

I N S T I T U T D E
S T A T I S T I Q U E

UNIVERSITÉ CATHOLIQUE DE LOUVAIN



D I S C U S S I O N
P A P E R

0818

**ON THE VALIDITY OF THE BOOTSTRAP
IN NONPARAMETRIC FUNCTIONAL
REGRESSION**

FERRATY, F., VAN KEILEGOM, I. and P. VIEU

This file can be downloaded from
<http://www.stat.ucl.ac.be/ISpub>

On the validity of the bootstrap in nonparametric functional regression

Frédéric Ferraty¹ Ingrid Van Keilegom² Philippe Vieu¹

May 16, 2008

Abstract

We consider the functional nonparametric regression model $Y = r(\mathcal{X}) + \varepsilon$, where the response Y is univariate, $\mathcal{X}$ is a functional covariate (i.e. valued in some infinite-dimensional space), and the error ε satisfies $E(\varepsilon|\mathcal{X}) = 0$. For this model, the pointwise asymptotic normality of a kernel estimator $\hat{r}(\cdot)$ of $r(\cdot)$ has been proved in the literature. In order to use this result for building pointwise confidence intervals for $r(\cdot)$, the asymptotic variance and bias of $\hat{r}(\cdot)$ need to be estimated. However, the functional covariate setting makes this task very hard. To circumvent the estimation of these quantities, we propose to use a bootstrap procedure to approximate the distribution of $\hat{r}(\cdot) - r(\cdot)$. Both a naive and a wild bootstrap procedure are studied, and their asymptotic validity is proved. The obtained consistency results are discussed from a practical point of view via a simulation study. Finally, the wild bootstrap procedure is applied to a food industry quality problem in order to compute pointwise confidence intervals.

Key words: Asymptotic normality, confidence intervals, functional data, kernel estimator, naive bootstrap, nonparametric regression, wild bootstrap.

¹ Institut de Mathématiques de Toulouse, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse, France, E-mail: ferraty@cict.fr, vieu@cict.fr

² Institute of Statistics, Université catholique de Louvain, Voie du Roman Pays 20, 1348 Louvain-la-Neuve, Belgium, E-mail: ingrid.vankeilegom@uclouvain.be

1 Introduction

Functional data analysis (FDA) is definitely an important field in statistics. It concerns any statistical method dealing with random variables valued in some infinite-dimensional space (called functional variables). One observation (i.e. functional data point) may be a curve (see e.g. Borggaard and Thodberg, 1992, for spectrometric curves in the near infrared, Hall *et al.*, 2001, Dabo-Niang *et al.*, 2007, for radar waveforms, and Hastie *et al.*, 1995, for speech recognition data), a 2D (or 3D)-image (see e.g. Sangalli *et al.*, 2007, for a sample of 3D-images of internal carotid artery), or any other object living in a functional space. Numerous recent work, from both theoretical and practical point of view, confirms the pertinence of investigating this area. For an overview of this topic, see the monographs of Bosq (2000), Ramsay and Silverman (2002, 2005), Ferraty and Vieu (2006) and the recent special issues in various statistical journals (Davidian *et al.*, 2004, González Manteiga and Vieu, 2007, and Valderrama, 2007).

This paper focuses on the functional nonparametric regression model when one considers a functional covariate and a scalar response. A functional covariate means that the explanatory variable is valued in some infinite-dimensional space $\mathcal{E}$. So, one focuses on the following functional nonparametric regression model:

$$Y = r(\mathcal{X}) + \varepsilon, \tag{1.1}$$

where Y is a scalar response variable, $\mathcal{X}$ is a functional covariate, r is an unknown but smooth regression function, and the error ε satisfies $E(\varepsilon|\mathcal{X}) = 0$. Let $E(\varepsilon^2|\mathcal{X}) = \sigma_\varepsilon^2(\mathcal{X})$ denote the second conditional moment of ε .

This functional nonparametric model (see Ferraty and Vieu, 2002, for a first study of this kind of model and Ferraty and Vieu, 2006, for an overview) was studied intensively in the last years and became more and more popular in the FDA community. It is an interesting and complementary alternative to the functional linear model (see Cardot *et al.*, 1999, Cai and Hall, 2006, and Crambes *et al.*, 2008). Lots of theoretical properties concerning a kernel estimator $\hat{r}(\cdot)$ of $r(\cdot)$ as well as its practical aspects have been investigated in this functional nonparametric regression setting (see Ferraty and Vieu, 2006, and references therein).

Asymptotic normality has been established in Masry (2005), Delsol (2008) and Ferraty *et al.* (2007). The great interest of getting the asymptotic normality is the opportunity of building pointwise confidence intervals for the unknown regression operator $r(\cdot)$. However, in this functional setting, it is a heavy task to compute the bias and variance involved

in the asymptotic normality. One way to overcome this problem is to use a bootstrap procedure.

In the context of nonparametric regression with scalar covariates, the bootstrap methodology has been extensively studied in the literature. See e.g. Hall (1990, 1992), Hall and Hart (1990), Härdle and Marron (1991) and Cao (1991), among many other papers.

In the functional variable setup, the bootstrap is much less developed. Theoretical results for the bootstrap have been established but in the empirical distributions spirit (see e.g. Dudley, 1990, Giné and Zinn, 1990, Politis and Romano, 1994, van der Vaart and Wellner, 1996, Cuevas and Fraiman, 2004), whereas recent practical advances can be found in Fernández de Castro *et al.* (2005) and Cuevas *et al.* (2006).

The aim of this paper is to propose a bootstrap methodology allowing to compute pointwise confidence intervals for the regression operator. Both theoretical and practical aspects are investigated. It is worth noting that this is the first time that the asymptotic behavior of the bootstrap procedure is established in a functional nonparametric regression setting.

This paper is organized as follows. In the next section, we explain how to bootstrap under the functional nonparametric regression model (1.1). Section 3 states the asymptotic consistency of the naive and the wild bootstrap under this model. The practical issue on how to build pointwise confidence intervals for the regression operator $r(\cdot)$ is also discussed. In Section 4, a simulation study is carried out in order to verify the behavior of the bootstrap procedure in practice, and pointwise confidence intervals are computed for a functional dataset coming from a food industry quality problem. Some general conclusions and ideas for further research are outlined in Section 5. Finally, the proofs of the asymptotic results are collected in the Appendix.

2 Bootstrap in functional regression

We will assume throughout that the functional space $\mathcal{E}$ is endowed with a semi-metric d . This setting is quite general, since it contains the space of continuous functions, L^p spaces as well as more complicated spaces like Sobolev or Besov spaces. Let $\mathcal{S} = \{(\mathcal{X}_1, Y_1), \dots, (\mathcal{X}_n, Y_n)\}$ be a sample of i.i.d. data drawn from the distribution of the pair $(\mathcal{X}, Y)$. The estimator of the regression operator r is given by

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

χ being a fixed element of $\mathcal{E}$. Here, K is a probability density function (kernel) and h is a smoothing parameter (bandwidth sequence), tending to zero when n tends to infinity. To fix the ideas, in the application part, $\mathcal{X}_1, \dots, \mathcal{X}_n$ are a sample of curves, for which the corresponding responses $Y_1, \dots, Y_n$ have been observed and χ is an additional fixed curve.

In such a functional setting, a natural way for implementing a bootstrap procedure consists in bootstrapping the errors which are real random variables. To do this, we propose to adapt the standard bootstrap scheme used in the multivariate context (i.e. when one considers a multivariate covariate) to the considered functional situation.

Naive bootstrap. We assume here that the model is homoscedastic, i.e. $\sigma_\varepsilon^2(\mathcal{X}) \equiv \sigma_\varepsilon^2$. Under this extra assumption, we resample as follows:

- Step 1: Let $\widehat{\varepsilon}_{i,b} = Y_i - \widehat{r}_b(\mathcal{X}_i)$, for all $i = 1, \dots, n$, where b is a second smoothing parameter,
- Step 2: Draw n i.i.d. random variables $\varepsilon_1^{boot}, \dots, \varepsilon_n^{boot}$ from the cumulative distribution of $(\widehat{\varepsilon}_{1,b} - \widehat{\varepsilon}_b, \dots, \widehat{\varepsilon}_{n,b} - \widehat{\varepsilon}_b)$, where $\widehat{\varepsilon}_b = n^{-1} \sum_{i=1}^n \widehat{\varepsilon}_{i,b}$,
- Step 3: Define $Y_i^{boot} = \widehat{r}_b(\mathcal{X}_i) + \varepsilon_i^{boot}$, for all $i = 1, \dots, n$, and let $\mathcal{S}^{boot} = (\mathcal{X}_i, Y_i^{boot})_{i=1, \dots, n}$,
- Step 4: Define $\widehat{r}_{hb}^{boot}(\chi) = \frac{\sum_{i=1}^n Y_i^{boot} K(h^{-1}d(\mathcal{X}_i, \chi))}{\sum_{i=1}^n K(h^{-1}d(\mathcal{X}_i, \chi))}$ (see Condition (C6) in Section 3 for the theoretical link between b and h).

Wild bootstrap. Only step 2 changes: define $\varepsilon_i^{boot} = \widehat{\varepsilon}_{i,b} V_i$, where $V_1, \dots, V_n$ are i.i.d. random variables that are independent of the data $(\mathcal{X}_i, Y_i)$ ($i = 1, \dots, n$) and that satisfy $E(V_1) = 0$ and $E(V_1^2) = 1$.

It is clear that these bootstrap procedures are easy to implement (keeping in mind that the kernel estimator is also very easy to compute). However, one can remark that two bandwidths, b and h , have to be fixed. If, from a theoretical point of view, one has an idea on their mutual asymptotic behavior (Condition (C6) implies, among others, that $h = o(b)$), from a practical point of view one can study the influence of both bandwidths on the behavior of the bootstrap procedure. This will be discussed in the simulation study (see Section 4.1).

