Econometrics and Statistics 1 (2017) 118–127

Contents lists available at ScienceDirect

Econometrics and Statistics

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

On the consistency of bootstrap methods in separable Hilbert spaces

Gil González-Rodríguez*, Ana Colubi

INDUROT/Department of Statistics and OR, Campus de Mieres. University of Oviedo, Mieres 3600, Spain

A R T I C L E I N F O

Article history:
Received 2 June 2016
Revised 14 November 2016
Accepted 14 November 2016
Available online 22 November 2016

Keywords:
Bootstrap methods
Consistency
Hilbert spaces
Functional data
Independent random elements
Functional sample mean
Functional regression models

A B S T R A C T

Hilbert spaces are frequently used in statistics as a framework to deal with general random elements, specially with functional-valued random variables. The scarcity of common parametric distribution models in this context makes it important to develop non-parametric techniques, and among them, bootstrap has already proved to be specially valuable. The aim is to establish a methodology to derive consistency results for some usual bootstrap methods when working in separable Hilbert spaces. Naive bootstrap, bootstrap with arbitrary sample size, wild bootstrap, and more generally, weighted bootstrap methods, including double bootstrap and bootstrap generated by deterministic weights with the particular case of delete $-h$ jackknife, will be proved to be consistent by applying the proposed methodology. The main results concern the bootstrapped sample mean, however since many usual statistics can be written in terms of means by considering suitable spaces, the applicability is notable. An illustration to show how to employ the approach in the context of a functional regression problem is included.

1. Introduction

The explosion of functional data analysis during the last decades underlines the need to develop statistical tools in general spaces (Hormann and Kokoszka, 2010; Ramsay and Silverman, 2005; Wang et al., 2016; Yao et al., 2005). Separable Hilbert spaces (Cardot et al., 2013; Gabrys and Kokoszka, 2007; González-Rodríguez et al., 2012) provide a natural and flexible framework. The good properties of the metric structure of separable Hilbert spaces make it intuitive to generalize classical concepts and results, such as the expectation, the covariance matrix, the linear regression models, etc. (Cardot et al., 1999; Kosorok, 2008; Ledoux and Talagrand, 1991). These general concepts and results then become applicable to functional and other complex data (Biglieri and Yao, 1989; González-Rodríguez et al., 2012; Li and Hsing, 2010).

Given the generality and the inherent high-dimensionality of this kind of spaces, there is a scarcity of parametric models that are used in practice to model Hilbert-valued random elements, although the Gaussian processes continue playing an important role via the CLT (Araujo and Giné, 1980). Thus, non-parametric tools are largely employed (Ferraty and Vieu, 2006) and, in this context, bootstrap techniques are very useful (Cuevas et al., 2006; Ferraty et al., 2010; Wang et al., 2016). In order to theoretically support the use of these techniques, their consistency should be analyzed. There exists a number of results in the literature devoted to prove the consistency of different types of bootstrap in certain general spaces for the

* Corresponding author.
E-mail addresses: gil@uniovi.es, gil.gonzalez.rodriguez@gmail.com (G. González-Rodríguez).

http://dx.doi.org/10.1016/j.ecosta.2016.11.001
2452-3062/

case of independent random elements, which is the one considered in this paper. Remarkably, Gine and Zinn (1990) proved the consistency of the naive bootstrap for the sample mean in a space of indexed empirical processes and derived the same bootstrap in separable Hilbert spaces as a corollary. Given the importance of such a space in dealing with empirical measures, other bootstrap methods have been studied, e.g. Kosorok (2003); Ledoux and Talagrand (1988); Praestgaard and Wellner (1993). The variety of techniques makes it complicated to analyze their counterpart in Hilbert spaces.

