Bootstrap for Functional Regression

Javier Zapata

Final Project - PSTAT 227
Prof. Alex Petersen

June 4, 2019

Overview

- In this presentation we will discuss the proofs in:
  Ferraty et al (2010) - On the Validity of the Bootstrap in Non-Parametric Functional Regression
- On a non-parametric regression model $Y = r(\mathcal{X}) + \epsilon$, we use bootstrap to build a pointwise confidence intervals for $r(\cdot)$.
- The paper gives three main results:
  1. A naive and wild bootstrap algorithms for a pointwise CI of $r(\cdot)$.
  2. Bootstrap consistency for both cases (**Theorem 1**)
  3. Functional spaces satisfying the assumptions of Theorem 1 (**Proposition 1**)

Functional Regression (Ferraty et. al. 2007)

- Let $\mathcal{E}$ functional space endowed with semi-metric d. General setting including space of continuous functions, L^p, Sobolev and Besov spaces.
- Let $\mathcal{S} = \{(\mathcal{X}_1, Y_1), \dots, (\mathcal{X}_n, Y_n)\}$ iid data, where $\mathcal{X}_i$ are a sample of curves for which corresponding responses Y_i have been observed.
- We solve a **non-parametric regression model**:

$$Y = r(\mathcal{X}) + \epsilon$$

for $r(\cdot)$ smooth and $\epsilon | \mathcal{X} \sim \left(0, \sigma_\epsilon^2(\chi)\right)$

- The estimator of the regression operator r at $\chi \in \mathcal{E}$ is given by:

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

for some $K(\cdot)$ smooth and bandwidth h s.t. $h \overset{n \to \infty}{\longrightarrow} 0$

Bootstrap Algorithm

Naive bootstrap: Assume homoscedastic model $\sigma_\epsilon^2(\mathcal{X}) \equiv \sigma_\epsilon^2$.

1. Let $\hat{\varepsilon}_{i,b} = Y_i - \hat{r}_b(\chi_i)$ for $i = 1, \ldots, n$ where b is a second smoothing parameter.
2. Draw n iid r.v.s $\varepsilon_1^{\text{boot}}, \ldots, \varepsilon_n^{\text{boot}}$ from the empirical cumulative distribution of $\left(\hat{\varepsilon}_{1,b} - \overline{\hat{\varepsilon}}_b, \ldots, \hat{\varepsilon}_{n,b} - \overline{\hat{\varepsilon}}_b\right)$, where $\overline{\hat{\varepsilon}}_b = n^{-1}\sum_{i=1}^n \hat{\varepsilon}_{i,b}$
3. Define $Y_i^{\text{boot}} = \hat{r}_b(\chi_i) + \varepsilon_i^{\text{boot}}$, for all $i = 1, \ldots, n$ and let $\mathcal{S}^{\text{boot}} = \left(\mathcal{X}_i, Y_i^{\text{boot}}\right)_{i=1}^n$
4. Define $\hat{r}_{hb}^{\text{boot}}(\chi) = \frac{\sum_{i=1}^n Y_i^{\text{boot}} K\left(h^{-1}d(\mathcal{X}_i,\chi)\right)}{\sum_{i=1}^n K(h^{-1}d(\mathcal{X}_i,\chi))}$

Wild bootstrap: Only step 2 changes

2. $\varepsilon_i^{\text{boot}} = \hat{\varepsilon}_{i,b} V_i$ where $V_1, \ldots, V_n$ are $\stackrel{iid}{\sim} (0,1)$ r.v.s that are independent of the data $(\mathcal{X}_i, Y_i)_{i=1}^n$.

Boostrap Confidence Intervals

For a fixed sample $\mathcal{S}$ and confidence level α:

1. Get bootstrapped estimators $\hat{r}_{hb}^{\text{boot } 1}(\chi), \hat{r}_{hb}^{\text{boot } 2}(\chi), \ldots$
2. Compute bootstrapped errors
$\hat{r}_{hb}^{\text{boot } 1}(\chi) - \hat{r}_b(\chi), \hat{r}_{hb}^{\text{boot } 2}(\chi) - \hat{r}_b(\chi), \ldots$
3. Obtain the α-quantile t^*_α of the bootstrapped errors, as an estimator of $t_\alpha(\chi)$ the α-quantile of the distribution of the true error (i.e. $P\left(\hat{r}_h(\chi) - r(\chi) < t_\alpha(\chi)\right) = \alpha$)
4. Compute $(1 - 2\alpha)$-confidence interval from the distribution of the bootstrapped errors:

$$\left[\hat{r}_h(\chi) + t^*_\alpha(\chi), \hat{r}_h(\chi) + t^*_{1-\alpha}(\chi)\right]$$

Key Elements

1. **Small ball probabilities**: Classic assumptions in multivariate nonparametric setting assume that the density of the multivariate predictor is strictly positive. The functional notion of it is:

   $$\text{For a given function } \chi: \quad \forall \epsilon > 0 \quad P(\mathcal{X} \in B(\chi, \epsilon)) > 0$$

   One way to tackle this is to change the notion of closeness from a metric to a **semi-metric** d satisfying:

   - $\forall \chi \in \mathcal{E}, d(\chi, \chi) = 0$
   - $\forall (\chi, f, g) \in \mathcal{E}, d(\chi, f) \leq d(\chi, g) + d(g, f)$

2. **Oversmoothing**: Bootstrap bandwidth b has to be "asymptotically greater" than h.

