

RESAMPLING METHODS FOR FUNCTIONAL DATA

Timothy MCMURRY^a and Dimitris N. POLITIS^b

^a Department of Mathematical Sciences
DePaul University, Chicago, IL, USA
tmcmurry@depaul.edu

^b Department of Mathematics
University of California, San Diego,
La Jolla, CA, USA
politis@math.ucsd.edu

Abstract

We discuss the state-of-the-art of methodological and practical developments for resampling inference techniques in situations where either the data and/or the parameters being estimated take values in a space of functions. These problems have been investigated from the perspective of non-parametric smoothing, purely functional data, and empirical processes. In all of these settings, bootstrap and subsampling are particularly attractive because traditional tools for inference are often unavailable, unwieldy, or difficult to implement.

1 Introduction

Denote by $\mathcal{X}_1, \dots, \mathcal{X}_n$ our observations that take values in an abstract space D ; for simplicity of presentation, our data will be assumed to be independent and identically distributed (i.i.d.) from probability distribution P . The goal of the analysis is typically inference for a parameter $\theta(P)$ that could also be denoted simply as θ when there is no ambiguity as to the underlying measure.

Depending on the application, the parameter θ may be finite dimensional or function valued; our primary focus will be on the latter. If θ is finite dimensional, traditional tools for establishing bootstrap consistency as well as central limit theorems often apply. In the situation where θ is functional, resampling often becomes one of the only practical tools for inference.

Our goal in the present work is to highlight the current state of resampling literature in the context of functional data and functional parameters. We use *resampling* as a generic term which encompasses inference techniques

based on recalculating the estimator of $\theta(P)$ either on subsets of the data (*subsampling*) or on samples chosen with replacement from the original data set (*bootstrap*).

The problem of resampling based inference for functional data and parameters has been explored from several perspectives, including nonparametric smoothing, bootstrap, subsampling, and empirical processes. Extensive theoretical results for functional data have been achieved in the context of subsampling and in the empirical process literature; the bootstrap is less developed in this respect, and numerical properties and finite sample performance have received only modest attention.

The remainder of this survey is laid out as follows: Section 2 contains basic background and notation which will be used throughout the chapter; Section 3 discusses results from nonparametric smoothing; Section 4 summarizes the major results in subsampling; Section 5 discusses what is known about the bootstrap; Section 6 presents a few recent real-data applications of bootstrapping with functional data; and Section 7 offers possible directions for further research and exploration.

2 Background and notation

We will consider a sequence of i.i.d. random elements $\mathcal{X}_1, \dots, \mathcal{X}_n$. The elements are assumed to be D -valued, where D is, unless otherwise noted, a metric space with probability measure P . If D is strictly finite-dimensional, then we will emphasize that by denoting the data by X_1, \dots, X_n (i.e., Roman fonts).

Bootstrap data will be denoted with asterisks, for example \mathcal{X}_1^* . Similarly, P^* and \mathbb{E}^* will refer to probability and expectation based on bootstrap randomness, conditional on the original random elements $\mathcal{X}_1, \dots, \mathcal{X}_n$.

Our primary goal is to estimate the sampling distribution of $\hat{\theta}_n$, a statistic estimating the parameter of interest $\theta(P)$. In our discussion of the functional data bootstrap in Section 5, the estimators considered will be limited to those of the form $\hat{\theta}_n = f(\bar{\mathcal{X}})$, where $\bar{\mathcal{X}}$ denotes the average of $\mathcal{X}_1, \dots, \mathcal{X}_n$, or $\hat{\theta}_n = \overline{f(\mathcal{X})}$ for a suitably smooth function f . Bootstrap consistency results in nonparametric smoothing have been developed in many specific applications; density estimation and regression are discussed in Section 3. By contrast, subsampling has been shown to apply in quite broad settings which we discuss in full generality in Section 4.

In each case the primary tool is that of weak convergence. To start with, the distribution of the root $\tau_n(\hat{\theta}_n - \theta(P))$ is shown (or assumed) to have a

well defined limit as $n \rightarrow \infty$. Here, τ_n is the rate of convergence of $\hat{\theta}_n - \theta(P)$; in regular situations it is of order $n^{1/2}$ but this is not always the case. Then, this asymptotic distribution is approximated by its resampled or subsampled analog which is basically an empirical distribution of *pseudo-statistics* that are appropriately centered and scaled; when this approximation is valid asymptotically, we say that the bootstrap mechanism *works*.

In general spaces, a sequence of probability measures P_n converges weakly to a measure P if $\int f dP_n \rightarrow \int f dP$ for all bounded continuous functions f ; on the real line this is equivalent to the usual notion of weak convergence by the Portmanteau theorem. In order to simplify notation, we define

$$Pf := \int f dP.$$

Weak convergence can be metrized by looking at the rate of the above convergence across a slightly restricted class of functions. Let BL_1 denote the class of Lipschitz continuous functions which are bounded by 1 and with Lipschitz constant bounded by 1. The bounded Lipschitz metric on a collection of measures is then given by

$$\rho_{BL}(P_1, P_2) = \sup_{f \in BL_1} \left| \int f dP_1 - \int f dP_2 \right|.$$

Convergence in bounded Lipschitz metric is equivalent to weak convergence.

The preceding discussion sweeps under the rug many issues of measurability. In many situations, our random elements and estimators in function space cannot be assumed to be integrable. We will require all estimators to be asymptotically measurable, and any expectations of nonmeasurable elements should be regarded as outer expectations in the sense discussed in van der Vaart and Wellner (1996).

3 Nonparametric smoothing

In this section, we give a brief discussion of bootstrap results as applied in nonparametric smoothing. The data in this setting have traditionally been finite dimensional, but the parameter of interest is functional, which ties these problems to related problems involving functional data. Very recent work, discussed at the end of this section, has begun to extend these results to functional data.

The goal of nonparametric smoothing is typically to use the the data to estimate an unknown function, such as a probability density or conditional

expectation (regression function), which is only assumed to possess some degree of smoothness. Unfortunately, the flexibility allowed by most estimators makes it difficult to ascertain whether observed features are physically meaningful or due to randomness in the data and the estimation process. These difficulties make effective inference for the entire estimated function vitally important, and the functional nature of the problem makes the bootstrap a natural approach.

In the remainder of this section we highlight some major results in density estimation and nonparametric regression.

3.1 Confidence bands in density estimation

Let X_1, \dots, X_n be i.i.d. real-valued observations from a density $f(x)$ which is unknown but assumed to be smooth in some sense. The classical kernel estimate of $f(x)$ proposed by Rosenblatt (1956) is given by

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right), \quad (1)$$

where $K(\cdot)$ is typically a symmetric probability density such as the normal distribution.