3 Asymptotic validity of the bootstrap

First, in Subsection 3.1 we state the conditions under which the asymptotic results are valid, and we discuss in detail when these conditions are satisfied. In Subsection 3.2

we give the main result of this paper, namely the asymptotic validity of the two bootstrap procedures by comparing the respective asymptotic distribution of $\widehat{r}_h(\chi) - r(\chi)$ and $\widehat{r}_{hb}^{boot}(\chi) - \widehat{r}_b(\chi)$. Finally, Subsection 3.3 describes how one can get pointwise confidence intervals from the main theoretical result. Before introducing the assumptions, let us give some notations. Let $B(\chi, t) = \{\chi_1 \in \mathcal{E} : d(\chi_1, \chi) \leq t\}$ be the ball in $\mathcal{E}$ with center χ and radius t , and let $F_\chi(t) = P(d(\mathcal{X}, \chi) \leq t)$. It is clear that $F_\chi(t) = P(\mathcal{X} \in B(\chi, t))$ and when t is a decreasing sequence to zero, such quantity is called small ball probability. Further, define $\varphi_\chi(s) = E[\{r(\mathcal{X}) - r(\chi)\} | d(\mathcal{X}, \chi) = s]$ and $\tau_{h\chi}(s) = F_\chi(hs)/F_\chi(h) = P(d(\mathcal{X}, \chi) \leq hs | d(\mathcal{X}, \chi) \leq h)$ for $0 < s \leq 1$, and let $\tau_{0\chi}(s) = \lim_{h \downarrow 0} \tau_{h\chi}(s)$. Finally, denote $M_{0\chi} = K(1) - \int_0^1 (sK(s))' \tau_{0\chi}(s) ds$, $M_{1\chi} = K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) ds$, and $M_{2\chi} = K^2(1) - \int_0^1 (K^2)'(s) \tau_{0\chi}(s) ds$.

3.1 Assumptions

Most of the conditions are expressed in a pointwise way (i.e. for a fixed element χ of $\mathcal{E}$).

Regularity conditions: Conditions (C1)-(C5) below are standard regularity conditions, related to the smoothness and finiteness of the functions r , σ_ε^2 , φ_χ , F_χ , $\tau_{0\chi}$ and K . Most of them have been already introduced in Ferraty *et al.* (2007) (see comments therein).

- (C1) The functions $r(\chi)$, $\sigma_\varepsilon^2(\chi)$ and $E(|Y| | \chi)$ are continuous in a neighborhood of χ , and $\sup_{d(\chi_1, \chi) < \varepsilon} E(|Y|^m | \chi_1) < \infty$ for some $\varepsilon > 0$ and for all $m \geq 1$.
- (C2) For all (χ_1, s) in a neighborhood of $(\chi, 0)$, $\varphi_{\chi_1}(0) = 0$, $\varphi'_{\chi_1}(s)$ exists, $\varphi'_{\chi_1}(0) \neq 0$, and $\varphi'_{\chi_1}(s)$ is uniformly Lipschitz continuous of order $0 < \alpha \leq 1$ in (χ_1, s) .
- (C3) For all $\chi_1 \in \mathcal{E}$, $F_{\chi_1}(0) = 0$ and $F_{\chi_1}(t)/F_\chi(t)$ is Lipschitz continuous of order α in χ_1 , uniformly in t in a neighborhood of 0 (with α defined in (C2)).
- (C4) For all $\chi_1 \in \mathcal{E}$ and all $0 \leq s \leq 1$, $\tau_{0\chi_1}(s)$ exists, $\sup_{\chi_1 \in \mathcal{E}, 0 \leq s \leq 1} |\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)| = o(1)$, $M_{0\chi} > 0$, $M_{2\chi} > 0$, $\inf_{d(\chi_1, \chi) < \varepsilon} M_{1\chi_1} > 0$ for some $\varepsilon > 0$, and $M_{k\chi_1}$ is Lipschitz continuous of order α for $k = 0, 1, 2$ (with α defined in (C2)).
- (C5) K is supported on $[0, 1]$, K has a continuous derivative on $[0, 1)$, $K'(s) \leq 0$ for $0 \leq s < 1$, and $K(1) > 0$.

Conditions on the bandwidths $h = h_n$ and $b = b_n$:

(C6) $h, b \rightarrow 0$, $h/b \rightarrow 0$, $nF_\chi(h) \rightarrow \infty$, $h(nF_\chi(h))^{1/2} = O(1)$, $b^{1+\alpha}(nF_\chi(h))^{1/2} = o(1)$, $F_\chi(b+h)/F_\chi(b) \rightarrow 1$, $[F_\chi(h)/F_\chi(b)] \log n = o(1)$, and $bh^{\alpha-1} = O(1)$ (with α defined in (C2)).

Note that b has to be asymptotically larger than h , so oversmoothing is needed to make the bootstrap procedure work, as is the case for nonparametric regression in finite dimension. Note also that the optimal choice of h , namely $h = O((nF_\chi(h))^{-1/2})$ is allowed.

Technical condition:

(C7) For each n , there exist $r_n \geq 1$, $\ell_n > 0$ and curves $t_{1n}, \dots, t_{r_n n}$ such that $B(\chi, h) \subset \cup_{k=1}^{r_n} B(t_{kn}, \ell_n)$, with $r_n = O(n^{b/h})$ and $\ell_n = o(b(nF_\chi(h))^{-1/2})$.

Note that it follows from condition (C6) that a possible choice for ℓ_n is $\ell_n = bh(\log n)^{-1}$. The following proposition shows two examples of spaces $(\mathcal{E}, d)$ for which condition (C7) is satisfied. The proof is given in the Appendix. The proof of the second part is based on the fact that the number of balls of radius ℓ_n to cover $B(\chi, h)$ is equal to the number of balls of radius ℓ_n/h to cover $B(\chi/h, 1)$. This property is valid for any space $(\mathcal{E}, d)$ with $d(\chi_1, \chi_2) = \|\chi_1 - \chi_2\|$ for some semi-norm $\|\cdot\|$. Van der Vaart and Wellner (1996) give upper bounds for the so-called covering number of the ball $B(\chi/h, 1)$ for a large variety of functional spaces.

Proposition 3.1

1. Suppose that $\mathcal{E}$ is a separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and with orthonormal basis $\{e_j : j = 1, \dots, \infty\}$, and let $k > 0$ be a fixed integer. Let d_k be the semi-metric defined by:

$$d_k(\chi_1, \chi_2) = \sqrt{\sum_{j=1}^k \langle \chi_1 - \chi_2, e_j \rangle^2},$$

for any $\chi_1, \chi_2 \in \mathcal{E}$. Then, the space $(\mathcal{E}, d_k)$ satisfies condition (C7) provided $n^{b/h} b^k (\log n)^{-k} \rightarrow \infty$.

2. Suppose that $\mathcal{E}$ is the space of all continuous functions $\chi : [a, b] \rightarrow \mathbb{R}$ with $\|\chi\|_\gamma \leq M$, where $-\infty < a < b < \infty$, $0 < \gamma < \infty$,

$$\|\chi\|_\gamma = \max_{k \leq \underline{\gamma}} \sup_t |\chi^{(k)}(t)| + \sup_{t_1, t_2} \frac{|\chi^{(\underline{\gamma})}(t_1) - \chi^{(\underline{\gamma})}(t_2)|}{\|t_1 - t_2\|_2^{\gamma - \underline{\gamma}}},$$

$\|\cdot\|_2$ is the Euclidean norm, and $\underline{\gamma}$ is the largest integer strictly smaller than γ . Then, the space $(\mathcal{E}, d_{L^p})$ satisfies condition (C7) provided $hb^{-(1+1/\gamma)}(\log n)^{-1+1/\gamma} = o(1)$, where d_{L^p} is the L^p -distance in $\mathcal{E}$ and $1 \leq p \leq \infty$.

3.2 Main result

We are now ready to state the main result of this paper:

Theorem 3.2 *Assume (C1)-(C7). Then, for both the naive and the wild bootstrap procedure, we have*

$$\sup_{y \in \mathbb{R}} \left| P^{\mathcal{S}} \left(\sqrt{nF_{\chi}(h)} \left\{ \hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi) \right\} \leq y \right) - P \left(\sqrt{nF_{\chi}(h)} \left\{ \hat{r}_h(\chi) - r(\chi) \right\} \leq y \right) \right| \xrightarrow{a.s.} 0,$$

where $P^{\mathcal{S}}$ denotes probability, conditionally on the sample $\mathcal{S}$ (i.e. $(\mathcal{X}_i, Y_i)$, $i = 1, \dots, n$).

The proof for the naive bootstrap is given in the Appendix. The proof for the wild bootstrap is very similar, and is therefore omitted. The only difference with the proof for the naive bootstrap is the calculation of the variance of the estimator $\hat{r}_{hb}^{boot}(\chi)$, given in Lemma A.3. We explain in Remark A.1 how this proof needs to be adapted for the wild bootstrap.

Remark. Because of Slutsky's Theorem and the fact that $nF_{\chi}(h)$ tends to $+\infty$ with n , it is easy to show that this result remains valid if $F_{\chi}(h)$ is replaced by its empirical version $n^{-1} \sum_{i=1}^n 1_{B(\chi, h)}(\mathcal{X}_i)$.