A unified framework will be provided to derive the consistency of the usual bootstrap approaches that can be employed for separable Hilbert spaces under independence. A number of bootstrap methods for the sample mean will be obtained as examples within the considered framework. Most notably, naive bootstrap, bootstrap with arbitrary sample size, wild bootstrap, and various weighted bootstrap, including double bootstrap, Bayesian bootstrap, urn model bootstrap and bootstrap generated by deterministic weights, with the particular case of delete $-h$ jackknife, can be addressed. Moreover, how to extend these results to other mean-based statistics will be illustrated. As an important example, a problem concerning a functional regression model will be considered. The novelty is based on the introduction of a linkage function that allows us to easily derive bootstrap results in Hilbert spaces from the well-developed theory of empirical processes indexed by a family of functions. This simplifies the task of the applied statisticians by translating elaborated probability results into a closer and more familiar framework.

The rest of the manuscript is organized as follows. In Section 2, the main results for separable Hilbert spaces are stated. That is, the linkage function to connect any separable Hilbert space with a particular space of functions indexed by a class of functions is introduced, and its main properties are derived. The consistency of weighted bootstrap approaches is guaranteed under general conditions. For illustrative purposes, Section 3 will be devoted to the development of a linear independence test between a functional response and any subset of scalar regressors involved in a multiple linear model. Various bootstrap procedures will be proposed, and they will be shown to be consistent by employing the results previously stated. In Section 4, technical details are gathered, namely, some notation and well-known results concerning bootstrap methods for empirical processes are recalled. This supporting theory is used in order to prove the main results stated in Section 2. The notation for the relevant case of L_2 spaces is clarified in a subsection. Some concluding remarks are collected in Section 5.

2. Bootstrap approaches in separable Hilbert spaces

Let $\mathcal{X}$ be an arbitrary space, and let $\mathcal{F}$ be any class of measurable functions $f : \mathcal{X} \to \mathbb{R}$. In the same way, let

$$l^\infty(\mathcal{F}) = \{g : \mathcal{F} \to \mathbb{R}, \|g\|_\mathcal{F} < \infty\}$$

with $\|g\|_\mathcal{F} = \sup_{f\in\mathcal{F}} |g(f)|$, which is a Banach space when the sum and the product are defined pointwise. This is a natural space for analyzing the behavior of the empirical processes indexed by the class of functions $\mathcal{F}$, which will be considered in Section 4 for technical purposes.

Let $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ be a separable Hilbert space and denote by $\|\cdot\|$ the norm associated with the inner product. Let $(\Omega, \mathcal{A}, P)$ be a probability space and let X be an $\mathcal{H}-$valued random element so that $E\|X\|^2 < \infty$. Z_X will denote a centred Gaussian $\mathcal{H}$-valued random element having the same covariance operator as X. Finally let $X_1, X_2, \ldots$ be a sequence of i.i.d. $\mathcal{H}-$valued random elements following the distribution of X.

The CLT for i.i.d. separable Hilbert-valued random elements (see, e.g., Laha and Rohatgi, 1979) ensures that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(X_i - E(X)) \to Z_X$$

weakly in $\mathcal{H}$.

In order to link the spaces $\mathcal{H}$ and $l^\infty(\mathcal{F})$, the index class of functions to be considered is the closed unit ball of the dual space of H, namely,

$$\mathcal{F} = \{f \in \mathcal{H}' \,|\, \|f\|' \le 1\}.$$

It should be noted that, by definition, $f \in \mathcal{H}'$ if and only if $f : \mathcal{H} \to \mathbb{R}$ is a continuous linear function and consequently is measurable.

The next theorem establishes a useful linkage between the spaces $\mathcal{H}$ and $l^\infty(\mathcal{F})$. This linkage will allow us to derive results for $\mathcal{H}-$valued random elements from their counterpart results stated for empirical process indexed by a class of functions. Although this paper focuses only on the consistency of bootstrap approaches, other interesting results within the well-known context of empirical processes could be adapted as well (e.g., Kosorok, 2008).

Theorem 1. *Let $D : \mathcal{H} \to l^\infty(\mathcal{F})$ be defined so that*

$$D(h)(f) = f(h)$$

for all $h \in \mathcal{H}$ and all $f \in \mathcal{F}$. Then

(a) *D is a continuous mapping.*
(b) *There exists $D^{-1} : l^\infty(\mathcal{F}) \to \mathcal{H}$ continuous with $D^{-1}(D(h)) = h\ \forall h \in \mathcal{H}$.*