Letting the bandwidth h be proportional to $n^{-1/5}$, the estimator (1) has a mean square error of order $n^{-4/5}$ provided that $f(x)$ is twice differentiable; see, e.g. Wand and Jones (1995). However, finite samples often produce estimates with irregular shapes and spurious modes which necessitates effective inference techniques even for exploratory analysis.

Hall (1993) proposes a resampling algorithm to produce confidence bands for f , and develops the Edgeworth expansions necessary to calculate their convergence rates. His approach is to use the bootstrap to estimate the distribution of the suprema of the approximately pivotal quantity

$$R(x) = \frac{\hat{f}(x) - \mathbb{E}\hat{f}(x)}{[\mathbb{E}\hat{f}(x)]^{1/2}}.$$

$R(x)$ is approximately pivotal, where a *pivot* is a statistic whose distribution does not depend on any unknown quantities. Working with pivots often improves bootstrap performance; see Efron and Tibshirani (1993).

The bootstrap sample is constructed by drawing X_1^*, \dots, X_n^* i.i.d. from the empirical distribution of X_1, \dots, X_n ; $\hat{f}^*(x)$ is constructed in the same way as $\hat{f}(x)$, just using the resampled data. It is easy to see that $\mathbb{E}^* \hat{f}^*(x) =$

$\hat{f}(x)$, which leads to the bootstrap pivot

$$R^*(x) = \frac{\hat{f}^*(x) - \hat{f}(x)}{\hat{f}(x)^{1/2}}.$$

A $(1 - \alpha)100\%$ confidence band for $\mathbb{E}\hat{f}(x)$ on the interval $[a, b]$ can be produced by choosing the lower limit L^* and upper limit U^* such that

$$P^* \left[L^* \leq \inf_{x \in [a, b]} R^*(x) \leq \sup_{x \in [a, b]} R^*(x) \leq U^* \right] = 1 - \alpha.$$

The bootstrap approximation to the distribution of the pivotal statistic would then imply that

$$P \left[L^* \leq \inf_{x \in [a, b]} R(x) \leq \sup_{x \in [a, b]} R(x) \leq U^* \right] \simeq 1 - \alpha.$$

The above yields an approximate $(1 - \alpha)100\%$ confidence band for $\mathbb{E}\hat{f}(\cdot)$; this is generally different from a band for $f(\cdot)$ since $\hat{f}(\cdot)$ is a biased estimate. To obtain a confidence band for $f(\cdot)$, either an explicit bias correction must be employed, or the bandwidth h must be chosen in a suboptimal way, i.e., of smaller order than $n^{-1/5}$; In the latter case, the confidence band is consequently being based on an under-smoothed estimator whose bias is negligible (as compared to its standard deviation).

The issues in this setting are similar to those in nonparametric regression, and the theory has been more developed in that setting. For these reasons, we defer discussion until the next section.

3.2 Confidence bands in nonparametric regression and autoregression

In nonparametric regression, data $(X_1, Y_1), \dots, (X_n, Y_n)$ are observed and the goal is to estimate the conditional mean

$$r(x) = \mathbb{E}(Y|X = x).$$

There are several well investigated estimators of $r(\cdot)$, with the most common being the Nadaraya-Watson estimator and the local linear estimator. Each roughly takes the form of a weighted average

$$\hat{r}(x) = \sum_{i=1}^n w(X_i)Y_i,$$

where the weights sum to 1, reflect the distance from X_i to x , the degree of smoothing chosen by the practitioner, and aspects of the local spacing of the X_i 's. A thorough exposition is given in many texts on kernel smoothing; see, for example, Wand and Jones (1995).

Several authors have investigated the construction of bootstrap confidence intervals and bands for these regression estimates. Algorithms differ greatly in their details and assumptions. One approach is to resample the pairs of data; this is discussed in Efron and Tibshirani (1986, 1993) and Mammen (1992). Claeskens and van Keilegom (2003) use the data pairs to estimate a bivariate density and then use the smoothed bootstrap of Silverman and Young (1987), which resamples from the estimated joint density of X and Y .

Other implementations perform the resampling on the residuals. In some cases, for example Härdle and Bowman (1988), the residuals themselves are bootstrapped. In others, such as Härdle and Marron (1991) and Neumann and Polzehl (1998), the wild bootstrap of Wu (1986) is used; the wild bootstrap is typically used with heteroskedastic errors and involves producing a simulated residual e_i^* from only the single observed residual e_i . A final possible approach is to use the local bootstrap of Shi (1991). For concreteness, we outline the approach used in the residual and wild bootstraps.

In this class of methods, an initial pilot estimator is used to smooth the data and estimate the residuals, giving

$$e_i = Y_i - \hat{r}_1(x).$$

Next, the residuals are bootstrapped, using either the residual or wild bootstrap and added to a possibly different estimate of the regression function $r(\cdot)$, producing bootstrap *pseudo*-data

$$Y_i^* = e_i^* + \hat{r}_2(x).$$

The pseudo-data are then smoothed again to produce a *pseudo*-statistic, i.e., a bootstrap regression curve. Repeating this procedure a number of times results in a collection of bootstrap curves that can be used in order to produce a confidence band for the true underlying curve. Neumann and Polzehl (1998) investigated confidence bands for regression, and these ideas were extended to include derivatives of the regression function by Claeskens and van Keilegom (2003). Auestad and Tjøstheim (1991) suggest using the bootstrap for confidence bands in time series *autoregression*, and the theoretical underpinnings are established by Neumann and Kreiss (1998)

who showed the formal equivalence of autoregression to regression in large samples.

The need for multiple estimators $\hat{r}_1(x), \hat{r}_2(x)$, *etc.* is one of the more undesirable aspects of the theory and is due to the fact that nonparametric regression estimators are inherently biased. Since the estimates are not centered at the correct value, confidence intervals need to be explicitly re-centered or constructed in a way that mimics the bias of the original estimator which can be achieved through multiple levels of smoothing. Härdle and Bowman (1988) used explicit bias correction; Hall (1992) advocates under-smoothing, which has guided most subsequent investigations. Interestingly, however, McMurry and Politis (2008) have showed that in some settings the need for bias correction is obviated through the use of infinite order, flat-top kernels.

3.3 Bootstrap in nonparametric functional regression

Recently, Ferraty, van Keilegom, and Vieu (2008) established the consistency of the bootstrap for confidence intervals in nonparametric functional regression. In this setting the observed data are assumed to be i.i.d. pairs of the form $(\mathcal{X}_1, Y_1), \dots, (\mathcal{X}_n, Y_n)$, where the \mathcal{X}_i take values in a function space with semi-metric $d(\cdot, \cdot)$; the Y_i are univariate. They are related by

$$Y_i = r(\mathcal{X}_i) + \epsilon_i,$$

where $r(\cdot)$ is the unknown regression functional and ϵ_i is a real-valued, mean zero error. The functional Nadaraya-Watson estimator is then given by

$$\hat{r}_h(\mathcal{X}) = \frac{\sum_{i=1}^n Y_i K(d(\mathcal{X}_i, \mathcal{X})/h)}{\sum_{i=1}^n K(d(\mathcal{X}_i, \mathcal{X})/h)}, \quad (2)$$

where K is a non-negative kernel, and h is the bandwidth. The asymptotic properties of (2) are discussed in Ferraty and Vieu (2006) and more recent developments are presented in Chapter 4.