3.3 Building confidence intervals

The theorem given just before shows, from a theoretical point of view, the good asymptotic behavior of the bootstrapped error $\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)$ with respect to the true error $\hat{r}_h(\chi) - r(\chi)$ in this functional nonparametric regression setting (i.e. when the explanatory variable is functional). Now, from a practical point of view, this result is useful for building a confidence interval for $r(\chi)$. Indeed, for a fixed sample $\mathcal{S}$, one can compute $\hat{r}_b(\chi)$ and one can repeat several times the bootstrap procedure in order to get bootstrapped estimators $\hat{r}_{hb}^{boot1}(\chi), \hat{r}_{hb}^{boot2}(\chi), \dots$. The pointwise α -quantile $t_{\alpha}(\chi)$ of the distribution of the *true* error (i.e. $P(\hat{r}_h(\chi) - r(\chi) < t_{\alpha}(\chi)) = \alpha$) can be approximated by $t_{\alpha}^*(\chi)$, the α -quantile computed from the distribution of the bootstrapped errors (i.e. $\hat{r}_{hb}^{boot1}(\chi) - \hat{r}_b(\chi), \hat{r}_{hb}^{boot2}(\chi) - \hat{r}_b(\chi), \dots$). In summary, the $100(1 - 2\alpha)\%$ -confidence interval $[\hat{r}_h(\chi) + t_{\alpha}(\chi), \hat{r}_h(\chi) + t_{1-\alpha}(\chi)]$ can be approximated by means of the percentile method by $[\hat{r}_h(\chi) + t_{\alpha}^*(\chi), \hat{r}_h(\chi) + t_{1-\alpha}^*(\chi)]$.

4 Wild bootstrap in action

4.1 Simulation study

The main goal of this section is to illustrate the theoretical result throughout simulated data. One way to do that consists in comparing the density f_{true} of the true error $\hat{r}_h(\chi) - r(\chi)$ with its bootstrapped version (i.e. the density f_{boot} of the bootstrapped error $\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)$). We also study the sensitivity of the bootstrapped error with respect to both bandwidths b and h .

Building the sample $(\mathcal{X}_i, Y_i)_{i=1,2,\dots,250}$. First of all, one builds simulated discretized curves:

$$\mathcal{X}_i(t_j) = a_i \cos(2t_j) + b_i \sin(4t_j) + c_i(t_j^2 - \pi t_j + \frac{2}{9}\pi^2), \quad i = 1, 2, \dots, 250,$$

where $0 = t_1 < t_2 < \dots < t_{99} < t_{100} = \pi$ are equispaced points, the a_i 's, b_i 's and c_i 's being independent observations uniformly distributed on $[0, 1]$. Figure 1 (left panel) gives an idea of their shape. Once the curves defined, one simulates a functional regression model in order to compute the responses (i.e. Y_i 's):

1. build a regression operator r :
$$\begin{cases} \text{model 1: } r(\mathcal{X}_i) = 10 \times (a_i^2 - b_i^2), \\ \text{model 2: } r(\mathcal{X}_i) = \int t \cos(t) \left(\mathcal{X}_i'(t) \right)^2 dt, \end{cases}$$
2. generate $\varepsilon_1, \varepsilon_2, \dots$, independent centered gaussians of variance equal to 0.1 times the empirical variance of $r(\mathcal{X}_1), r(\mathcal{X}_2), \dots$ (i.e. signal-to-noise ratio= 0.1),
3. compute the corresponding responses: $Y_i = r(\mathcal{X}_i) + \varepsilon_i$.

Estimating the density of the true error. For a fixed curve χ and a fixed bandwidth h , one has to compute the density f_{true} of $\hat{r}_h(\chi) - r(\chi)$. To do that, one uses the following Monte-Carlo scheme:

1. build 100 datasets: $\left\{ (\mathcal{X}_i^s, Y_i^s)_{i=1,\dots,250} \right\}_{s=1,\dots,100}$,
2. compute 100 estimates $\{\hat{r}_h^s(\chi) - r(\chi)\}_{s=1,\dots,100}$, where $\hat{r}_h^s$ is the functional kernel estimator of the regression operator r computed over the s th dataset,
3. estimate the density f_{true} by using any standard density estimator over the 100 values $\{\hat{r}_h^s(\chi) - r(\chi)\}_{s=1,\dots,100}$.

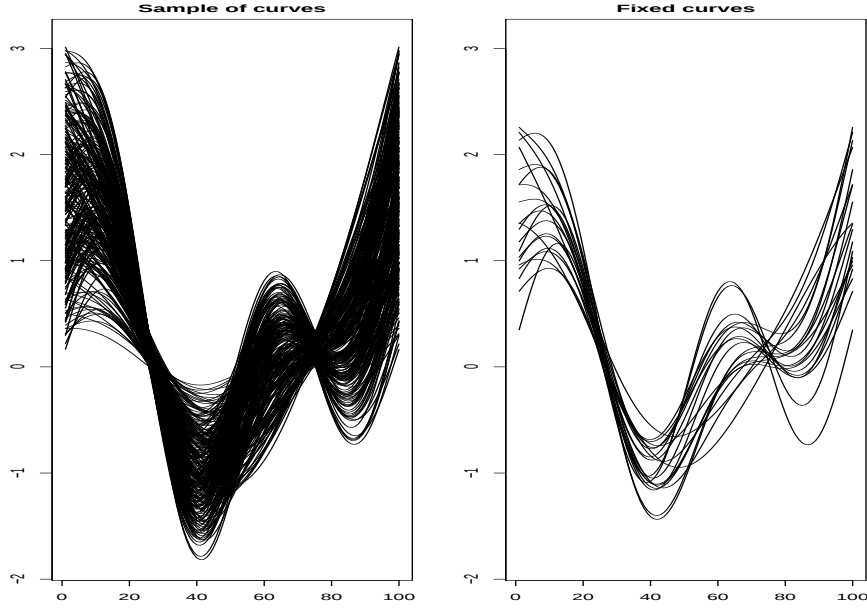


Figure 1: Left panel: 100 simulated curves from the learning sample; right panel: 20 additional fixed curves (i.e. 20 curves at which the regression operator is evaluated).

Computing the density of the bootstrapped error. For a fixed curve χ and two fixed bandwidths h and b , one estimates the density f_{boot} of $\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)$ by considering one dataset $\mathcal{S} = (\mathcal{X}_i, Y_i)_{i=1, \dots, 250}$ and the wild bootstrap procedure (see Section 2). More precisely, we use the same procedure described in Härdle and Marron (1991): $V_1, V_2, \dots, V_{250}$ are i.i.d. random variables drawn from the sum of the following two Dirac distributions: $0.1(5 + \sqrt{5})\delta_{(1-\sqrt{5})/2} + 0.1(5 - \sqrt{5})\delta_{(1+\sqrt{5})/2}$. Such a distribution was introduced because it ensures that $E(V_1) = 0$ and $E(V_1^2) = E(V_1^3) = 1$, which implies, for $i = 1, \dots, 250$ that ε_i^{boot} and $\hat{\varepsilon}_{i,b}$ have the same first three moments. The three following steps allow us to estimate the density f_{boot} :

1. compute $\hat{r}_b(\chi)$ over the dataset $\mathcal{S}$,
2. repeat 100 times the bootstrap algorithm over the same dataset $\mathcal{S}$,
3. estimate the density f_{boot} by using any standard density estimator over the 100 values $\hat{r}_{hb}^{boot1}(\chi) - \hat{r}_b(\chi), \hat{r}_{hb}^{boot2}(\chi) - \hat{r}_b(\chi), \dots, \hat{r}_{hb}^{boot100}(\chi) - \hat{r}_b(\chi)$.

Results. 20 additional curves $\chi_1, \dots, \chi_{20}$ are simulated (see Figure 1: right panel) which allows us to compare $\hat{r}_h(\chi) - r(\chi)$ and $\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)$ for $\chi \in \{\chi_1, \dots, \chi_{20}\}$. These are what we call *fixed* curves. For the kernel estimator of r , in Model 1 the selected

semi-metric is $d_3(\cdot, \cdot)$ and in Model 2, we took $d_1(\cdot, \cdot)$ where, for all $(\chi_1, \chi_2) \in \mathcal{E} \times \mathcal{E}$:

$$d_k(\chi_1, \chi_2) = \sqrt{\int_0^\pi \left(\chi_1^{(k)}(t) - \chi_2^{(k)}(t) \right)^2 dt},$$

$\chi_1^{(k)}$ being the k th derivative of χ_1 . These particular choices are discussed later on.

For the comparison between the true error and the bootstrapped one, the bandwidth h is defined in terms of k nearest neighbors and the number k is selected through a cross-validation procedure (here $k = 9$). We also take the same bandwidth for b . Figure 2 (resp. 3) compares these densities at 20 fixed curves $\chi_1, \dots, \chi_{20}$ in Model 1 (resp. 2). For each fixed curve at which the regression operator is evaluated, we indicate between brackets the relative squared L^2 distance between the true and the bootstrapped densities (i.e. $\delta_{b,h} = \int (f_{true} - f_{boot})^2 / \int f_{true}^2$). The first feeling when one looks at these densities is that the wild bootstrap procedure works well for both models for h and b fixed as indicated. Now, an interesting question is the following: does it still work well when both b and h vary? This is the purpose of the next paragraph.

From a practical point of view, it is interesting to study the sensitivity of the density of the bootstrapped errors with respect to the bandwidths b and h . Even if a standard choice consists in choosing the bandwidths throughout a cross-validation procedure (which was the case for the previous results), it can be important to see what happens when both bandwidths b and h vary. Therefore, we compute, at each fixed curve $\chi_1, \dots, \chi_{20}$, the relative squared distance $\delta_{b_k, h_{k'}}$ where b_k (resp. $h_{k'}$) corresponds to the bandwidths taking into account k (resp. k') nearest neighbors. Now, let k (resp. k') vary in $\{8, 10, \dots, 20\}$. At each fixed curve, we get the $7 \times 7 = 49$ quantities $\delta_{b_k, h_{k'}}$ ($k = 8, 10, \dots, 20$ and $k' = 8, 10, \dots, 20$). Figure 4 displays, for both models and at each fixed curve, the box-and-whiskers summarizing the distribution of the $\delta_{b_k, h_{k'}}$'s. The horizontal dashed line indicates an empirical bound (determined from Figures 2 and 3) under which one can consider that f_{true} and f_{boot} are close. Model 1 shows that the bootstrapped errors are not sensitive with respect to both bandwidths h and b (i.e. the relative L^2 norm between bootstrapped/true error density remains small in any case). Model 2 is more complicated than Model 1 and seems to be more sensitive with respect to the bandwidths. However, for most of the 20 fixed curves, the results confirm the small sensitivity with respect to h and b .

