3.7.2 Exchangeable Bootstrap

Lu Yang

October 26, 2020

Recap

- Given an i.i.d. sample $X_1, \ldots, X_n$, let $\hat{X}_1, \ldots, \hat{X}_n$ be an i.i.d. sample from $\mathbb{P}_n$

Recap

- Given an i.i.d. sample $X_1, \ldots, X_n$, let $\hat{X}_1, \ldots, \hat{X}_n$ be an i.i.d. sample from $\mathbb{P}_n$
- Bootstrap empirical distribution

$$\hat{\mathbb{P}}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{X}_i} = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\delta_{X_i}$$

  where M_{ni} is the number of times that X_i is "redrawn" from the original sample
- $(M_{n1}, M_{n2}, \ldots, M_{nn})$ and multinomially distributed with parameters n and probabilities $1/n, \ldots, 1/n$

Recap

- Given an i.i.d. sample $X_1, \ldots, X_n$, let $\hat{X}_1, \ldots, \hat{X}_n$ be an i.i.d. sample from $\mathbb{P}_n$
- Bootstrap empirical distribution

$$\hat{\mathbb{P}}_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{\hat{X}_i} = \frac{1}{n}\sum_{i=1}^{n} M_{ni}\delta_{X_i}$$

  where M_{ni} is the number of times that X_i is “redrawn" from the original sample
- $(M_{n1}, M_{n2}, \ldots, M_{nn})$ and multinomially distributed with parameters n and probabilities $1/n, \ldots, 1/n$
- The corresponding bootstrap empirical process is

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

Exchangeable bootstrap

- For each n, let $(W_{n1,} \ldots, W_{nn})$ be an exchangeable, nonnegative random vector

Exchangeable bootstrap

- For each n, let $(W_{n1,}, \ldots, W_{nn})$ be an exchangeable, nonnegative random vector
- Weighted bootstrap empirical measure

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^{n} W_{ni} \delta_{X_i}$$

Exchangeable bootstrap

- For each n, let $(W_{n1}, \ldots, W_{nn})$ be an exchangeable, nonnegative random vector
- Weighted bootstrap empirical measure

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^{n} W_{ni} \delta_{X_i}$$

- The corresponding weighted bootstrap empirical process

$$\begin{aligned}
\hat{\mathbb{G}}_n =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(W_{ni} - \bar{W}_n \right) \delta_{X_i} = \sqrt{n} \left(\hat{\mathbb{P}}_n - \bar{W}_n \mathbb{P}_n \right) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni} \left(\delta_{X_i} - \mathbb{P}_n \right)
\end{aligned}$$

Exchangeable bootstrap

- For each n, let $(W_{n1,}, \dots, W_{nn})$ be an exchangeable, nonnegative random vector
- Weighted bootstrap empirical measure

$$\hat{\mathbb{P}}_n = \frac{1}{n} \sum_{i=1}^{n} W_{ni} \delta_{X_i}$$

- The corresponding weighted bootstrap empirical process

$$\begin{aligned}
\hat{\mathbb{G}}_n =& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(W_{ni} - \bar{W}_n \right) \delta_{X_i} = \sqrt{n} \left(\hat{\mathbb{P}}_n - \bar{W}_n \mathbb{P}_n \right) \\
=& \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni} \left(\delta_{X_i} - \mathbb{P}_n \right)
\end{aligned}$$

- This corresponds to resampling W_{ni} times the variable X_i, without requiring that the weights W_{ni} to be integer-valued

- It is assumed that

$$\begin{aligned}
\sup_n ||W_{n1} - \bar{W}_n||_{2,1} < \infty, \\
n^{-1/2}\mathrm{E} \max_{1 \le i \le n} |W_{ni} - \bar{W}_n| \to 0, \\
n^{-1} \sum_{i=1}^{n} (W_{ni} - \bar{W}_n)^2 \xrightarrow{P} c^2 > 0.
\end{aligned} \tag{3.7.8}$$

---

[0] $\|\xi\|_{2,1} = \int_0^\infty \sqrt{P(|\xi| > t)}\mathrm{d}t$

- It is assumed that

$$\sup_n ||W_{n1} - \bar{W}_n||_{2,1} < \infty,$$
$$n^{-1/2}\mathrm{E} \max_{1\le i\le n} |W_{ni} - \bar{W}_n| \to 0, \qquad (3.7.8)$$
$$n^{-1}\sum_{i=1}^{n}(W_{ni} - \bar{W}_n)^2 \xrightarrow{P} c^2 > 0.$$