Ferraty et al. (2008) prove that both the residual bootstrap and wild bootstrap are consistent, and they use their results to construct asymptotically valid pointwise confidence intervals. Their results are strikingly similar to those achieved in the traditional nonparametric regression setting; once the estimator (2) is defined, the procedures remain substantially unchanged. Multiple levels of smoothing are used for bias correction, and the usefulness of the approach is demonstrated through a simulation study and a real data example.

4 Subsampling

The theory of subsampling for high dimensional and even functional data has been developed in Politis, Romano and Wolf (1999). More recently, a simplified and unified way of looking at functional data and parameters is presented in a K -sample setting in Politis and Romano (2008, 2009). In both settings the theory is fully developed for functional stationary time series data satisfying strong mixing conditions. While such generality is desirable in many applications, for simplicity of exposition we focus on the set-up of an i.i.d. sequence that can be generalized to triangular array data whose rows are i.i.d.

Throughout, let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be an i.i.d. sequence with distribution P , and assume the goal of the analysis is inference for the parameter $\theta(P)$, which takes values in a metric space Θ . To this end, we study the probabilistic distribution of a *root*, which is simply a Θ -valued function of the data $\mathcal{X}_1, \dots, \mathcal{X}_n$, the parameter $\theta(P)$, and the sample size n , with the property that if the distribution of the root were known, the desired inference could be performed.

An easy example of a root can be found in the classical procedure for estimating the mean, μ , of real valued i.i.d. data X_1, \dots, X_n , in which case the quantity $R_n := n^{1/2}(\bar{X} - \mu)$ has the desired property under quite general conditions. In the preceding example, $n^{1/2}$ is the normalizing factor which gives R_n a non-degenerate limiting distribution; while $n^{1/2}$ is the most typical rate, in order to accommodate different estimators, we denote a general normalizing rate by τ_n .

We let $R_n(\mathcal{X}_1, \dots, \mathcal{X}_n, \theta(P))$ denote the root and $J_n(P)$ denote its probability law, where the root is regarded as a random element of Θ . We will show that the probability law $J_n(P)$ can be estimated through use of the appropriate subsampling procedure.

The primary idea behind subsampling is that the distribution of the root can be approximated by an appropriately scaled version of the empirical distribution of the root calculated on subsets of the data of size b , where $b < n$. For i.i.d. data we typically examine all $\binom{n}{b}$ subsamples, or if this is too large, a random sample thereof.

We begin with the primary result which establishes conditions under which the empirical distribution for random elements from a triangular array converges to their marginal distribution. Throughout, we assume random elements, denoted here by \mathcal{Y} , take values in a metric space D , equipped with metric $d(\cdot, \cdot)$. As discussed previously, the random elements are not necessarily assumed to be either separable or measurable.

Theorem 1 (*Politis et al., 1999*) *Let $\mathcal{Y}_1, \dots, \mathcal{Y}_n$ be independent D -valued observations from distribution P . Let \hat{P}_n denote the empirical measure of the \mathcal{Y}_i , $1 \leq i \leq n$. In addition, assume P is tight; that is, for every $\epsilon > 0$ there exists a compact set K so that $P(K) > 1 - \epsilon$. Then $\rho_{BL}(\hat{P}_n, P) \rightarrow_P 0$, which implies weak convergence.*

The above result can be used to establish the validity of subsampling for a wide range of applications, of which we now explore several.

4.1 Parameter estimation

As before, we denote the distribution of the root by $J_n(P)$. Let $L_{b,n}$ be the empirical distribution of $R_b(\mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_b}; \hat{\theta}_n)$, which puts mass $1/\binom{n}{b}$ at each of its $\binom{n}{b}$ possible values. Finally denote the empirical distribution of the subsampled values of $R_b(\mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_b}; \theta(P))$ by $H_{b,n}$. Consistency of the subsampling distribution can be proved by showing first that $H_{b,n}$ converges weakly to $J_n(P)$, and then extending the result to the typical situation where $\theta(P)$ is unknown.

Lemma 1 (*Politis et al., 1999*) *Let $\mathcal{X}_{n,1}, \dots, \mathcal{X}_{n,n}$, for $n \in \mathbb{N}$, be a triangular array of row-wise independent random observations. Assume $J_n(P)$ converges weakly to a limit law $J(P)$ which concentrates on a separable subset of S , $b/n \rightarrow 0$, and $b \rightarrow \infty$ as $n \rightarrow \infty$. Then $\rho_{BL}(H_{n,b}, J_n(P)) \rightarrow_P 0$.*

Lemma 1 is a consequence of a generalization of Theorem 1. The theorem is stated for independent \mathcal{Y} 's but it can be extended to the setting where the \mathcal{Y} 's are the $\binom{n}{b}$ distinct subsampled values of the root rather than the original data.

With this generalization in hand, let the i 'th subsample consist of observations i_1, \dots, i_b . We let $\mathcal{Y}_i = R_b(\mathcal{X}_{i_1}, \dots, \mathcal{X}_{i_b}; \theta(P))$, and then apply the theorem. In this case $\hat{P}_n = H_{n,b}$, and $P_n = J_n$, and the immediate result is $\rho_{BL}(H_{n,b}, J_n(P)) \rightarrow_P 0$ as desired.

Of course, the primary concern is inference for $\theta(P)$ which, except in the case of hypothesis testing, is unknown. The following theorem shows that under the additional mild assumption $\tau_b/\tau_n \rightarrow 0$, Lemma 1 remains true when $\theta(P)$ is replaced by $\hat{\theta}_n$.

Theorem 2 (*Politis et al., 1999*) *Under the assumptions of Theorem 1, and the additional assumption $\tau_b/\tau_n \rightarrow 0$ as $n \rightarrow \infty$, $\rho_{BL}(L_{n,b}, J_n(P)) \rightarrow_P 0$.*

It is interesting to see how easily Theorem 2 follows from Lemma 1. By the triangle inequality, it suffices to show $\rho_{BL}(L_{n,b}, H_{n,b}) \rightarrow_P 0$. We note that $L_{n,b}$ is the empirical distribution of the values

$$\tau_b[\hat{\theta}_{b,i,n} - \hat{\theta}_n] = \tau_b[\hat{\theta}_{b,i,n} - \theta(P)] + \tau_b[\theta(P) - \hat{\theta}_n],$$

which is exactly the distribution of $H_{n,b}$ shifted by

$$\tau_b[\theta(P) - \hat{\theta}_n] = (\tau_b/\tau_n) \left(\tau_n[\theta(P) - \hat{\theta}_n] \right).$$

Since $\tau_n[\theta(P) - \hat{\theta}_n]$ has a non-degenerate limiting distribution, and $\tau_b/\tau_n \rightarrow 0$, it is immediate that this shift is asymptotically negligible.