About the semi-metric d : a tool for vanishing nuisance parameter. One can remark that the coefficients c_i 's (allowing to build the simulated curves) can be considered as “nuisance” parameters in Model 1. Indeed, the considered regression model does not

Model 1

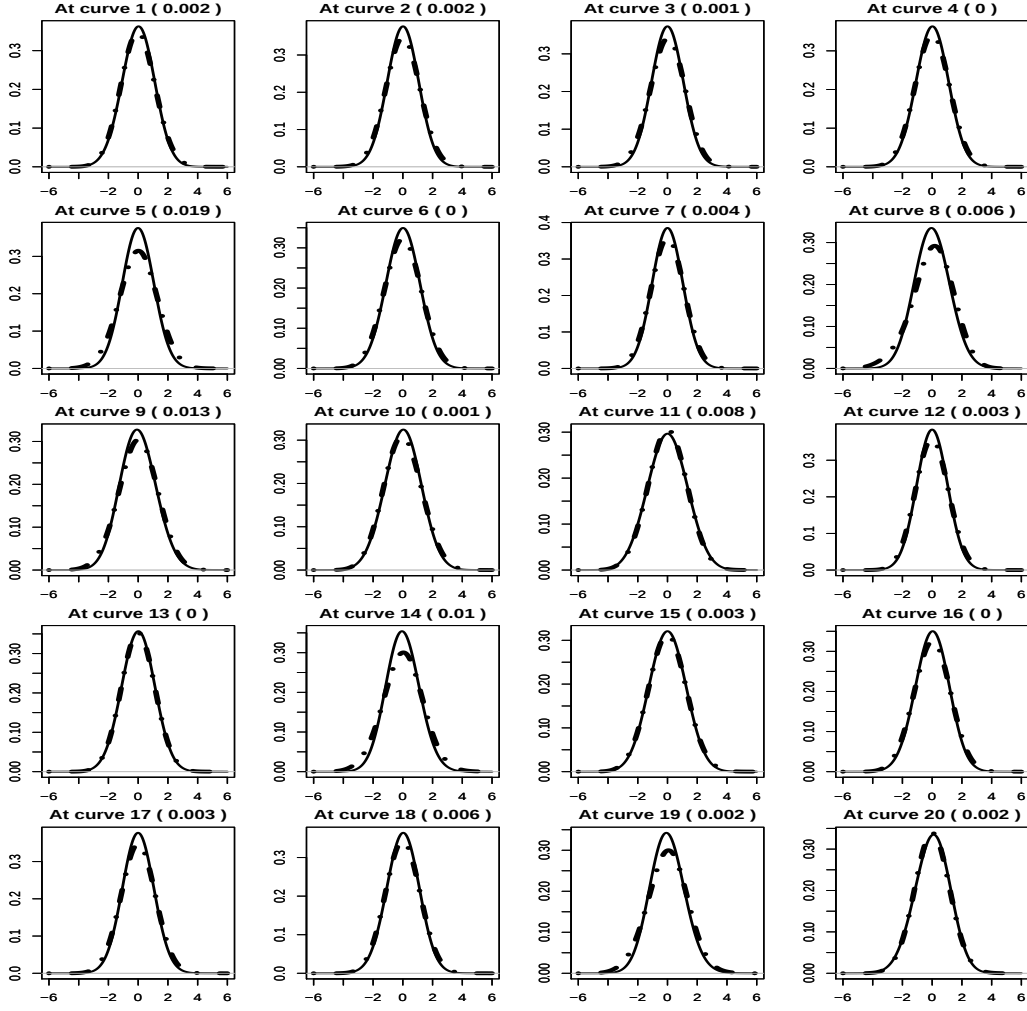


Figure 2: The solid curve represents the density of the bootstrapped error, the dashed curve the density of the true error.

use the c_i 's. Equivalently, one can say that the polynomial part present in the curves is non-informative in this model. So, one way to cancel the nuisance-effect of the c_i 's (resp. polynomial part) is to use the third derivative of the curves instead of the curves themselves which amounts to use d_3 . Note also that the third derivative leads to the best mean square error between the observed responses and their estimations when one uses other derivatives (including the standard L^2 norm).

Concerning model 2, the first derivative of the curves is directly used in the regression operator. So, a natural choice of semi-metric is d_1 .

Model 2

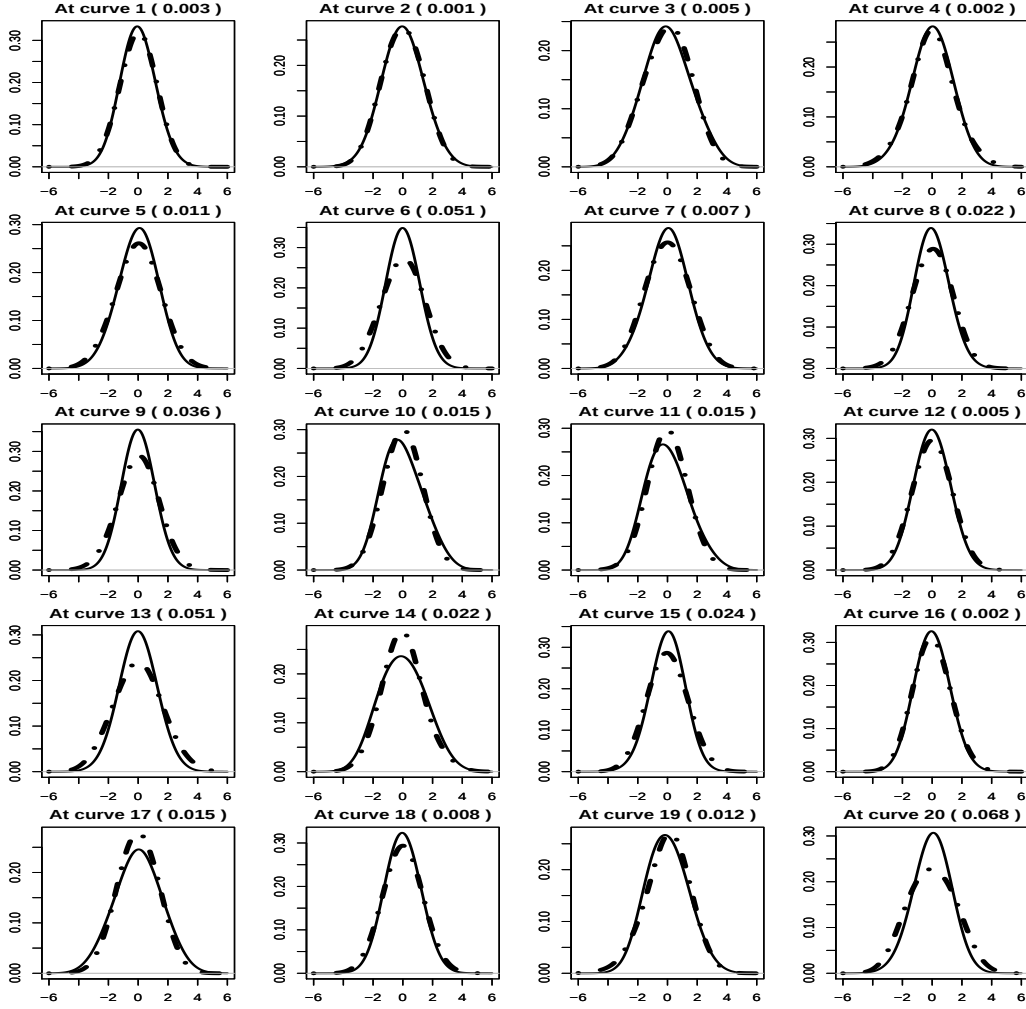


Figure 3: The solid curve represents the density of the bootstrapped error, the dashed curve the density of the true error.