- (3.7.8) alone imply the conditional weak convergence of the weighted bootstrap empirical process

[0] $\|\xi\|_{2,1} = \int_0^\infty \sqrt{P(|\xi| > t)}\mathrm{d}t$

- It is assumed that

$$\sup_n ||W_{n1} - \bar{W}_n||_{2,1} < \infty,$$
$$n^{-1/2}\mathrm{E} \max_{1 \le i \le n} |W_{ni} - \bar{W}_n| \to 0, \qquad (3.7.8)$$
$$n^{-1} \sum_{i=1}^{n} (W_{ni} - \bar{W}_n)^2 \xrightarrow{P} c^2 > 0.$$

- (3.7.8) alone imply the conditional weak convergence of the weighted bootstrap empirical process
- Sufficient for the second condition is that the variables $W_{ni} - \bar{W}_n$ are uniformly weak-L_2: $\sup_{x > \varepsilon\sqrt{n}} n^{-1} \sum_{i=1}^{n} x^2 P(|W_{ni} - \bar{W}_n| > x) \to 0$ for every $\varepsilon > 0$ (Problem 2.3.3)

---

[0] $\|\xi\|_{2,1} = \int_0^\infty \sqrt{P(|\xi| > t)}\mathrm{d}t$

- It is assumed that

$$
\begin{aligned}
\sup_n ||W_{n1} - \bar{W}_n||_{2,1} &< \infty, \\
n^{-1/2} \mathrm{E} \max_{1 \le i \le n} |W_{ni} - \bar{W}_n| &\to 0, \qquad (3.7.8) \\
n^{-1} \sum_{i=1}^{n} (W_{ni} - \bar{W}_n)^2 &\xrightarrow{P} c^2 > 0.
\end{aligned}
$$

- (3.7.8) alone imply the conditional weak convergence of the weighted bootstrap empirical process
- Sufficient for the second condition is that the variables $W_{ni} - \bar{W}_n$ are uniformly weak-L_2: $\sup_{x > \varepsilon\sqrt{n}} n^{-1} \sum_{i=1}^{n} x^2 P(|W_{ni} - \bar{W}_n| > x) \to 0$ for every $\varepsilon > 0$ (Problem 2.3.3)
- Both the first and second conditions are valid if $||W_{n1}||_{2+\epsilon}$ is uniformly bounded for some $\epsilon > 0$

---

[0] $\|\xi\|_{2,1} = \int_0^\infty \sqrt{P(|\xi| > t)}\mathrm{d}t$

- **3.7.9 Example: Bayesian bootstrap.** If $Y_1, \ldots, Y_n$ are i.i.d. nonnegative random variables with $||Y_1||_{2,1} < \infty$, then the weights $W_{ni} = Y_i / \bar{Y}_n$ satisfy (3.7.8) with $c = \sigma(Y_1)/\mathrm{E} Y_1$.

- **3.7.9 Example: Bayesian bootstrap.** If $Y_1, \ldots, Y_n$ are i.i.d. nonnegative random variables with $||Y_1||_{2,1} < \infty$, then the weights $W_{ni} = Y_i / \bar{Y}_n$ satisfy (3.7.8) with $c = \sigma(Y_1)/\mathrm{E}Y_1$.

- If the variables Y_i are exponentially distributed with mean 1, then the weight vector $(W_{n1}, \ldots, W_{nn})$ follows the n-dimensional Dirichlet distribution with parameter vector $(1/n, \ldots, 1/n)$

- **3.7.10 Example: empirical bootstrap.** Multinomial vectors $(W_{n1}, \dots, W_{nn})$ with parameters n and (probabilities) $(1/n, \dots, 1/n)$ satisfy (3.7.8) with $c = 1$

- **3.7.10 Example: empirical bootstrap.** Multinomial vectors $(W_{n1}, \ldots, W_{nn})$ with parameters n and (probabilities) $(1/n, \ldots, 1/n)$ satisfy (3.7.8) with $c = 1$

- **3.7.11 Example: empirical bootstrap with k replicates.** The vectors $(W_{n1}, \ldots, W_{nn})$ equal to $\sqrt{n/k}$ times multinomial vectors with parameters k and probabilities $(1/n, \ldots, 1/n)$ satisfy (3.7.8) with $c = 1$ provided $k \to \infty$.

