Universiteit Leiden
Mathematical Institute

The empirical bootstrap

Chapter 3.7.1

Thomas Nagler <t.w.nagler@math.leidenuniv.nl>
Leiden University

Monday 19th October, 2020

The empirical bootstrap

- $X_1, \dots, X_n \sim P$ *iid* with empirical measure and process

$$\mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{X_i}, \qquad \mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P).$$

- Suppose $\mathbb{G}_n \rightsquigarrow \mathbb{G}$, but asymptotic distribution is complicated.

Bootstrap

1. Draw *iid* samples $\hat{X}_1, \dots, \hat{X}_n \sim \mathbb{P}_n$.
2. Define bootstrap empirical measure/process

$$\hat{\mathbb{P}}_n = n^{-1} \sum_{i=1}^{n} \delta_{\hat{X}_i}, \qquad \hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n).$$

3. Approximate law of $\mathbb{G}$ by law of $\hat{\mathbb{G}}_n$.

The main question

When is the approximation "valid"?

- Traditionally: Valid if

$$\hat{\mathbb{G}}_n \rightsquigarrow \mathbb{G} \text{ conditionally on } X_1, X_2, \ldots$$

  in probability/a.s.

- Segers/Dette/Bücher: Valid if

$$\forall N \in \mathbb{N}: \quad (\hat{\mathbb{G}}_n^{(1)}, \ldots, \hat{\mathbb{G}}_n^{(N)}) \rightsquigarrow (\mathbb{G}^{(1)}, \ldots, \mathbb{G}^{(N)}),$$

  with $\mathbb{G}^{(1)}, \ldots, \mathbb{G}^{(N)}$ iid copies of $\mathbb{G}$.

Spoiler: We will show that the bootstrap is valid for Donsker classes.

The strategy

Multiplier CLTs, Chapter 2.9

For *iid* $\xi_1, \ldots, \xi_n$ with $\mathrm{E}[\xi_1] = 0$, $\mathrm{var}[\xi_i] = 1$, $\|\xi_i\|_{2,1} < \infty$ and $\mathcal{F}$ Donsker,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P) \rightsquigarrow \mathbb{G} \quad \text{in } \ell^\infty(\mathcal{F}),$$

conditionally on $X_1, X_2, \ldots$, in probability (a.s. if $P\|f - Pf\|_{\mathcal{F}}^2 < \infty$).

- Define $M_{n,i} = \#\{j \colon \hat{X}_j = X_i\}$ and write

$$\hat{\mathbb{G}}_n = \sqrt{n}(\hat{\mathbb{P}}_n - \mathbb{P}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{\hat{X}_i} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_{X_i} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (M_{n,i} - 1)\delta_{X_i}.$$

- Problem: $(M_{n,1}, \ldots, M_{n,n}) \sim \mathrm{Multinomial}(n, n^{-1}, \ldots, n^{-1})$

  $\Rightarrow \xi_i = (M_{n,i} - 1)$ are dependent.

The strategy

Strategy

1. Approximate ξ_i's by independent multipliers through Poissonization.
2. Use multiplier CLT on approximated bootstrap process.
3. Show that approximation error is negligible.
4. Make everything super general (and technical).

Poissonization

- Define bootstrap process based on k replicates $\hat{X}_1, \dots, \hat{X}_k \sim \mathbb{P}_n$ as

$$\hat{\mathbb{G}}_{n,k} = \sqrt{k}(\hat{\mathbb{P}}_k - \mathbb{P}_n) = \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \left(M_{k,i} - \frac{k}{n} \right) \delta_{X_i}.$$

  and note that $\hat{\mathbb{G}}_{n,n} = \hat{\mathbb{G}}_n$.

- Instead of n, take $N_n \sim \mathrm{Poisson}(n)$ replicates. Chapter 3.6.2:

$$M_{N_n,1} = \sum_{i=1}^{N_n} \mathbb{1}\{\hat{X}_i = X_1\}, \quad \dots, \quad M_{N_n,n} = \sum_{i=1}^{N_n} \mathbb{1}\{\hat{X}_i = X_n\}$$

  are *iid* $\mathrm{Poisson}(1)$.