4.2 About a food industry quality control problem

Food industry, in relation with chemometrics, is a good example where functional datasets occur frequently. In particular, when one wishes to determine the fat content of pieces of meat, an economical way is to use a spectrometer in the near infrared (NIR). For each piece of meat, the spectrometer produces one spectrometric curve. Finally, one has at hand 215 pieces of meat which produce 215 spectrometric curves $\mathcal{X}_1, \dots, \mathcal{X}_{215}$ (see Figure 5). The first step is to study the relationship between the fat content and the corresponding spectrometric curve. Sometimes, this kind of statistical problem is called “calibration problem” in the literature (see, for instance, Martens and Naes, 1988). To this end, a

Stability with respect to both bandwidths (h and b)

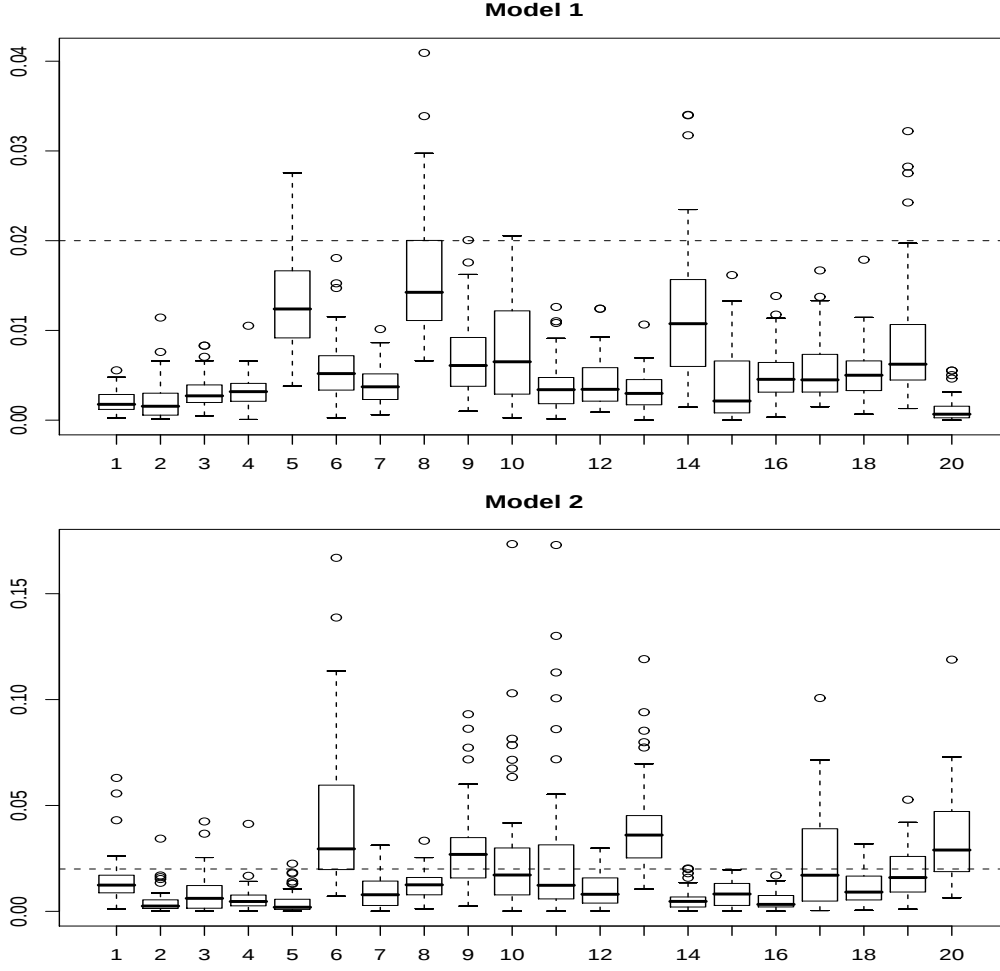


Figure 4: Top panel: results from Model 1; bottom panel: results from Model 2.

chemical process allows to collect the fat content for each piece of meat $(Y_1, \dots, Y_{215})$. Finally, one has at hand $(\mathcal{X}_i, Y_i)_{i=1, \dots, 215}$, which is a collection of 215 i.i.d. pairs. So the usual challenge for any statistician consists in estimating the relationship between a scalar response (fat content) from a functional explanatory variable (spectrometric curve). From a regression point of view, this situation amounts to estimating the regression operator r in the model $Y = r(\mathcal{X}) + \text{error}$ and a deep study can be found in Ferraty and Vieu (2006).

Here, we are interested in computing the confidence intervals. To this end, we split the functional dataset into two samples: the learning sample $(\mathcal{X}_i, Y_i)_{i=1, \dots, 160}$ allowing to build the estimator and the 55 fixed curves $\mathcal{X}_{161}, \dots, \mathcal{X}_{215}$ at which the estimated regression operator is evaluated. Figure 6 gives the 95% confidence intervals $[\hat{r}_h(\chi) \pm t_{0.025}^*(\chi)]$,

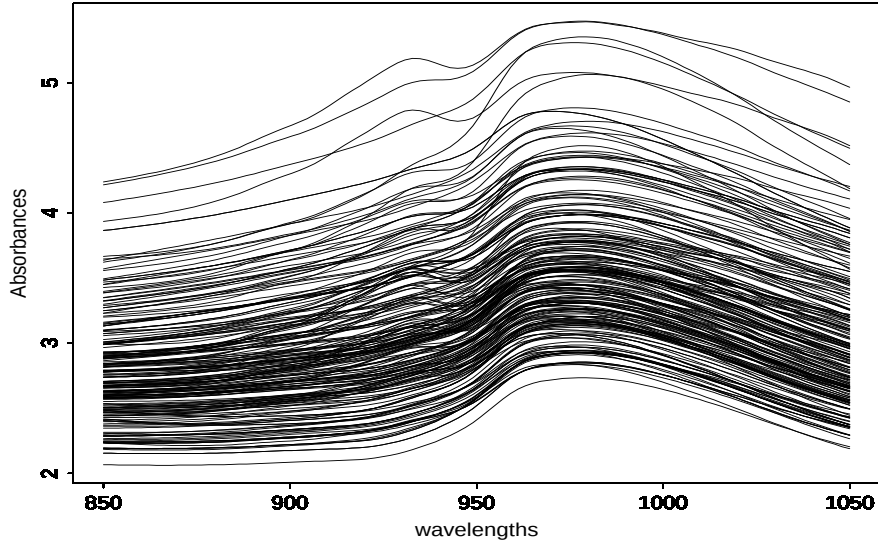


Figure 5: Sample of spectrometric curves.

$\hat{r}_h(\chi) + t_{0.975}^*(\chi)]$ for $\chi \in \{\mathcal{X}_{161}, \dots, \mathcal{X}_{215}\}$, which approximate the confidence intervals of the true error (i.e. $[\hat{r}_h(\chi) + t_{0.025}(\chi), \hat{r}_h(\chi) + t_{0.975}(\chi)]$ for $\chi \in \{\mathcal{X}_{161}, \dots, \mathcal{X}_{215}\}$). The errors were bootstrapped 500 times (i.e. one gets $\hat{r}_{hb}^{boot1}(\chi), \dots, \hat{r}_{hb}^{boot500}(\chi)$). The scatterplot corresponds to the display of observed responses versus the predicted ones. The bandwidth h is selected by a cross-validation procedure and we set $b = h$. It is worth noting that the confidence intervals in Figure 6 are not prediction intervals, and this is why they are not necessarily symmetric.

5 Conclusions/Perspectives

This first theoretical study validates the use of the bootstrap procedure in the nonparametric regression setting when one has at hand a functional covariate. Its good asymptotic behavior as well as its easy implementation make this functional bootstrap method very attractive. From a general point of view, this work proposes a general methodology for showing the validity of the bootstrap in nonparametric regression when the covariate is functional. In this paper we have focused on the mean, but one could also show the validity of the bootstrap when one uses the median or the mode as predictor. An open question for further investigation concerns the situation where the response is also a functional variable. In that case, the errors are functional variables, and one needs to bootstrap these functional errors. Can we expect to get similar results?

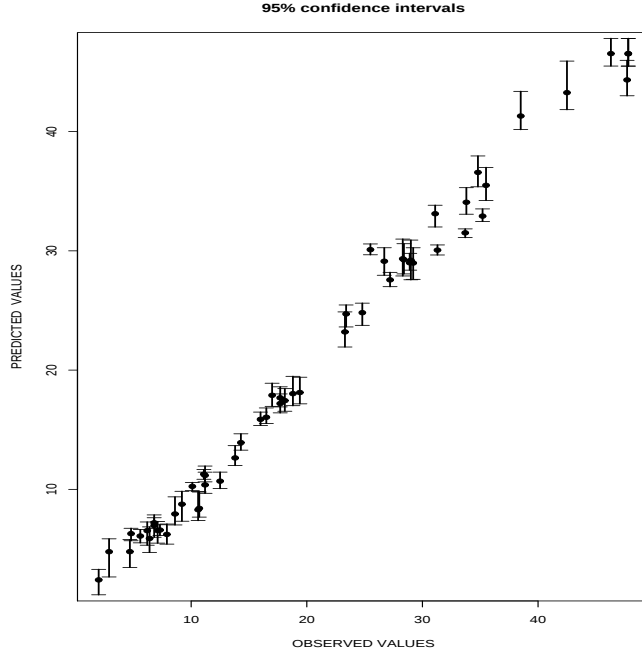


Figure 6: Pointwise confidence intervals computed by means of the bootstrap procedure.

Acknowledgments. All the participants of the STAPH group on Functional Statistics in the Mathematical Institute of Toulouse are greatly acknowledged for their helpful comments on this work (activities of this group are available at the webpage <http://www.lsp.univ-tlse/staph>). In addition, the second author acknowledges support from IAP research network nr. P6/03 of the Belgian government (Belgian Science Policy) and the financial support delivered by the Toulouse University as invited Professor.

Appendix

Proof of Proposition 3.1. We start with proving the first part of the proposition. First note that it follows from Lemma 13.6 in Ferraty and Vieu (2006) that d_k is a semi-metric. The result follows from Lemma A.1 below, by choosing $\mathcal{C} = B(\chi, h)$, $\ell = \ell_n$ and $\tau = r_n$, which implies that $r_n \leq K(h/\ell_n)^k$ for some $K > 0$, and hence condition (C7) is satisfied by choosing $\ell_n = bh(\log n)^{-1}$ and provided $n^{b/h}b^k(\log n)^{-k} \rightarrow \infty$.