4.2 K -sample subsampling and a unified approach to functional parameters

The results of the previous section can be extended to the situation where rather than estimating a single parameter, the goal is to estimate a parameter which depends on several independent samples; this is the standard set-up in a classical one-way ANOVA, but the generality of Theorem 1 makes it possible to explore more general situations as well.

In the present subsection, we assume the data consist of K independent samples with $K \geq 1$; the k 'th sample will be denoted $\mathcal{X}^{(k)} = (\mathcal{X}_1^{(k)}, \dots, \mathcal{X}_{n_k}^{(k)})$. The random variables $\mathcal{X}_j^{(k)}$ are assumed to take values in a metric space D , which may be a functional space or a simpler space like \mathbb{R}^d .

It will be assumed throughout that $\mathcal{X}_1^{(k)}, \dots, \mathcal{X}_{n_k}^{(k)}$ is an i.i.d. sample from distribution $P^{(k)}$ and that $\mathcal{X}^{(k)}$ and $\mathcal{X}^{(k')}$ are independent for $k \neq k'$; it is possible to relax the assumption of independence within samples. Taken together, the probability law for the entire experiment is $P = (P^{(1)}, \dots, P^{(k)})$, and the goal will be inference for some parameter $\theta(P)$ that takes values in a normed linear space Θ with norm denoted by $\|\cdot\|$.

Since the goal is typically to quantify the distance between $\theta(P)$ and its estimator $\hat{\theta}_n$, rather than working with them directly, we consider a real valued function of the root, $g[\tau_n(\hat{\theta}_n - \theta(P))]$, where $g : \Theta \rightarrow \mathbb{R}$ is a continuous function; this provides a unified and simplified approach which remains valid even when $K = 1$. Let $J_n(x, P)$ denote the probability law of $g[\tau_n(\hat{\theta}_n - \theta(P))]$. The function g will typically be a norm on Θ or a projection of Θ onto \mathbb{R} . Allowing it to remain unspecified allows the present work to be immediately applicable to a wide range of interesting situations.

As in the previous section, the main idea is to reconstruct the statistic of interest on subsamples of the original data. We let $\mathbf{n} = (n_1, \dots, n_K)$ denote sizes for each of the K samples and $\mathbf{b} = (b_1, \dots, b_K)$ denote the subsamples sizes, which are chosen so that $b_k/n_k \rightarrow 0$ for all k .

Since the K samples are assumed to be independent of each other, we calculate all combinations of subsamples on each of the samples. For simplicity of notation, let q_k denote the number of possible subsamples from the k 'th sample, and let $i_k = 1, \dots, q_k$ index these samples, and $q = \prod_{k=1}^K q_k$. With this notation, we can denote the empirical subsampling approximation to J_n by the empirical distribution of the subsampled values, given by

$$L_{n,b}(x) = \frac{1}{q} \sum_{i_1=1}^{q_1} \sum_{i_2=1}^{q_2} \dots \sum_{i_K=1}^{q_K} 1\{g[\tau_{\mathbf{b}}(\hat{\theta}_{i,b} - \hat{\theta}_n)] \leq x\}.$$

Under the modest conditions described in the following theorem, $L_{n,b}$ provides a consistent estimate for the true law J_n .

Theorem 3 (*Politis and Romano, 2008*) *Assume there exists a nondegenerate limiting law $J(P)$ such that $J_n(P)$ converges weakly to $J(P)$ as $\min_k n_k \rightarrow \infty$; as $\min_k n_k \rightarrow \infty$, $\tau_b \|\hat{\theta}_n - \theta(P)\| = o_P(1)$; $b_k/n_k \rightarrow 0$, $\tau_b/\tau_n \rightarrow 0$, and $b_k \rightarrow \infty$ as $\min_k n_k \rightarrow \infty$; and that g is uniformly continuous. Then*

- (i) *Then $L_{n,b}(\cdot) \rightarrow_P J(x, P)$ for all points of continuity of $J(\cdot, P)$.*
- (ii) *If $J(\cdot, P)$ is continuous at $J^{-1}(1 - \alpha, P)$, then the event*

$$g[\tau_n(\hat{\theta}_n - \theta(P))] \leq L_{n,b}^{-1}(1 - \alpha)$$

has asymptotic probability equal to $1 - \alpha$.

Here, $J^{-1}(s, P) \equiv \inf\{x : J(x, P) \geq s\}$, with a similar definition for $L_{n,b}^{-1}(s)$.

4.3 Subsampling the empirical process

Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be i.i.d. D -valued observations with marginal distribution P . Let \hat{P}_n denote the empirical measure. The empirical process indexed by a class of functions $f \in \mathcal{F}$ is given by

$$Z_n(f) = n^{1/2}[\hat{P}_n f - P f], \quad (3)$$

where $f : D \rightarrow \mathbb{R}$. Empirical processes have received considerable attention in statistics and probability literature; an excellent overview can be found

in van der Vaart and Wellner (1996). Very often, interest focuses on uniform properties of the process given by equation (3) across \mathcal{F} . Necessary and sufficient conditions for uniform consistency of the bootstrap have been studied extensively for i.i.d. functional and finite dimensional data; these will be discussed in Section 5.3. Comparable results for real-valued time series are found in Bühlmann (1994), Naik-Nimbalkar and Rajarshi (1994), and Radulović (2002).

With regards to subsampling, Politis, Romano, and Wolf (1999) have developed convergence criteria in a general time series framework which includes both finite dimensional and functional data. Here we discuss the i.i.d. version of their results. In the following, let $L^\infty(\mathcal{F})$ denote the metric space of real-valued bounded functions on \mathcal{F} equipped with the sup norm, $\|\cdot\|_\infty$. The following theorem is a direct consequence of Theorem 1 using $R_n = n^{1/2}[\hat{P}_n(\cdot) - P_1(\cdot)]$ and $D = L^\infty(\mathcal{F})$.

Theorem 4 (*Politis et al., 1999*) *Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be i.i.d. D -valued observations. Assume the law of Z_n converges weakly to a limiting process Z that concentrates on a separable subset of $L^\infty(\mathcal{F})$. Finally, assume $b/n \rightarrow 0$ as $n \rightarrow \infty$. Then $\rho_{BL}(J_n(P), L_{n,b}) \rightarrow_P 0$.*

Theorem 4 is not as general as the one presented in Section 5.3 but it requires substantially fewer conditions; this is typical since, in contrast with the bootstrap, subsampling works under almost minimal assumptions.