Poissonization

- Write

$$\hat{\mathbb{G}}_{n,N_n} = \frac{1}{\sqrt{N_n}} \sum_{i=1}^{n} (M_{N_n,i} - 1)(\delta_{X_i} - P) - \frac{N_n - n}{\sqrt{N_n}} (\mathbb{P}_n - P)$$

- We can use multiplier CLTs for first term.
- Second term vanishes almost surely if $\mathcal{F}$ is Glivenko-Cantelli.
- Must show:

$$\mathrm{P}\left(\|\hat{\mathbb{G}}_{n,N_n} - \hat{\mathbb{G}}_n\|_{\mathcal{F}} > \epsilon \mid X_1, \ldots X_n \right) \to 0$$

in probability/almost surely.

Clarification

Multiplier CLTs, Chapter 2.9

For *iid* $\xi_1, \ldots, \xi_n$ with $\mathrm{E}[\xi_1] = 0$, $\mathrm{var}[\xi_i] = 1$, $\|\xi_i\|_{2,1} < \infty$ and $\mathcal{F}$ Donsker,

$$\tilde{\mathbb{G}}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i(\delta_{X_i} - P) \rightsquigarrow \mathbb{G} \quad \text{in } \ell^\infty(\mathcal{F}),$$

conditionally on $X_1, X_2, \ldots$, in probability (a.s. if $P\|f - Pf\|_{\mathcal{F}}^2 < \infty$).

- Mathematically the convergence statement means

$$\sup_{h \in BL_1} \left| \mathrm{E}_\xi h(\tilde{\mathbb{G}}_n) - \mathrm{E} h(\mathbb{G}) \right| \to 0.$$

in probability/almost surely (w.r.t. P^*).

Main results

Theorem 3.7.1 (w/o measurability)

Let $\mathcal{F}$ have finite envelope and define

$$\hat{\mathbb{Y}}_n = n^{-1/2} \sum_{i=1}^{n} (M_{N_n,i} - 1)(\delta_{X_i} - P).$$

The following are equivalent:

(i) $\mathcal{F}$ is Donsker.

(ii) $\sup_{h \in BL_1} \left| \mathrm{E}_{M,N} h(\hat{\mathbb{Y}}_n) - \mathrm{E} h(\mathbb{G}) \right| \xrightarrow{\mathrm{P}} 0.$

(iii) $\sup_{h \in BL_1} \left| \mathrm{E}_{M} h(\hat{\mathbb{G}}_n) - \mathrm{E} h(\mathbb{G}) \right| \xrightarrow{\mathrm{P}} 0.$

Main results

Theorem 3.7.2 (w/o measurability)

Let $\mathcal{F}$ have finite envelope and define

$$\hat{\mathbb{Y}}_n = n^{-1/2} \sum_{i=1}^{n} (M_{N_n,i} - 1)(\delta_{X_i} - P).$$

The following are equivalent:

(i) $\mathcal{F}$ is Donsker and $P\|f - Pf\|_{\mathcal{F}}^2 < \infty$.

(ii) $\sup_{h \in BL_1} \left| \mathrm{E}_{M,N} h(\hat{\mathbb{Y}}_n) - \mathrm{E} h(\mathbb{G}) \right| \xrightarrow{a.s.} 0.$

(iii) $\sup_{h \in BL_1} \left| \mathrm{E}_{M} h(\hat{\mathbb{G}}_n) - \mathrm{E} h(\mathbb{G}) \right| \xrightarrow{a.s.} 0.$

Proof of main results

- (i) $\Leftrightarrow$ (ii) follows from Theorems 2.9.6 and 2.9.7.
- For (i) + (ii) $\Rightarrow$ (iii), we must show that

$$\mathrm{P}\left(\|\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n\|_{\mathcal{F}} > \epsilon \mid X_1, \dots X_n \right) \to 0 \qquad (i.p./a.s.)$$

- We will do this by constructing a coupling of the two processes.

Proof of main results

Coupling

- Let $m_n^{(1)}, m_n^{(2)}, \ldots \in \{0,1\}^n$ be *iid* $\mathrm{multinom}(1, n^{-1}, \ldots, n^{-1})$ independent of N_n.

- Set
$$M_n = \sum_{j=1}^{n} m_n^{(j)}, \qquad M_{N_n} = \sum_{j=1}^{N_n} m_n^{(j)}.$$

- Defining $\hat{\mathbb{G}}_n$ using M_n and $\hat{\mathbb{Y}}_n$ using M_{N_n},
$$\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (M_{N_n,i} - M_{n,i})(\delta_{X_i} - P).$$

- We shall first show that $\max_i |M_{N_n,i} - M_{n,i}|$ is bounded with high probability.