Next, we prove the second part. First note that $B(\chi, h) = \{h\chi_1 : \chi_1 \in B(\chi/h, 1)\}$. From Theorem 2.7.1 in Van der Vaart and Wellner (1996) we know that the number of balls of radius ℓ_n/h to cover $B(\chi/h, 1)$ with respect to the L^p ($1 \leq p \leq \infty$) metric, is

bounded by

$$r_n \equiv \exp \left(K \left(\frac{h}{\ell_n} \right)^{1/\gamma} \right)$$

for some $K > 0$. Let $z_1, \dots, z_{r_n}$ be the centers of these r_n balls. Fix $\chi_2 \in B(\chi, h)$. Then, there exists a $\chi_1 \in B(\chi/h, 1)$ such that $\chi_2 = h\chi_1$. Suppose $\chi_1 \in B(z_j, \ell_n/h)$ for some $j = 1, \dots, r_n$. Calculate

$$d_{L^p}(\chi_2, h z_j) = h d_{L^p}(\chi_1, z_j) \leq \ell_n,$$

and hence $\chi_2 \in B(h z_j, \ell_n)$. It now follows that the balls $B(h z_j, \ell_n)$, $j = 1, \dots, r_n$ cover $B(\chi, h)$.

It remains to verify the conditions on ℓ_n and r_n . Take $\ell_n = b h (\log n)^{-1}$, which is allowed by condition (C6). Then, we need to check whether $r_n = \exp(K b^{-1/\gamma} (\log n)^{1/\gamma})$ satisfies $r_n = O(n^{b/h})$. An easy calculation shows that this is true under the stated conditions on h and b . $\square$

Lemma A.1 *Assume that $\mathcal{H}$ is a separable Hilbert space, with inner product $\langle \cdot, \cdot \rangle$ and with orthonormal basis $\{e_j : j = 1, \dots, \infty\}$. Let $k > 0$ be a fixed integer, and d_k be the semi-metric defined in the statement of Proposition 3.1. Then, for any closed and bounded subset $\mathcal{C}$ in $(\mathcal{H}, d_k)$ and any $\ell > 0$, there exist $K > 0$, $\tau \geq 1$ and $t_1, \dots, t_\tau$ such that*

$$\mathcal{C} \subset \cup_{i=1}^\tau B_k(t_i, \ell) \quad \text{and} \quad \tau \leq K \left(\frac{\Delta}{\ell} \right)^k,$$

where Δ is the diameter of $\mathcal{C}$, i.e. $\Delta = \sup\{d_k(\chi_1, \chi_2) : \chi_1, \chi_2 \in \mathcal{C}\}$, and $B_k(\cdot, \cdot)$ is an open ball in $\mathcal{H}$ for the semi-metric d_k .

Proof. Let ϕ be the operator from $\mathcal{H}$ into $\mathbb{R}^k$ defined by

$$\phi(\chi) = (\langle \chi, e_1 \rangle, \dots, \langle \chi, e_k \rangle).$$

Let d_{eucl} be the Euclidean distance in $\mathbb{R}^k$, and let us denote by $B_{eucl}(\cdot, \cdot)$ an open ball in $\mathbb{R}^k$ for the associated topology. By construction of the map ϕ between the topological spaces $(\mathcal{H}, d_k)$ and $(\mathbb{R}^k, d_{eucl})$, we have that $\phi(\mathcal{C})$ is a compact subset of $\mathbb{R}^k$. Therefore, by using standard results for finite dimensional Euclidean topologies, for each $\ell > 0$ we can cover $\phi(\mathcal{C})$ by balls of centers $z_i \in \mathbb{R}^k$:

$$\phi(\mathcal{C}) \subset \cup_{i=1}^\tau B_{eucl}(z_i, \ell), \text{ with } \tau \leq K \left(\frac{\Delta}{\ell} \right)^k.$$

Note that this is possible because by construction $\mathcal{C}$ and $\phi(\mathcal{C})$ have the same diameter.

For $i = 1, \dots, \tau$, let χ_i be an element of $\mathcal{H}$ such that $\phi(\chi_i) = z_i$. The solution of the equation $\phi(\chi) = z_i$ is not unique in general, but just take χ_i to be one of these solutions. From the definition of d_k we have that:

$$\phi^{-1}(B_{eucl}(z_i, \ell)) = B_k(\chi_i, \ell).$$

Hence, $\mathcal{C}$ can be covered by the balls $B_k(\chi_i, \ell)$, $i = 1, \dots, \tau$. $\square$

Proof of Theorem 3.2. The expression between absolute values in the statement of the theorem can be written as

$$\begin{aligned} & P^{\mathcal{S}}\left(\sqrt{nF_{\chi}(h)}\left\{\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)\right\} \leq y\right) - \Phi\left(\frac{y - \sqrt{nF_{\chi}(h)}\{E^{\mathcal{S}}\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)\}}{\sqrt{nF_{\chi}(h)\text{Var}^{\mathcal{S}}(\hat{r}_{hb}^{boot}(\chi))}}\right) \\ & + \Phi\left(\frac{y - \sqrt{nF_{\chi}(h)}\{E^{\mathcal{S}}\hat{r}_{hb}^{boot}(\chi) - \hat{r}_b(\chi)\}}{\sqrt{nF_{\chi}(h)\text{Var}^{\mathcal{S}}(\hat{r}_{hb}^{boot}(\chi))}}\right) - \Phi\left(\frac{y - \sqrt{nF_{\chi}(h)}\{E\hat{r}_h(\chi) - r(\chi)\}}{\sqrt{nF_{\chi}(h)\text{Var}(\hat{r}_h(\chi))}}\right) \\ & + \Phi\left(\frac{y - \sqrt{nF_{\chi}(h)}\{E\hat{r}_h(\chi) - r(\chi)\}}{\sqrt{nF_{\chi}(h)\text{Var}(\hat{r}_h(\chi))}}\right) - P\left(\sqrt{nF_{\chi}(h)}\left\{\hat{r}_h(\chi) - r(\chi)\right\} \leq y\right) \\ & = T_1(y) + T_2(y) + T_3(y), \end{aligned}$$

where $E^{\mathcal{S}}$ and $\text{Var}^{\mathcal{S}}$ denote expectation and variance, conditionally on the sample $\mathcal{S}$, and where Φ is the standard normal distribution function. The a.s. convergence to zero of $T_1(y)$ and $T_3(y)$ for a fixed value of y is given by Lemma A.2 below. Hence, the uniform convergence over all $y \in \mathbb{R}$ follows from Polya's Theorem (see e.g. Serfling, 1980, p. 18) together with the continuity of the function Φ . It remains to consider $T_2(y)$. Note that for any $a \in \mathbb{R}$ and $b > 0$,

$$\sup_y |\Phi(a + by) - \Phi(y)| \leq |a| + \max(b, b^{-1}) - 1,$$

and hence the convergence of $\sup_y |T_2(y)|$ follows from Lemmas A.3 and A.4 below. $\square$

For the proofs of the lemmas below, we need to introduce the following estimators:

$$\hat{g}_{hb}^{boot}(\chi) = (nF_{\chi}(h))^{-1} \sum_{i=1}^n Y_i^{boot} K\left(h^{-1}d(\mathcal{X}_i, \chi)\right)$$

and

$$\hat{f}_h(\chi) = (nF_{\chi}(h))^{-1} \sum_{i=1}^n K\left(h^{-1}d(\mathcal{X}_i, \chi)\right). \quad (\text{A.1})$$

It is clear that $\hat{r}_{hb}^{boot}(\chi) = \hat{g}_{hb}^{boot}(\chi)/\hat{f}_h(\chi)$.

Lemma A.2 Assume (C1)-(C6). Then,

$$\frac{\widehat{r}_h(\chi) - E[\widehat{r}_h(\chi)]}{\sqrt{\text{Var}[\widehat{r}_h(\chi)]}} \xrightarrow{d} N(0, 1)$$

and

$$\frac{\widehat{r}_{hb}^{boot}(\chi) - E^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)]}{\sqrt{\text{Var}^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)]}} \xrightarrow{d} N(0, 1),$$

a.s., conditionally on the sample $\mathcal{S}$.

Proof. The first statement follows from Theorem 2 in Ferraty *et al.* (2007). For the second one, note that

$$\frac{\widehat{r}_{hb}^{boot}(\chi) - E^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)]}{\sqrt{\text{Var}^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)]}} = \frac{\widehat{g}_{hb}^{boot}(\chi) - E^{\mathcal{S}}[\widehat{g}_{hb}^{boot}(\chi)]}{\sqrt{\text{Var}^{\mathcal{S}}[\widehat{g}_{hb}^{boot}(\chi)]}}.$$

Since $\widehat{g}_{hb}^{boot}(\chi)$ is a sum of independent terms, the asymptotic normality follows after checking Liapunov's condition (see e.g. Serfling, 1980, p. 32):

$$\frac{\sum_{i=1}^n E^{\mathcal{S}} \left| (nF_{\chi}(h))^{-1} \left\{ Y_i^{boot} - E^{\mathcal{S}}(Y_i^{boot}) \right\} K \left(h^{-1}d(\mathcal{X}_i, \chi) \right) \right|^3}{\left\{ \text{Var}^{\mathcal{S}} \left(\widehat{g}_{hb}^{boot}(\chi) \right) \right\}^{3/2}} \xrightarrow{a.s.} 0. \quad (\text{A.2})$$

The denominator in (A.2) is $O((nF_{\chi}(h))^{-3/2})$ a.s. by Lemma A.3 below and by the fact that $\text{Var}(\widehat{r}_h(\chi)) = O((nF_{\chi}(h))^{-1})$, whereas it can be easily shown that the numerator is $O((nF_{\chi}(h))^{-2})$ a.s. Hence, the Liapunov ratio in (A.2) is $((nF_{\chi}(h))^{-1/2}) = o(1)$ a.s. $\square$

Lemma A.3 Assume (C1)-(C6). Then,

$$\frac{\text{Var}^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)]}{\text{Var}[\widehat{r}_h(\chi)]} \xrightarrow{a.s.} 1.$$