5 Bootstrap

Bootstrap consistency results for functional data can be split into two major categories depending on whether or not the statistic of interest is also functional. Politis and Romano (1994a) examine the bootstrap for the mean of Hilbert space valued time series; Section 5.1 follows their work. Cuevas, Febrero, and Fraiman (2006) conducted a substantial simulation experiment which is described in Section 5.2. Theorems 2.4 and 3.1 of Giné and Zinn (1990) establish consistency for many real valued statistics and, as a corollary, for the bootstrap for the mean of Banach space valued random variables; these results are presented in Section 5.3.

5.1 Bootstrap for functional statistics

The results of Politis and Romano (1994a) were motivated by the study of estimators which are smooth functionals of the empirical process for time

series data. The empirical process takes values in an appropriate Hilbert space, which leads to a general theory for a class of Hilbert-space valued random variables. Their results apply very broadly and are stated in terms of triangular arrays of random elements satisfying strong mixing conditions.

The dependence somewhat complicates matters because the bootstrap must be modified so that the resampled data has an asymptotically similar dependence structure to the original data; this is accomplished through the use of the stationary bootstrap introduced in Politis and Romano (1994b). The idea is that the resampled data contains random length strings of consecutive observations from the original data, where the average length of the strings grows as n becomes large.

For simplicity, as in preceding sections, we focus on the i.i.d. setting; notably, the stationary bootstrap reduces to the usual Efron's (1979) i.i.d. bootstrap simply by letting the average length of the above mentioned strings equal to one. In the following, let H be a Hilbert space. The core of the theory rests on the following theorem, which establishes consistency of the bootstrap for the mean of the functional data. In the present form the theorem is weaker than Corollary 2 (below), but as shown in Politis and Romano (1994a), it also holds for data satisfying dependence conditions.

Theorem 5 (*Politis and Romano, 1994a*) *Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be an i.i.d. sequence of essentially bounded H -valued random variables with mean θ . Let $\bar{\mathcal{X}}_n = n^{-1} \sum_{i=1}^n \mathcal{X}_i$ and $\mathcal{Z}_n = \sqrt{n}(\bar{\mathcal{X}}_n - \theta)$. Denote the law of \mathcal{Z}_n by $L(\mathcal{Z}_n)$. Let $\mathcal{Z}_n^* = n^{1/2}(\mathcal{X}_n^* - \bar{\mathcal{X}}_n)$, and denote its law by $L(\mathcal{Z}_n^* | \mathcal{X}_1, \dots, \mathcal{X}_n)$. Then $\rho(L(\mathcal{Z}_n), L(\mathcal{Z}_n^* | \mathcal{X}_1, \dots, \mathcal{X}_n)) \rightarrow_P 0$, where ρ is any metric metrizing weak convergence on H .*

If Y_1, \dots, Y_n are i.i.d. real valued random variables with distribution F , then $\mathcal{X}_i(t) = 1_{[Y_i \leq t]} - F(t)$ can be regarded as a functional observation in the Hilbert space $L^2(\nu)$, where ν is a finite measure on \mathbb{R} . If we combine this reasoning with the previous theorem and the notion of Fréchet differentiability, we obtain an immediate corollary extending the consistency to smooth functionals of the empirical distribution.

In the following, let ν denote a finite measure on the real line, F and G denote distributions in some class $\mathcal{F} \subset L^2(\nu)$, and $T(\cdot)$ a functional on \mathcal{F} which is Fréchet differentiable in the sense that for fixed F and as $\|G - F\| \rightarrow 0$, $T(G) - T(F) = \langle \psi, G - F \rangle + o(\|G - F\|)$, for some $\psi \in L^2(\nu)$; here $\langle \cdot, \cdot \rangle$ denotes the inner product on H .

Corollary 1 (*Politis and Romano, 1994a*) *Let L_n denote the true distribution of $n^{1/2}(T(\hat{F}_n) - T(F))$ and let \hat{L}_n denote the bootstrap distribution of*

$n^{1/2}(T(\hat{F}_n^*) - T(\hat{F}_n))$, then $\rho_1(\hat{L}_n, L_n) \rightarrow 0$, where ρ_1 is any metric metrizing weak convergence on the real line. Moreover, L_n converges weakly to a Gaussian distribution with mean 0 and variance σ_ψ^2 , where

$$\sigma_\psi^2 = \text{Var}(\langle X_1, \psi \rangle).$$

In the context of functional data, the parameters of interest are often also functional, so it is desirable to study the case where T takes on values in a function space. Corollary 1 can be extended to this setting using an appropriate differentiability condition and a “functional delta method.” We briefly outline the approach; technical details can be found in Cuevas and Fraiman (2004), and Cuevas et al. (2006) provides additional insights.

Let T be an operator defined on \mathcal{P} , a suitable space of probability measures, with values in a Banach space F . Assume that T is differentiable at $P \in \mathcal{P}$ in the sense that

$$T(Q) = T(P) + T'_P(Q - P) + o(\rho_{BL}(P, Q)), \quad (4)$$

where T'_P is a linear operator on an appropriate space of finite signed measures. In order to simplify notation, let $S_n = n^{1/2}(T(P_n) - T(P))$ and $S_n^* = n^{1/2}(T(P_n^*) - T(P_n))$, and let $L(S_n)$ and $L(S_n^*|\mathcal{X}_1, \dots, \mathcal{X}_n)$ respectively denote their laws. In order to establish bootstrap consistency we need to show

$$\rho_{BL}(L(S_n^*|\mathcal{X}_1, \dots, \mathcal{X}_n), L(S_n)) \rightarrow 0 \quad (5)$$

either in probability or almost surely. The convergence (5) can be established via the linearization which results from differentiability condition (4).

$$\begin{aligned} S_n &= n^{1/2}T'_P(P_n - P) + n^{1/2}o(\rho_{BL}(P_n, P)) \\ &= L_n + n^{1/2}o(\rho_{BL}(P_n, P)), \end{aligned}$$

where L_n denotes the linearization of S_n . Let L_n^* denote the corresponding linearization of S_n^* . The linearizations L_n and L_n^* tend to the same Gaussian limit by the Central Limit Theorem and the bootstrap consistency results given in Theorem 5 (above) or Corollary 2 (below).

The remainder terms, $n^{1/2}o(\rho_{BL}(P_n, P))$ and $n^{1/2}o(\rho_{BL}(P_n^*, P_n))$, can be handled provided we are able to establish boundedness in probability, uniform on \mathcal{P} , of the sequences $n^{1/2}\rho_{BL}(P_n, P)$.

5.2 Simulation studies

Cuevas, Febrero, and Fraiman (2006) performed an extensive simulation study to assess the performance of the functional data bootstrap. To our

knowledge this is the only large scale investigation to date. We briefly summarize their experiments and findings. They investigated bootstraps for several functional measures of center and one measure of spread for data in $C[0, 1]$. The data were generated by the following two processes:

- A Wiener process on $[0, 1]$ given by $\mathcal{X}(t) = 10t(1 - t) + B(t)$ where $B(t)$ is a standard Brownian motion.
- A Gaussian process $\mathcal{X}(t)$ with mean $10t(1 - t)$ and covariance function $Cov(\mathcal{X}(s), \mathcal{X}(t)) = \exp(-|s - t|/0.3)$.