Proof of main results

Showing that $\max_i |M_{N_n,i} - M_{n,i}|$ is bounded with high probability

- $|M_{N_n} - M_n|$ is a sum of $|N_n - n|$ of the $m_n^{(j)} \overset{iid}{\sim} \mathrm{multinom}(1, n^{-1}, \dots)$.
- Given $N_n = k$, $|M_{N_n,i} - M_{n,i}| \sim \mathrm{binom}(|k-n|, n^{-1})$.
- Using $N_n \sim \mathrm{Poisson}(n)$ and Markov's inequality,

$$\exists\, K < \infty: \quad \mathrm{P}(|N_n - n| \geq K\sqrt{n}) \leq \epsilon \quad \forall n.$$

- Thus,

$$\begin{aligned}
\mathrm{P}\left(\max_{1 \leq i \leq n} |M_{N_n,i} - M_{n,i}| > 2\right) &\leq \epsilon + n \times \mathrm{P}\left(\mathrm{binom}(\lceil K\sqrt{n}\rceil, n^{-1}) > 2\right) \\
&\leq \epsilon + n \times [\text{Bernstein/Chernoff bound}] \\
&\to \epsilon.
\end{aligned}$$

Proof of main results

Bounding $\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n$

- Define $I_n^j = \{i \colon M_{N_n,i} - M_{n,i} = j\}$.

- On the set where $\max_{1 \le i \le n} |M_{N_n,i} - M_{n,i}| \le 2$, we have

$$
\begin{aligned}
\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (M_{N_n,i} - M_{n,i})(\delta_{X_i} - P) \\
&= \sum_{j=-2}^{2} \frac{j}{\sqrt{n}} \sum_{i \in I_n^j} (\delta_{X_i} - P) \\
&= \sum_{j=-2}^{2} \frac{j|I_n^j|}{\sqrt{n}} \left(\frac{1}{|I_n^j|} \sum_{i \in I_n^j} (\delta_{X_i} - P) \right)
\end{aligned}
$$

Proof of main results

Bounding $\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n$

$$\hat{\mathbb{Y}}_n - \hat{\mathbb{G}}_n = \sum_{j=-2}^{2} \frac{j|I_n^j|}{\sqrt{n}} \left(\frac{1}{|I_n^j|} \sum_{i \in I_n^j} (\delta_{X_i} - P) \right).$$

- $|I_n^j| \leq |N_n - n| = O_p(\sqrt{n})$.
- Lemma 3.7.16 (next time): if $\mathcal{F}$ is Glivenko-Cantelli,

$$\left\| \frac{1}{|I_n^j|} \sum_{i \in I_n^j} (\delta_{X_i} - P) \right\|_{\mathcal{F}} \xrightarrow{a.s.} 0.$$

It follows that (i)+(ii) and (iii) are equivalent (in both theorems) provided $\mathcal{F}$ is Glivenko-Cantelli. If (i)+(ii) holds, then $\mathcal{F}$ is Donsker and certainly Glivenko-Cantelli. Thus, the proof of the theorem in the most interesting direction is complete.

$\square$

Two lemmas

Lemma 3.7.6

For fixed elements $x_1, \ldots, x_n$ of a set $\mathcal{X}$, let

$$\hat{X}_1, \ldots, \hat{X}_k \overset{iid}{\sim} \mathbb{P}_n = n^{-1} \sum_{i=1}^{n} \delta_{x_i}$$

and

$$N_1, N_1', \ldots, N_n, N_n' \overset{iid}{\sim} \mathrm{Poisson}(0.5k/n).$$

Then for every class $\mathcal{F}$

$$\mathrm{E}_{\hat{X}} \left\| \sum_{j=1}^{k} (\delta_{\hat{X}_j} - \mathbb{P}_n) \right\|_{\mathcal{F}} \leq 4 \mathrm{E}_{N,N'} \left\| \sum_{i=1}^{n} (N_i - N_i') \delta_{x_i} \right\|_{\mathcal{F}}.$$

Two lemmas

Proof of Lemma 3.7.6

- Using symmetrization (Lemma 2.3.1) and Le Cam's Poissonization lemma (3.6.4),

$$\mathrm{E}\left\|\sum_{j=1}^{k}(\delta_{\hat{X}_j} - \mathbb{P}_n)\right\|_{\mathcal{F}} \leq 2\mathrm{E}\left\|\sum_{j=1}^{k}\varepsilon_j\delta_{\hat{X}_j}\right\|_{\mathcal{F}} \leq 4\mathrm{E}\left\|\sum_{j=1}^{N}\varepsilon_j\delta_{\hat{X}_j}\right\|_{\mathcal{F}},$$

  where $N \sim \mathrm{Poisson}(k)$ independent of everything.

