Journal of Multivariate Analysis 155 (2017) 344–368

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Change-point detection and bootstrap for Hilbert space valued random fields

Béatrice Bucchia[a,*], Martin Wendler[b]

[a] Mathematical Institute, University of Cologne, Weyertal 86–90, 50931 Köln, Germany

[b] Department of Mathematics and Computer Science, Ernst Moritz Arndt University Greifswald, Walther-Rathenau-Straße 47, 17487 Greifswald, Germany

HIGHLIGHTS

- Functional central limit theorem for Hilbert space valued random fields.
- Dependent wild bootstrap generalized to Hilbert spaces and random fields.
- Test for epidemic changes in the distribution function.

ARTICLE INFO

Article history:
Received 10 November 2015
Available online 16 January 2017

AMS 2010 subject classifications:
62H15
62E20
62M99
60G60

Keywords:
Change-point detection
Dependent wild bootstrap
FCLT for Hilbert space valued r.v.
Random fields

ABSTRACT

The problem of testing for the presence of epidemic changes in random fields is investigated. In order to be able to deal with general changes in the marginal distribution, a Cramér–von Mises type test is introduced which is based on Hilbert space theory. A functional central limit theorem for ρ-mixing Hilbert space valued random fields is proven. In order to avoid the estimation of the long-run variance and obtain critical values, Shao's dependent wild bootstrap method is adapted to this context. For this, a joint functional central limit theorem for the original and the bootstrap sample is shown. Finally, the theoretic results are supplemented by a short simulation study.

1. Introduction

1.1. Change-point tests for random fields

The focus of this paper lies in the problem of epidemic change in the mean for Hilbert space valued random fields. Given a data set, a classical problem in change-point analysis consists of testing whether all the observations have the same stochastic structure (i.e., marginal distribution) or whether there is a subset (the change-set) of the data in which the structure is different. For data corresponding to a time series, the split into different data subsets can be characterized by the points in time (the change-points) at which there is a structural break. In the epidemic change model, there are two possible change-points (the start and end of an "epidemic") and the structure of the data changes after the first change-point but

* Corresponding author.
E-mail addresses: bbucchia@math.uni-koeln.de (B. Bucchia), martin.wendler@uni-greifswald.de (M. Wendler).

http://dx.doi.org/10.1016/j.jmva.2017.01.007
0047-259X/

reverts back to its original state after the second change-point. Extended to random fields, this becomes the problem of testing for rectangular change-sets. Epidemic changes are of interest not only in medicine (see, e.g., [34]) but also, e.g., in signal detection and textile fabric quality control; see, e.g., [57]. The epidemic change-point problem was introduced by Levin and Kline [34] and has since been the subject of numerous publications; see, e.g., [1,13,28,44] and the publications listed therein. For random fields with a change in the mean, a nonparametric approach for this type of problem was considered in [29,55] for i.i.d. observations and in [7,8] for weakly dependent data. The test statistics considered in these publications are a special type of scan statistic, variants of which could – under the assumption that the distributions of the observations belong to a parametric family – also be used to test for changes in other parameters of a distribution; see, e.g., [29,36,49]. For the nonparametric problem of a change in the distributions without any prior information on the family of distributions, however, a test based on the empirical distribution function F_n with

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i \le t\}}$$

might be more useful. Equipped with the appropriate norm, one can regard these as sums of Hilbert space valued random variables, where the true distribution function of X_i is the expected value (in the Hilbert space) of $\mathbb{1}_{\{X_i \le \cdot\}}$. Therefore, the change in distribution problem can be translated into a change in mean problem for Hilbert space valued random variables.

The analysis of functional data over a spatial region is of independent interest. As a special case of spatio-temporal data, where measurements over time are taken at different locations in space, functional data may arise for instance in brain imaging or in space physics; see [22].

For weakly dependent time series of functional data, the epidemic change model was investigated by Aston and Kirch [1], who constructed test statistics based on projections on the principal components. By contrast, we aim to apply the approach used by Sharipov et al. [48], who take the full functional structure into account. To the best of our knowledge, there are no results on asymptotic change-point tests for the specific setting considered here.

A popular approach for the construction of asymptotic tests for change in mean problems are so-called CUSUM-type tests, where the mean is estimated using cumulative sums of the observations. This leads to test statistics that can be written as functionals of the partial sum process of the data. Therefore, the first aim of this paper is to give a functional central limit theorem (FCLT) for weakly dependent Hilbert space valued random fields that can be used for change-point tests. The continuous mapping theorem can then be applied to obtain the limit distribution of a CUSUM-type test statistic.

Although the central limit theorem is known for multivariate weakly dependent random fields (see [10,50]) and even random fields with values in a Hilbert space (see [51]), most of the literature on FCLTs for random fields has focused on real-valued fields. For this one-dimensional setting, numerous results have been given not only for independent observations (see [52]) but also for weakly dependent fields. For instance, the monographs by Bulinski and Shashkin [11] and Lin and Lu [35] give examples of FCLTs under conditions related to association and mixing conditions respectively. For mixing fields of real-valued random variables, Deo [16,17] proved FCLTs under φ-mixing conditions and Kim and Seok [31] extended the ideas of Deo's proofs to obtain FCLTs for ρ-mixing random fields. For i.i.d. Hilbert space valued random fields, Zemlys [55] introduced a Hölderian FCLT. The FCLT presented here can be viewed as an extension of the approach by Deo [16] first to vector-valued fields and then to Hilbert space valued fields.

After describing the bootstrap method considered here (Section 1.2), we introduce the notations used throughout this article (Section 1.3). We then present our main results in Section 2. To illustrate our theoretical findings, Section 3 reports some simulation results. Proofs of our main results are relegated to Section 4.

1.2. Bootstrap for Hilbert space valued processes

Nonparametric resampling methods like bootstrap are especially useful when dealing with stochastic processes, as the asymptotic distribution typically depends on a parameter function, which is hard to estimate. The bootstrap of the empirical distribution function has been well studied, starting with [2] in the independent case. This was extended to time series data by Naik-Nimbalkar and Rajarshi [38], Peligrad [41] and Radulović [45] using block bootstrap methods adjusted for dependence. For an overview of the block bootstrap methods, see the book by Lahiri [32]. Shao [47] introduced a different resampling method for time series: the dependent wild bootstrap, which generalizes Wu's [53] wild bootstrap. Recently, Doukhan et al. [19] extended the dependent wild bootstrap to empirical distribution functions and were able to show its validity. As seen above, the empirical distribution function can be interpreted as a function of Hilbert space valued random variables.

For more general Hilbert spaces, the bootstrap has been investigated in [15,43].

For the application to change-point detection, one needs a sequential bootstrap to mimic the behavior of the partial sum process. The consistency of the sequential multiplier bootstrap for the empirical distribution function under independence was shown in [21,25] for the sequential empirical process indexed by functions. For dependent data, Inoue [27] proposed a block multiplier bootstrap for the sequential empirical distribution function. Sharipov et al. [48] studied block bootstrap for the partial sum process of Hilbert space valued random variables.

While there is a broad range of results for different bootstrap methods in the time series setting, much less work has been done for random fields, although ideas for this can be traced back thirty years to [24]. Politis and Romano [42] studied