3.7.13 Theorem Let $\mathcal{F}$ be a Donsker class of measurable functions such that $\mathcal{F}_\delta$ is measurable for every $\delta > 0$. For each n let $(W_{n1}, \ldots, W_{nn})$ be an exchangeable, nonnegative random vector independent of $X_1, X_2, \ldots$ such that the conditions (3.7.8) are satisfied. Then as $n \to \infty$

$$\sup_{h \in BL_1} \left| \mathrm{E}_W h(\hat{\mathbb{G}}_n) - \mathrm{E} h(c\mathbb{G}) \xrightarrow{P^*} \right| 0.$$

Furthermore, the sequence $\mathrm{E}_W h(\hat{\mathbb{G}}_n)^* - \mathrm{E}_W h(\hat{\mathbb{G}}_n)_*$ converges to zero in outer probability. If $P^* ||f - Pf||^2_{\mathcal{F}} < \infty$, then the convergence is also outer almost surely.

[1] $\mathcal{F}_\delta = \{f - g : f, g \in \mathcal{F}, \rho_P(f - g) < \delta\}$

Proof of Theorem 3.7.13

- Without loss of generality, assume that $\bar{W}_n = 0$ and that $Pf = 0$ for every n and f

Proof of Theorem 3.7.13

- Without loss of generality, assume that $\bar{W}_n = 0$ and that $Pf = 0$ for every n and f
- It suffices to prove
  - Conditional almost-sure weak convergence of every marginal
  - Conditional asymptotic equicontinuity in probability or almost surely

Proof of Theorem 3.7.13: finite-dimensional convergence

Lemma 3.7.15 For each n, let $(a_{n1}, \ldots, a_{nn})$ and $(W_{n1}, \ldots, W_{nn})$ be a vector of numbers and an exchangeable random vector such that

$$\bar{a}_n = 0; \ \frac{1}{n}\sum_{i=1}^{n} a_{ni}^2 \to \sigma^2 > 0; \quad \lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} a_{ni}^2\{|a_{ni}| > M\} = 0;$$

$$\bar{W}_n = 0; \ \frac{1}{n}\sum_{i=1}^{n} W_{ni}^2 \xrightarrow{P} \tau^2 > 0; \qquad \frac{1}{n}\max_{1\le i\le n} W_{ni}^2 \xrightarrow{P} 0.$$

Then the sequence $n^{-1/2}\sum_{i=1}^{n} a_{ni}W_{ni}$ converges weakly to a $N(0, \sigma^2\tau^2)$-distribution.

Proof of Theorem 3.7.13: finite-dimensional convergence

- The variables W_{ni} satisfy the conditions of Lemma 3.7.15 by (3.7.8)

Proof of Theorem 3.7.13: finite-dimensional convergence

- The variables W_{ni} satisfy the conditions of Lemma 3.7.15 by (3.7.8)
- Let $a_{ni} = f(X_i) - \mathbb{P}_n f$. Since $Pf^2 < \infty$, almost surely
  - $\mathbb{P}_n(f - \mathbb{P}_n f)^2 \to Pf^2$
  - $\mathbb{P}_n(f - \mathbb{P}_n)^2\{|f - \mathbb{P}_n f| > M\} \to 0$

  as $n \to \infty$ followed by $M \to \infty$

Proof of Theorem 3.7.13: finite-dimensional convergence

- The variables W_{ni} satisfy the conditions of Lemma 3.7.15 by (3.7.8)
- Let $a_{ni} = f(X_i) - \mathbb{P}_n f$. Since $Pf^2 < \infty$, almost surely
  - $\mathbb{P}_n(f - \mathbb{P}_n f)^2 \to Pf^2$
  - $\mathbb{P}_n(f - \mathbb{P}_n)^2\{|f - \mathbb{P}_n f| > M\} \to 0$

  as $n \to \infty$ followed by $M \to \infty$
- Conditionally on the sequence $X_1, X_2, \ldots$, the sequence $\hat{\mathbb{G}}_n f = n^{-1/2} \sum_{i=1}^n W_{ni} \left(f(X_i) - \mathbb{P}_n f\right)$ is asymptotically normal $N(0, c^2 Pf^2)$