- Set

$$N_i = \#\{j \leq N : \hat{X}_j = x_i, \varepsilon_j = 1\}, \quad N_i' = \#\{j \leq N : \hat{X}_j = x_i, \varepsilon_j = -1\}.$$

- Then $N_1, N_1', \ldots, \overset{iid}{\sim} \mathrm{Poisson}(0.5k/n)$ and

$$\sum_{j=1}^{N}\epsilon_j\delta_{\hat{X}_j} = \sum_{j=1}^{n}(N - N_i')\delta_{x_i}. \qquad \square$$

Two lemmas

Lemma 3.7.7

For arbitrary stochastic processes $Z_1, \dots, Z_n$, every exchangeable random vector $(\xi_1, \dots, \xi_n)$ independent of $Z_1, \dots, Z_n$ and any $1 \le n_0 \le n$,

$$\mathrm{E}_\xi \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_i Z_i \right\|_{\mathcal{F}}^* \le 2(n_0 - 1) \frac{1}{n} \sum_{i=1}^n \|Z_i\|_{\mathcal{F}}^* \mathrm{E}_\xi \max_{1 \le i \le n} \frac{\xi_i}{\sqrt{n}} + 2\|\xi_1\|_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^k Z_{R_i} \right\|_{\mathcal{F}}^*,$$

where $(R_1, \dots, R_n)$ is uniformly distributed on the set of all permutations of $\{1, \dots, n\}$ and independent of $Z_1, \dots, Z_n$.

Two lemmas

Proof of Lemma 3.7.7

- Since ξ_i can be split into positive/negative parts, take $\xi_i > 0$ wlog.
- Let $\xi_{(1)} \geq \cdots \geq \xi_{(n)}$. By exchangeability,

$$\mathrm{E}_\xi \left\| \sum_{i=1}^n \xi_i Z_i \right\|_{\mathcal{F}}^* = \mathrm{E}_{\xi,R} \left\| \sum_{i=1}^n \xi_{R_i} Z_i \right\|_{\mathcal{F}}^* = \mathrm{E}_{\xi,R} \left\| \sum_{i=1}^n \xi_{(i)} Z_{S_{R_i}} \right\|_{\mathcal{F}}^*,$$

  where S is another permutation such that $\xi_{(i)} = \xi_{S_i}$.
- It holds $R \circ S \stackrel{d}{=} R$ independent of S and $\xi_1, \ldots, \xi_n$.
- Continue as in Lemma 2.9.1. $\square$

Result for $k \neq n$ samples

Theorem 3.7.3 (w/o measurability)

If $\mathcal{F}$ is Donsker, then for every $k_n \to \infty$,

$$\sup_{h \in BL_1} \left| \mathrm{E}_M h(\hat{\mathbb{G}}_{n,k_n}) - \mathrm{E} h(\mathbb{G}) \right| \xrightarrow{\mathrm{P}} 0.$$

If $P\|f - PF\|_{\mathcal{F}}^2 < \infty$, convergence is almost sure.

Result for $k \neq n$ samples

Proof of fidi-convergence

- Recall that

$$\hat{\mathbb{G}}_{n,k} f = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} f(\hat{X}_i) - \mathbb{P}_n f\right) = \sqrt{k}\left(\frac{1}{k}\sum_{i=1}^{k} f(\hat{X}_i) - \mathrm{E}_{\hat{X}} f(\hat{X}_1)\right).$$

- By the LLN,

$$\mathrm{E}_{\hat{X}} f^2(\hat{X}_1) = \mathbb{P}_n f^2 \overset{a.s.}{\to} Pf^2,$$
$$\mathrm{E}_{\hat{X}} f^2(\hat{X}_1)\{|f(\hat{X}_1)| > \epsilon\sqrt{k}\} = \mathbb{P}_n f^2\{|f| > \epsilon\sqrt{k}\} \overset{a.s.}{\to} 0.$$

- Lindeberg CLT: $\hat{\mathbb{G}}_{n,k} f \to_d \mathcal{N}(0, Pf^2)$ for almost every $X_1, X_2, \ldots$

Result for $k \neq n$ samples

Proof of tightness (almost sure part)

- Lemma 3.7.6:

$$\mathrm{E}_{\hat{X}} \left\| \hat{\mathbb{G}}_{n,k} \right\|_{\mathcal{F}_\delta} \leq 4 \mathrm{E}_{\tilde{N}} \left\| \frac{1}{\sqrt{k}} \sum_{i=1}^{n} \tilde{N}_i \delta_{X_i} \right\|_{\mathcal{F}_\delta}.$$

- $\varepsilon_1, \varepsilon_2, \ldots,$ *iid* Rademacher independent of everything $\Rightarrow \tilde{N}_i \stackrel{d}{=} \varepsilon_i |\tilde{N}_i|$.