They constructed bootstrap confidence intervals based on the following estimators:

- The sample mean $\bar{\mathcal{X}}$.
- The sample variance.
- The FM -median.
- The α -trimmed FM mean with $\alpha = 0.25$.
- The “kernel mode” estimate for the underlying density using a Gaussian kernel.

The FM trimmed mean and median were based on Fraiman and Muniz’s (2001) notion of data depth which is defined as follows. For each $t \in [0, 1]$ let $F_{n,t}$ be the empirical distribution of the sample $x_1(t), \dots, x_n(t)$. Let $Z_i(t) = 1 - |1/2 - F_{n,t}(x_i(t))|$ denote the univariate depth of the datum $x_i(t)$. These values are aggregated across time by

$$I_i = \int_0^1 Z_i(t) dt. \quad (6)$$

Observations x_i can then be ranked according to the corresponding value of I_i , with the “deepest” data having the largest values of I . Using this measure, the α -trimmed mean becomes the average of the $100(1 - \alpha)\%$ deepest functions in the sample, and the FM -median is the deepest function in the sample.

The “kernel mode” estimator, an exploratory data analysis tool, is given by

$$\operatorname{argmax}_{x \in E} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\|x_i - x\|}{h}\right),$$

which is approximated for computational efficiency by

$$\operatorname{argmax}_{x \in \{x_1, \dots, x_n\}} \frac{1}{nh} \sum_{i=1}^n K\left(\frac{\|x_i - x\|}{h}\right).$$

Cuevas et al. (2006) also tested three different bootstraps, the standard bootstrap, a smoothed bootstrap, and for the mean, a parametric bootstrap. Confidence intervals were constructed using both the L^2 and L^∞ norms. Sample sizes were $n = 25$ and $n = 100$. Each combination of estimator, norm, bootstrap procedure, and sample size was run 500 times.

Their simulations show substantial promise. To summarize their conclusions briefly:

- The standard bootstrap proved unreliable for the sample median and kernel mode; this appears to be because the resampling produces a very small number of curves. These results were substantially improved by using a smoothed bootstrap.
- With the exception of the standard bootstrap failure for the median and kernel mode, all other estimators performed quite well, but the parametric bootstrap did not provide any substantial advantage.
- The performance improved substantially from $n = 25$ to $n = 100$. For almost all estimators at the larger sample size, observed coverage was close to expected coverage; interval width also decreased as expected.
- Both the L^2 and L^∞ norms performed satisfactorily.

5.3 Bootstrapping the empirical process

The bootstrap for empirical processes has been extensively investigated by several authors including Giné and Zinn (1990) and Dudley (1990), and nicely summarized in Section 3.6 of van der Vaart and Wellner (1996). The results are extremely general and apply equally well to functional and finite dimensional data. The results in this setting are substantially more complex than results in the preceding sections, so we focus on the highlights of the theory and refer readers to the preceding references for additional details.

Let $\mathcal{X}_1, \dots, \mathcal{X}_n$ be a D -valued i.i.d. sample with probability distribution P , where D is again a metric space with metric $d(\cdot, \cdot)$. For consistency with notation elsewhere, we let \hat{P}_n denote the empirical distribution. The bootstrap sample $\mathcal{X}_1^*, \dots, \mathcal{X}_n^*$ is an i.i.d. sample from \hat{P}_n , and its empirical distribution will be denoted \hat{P}_n^* . The process is typically viewed as a random

functional on a class of functions $\mathcal{F} \subset \{f : D \rightarrow \mathbb{R}\}$, for which it is assumed that $\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty$.

As in Section 4.3, we define the empirical process as follows

$$Z_n(\cdot) = n^{1/2}[\hat{P}_n(\cdot) - P(\cdot)]. \quad (7)$$

and its bootstrap version

$$Z_n^*(\cdot) = n^{1/2}[\hat{P}_n^*(\cdot) - \hat{P}_n(\cdot)] = n^{1/2} \sum_{i=1}^n (M_{ni} - 1) \delta_{\mathcal{X}_i}(\cdot) \quad (8)$$

where M_{ni} is the number of times \mathcal{X}_i appeared in the resample. For any particular $f \in \mathcal{F}$ the consistency of the bootstrap is not particularly surprising because $\hat{P}_n f - Pf = (1/n) \sum_i f(\mathcal{X}_i) - \mathbb{E}(f(X_i))$, an average of real valued random variables, and a situation in which the bootstrap is typically consistent.

The major success of empirical process bootstrap has been to establish conditions on \mathcal{F} under which a sort of uniform convergence occurs that implies the bootstrap validity for all $f \in \mathcal{F}$ simultaneously. Roughly speaking, these conditions require that \mathcal{F} not be too “big.” We first start with a definition that describes a class of functions for which a uniform central limit theorem holds. We then discuss how this is sufficient for bootstrap consistency in probability, and along with an additional moment assumption implies almost sure bootstrap consistency.

Definition 1 *Let \mathcal{F} be a collection of square integrable functions $f : D \rightarrow \mathbb{R}$ for which $\sup_{f \in \mathcal{F}} |f(x) - Pf| < \infty$ and $Z_n = n^{1/2}(\hat{P}_n - P) \Rightarrow G$ in $L^\infty(\mathcal{F})$, where G is a tight Borel measurable element in $L^\infty(\mathcal{F})$. Then \mathcal{F} is called a Donsker class.*

Let \mathcal{F} be a Donsker class. The multivariate central limit theorem shows that for any functions $f_1, \dots, f_k \in \mathcal{F}$, $(Z_n f_1, \dots, Z_n f_k) \Rightarrow N(0, \Sigma_k)$, where $[\Sigma_k]_{i,j} = P(f_i - Pf_i)(f_j - Pf_j)$; this can be used to show that the limiting process G is by necessity Gaussian.

The Donsker assumption is almost enough to ensure uniform bootstrap consistency; we discuss these results first and then return to conditions that imply a class is Donsker.

Definition 2 *An envelope function for a class of functions \mathcal{F} is any function $F(x)$ such that $|f(x)| \leq F(x)$ for all $x \in D$ and $f \in \mathcal{F}$.*

The two preceding conditions imply convergence to a mean zero Gaussian process G in outer probability.

Theorem 6 (*van der Vaart and Wellner, 1996*) *Let \mathcal{F} be a class of measurable functions with finite envelope function. Then \mathcal{F} is Donsker if and only if $\rho_{BL}(Z_n^*, G) \rightarrow_P 0$ and Z_n^* is asymptotically measurable, where G is the same limiting Gaussian process as in Definition 1.*

With an additional moment assumption, the preceding result can be strengthened to outer almost sure convergence.