Proof. Define $\widehat{\sigma}_{\varepsilon}^2 = n^{-1} \sum_{i=1}^n (\widehat{\varepsilon}_{i,b} - \bar{\varepsilon}_b)^2$. Then,

$$\begin{aligned} \text{Var}^{\mathcal{S}}[\widehat{r}_{hb}^{boot}(\chi)] &= \frac{\widehat{\sigma}_{\varepsilon}^2}{\widehat{f}_h(\chi)^2} (nF_{\chi}(h))^{-2} \sum_{i=1}^n K^2(h^{-1}d(\mathcal{X}_i, \chi)) \\ &= \frac{\sigma_{\varepsilon}^2}{E[\widehat{f}_h(\chi)]^2} (nF_{\chi}(h))^{-1} E[K^2(h^{-1}d(\mathcal{X}, \chi))] (1 + o(1)) \\ &= \frac{\sigma_{\varepsilon}^2}{M_{1\chi}^2} \frac{M_{2\chi}}{nF_{\chi}(h)} (1 + o(1)) \\ &= \text{Var}[\widehat{r}_h(\chi)] + o((nF_{\chi}(h))^{-1}) \quad \text{a.s.,} \end{aligned}$$

where the last and the one but last equalities follow from Lemma 4, Lemma 5 and Theorem 1 in Ferraty *et al.* (2007). Hence, $\text{Var}^S[\widehat{r}_{hb}^{boot}(\chi)]/\text{Var}[\widehat{r}_h(\chi)]$ converges to one, a.s., since $\text{Var}[\widehat{r}_h(\chi)] = O((nF_\chi(h))^{-1})$. $\square$

Remark A.1. When the wild bootstrap is used, the proof of the above Lemma A.3 needs to be adapted. We have in that case:

$$\begin{aligned}\text{Var}^S[\widehat{r}_{hb}^{boot}(\chi)] &= \frac{1}{\widehat{f}_h(\chi)^2} (nF_\chi(h))^{-2} \sum_{i=1}^n \widehat{\varepsilon}_{i,b}^2 K^2(h^{-1}d(\mathcal{X}_i, \chi)) \\ &= \frac{1}{E[\widehat{f}_h(\chi)]^2} (nF_\chi^2(h))^{-1} E[\varepsilon_i^2 K^2(h^{-1}d(\mathcal{X}, \chi))] (1 + o(1)) \\ &= \frac{\sigma_\varepsilon^2(\chi)}{M_{1\chi}^2} \frac{M_{2\chi}}{nF_\chi(h)} (1 + o(1)) \\ &= \text{Var}[\widehat{r}_h(\chi)] + o((nF_\chi(h))^{-1}) \quad \text{a.s.}\end{aligned}$$

All the other proofs in the Appendix remain valid for the wild bootstrap. $\square$

Lemma A.4 Assume (C1)-(C7). Then,

$$\sqrt{nF_\chi(h)} \left\{ E^S[\widehat{r}_{hb}^{boot}(\chi)] - \widehat{r}_b(\chi) - E[\widehat{r}_h(\chi)] + r(\chi) \right\} \xrightarrow{\text{a.s.}} 0.$$

Proof. Write

$$\begin{aligned}& E^S[\widehat{r}_{hb}^{boot}(\chi)] - \widehat{r}_b(\chi) \\ &= \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^n \{ \widehat{r}_b(\mathcal{X}_i) - \widehat{r}_b(\chi) - E[\widehat{r}_b(\mathcal{X}_i)] + E[\widehat{r}_b(\chi)] \} K(h^{-1}d(\mathcal{X}_i, \chi)) \\ &\quad + \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^n \{ E[\widehat{r}_b(\mathcal{X}_i)] - E[\widehat{r}_b(\chi)] - r(\mathcal{X}_i) + r(\chi) \} K(h^{-1}d(\mathcal{X}_i, \chi)) \\ &\quad + \frac{(nF_\chi(h))^{-1}}{\widehat{f}_h(\chi)} \sum_{i=1}^n \{ r(\mathcal{X}_i) - r(\chi) \} K(h^{-1}d(\mathcal{X}_i, \chi)) \\ &= U_1 + U_2 + U_3.\end{aligned}$$

It is easily seen that $U_3 = E[\widehat{r}_h(\chi)] - r(\chi) + o((nF_\chi(h))^{-1/2})$ a.s., whereas U_1 and U_2 are $o((nF_\chi(h))^{-1/2})$ a.s. by Lemmas A.5 and A.6 below. $\square$

Lemma A.5 Assume (C1)-(C6). Then,

$$\sup_{d(\chi_1, \chi) \leq h} \left| E[\widehat{r}_b(\chi_1)] - E[\widehat{r}_b(\chi)] - r(\chi_1) + r(\chi) \right| = o((nF_\chi(h))^{-1/2}) \text{ a.s.}$$

Proof. From Theorem 1 in Ferraty *et al.* (2007) it follows that for each fixed χ, χ_1 ,

$$\begin{aligned} & \left| E[\widehat{r}_b(\chi_1)] - r(\chi_1) - E[\widehat{r}_b(\chi)] + r(\chi) \right| \\ &= \left| \varphi'_{\chi_1}(0) \frac{M_{0\chi_1}}{M_{1\chi_1}} - \varphi'_\chi(0) \frac{M_{0\chi}}{M_{1\chi}} \right| b + O((nF_\chi(b))^{-1}) + O((nF_{\chi_1}(b))^{-1}) + O(b^{1+\alpha}). \end{aligned}$$

Note that the rate of the last remainder term follows from the fact that $\varphi'_{\chi_1}(s)$ is Lipschitz of order α in s uniformly in χ_1 . To show that the order of the remainder terms holds uniformly for all χ_1 for which $d(\chi_1, \chi) \leq h$, it follows from the proof of the above theorem that we need to show that

$$\sup_{d(\chi_1, \chi) \leq h} \left| \int_0^1 [\varphi_{\chi_1}(ht) - h\varphi'_{\chi_1}(0)t] K(t) dF_{\chi_1}(ht) \right| = o(h), \quad (\text{A.3})$$

$$\sup_{d(\chi_1, \chi) \leq h} \left| \int_0^1 [sK(s)]' [\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)] ds \right| = o(1), \quad (\text{A.4})$$

$$\sup_{d(\chi_1, \chi) \leq h} \frac{F_{\chi_1}(b)}{F_\chi(b)} = O(1). \quad (\text{A.5})$$

First consider (A.3). From a first order Taylor expansion of $\varphi_{\chi_1}(ht)$ and the continuity of $\varphi'_{\chi_1}(s)$ in s uniformly for (χ_1, s) in a neighborhood of $(\chi, 0)$, it follows that (A.3) is bounded by

$$h \sup_{d(\chi_1, \chi) \leq h, |s| \leq h} |\varphi'_{\chi_1}(s) - \varphi'_{\chi_1}(0)| \left| \int t K(t) dF_{\chi_1}(ht) \right| = o(h),$$

uniformly in χ_1 , since $\sup_t |tK(t)| < \infty$. Next, from condition (C4) it follows that (A.4) is bounded by

$$\sup_{\chi, s} |\tau_{h\chi}(s) - \tau_{0\chi}(s)| \int_0^1 |[sK(s)]'| ds = o(1).$$

Finally, consider (A.5):

$$\frac{F_{\chi_1}(b)}{F_\chi(b)} = 1 + \frac{F_{\chi_1}(b) - F_\chi(b)}{F_\chi(b)} \leq 1 + Md(\chi_1, \chi)^\alpha = O(1)$$

for some $M > 0$, since $F_{\chi_1}(b)/F_\chi(b)$ is Lipschitz of order α in χ_1 . Hence, the order of the left hand side in the statement of the lemma is $O(b^{1+\alpha}) + O((nF_\chi(b))^{-1}) = o((nF_\chi(b))^{-1/2})$ uniformly in $d(\chi_1, \chi) \leq h$, using the Lipschitz continuity of the functions $\varphi'_{\chi_1}(0)$, $M_{0\chi_1}$ and $M_{1\chi_1}$, and using that $\inf_{d(\chi_1, \chi) \leq h} M_{1\chi_1} > 0$ and $h/b = o(1)$. $\square$

Lemma A.6 *Assume (C1)-(C7). Then,*

$$\sup_{d(\chi_1, \chi) \leq h} \left| \widehat{r}_b(\chi_1) - \widehat{r}_b(\chi) - E[\widehat{r}_b(\chi_1)] + E[\widehat{r}_b(\chi)] \right| = o((nF_\chi(b))^{-1/2}) \text{ a.s.}$$

Proof. Write $\widehat{r}_b(\chi) = \widehat{g}_b(\chi)/\widehat{f}_b(\chi)$, where

$$\widehat{g}_b(\chi) = (nF_\chi(b))^{-1} \sum_{i=1}^n Y_i K(b^{-1}d(\mathcal{X}_i, \chi)),$$

and $\widehat{f}_b(\chi)$ is given in (A.1) with h replaced by b . Note that it follows from Lemma 3 in Ferraty *et al.* (2007) that $E[\widehat{r}_b(\chi_1)] = E[\widehat{g}_b(\chi_1)]/E[\widehat{f}_b(\chi_1)] + O((nF_\chi(b))^{-1})$ uniformly in $d(\chi_1, \chi) \leq h$, where the uniformity can be shown in a similar way as in the proof of Lemma A.5. Then, the expression $\widehat{r}_b(\chi_1) - \widehat{r}_b(\chi) - E[\widehat{r}_b(\chi_1)] + E[\widehat{r}_b(\chi)]$ can be decomposed into several terms, the most difficult one to handle being $\widehat{g}_b(\chi_1) - \widehat{g}_b(\chi) - E[\widehat{g}_b(\chi_1)] + E[\widehat{g}_b(\chi)]$. We will therefore concentrate on the latter expression.