Proof of Theorem 3.7.13: finite-dimensional convergence

- The variables W_{ni} satisfy the conditions of Lemma 3.7.15 by (3.7.8)
- Let $a_{ni} = f(X_i) - \mathbb{P}_n f$. Since $Pf^2 < \infty$, almost surely
  - $\mathbb{P}_n(f - \mathbb{P}_n f)^2 \to Pf^2$
  - $\mathbb{P}_n(f - \mathbb{P}_n)^2\{|f - \mathbb{P}_n f| > M\} \to 0$

  as $n \to \infty$ followed by $M \to \infty$
- Conditionally on the sequence $X_1, X_2, \ldots$, the sequence $\hat{\mathbb{G}}_n f = n^{-1/2} \sum_{i=1}^n W_{ni}\,(f(X_i) - \mathbb{P}_n f)$ is asymptotically normal $N(0, c^2 Pf^2)$
- This establishes finite-dimensional convergence combined with the Cramér-Wold device ($\mathbf{X}_n \xrightarrow{d} \mathbf{X}$ iff $\mathbf{l}'\mathbf{X}_n \xrightarrow{d} \mathbf{l}'\mathbf{X}$)

Proof of Theorem 3.7.13: equicontinuity

- **Lemma 3.7.7 (Multiplier inequalities).** For arbitrary stochastic processes $Z_1, \ldots, Z_n$, every nonnegative exchangeable random vector $(\xi_1, \ldots, \xi_n)$ that is independent of $Z_1, \ldots, Z_n$, and any $1 \le n_0 \le n$,

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

Here $(R_1, \ldots, R_n)$ is uniformly distributed on the set of all permutations of $\{1, 2, \ldots, n\}$ and independent of $Z_1, \ldots, Z_n$

Proof of Theorem 3.7.13: equicontinuity

- By Lemma 3.7.7 applied with $Z_i = \delta_{X_i} - \mathbb{P}_n$

$$
\begin{aligned}
&\mathrm{E}_W ||\hat{\mathbb{G}}_n||_{\mathcal{F}_\delta} \\
=&\mathrm{E}_W \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni} \left(\delta_{X_i} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*} \\
\lesssim &\frac{n_0 - 1}{n} \sum_{i=1}^{n} ||\delta_{X_i} - \mathbb{P}_n||_{\mathcal{F}_\delta}^{*} \mathrm{E} \max_{1 \le i \le n} \frac{|W_{ni}|}{\sqrt{n}} \\
&+ ||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*}
\end{aligned}
$$

Proof of Theorem 3.7.13: equicontinuity

- By Lemma 3.7.7 applied with $Z_i = \delta_{X_i} - \mathbb{P}_n$

$$
\begin{aligned}
&\mathrm{E}_W ||\hat{\mathbb{G}}_n||_{\mathcal{F}_\delta} \\
=&\mathrm{E}_W \left|\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} W_{ni} \left(\delta_{X_i} - \mathbb{P}_n\right) \right|\right|^*_{\mathcal{F}_\delta} \\
\lesssim& \frac{n_0 - 1}{n} \sum_{i=1}^{n} ||\delta_{X_i} - \mathbb{P}_n||^*_{\mathcal{F}_\delta} \mathrm{E} \max_{1 \leq i \leq n} \frac{|W_{ni}|}{\sqrt{n}} \\
&+ ||W_{n1}||_{2,1} \max_{n_0 \leq k \leq n} \mathrm{E}_R \left|\left| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n\right) \right|\right|^*_{\mathcal{F}_\delta}
\end{aligned}
$$

- The first term converges to zero outer almost surely for every fixed n_0 by (3.7.8)

Proof of Theorem 3.7.13: equicontinuity

- The second term

$$||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*}$$

Proof of Theorem 3.7.13: equicontinuity

- The second term

$$||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^*$$

- $(X_{R_1}, \ldots, X_{R_n})$ is a random sample without replacement from $X_1, \ldots, X_n$

Proof of Theorem 3.7.13: equicontinuity

- The second term

$$||W_{n1}||_{2,1} \max_{n_0 \leq k \leq n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*}$$

- $(X_{R_1}, \ldots, X_{R_n})$ is a random sample without replacement from $X_1, \ldots, X_n$
- By Hoeffding's inequality (Lemma A.1.10), this term increases if the vector is replaced by a sample with replacement $\hat{X}_1, \ldots, \hat{X}_n$