Theorem 7 (*van der Vaart and Wellner, 1996*) *Let \mathcal{F} be a class of measurable functions with finite envelope function. Then \mathcal{F} is Donsker and $P\|f - Pf\|_{\mathcal{F}}^2 < \infty$ if and only if $\rho_{BL}(Z_n^*, G) \rightarrow_{as} 0$ and Z_n^* is asymptotically measurable, where G is the same limiting Gaussian process as in Definition 1.*

Theorem 6 is sufficient for bootstrap consistency. An interesting consequence of Theorem 7 is the following corollary which implies bootstrap consistency for functional data and is discussed in Remark 2.5 in Giné and Zinn (1990).

Corollary 2 (*Giné and Zinn, 1990*) *If $\mathcal{X}_1, \dots, \mathcal{X}_n$ are i.i.d. B -valued random variables, where B is a separable Banach space, then $\mathbb{E}(\|\mathcal{X}_i\|^2) < \infty$ and central limit theorem holds for the \mathcal{X}_i if and only if $n^{-1/2} \sum_{i=1}^n (\mathcal{X}_i^* - \bar{\mathcal{X}}) \Rightarrow G_X$ a.s.*

The preceding results require that \mathcal{F} is a Donsker class. There are numerous sufficient conditions for a class being Donsker, and a set of necessary and sufficient conditions in Giné and Zinn (1990). Here we present one simple sufficient condition from van der Vaart and Wellner (1996).

All of the criteria rest on ways to quantify the size of a class of functions \mathcal{F} . Given two functions l and u we denote by the bracket $[l, u]$ the set of functions f such that $l \leq f \leq u$. An ϵ -bracket is a bracket $[l, u]$ where $\|u - l\| < \epsilon$. The bracketing number $N_{[]}(\epsilon, \mathcal{F}, \|\cdot\|)$ is the minimum number of ϵ -brackets needed to cover \mathcal{F} . A sufficient condition for \mathcal{F} to be Donsker is

$$\int_0^\infty \sqrt{\log N_{[]}(\epsilon, \mathcal{F}, L_2(P))} d\epsilon < \infty.$$

There are other less restrictive criteria. Interested readers should consult the two previously mentioned references.

6 Real-data applications

In this section we briefly discuss three recent applications of the bootstrap in functional data settings.

6.1 fMRI data

Locascio et al. (1997) use a residual resampling approach to detect areas of brain activation in fMRI brain imaging experiments. In their experiments, areas of brain activity are measured over time, while the subject is exposed to various experimental conditions, with the goal of estimating the areas of the brain which are activated during certain tasks. Traditional tools are substantially inadequate in this situation because the problem is in essence that of performing a hypothesis test on every pixel with data that exhibit both spatial and temporal dependence. Their approach is roughly to use an ARMA model to estimate the effects of the experimental condition at each pixel. Once the pixel-by-pixel data has been modeled, a classical technique, such as a p -value from a t -test, is used to test the significance of the activation level at each pixel. A test based on permutation of the residuals across time and experimental conditions, but not space, is then used to make the necessary adjustments for multiple testing.

6.2 Sulfur dioxide concentrations

Fernández de Castro, Guillas, and González Manteiga (2005) use a functional bootstrap approach to make 30 minute in advance predictions of sulfur dioxide concentrations near a coal-fired power plant. The power plant had regulatory obligations to eliminate periods of high sulfur dioxide concentration in its vicinity, so its management was hoping to get some warning of coming spikes so they could take preventative measures, such as switching types of coal.

The data were in the form of a time series $\mathcal{X}_1, \mathcal{X}_2, \dots$, and were grouped in pairs $(\mathcal{X}_i, \mathcal{X}_{i+1})$ where the first coordinate function is the sulfur dioxide over the current half-hour period and the second coordinate function is the concentration over the subsequent half hour, which is to be predicted. Two functional approaches were used to model \mathcal{X}_{i+1} as a function of \mathcal{X}_i , a Nadaraya-Watson type functional kernel estimator, and an autoregressive Hilbertian model; see Bosq (2000). The bootstrap was used to construct prediction bands for each estimator.

The bootstrap procedure for the kernel estimator works as follows. Denote the current time period curve by \mathcal{Y}_0 , with the goal of forecasting \mathcal{Y}_1 , the next half hour period. Pairs of data $(\mathcal{X}_i, \mathcal{X}_{i+1})$ in the historical database are then resampled with probabilities depending on a distance from \mathcal{X}_i to \mathcal{Y}_0 ; this procedure is similar to the local bootstrap of Paparoditis and Politis (2002). The second coordinates of the bootstrapped pairs then give an estimate of the range of possible outcomes through the Fraiman and Muniz notion of data depth, given in equation (6).

The bootstrap procedure for the autoregressive Hilbertian model follows a more traditional residual bootstrap. One step forward residuals are calculated for all forecasts using pairs of data in the database. The residuals are then decomposed in terms of their principal components, and the pseudo-data is constructed by separately bootstrapping each of the principal components coordinates.

6.3 Language classification

Cuesta-Albertos et al. (2007) use a functional bootstrap hypothesis test to cluster languages according to rhythmic features. It was conjectured in the linguistics literature that languages can be classified as *stress-timed*, *syllable-timed*, or *mora-timed*, and the goal was to perform this classification empirically. Their approach was to look at *sonority*, a local index of regularity of the speech signal which takes values close to 1 for sonorant portions of the signal and 0 for obstruent portions.

The sonority of a speech sample can be viewed as a function in $L^2([0, T])$, where T is the length of the sample. Each individual sample sonority $S_{l,i}$, where l indexes language and i indexes sample within a language, was projected onto a (fixed) Brownian motion, through the L^2 inner product

$$\langle S_{l,i}, B \rangle = \int_0^T S_{l,i}(t)B(t)dt,$$

which reduces the curve to a univariate datum. A Kolmogorov-Smirnov type test is then used to compare the distribution of samples from two languages to see if they appear to group in the hypothesized way. Since the preceding test depends on the random Brownian path, its stability is increased by repeating the test for several independent Brownian motions.

The bootstrap is used to estimate the p -value for the preceding tests. The samples from the two languages l and l' are pooled and the bootstrap is used to draw independent samples from the pooled data set, thus mimicking a sampling environment that satisfies the null hypothesis.

Finally the authors employ a Monte Carlo simulation to empirically assess the ability of their procedure to distinguish between different sonority patterns. They construct Markov chains which mimic the features which are expected to be observed in sonority patterns in the different types of languages, and they use the modified Kolmogorov-Smirnov test to try to distinguish the different chains; the results were encouraging.