From condition (C7) it follows that the ball $B(\chi, h)$ can be covered by r_n small balls of radius ℓ_n and center t_{kn} , $k = 1, \dots, r_n$. Hence,

$$\begin{aligned} & \sup_{d(\chi_1, \chi) \leq h} |\widehat{g}_b(\chi_1) - \widehat{g}_b(\chi) - E[\widehat{g}_b(\chi_1)] + E[\widehat{g}_b(\chi)]| \\ & \leq \max_{1 \leq k \leq r_n} |\widehat{g}_b(t_{kn}) - \widehat{g}_b(\chi) - E[\widehat{g}_b(t_{kn})] + E[\widehat{g}_b(\chi)]| \\ & \quad + \max_{1 \leq k \leq r_n} \sup_{\chi_1 \in B(t_{kn}, \ell_n)} |\widehat{g}_b(t_{kn}) - \widehat{g}_b(\chi_1) - E[\widehat{g}_b(t_{kn})] + E[\widehat{g}_b(\chi_1)]| \\ & = V_1 + V_2. \end{aligned}$$

We start with V_1 . First note that the factor $F_{t_{kn}}(b)$ in the estimator $\widehat{g}_b(t_{kn})$ can be easily replaced by $F_\chi(b)$. Define for $i = 1, \dots, n$ and $k = 1, \dots, r_n$,

$$Z_{ik} = F_\chi(b)^{-1} Y_i \left\{ K(b^{-1}d(\mathcal{X}_i, \chi)) - K(b^{-1}d(\mathcal{X}_i, t_{kn})) \right\}.$$

Then, using the Lipschitz continuity in χ_1 of $F_{\chi_1}(b)/F_\chi(b)$ and $M_{2\chi_1}$, and using that $E[K(b^{-1}d(\mathcal{X}_i, \chi))] = F_\chi(b)\{M_{1\chi} + o(1)\}$ and $E[K^2(b^{-1}d(\mathcal{X}_i, \chi))] = F_\chi(b)\{M_{2\chi} + o(1)\}$ (see e.g. the proof of Lemma 5 in Ferraty *et al.*, 2007), some straightforward calculations show that for any $m \geq 2$,

$$E(|Z_{ik}|^m) = O\left(\left[F_\chi(b)^{-1} \frac{h}{b}\right]^{m-1}\right).$$

Hence, we can apply Corollary A.8 (p. 234) in Ferraty and Vieu (2006), which implies that for $\varepsilon_n = (nF_\chi(b))^{-1/2}(\log n)^{1/2}$, some $c > 0$ and $a_n = (F_\chi(b)^{-1}h/b)^{1/2}$,

$$\begin{aligned} & \sum_n P\left(\max_{1 \leq k \leq r_n} n^{-1} \left| \sum_{i=1}^n \{Z_{ik} - E[Z_{ik}]\} \right| > c\varepsilon_n\right) \\ & \leq 2 \sum_n r_n \exp\left\{-\frac{c^2 \varepsilon_n^2 n}{2a_n^2(1 + c\varepsilon_n)}\right\} \leq 2 \sum_n r_n n^{-\frac{c^2}{2} \frac{b}{h}} < \infty, \end{aligned}$$

for some $c > 0$, since $r_n = O(n^{b/h})$ and $b/h \rightarrow \infty$. Hence, $V_1 = O(\varepsilon_n) = o((nF_\chi(h))^{-1/2})$ a.s.

Consider now the term V_2 . Again, we can easily replace $F_{t_{kn}}(b)$ by $F_\chi(b)$. Hence,

$$\begin{aligned}
& \max_{1 \leq k \leq r_n} \sup_{\chi_1 \in B(t_{kn}, \ell_n)} |\widehat{g}_b(t_{kn}) - \widehat{g}_b(\chi_1)| \\
& \leq (nF_\chi(b))^{-1} \frac{\ell_n}{b} \max_k \sup_{\chi_1} \left\{ \sum_{i=1}^n |Y_i| I(\mathcal{X}_i \in B(t_{kn}, b) \cup B(\chi_1, b)) \right\} + o((nF_\chi(h))^{-1/2}) \\
& \leq \frac{F_\chi(b+h)}{F_\chi(b)} \frac{\ell_n}{b} \left\{ (nF_\chi(b+h))^{-1} \sum_{i=1}^n |Y_i| I(\mathcal{X}_i \in B(\chi, b+h)) \right\} + o((nF_\chi(h))^{-1/2}) \\
& = \frac{F_\chi(b+h)}{F_\chi(b)} \frac{\ell_n}{b} \left\{ E[|Y| | \chi] + o(1) \right\} + o((nF_\chi(h))^{-1/2}) \\
& = O\left(\frac{\ell_n}{b}\right) + o((nF_\chi(h))^{-1/2}) = o((nF_\chi(h))^{-1/2}) \text{ a.s.}
\end{aligned}$$

In a similar way, $\max_k \sup_{\chi_1} |E[\widehat{g}_b(t_{kn})] - E[\widehat{g}_b(\chi_1)]|$ can be taken care of. $\square$

References

- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*. Lecture Notes in Statistics **149**, Springer, New York.
- Borggaard, C. and Thodberg, H.H. (1992). Optimal minimal neural interpretation of spectra. *Anal. Chemistry*, **64**, 545-551.
- Cai, T.T. and Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.*, **34**, 2159-2179.
- Cao, R. (1991). Rate of convergence for the wild bootstrap in nonparametric regression. *Ann. Statist.*, **19**, 2226-2231.
- Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Prob. Letters*, **45**, 11-22.
- Crambes, C., Kneip, A. and Sarda, P. (2008). Smoothing splines estimators for functional linear regression. *Ann. Statist.* (to appear).
- Cuevas, A. and Fraiman, R. (2004). On the bootstrap methodology for functional data: In: Antoch, J. (Ed.), *Proceedings in Computational Statistics, COMPSTAT 2004*. Physica-Verlag, Heidelberg, 127-135.
- Cuevas, A., Febrero, M. and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Comput. Statist. Data Anal.*, **51**, 1063-1074.

- Dabo-Niang, S., Ferraty, F. and Vieu, P. (2007). On the using of modal curves for radar waveforms classification. *Comput. Statist. Data Anal.*, **51**, 4878-4890.
- Davidian, M., Lin, X. and Wang, J.L. (2004). Introduction to the emerging issues in longitudinal and functional data analysis (with discussion). *Statist. Sinica*, **14**, 613-629.
- Delsol, L. (2008). Advances on asymptotic normality in nonparametric functional time series analysis. *Statistics* (to appear).
- Dudley, R.M. (1990). Nonlinear functionals of empirical measures and the bootstrap. In: Eberlein, E., Kuelbs, J., Marcus, M.B. (Eds.), *Probability in Banach Spaces*, vol. 7 (Oberwolfach, 1988). Birkhäuser, Boston, 63-82.
- Fernández de Castro, B., González Manteiga, W. and Guillas, S. (2005). Functional samples and bootstrap for predicting sulfur dioxide levels. *Technometrics*, **47**, 212-222.
- Ferraty, F., Mas, A. and Vieu, P. (2007). Nonparametric regression on functional data: inference and practical aspects. *Austr. N. Z. J. Statist.*, **49**, 267-286.
- Ferraty, F. and Vieu, P. (2002). The functional nonparametric model and application to spectrometric data. *Computat. Statist.*, **17**, 545-564.
- Ferraty, F. and Vieu, P. (2006). *Nonparametric Functional Data Analysis: Theory and Practice*. Springer, New York.
- Giné, E. and Zinn, J. (1990). Bootstrapping general empirical measures. *Ann. Probab.*, **18**, 851-869.
- González Manteiga, W. and Vieu, P. (2007). Introduction to the Special Issue on Statistics for Functional Data. *Computat. Statist. Data Anal.*, **51**, 4788-4792.
- Hall, P. (1990). Using the bootstrap to estimate mean squared error and select smoothing parameter in nonparametric problems. *J. Multivariate Anal.*, **32**, 177-203.
- Hall, P. (1992). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.*, **20**, 695-711.
- Hall, P. and Hart, J. (1990). Bootstrap test for difference between means in nonparametric regression. *J. Amer. Statist. Assoc.*, **85**, 1039-1049.
- Hall, P., Poskitt, P and Presnell, D. (2001). A functional data-analytic approach to signal discrimination. *Technometrics*, **43**, 1-9.
- Härdle, W. and Marron, J. S. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, **19**, 778-796.
- Hastie, T., Buja, A. and Tibshirani, R. (1995). Penalized discriminant analysis. *Ann.*

- Statist.*, **13**, 435-475.
- Martens, H. and Naes, T. (1989). *Multivariate Calibration*, Wiley, New York.
- Masry, E. (2005). Nonparametric regression estimation for dependent functional data: asymptotic normality. *Stochastic Process. Appl.*, **115**, 155-177.
- Politis, D.N. and Romano, J.P. (1994). Limit theorems for weakly dependent Hilbert space valued random variables with application to the stationarity bootstrap. *Statist. Sinica*, **4**, 461-476.
- Ramsay, J. and Silverman, B. (2002). *Applied Functional Data Analysis: Methods and Case Studies*. Springer, New York.
- Ramsay, J. and Silverman, B. (2005). *Functional Data Analysis (second edition)*. Springer, New York.
- Sangalli, L.M., Secchi, P., Vantini, S., and Veneziani, A. (2007). Efficient estimation of 3D centerlines of inner carotid arteries and their curvature profiles by free knot regression splines. Technical Report 23/2007, MOX, Dipartimento di Matematica, Politecnico di Milano (<http://mox.polimi.it/it/progetti/pubblicazioni>).
- Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.
- Valderrama, M. (2007). Introduction to the Special Issue on Modelling Functional Data in Practice. *Computational Statist.*, **22**, 331-334.
- Van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes*. Springer, New York.