Notation

- Terms related to the small ball probability:
  $B(\chi, t) = \{\chi_1 \in \mathcal{E} : d(\chi_1, \chi) \leq t\}$ (small ball)
  $F_\chi(t) = P(d(\boldsymbol{\chi}, \chi) \leq t) = P(\boldsymbol{\chi} \in B(\chi, t))$ (small ball probability)
  $\varphi_\chi(s) = E[\{r(\boldsymbol{\chi}) - r(\chi)\} | d(\boldsymbol{\chi}, \chi) = s]$
  $\tau_{h\chi}(s) = F_\chi(hs)/F_\chi(h) = P(d(\boldsymbol{\chi}, \chi) \leq hs | d(\boldsymbol{\chi}, \chi) \leq h)$ for $0 \leq s \leq 1$
  $\tau_{0_\chi}(s) = \lim_{h \downarrow 0} \tau_{h\chi}(s)$
- Terms used for asymptotic convergence:
  $M_{0\chi} = K(1) - \int_0^1 (sK(s))' \tau_{0\chi}(s) \mathrm{d}s$
  $M_{1\chi} = K(1) - \int_0^1 K'(s) \tau_{0\chi}(s) \mathrm{d}s$
  $M_{2\chi} = K^2(1) - \int_0^1 \left(K^2\right)'(s) \tau_{0\chi}(s) \mathrm{d}s$
- Other terms
  $\hat{g}_{hb}^{\mathrm{boot}}(\chi) = (nF_\chi(h))^{-1} \sum_{i=1}^n Y_i^{\mathrm{boot}} K\left(h^{-1} d(\chi_i, \chi)\right)$
  $\hat{f}_h(\chi) = (nF_\chi(h))^{-1} \sum_{i=1}^n K\left(h^{-1} d(\chi_i, \chi)\right)$
  $\hat{r}_{hb}^{\mathrm{boot}}(\chi) = \hat{g}_{hb}^{\mathrm{boot}}(\chi) / \hat{f}_h(\chi)$

Assumptions

Regularity conditions: They relate to the smoothness and finiteness of $r, \sigma^2_{\mathcal{E}}, \varphi_\chi, F_\chi$ and τ_{0_χ}.

- (C1) Functions $r(.), \sigma^2_\varepsilon(.)$ and $E(|Y||\chi = .)$ are cntns in a neighborhood of χ:

$$\sup_{d(\chi_1,\chi)<\varepsilon} E\left(|Y|^m|\chi = \chi_1\right) < \infty \quad \text{for some } \varepsilon > 0 \text{ and } m \geq 1$$

- (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 continuos 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.
- (C4) For all $\chi_1 \in \mathcal{E}$ and all $0 \leq s \leq 1$ $\tau_{0\chi_1}(s)$ exists and:

$$\sup_{\chi_1 \in \mathcal{E}, 0 \leq s \leq 1} \left|\tau_{h\chi_1}(s) - \tau_{0\chi_1}(s)\right| = o(1)$$

Assumptions

Estimator's conditions: Conditions on the kernel K as well as on the asymp. relationship between h and b. In short, b has to be asymp. larger than h (oversmoothing)

- (C5) K is supported in $[0,1]$, K has a continuous derivative on $[0,1), K'(s) \leq 0$ for $0 \leq s < 1$ and $K(1) > 0$.
- (C6) $h, b \to 0, h/b \to 0, nF_\chi(h) \to \infty$, $h\left(nF_\chi(h)\right)^{1/2} = O(1)$, $b^{1+\alpha}\left(nF_\chi(h)\right)^{1/2} = o(1)$, $bh^{\alpha-1} = O(1)$
  $\frac{F_\chi(b+h)}{F_\chi(b)} \to 1$ and $\frac{F_\chi(h)}{F_\chi(b)} \log n = o(1)$

Asymptotics of $\hat{r}$ (Ferraty et al 2007)

Assume (C1)-(C6) hold and let $\widehat{F}(h) = \frac{\#\{i: d(\mathcal{X}_i, \chi) \leq h\}}{n}$. Then:

$$\mathrm{E}(\widehat{r}_h(\chi)) - r_h(\chi) = \underbrace{\varphi'(0)\frac{M_{0_\chi}}{M_{1_\chi}} h}_{B_n} + O\left((nF_\chi(h))^{-1}\right) + o(h)$$

$$\mathrm{Var}(\widehat{r}_h(\chi)) = \frac{1}{nF_\chi(h)} \frac{M_{2_\chi}}{M_{1_\chi}^2} \sigma_\varepsilon^2 + o\left(\frac{1}{nF_\chi(h)}\right)$$

From (C6): $h\big(nF_\chi(h)\big)^{1/2} \to 0$ allows to cancel the bias B_n hence:

$$\left(n\widehat{F}_\chi(h)\right)^{1/2} \left(\widehat{r}_h(\chi) - r_h(\chi)\right) \frac{M_{1_\chi}}{\sqrt{M_{2_\chi} \sigma_\varepsilon^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

$$\frac{\mathrm{E}(\widehat{r}_h(\chi)) - r_h(\chi)}{\sqrt{\mathrm{Var}(\widehat{r}_h(\chi))}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

Assumptions

Technical condition: Used to create an upper bound in the proof of Lemma 6.