7 Discussion

As functional data proliferate, we expect bootstrap and subsampling to become very popular approaches to inference since they are straightforward to implement and have broad applicability. Nonetheless, the field remains young and offers rich avenues for further development; we mention some areas of particular interest.

In finite dimensional problems, approaches such as studentization have been shown to produce bootstrap confidence intervals with higher order accuracy, i.e., coverage that is very close to the nominal. While it seems likely that analogous properties hold for functional data, the question of convergence rates for the bootstrap has not been addressed in any setting, and there has been little discussion as to whether higher order accuracy might be feasible. Furthermore, it is not trivial to see how the studentization can be optimally implemented with functional statistics.

Which method is preferable, bootstrap or subsampling? Of course, subsampling has in its favor its very wide applicability; by contrast, the bootstrap has thus far only been shown to be valid for the mean of functional data, the empirical process, and smooth functionals thereof, as outlined in Section 5, and regression as discussed in Section 3. Results establishing the validity of the bootstrap in other settings are highly desirable.

Focusing on the sample mean under regularity, both methods (bootstrap and subsampling) are consistent for a wide range of functional data but no work has been done to assess their performance relative to each other. In addition, very few simulation studies have explored the finite sample performance of these techniques.

In the coming years we expect these and other related problems to offer fruitful areas for theoretical development and practical implementation of resampling methods for functional data and/or statistics.

References

- AUESTAD, B., TJØSTHEIM, D. (1991). Functional identification in non-linear time series. In *Nonparametric Functional Estimation and Related Topics*, (G. Roussas, Ed.), 493-507. Kluwer Academic Publishers.
- BOSQ, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Springer.
- BÜHLMANN, P. (1994). Blockwise bootstrapped empirical process for stationary sequences. *Annals of Statistics*, **22**, 995-1012.
- CLAESKENS, G., VAN KEILEGOM, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Annals of Statistics*, **31**, 1852-1884.
- CUESTA-ALBERTOS, J., FRAIMAN, R., GALVES, A., GARCIA, J., SVARC, M. (2007). Classifying speech sonority functional data using a projected Kolmogorov-Smirnov approach. *Journal of Applied Statistics*, **34**, 749-761.
- CUEVAS, A., FEBRERO, M., FRAIMAN, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics & Data Analysis*, **51**, 1063-1074.
- CUEVAS, A., FRAIMAN, R. (2004). On the bootstrap methodology for functional data. In *COMPSTAT 2004 - Proceedings in Computational Statistics*, (Ed. J. Antoch), 127-135. Physica-Verlag.
- DUDLEY, R. M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In *Probability in Banach Spaces 7*, (Ed. E. Eberlein, J. Kuelbs, M. Marcus), 63-82. Birkhäuser.
- EFRON, B. (1979). Bootstrap methods: another look at the jackknife. *Annals of Statistics*, **7**, 1-26.
- EFRON, B., TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy (with discussion). *Statistical Science*, **1**, 54-96.
- EFRON, B., TIBSHIRANI, R. (1993). *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- FERNÁNDEZ DE CASTRO, B., GUILLAS, S., GONZÁLEZ MANTEIGA, W. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47**, 212-222.
- FERRATY, F., VAN KEILEGOM, I., AND VIEU, P. (2008). On the validity of the bootstrap in nonparametric functional regression. IAP Statistics Network, Technical Report 08027.
- FERRATY, F., VIEU, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.
- FRAIMAN, R., MUNIZ, G. (2001). Trimmed means for functional data. *Test*, **10**, 419-440.

- GINÉ, E., ZINN, J. (1990). Bootstrapping general empirical measures. *Annals of Probability*, **18**, 851-869.
- HALL, P. (1992). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density. *Annals of Statistics*, **20**, 675-694.
- HALL, P. (1993). On Edgeworth expansion and bootstrap confidence bands in nonparametric curve estimation. *Journal of the Royal Statistical Society, Series B*, **55**, 291-304.
- HÄRDLE, W., BOWMAN, A. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *Journal of the American Statistical Association*, **83**, 102-110.
- LOCASCIO, J., JENNINGS, P., MOORE, C., CORKIN, S. (1997). Time series analysis in the time domain and resampling methods for studies of functional magnetic resonance brain imaging. *Human Brain Mapping*, **5**, 168-193.
- MAMMEN, E. (1992). *When does bootstrap work? Asymptotic results and simulations*, Springer Lecture Notes in Statistics No. 77, New York.
- MCMURRY, T., POLITIS, D.N. (2008). Bootstrap confidence intervals in nonparametric regression with built-in bias correction. *Statistics and Probability Letters*, **78**, 2463-2469.
- NAIK-NIMBALKAR, U. V., RAJARSHI, M. B. (1994). Validity of blockwise bootstrap for empirical processes with stationary observations. *Annals of Statistics*, **22**, 980-994.
- NEUMANN, M., KREISS, J.-P. (1998). Regression-type inference in nonparametric autoregression. *Annals of Statistics*, **26**, 1570-1613.
- NEUMANN, M., POLZEHL, J. (1998). Simultaneous bootstrap confidence bands in nonparametric regression. *Journal of Nonparametric Statistics*, **9**, 307-333.
- PAPARODITIS, E., POLITIS, D.N. (2002). The local bootstrap for Markov processes. *Journal of Statistical Planning and Inference*, **108**, 301-328.
- POLITIS, D. N., ROMANO, J. P. (1994a). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationary bootstrap. *Statistica Sinica*, **4**, 461-476.
- POLITIS, D. N., ROMANO, J. P. (1994b). The stationary bootstrap. *Journal of the American Statistical Association*, **89**, 1303-1313.
- POLITIS, D. N., ROMANO, J. P. (2008). K -sample subsampling. In *Functional and Operatorial Statistics*, (Ed. S. Dabo-Niang, F. Ferraty), 247-253. Physica-Verlag, Heidelberg.
- POLITIS, D. N., ROMANO, J. P. (2009). K -sample subsampling in general spaces: the case of independent time series. To appear in *Journal of Multivariate Analysis*.

- POLITIS, D. N., ROMANO, J. P., WOLF, M. (1999). *Subsampling*, Springer, New York.
- RADULOVIĆ, D. (2002). On the bootstrap and empirical processes for dependent sequences. In *Empirical Process Techniques for Dependent Data*, (Ed. H. Dehling, T. Mikosch, M. Sørensen). Birkhäuser.
- ROSENBLATT, M. (1956). Remarks on some nonparametric estimates of a density function, *Annals of Mathematical Statistics*, **27**, 832-837.
- SHI, S. (1991). Local bootstrap. *Annals of the Institute of Statistical Mathematics*, **43**, 667-676.
- VAN DER VAART, A. W., WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes*, Springer Series in Statistics.
- WAND, M., JONES, M. (1995). *Kernel Smoothing*. Chapman and Hall/CRC Press.
- WU, C.F.J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Annals of Statistics*, **14**, 1262-1350.