations of dependence. In Section 5 the practical implementation of the test procedure is discussed. Sections 6 and 7 propose a short simulation study and an application to real financial data. Finally, a conclusion is given in Section 8 and Section 9 is devoted to the proofs.

2. A Neyman's type test

2.1. Construction of the test statistic

Let $Z = (Z_t)_{t\in\mathbb{Z}}$ be a discrete time process defined on some probability space $(\Omega, \mathcal{A}, \mathbb{P})$ and taking values in $\mathbb{R}^2$. We assume throughout that Z is strictly stationary. Let ν be a given probability measure with density h with respect to some reference measure λ (Lebesgue's or counting measure for instance). We set $Z = (X, Y)$, where X and Y are possibly cross-dependent. We denote by f_X and f_Y the respective unknown marginal densities of the X_t's and Y_t's with respect to λ and assume that they belong to the space $\mathbb{L}_2(\nu)$ of square integrable functions with respect to ν.

In this setup, we wish to test (1) based on the observation of a sample $(Z_1, \dots, Z_s, \dots, Z_n)$ of Z. For that task, we consider the expansions of f_X and f_Y along a dense family $(Q_j)_{j\in\mathbb{N}}$ of orthonormal functions in $\mathbb{L}_2(\nu)$:

$$f_X = \sum_{j\geqslant 0} a_j Q_j \quad \text{and} \quad f_Y = \sum_{j\geqslant 0} b_j Q_j, \tag{2}$$

with

$$a_j = \int_{\mathbb{R}} Q_j(t) f_X(t) d\nu(t) = \mathbb{E}(\widetilde{Q}_j(X_1)), \qquad b_j = \int_{\mathbb{R}} Q_j(t) f_Y(t) d\nu(t) = \mathbb{E}(\widetilde{Q}_j(Y_1)).$$

Here, $\mathbb{E}$ denotes the expectation and we set $\widetilde{Q}_j = hQ_j$, for all $j \in \mathbb{N}$ (recall that $\nu = h\lambda$). Then, (1) can be rewritten as

$$H_0 : a_j = b_j, \quad \text{for all } j > 0 \quad \text{versus} \quad H_1 : \text{there exists } j > 0,\ a_j \neq b_j. \tag{3}$$

For testing (3), we shall consider a sequence $(N_n(k))_{k\geqslant 1}$ of Neyman's type test statistics. They are defined, up to a renormalization factor, as the sum of the squared differences between the empirical estimators of the k first coefficients in (2). More formally, setting for all $k \geqslant 1$

$$\begin{aligned} &\Phi_j(x, y) = \widetilde{Q}_j(x) - \widetilde{Q}_j(y), \qquad V_t^{(j)} = \Phi_j(X_t, Y_t), \\ &V_t(k) = (V_t^{(j)})_{1\leqslant j\leqslant k}, \qquad V^{(k)} = (V_t^{(k)})_{t\in\mathbb{Z}}, \end{aligned} \tag{4}$$