- (C7) For each n, $\exists r_n \geq 1, \ell_n > 0$ and curves $t_{1n}, \ldots, t_{r_n n}$ s.t.

$$B(\chi, h) \subset \bigcup_{k=1}^{r_n} B\left(t_{kn}, \ell_n\right)$$

with $r_n = O\left(n^{b/h}\right)$ and $\ell_n = o\left(b\left(nF_\chi(h)\right)^{-1/2}\right)$

In the proof of Lemma 6, let $\hat{g}_b(\chi) = \left(nF_\chi(b)\right)^{-1} \sum_{i=1}^n Y_i K\left(b^{-1} d\left(\chi_i, \chi\right)\right)$

$$\begin{aligned}
&\sup_{d(\chi_1,\chi)\leq h} \left|\hat{g}_b\left(\chi_1\right) - \hat{g}_b(\chi) - E\left[\hat{g}_b\left(\chi_1\right)\right] + E\left[\hat{g}_b(\chi)\right]\right| \\
&\leq \max_{1\leq k\leq r_n} \left|\hat{g}_b\left(t_{kn}\right) - \hat{g}_b(\chi) - E\left[\hat{g}_b\left(t_{kn}\right)\right] + E\left[\hat{g}_b(\chi)\right]\right| \\
&+ \max_{1\leq k\leq r_n} \sup_{\chi_1 \in B(t_{kn},\ell_n)} \left|\hat{g}_b\left(t_{kn}\right) - \hat{g}_b\left(\chi_1\right) - E\left[\hat{g}_b\left(t_{kn}\right)\right] + E\left[\hat{g}_b\left(\chi_1\right)\right]\right| \\
&= \ldots = o((nF_\chi(h))^{-1/2})
\end{aligned}$$

Main Result

Theorem 1

Assume (C1)-(C7). Then, for the wild bootstrap procedure we have:

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

where $P^{\mathcal{S}}$ denote probability, conditionally on the sample $\mathcal{S}$ (i.e. $\chi_i, Y_i), i=1,\ldots,n$ In addition, if $\sigma_{\epsilon}^{2}(\chi)=\sigma_{\epsilon}^{2}$, then the same result holds for the naive bootstrap.

- The most important element in this proof is Lemma 2. Other lemmas show asymptotic convergence to constant terms or small probability upper bounds.

Lemma 2

Assume (C1)-(C6) hold. Then:

$$\frac{\hat{r}_h(\chi)-E[\hat{r}_h(\chi)]}{\sqrt{\mathrm{Var}[\hat{r}_h(\chi)]}}\xrightarrow{d}N(0,1)\quad\text{and}\quad\frac{\hat{r}_{hb}^{\mathrm{boot}}(\chi)-E^{\mathcal{S}}\left[\hat{r}_{hb}^{\mathrm{boot}}(\chi)\right]}{\sqrt{\mathrm{Var}^{\mathcal{S}}\left[\hat{r}_{hb}^{\mathrm{boot}}(\chi)\right]}}\xrightarrow{d}N(0,1)$$

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

Which spaces satisfy (C7)?

Two spaces $(\mathcal{E}, d)$ for which condition (C7) holds.

Proposition 1

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

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

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

Which spaces satisfy (C7)?

Two spaces $(\mathcal{E}, d)$ for which condition (C7) holds.

Proposition 1 (...continued...)

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

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

$\chi^{(k)}$ denotes the *kth* derivative of the function χ, $\|\cdot\|_2$ is the Euclidean norm and $\underline{\gamma} = \sup\{k \in Z : k < \gamma\}$. Then the space $(\mathcal{E}, d_{Lp})$ satisfies (C7) provided that $hb^{-(1+1/\gamma)}(\log n)^{-1+1/\gamma} = o(1)$ where d_{L_p} is the $L^p - distance$ in $\mathcal{E}$ and $1 \leq p \leq \infty$

Proof of Proposition 1 uses Lemma 1 that involves covering number of balls as seen in (C7)

The End

References

- Ferraty et al (2007) - Nonparametric Regression of Functional Data, Inference and Practical Aspects
- Ferraty et al (2010)- On the Validity of the Bootstrap in Non-Parametric Functional Regression
- Ferraty, Vieu (2006) - Nonparametric Functional Data Analysis