- Lemma 3.7.7 with $Z_i = \varepsilon_i \delta_{X_i}$: RHS above bounded by multiple of

$$(n_0 - 1) \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}_\varepsilon \|\varepsilon_i \delta_{X_i}\|_{\mathcal{F}_\delta}^* \mathrm{E} \max_{1 \leq i \leq n} \frac{|\tilde{N}_i|}{\sqrt{k}}$$
$$+ \sqrt{\frac{n}{k}} \|\tilde{N}_1\|_{2,1} \max_{n_0 \leq k \leq n} \mathrm{E}_{\varepsilon,R} \left\| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{X_{R_i}} \right\|_{\mathcal{F}_\delta}^*$$

- We bound the two terms separately.

Result for $k \neq n$ samples

Proof of tightness (almost sure part)

- First term: $T_1 = (n_0 - 1)\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \mathrm{E}_\varepsilon \|\varepsilon_i \delta_{X_i}\|^*_{\mathcal{F}_\delta} \mathrm{E} \max_{1 \le i \le n} \dfrac{|\tilde{N}_i|}{\sqrt{k}}$
- Problem 3.7.4: $\mathrm{E} \max_{1 \le i \le n} |\tilde{N}_i|/\sqrt{k} = O\big((n \wedge k)^{-1/4}\big)$.
- Hence, $T_1 = o(1) \times \mathbb{P}_n F \overset{a.s.}{\to} 0$.

Result for $k \neq n$ samples

Proof of tightness (almost sure part)

- Second term: $T_2 = \sqrt{\dfrac{n}{k}}\|\tilde{N}_1\|_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_{\varepsilon,R}\left\| \dfrac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{X_{R_i}} \right\|_{\mathcal{F}_\delta}^{*}$

- Problem 3.7.3:
$$\sqrt{n/k}\|\tilde{N}_1\|_{2,1} = O(1).$$

- Hence,
$$T_2 \lesssim \max_{n_0 \le k \le n} \mathrm{E}_{\varepsilon,R}\left\| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{X_{R_i}} \right\|_{\mathcal{F}_\delta}^{*}.$$

Result for $k \neq n$ samples

Proof of tightness (almost sure part)

- Define $U_j = \mathrm{E}_\varepsilon \left\| \frac{1}{\sqrt{j}} \sum_{i=1}^{j} \varepsilon_i \delta_{X_i} \right\|_{\mathcal{F}_\delta}^*$
- By Jensen's inequality,

$$\max_{n_0 \le k \le n} \mathrm{E}_{\varepsilon,R} \left\| \frac{1}{\sqrt{j}} \sum_{i=n_0}^{j} \varepsilon_i \delta_{X_{R_i}} \right\|_{\mathcal{F}_\delta}^* \le \mathrm{E}\left(\max_{n_0 \le j} U_j \mid \mathcal{S}_n \right),$$

  where $\mathcal{S}_n$ is σ-field generated by functions symmetric in first n coordinates.

- By the Hewitt-Savage zero-one law $\mathcal{S}_n \searrow \mathcal{S}$ where the latter consists of sets of probability 0 or 1 only.

Result for $k \neq n$ samples

Proof of tightness (almost sure part)

- Thus,

$$\mathrm{E}(\max_{n_0 \le j} U_j \mid \mathcal{S}_n) \overset{a.s.}{\to} \mathrm{E}\left(\max_{n_0 \le j} U_j \mid \mathcal{S}\right) = \mathrm{E}\left(\max_{n_0 \le j} U_j\right).$$

  which is bounded due to Corollary 2.9.9.

- By Lemma 2.9.8, $\limsup_{j \to \infty} U_j \lesssim \mathrm{E}\|\mathbb{G}\|_{\mathcal{F}_\delta}$ almost surely.

- Hence

$$\lim_{\delta \to 0} \lim_{n_0 \to \infty} \max_{n_0 \le j} U_j \overset{a.s.}{\to} 0,$$

  and $\mathrm{E}(\max_{n_0 \le j} U_j) \to 0$ by dominated convergence. $\square$