Proof of Theorem 3.7.13: equicontinuity

- The second term

$$||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*}$$

- $(X_{R_1}, \ldots, X_{R_n})$ is a random sample without replacement from $X_1, \ldots, X_n$
- By Hoeffding's inequality (Lemma A.1.10), this term increases if the vector is replaced by a sample with replacement $\hat{X}_1, \ldots, \hat{X}_n$
- The sum can be extended to the range from 1 to k

Proof of Theorem 3.7.13: equicontinuity

- The second term

$$||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right) \right\|_{\mathcal{F}_\delta}^{*}$$

- $(X_{R_1}, \ldots, X_{R_n})$ is a random sample without replacement from $X_1, \ldots, X_n$
- By Hoeffding's inequality (Lemma A.1.10), this term increases if the vector is replaced by a sample with replacement $\hat{X}_1, \ldots, \hat{X}_n$
- The sum can be extended to the range from 1 to k
- $\frac{1}{\sqrt{k}} \sum_{i=n_0}^{k} \left(\delta_{X_{R_i}} - \mathbb{P}_n \right)$ is bounded by $\frac{1}{\sqrt{k}} \sum_{i=1}^{k} \left(\delta_{\hat{X}_i} - \mathbb{P}_n \right)$

Proof of Theorem 3.7.13: equicontinuity

- Conclude that the second term is bounded by

$$\sup_n ||W_{n1}||_{2,1} \max_{n_0 \leq k \leq n} \mathrm{E}_{\hat{X}} \left|\left|\hat{\mathbb{G}}_{n,k}\right|\right|^*_{\mathcal{F}_\delta}$$

  where $\hat{\mathbb{G}}_{n,k}$ is the multinomial bootstrap process

Proof of Theorem 3.7.13: equicontinuity

- Conclude that the second term is bounded by

$$\sup_n ||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_{\hat{X}} \left|\left|\hat{\mathbb{G}}_{n,k}\right|\right|^*_{\mathcal{F}_\delta}$$

  where $\hat{\mathbb{G}}_{n,k}$ is the multinomial bootstrap process
- This expression converges to zero in outer probability as $n_0, n \to \infty$ followed by $\delta \downarrow 0$ (proof of Theorem 3.7.3)

Proof of Theorem 3.7.13: equicontinuity

- Conclude that the second term is bounded by

$$\sup_n ||W_{n1}||_{2,1} \max_{n_0 \le k \le n} \mathrm{E}_{\hat{X}} \left|\left|\hat{\mathbb{G}}_{n,k}\right|\right|^*_{\mathcal{F}_\delta}$$

  where $\hat{\mathbb{G}}_{n,k}$ is the multinomial bootstrap process
- This expression converges to zero in outer probability as $n_0, n \to \infty$ followed by $\delta \downarrow 0$ (proof of Theorem 3.7.3)
- Under the additional condition on the envelope function, the convergence is outer almost surely

- **3.7.14 Example: Bootstrap without replacement.** The bootstrap without replacement is based on resampling $k < n$ observations from $X_1, \ldots, X_n$ without replacement

- **3.7.14 Example: Bootstrap without replacement.** The bootstrap without replacement is based on resampling $k < n$ observations from $X_1, \ldots, X_n$ without replacement

- This can be incorporated in the scheme of the theorem by letting $(W_{n1}, \ldots, W_{nn})$ be a row of k times the number $n(n-k)^{-1/2}k^{-1/2}$ and $n-k$ times the number 0, ordered at random, independent of the X's

- **3.7.14 Example: Bootstrap without replacement.** The bootstrap without replacement is based on resampling $k < n$ observations from $X_1, \ldots, X_n$ without replacement

- This can be incorporated in the scheme of the theorem by letting $(W_{n1}, \ldots, W_{nn})$ be a row of k times the number $n(n-k)^{-1/2}k^{-1/2}$ and $n-k$ times the number 0, ordered at random, independent of the X's

- Then the conditions (3.7.8) on the weights are satisfied for $c = 1$, provided both $k \to \infty$ and $n - k \to \infty$

3.7.14: Bootstrap without replacement

- In this case the assertion of the theorem can be phrased in terms of the empirical measures

$$\tilde{\mathbb{P}}_{k,n} = \frac{1}{k}\sum_{i=1}^{k}\delta_{X_{R_{ni}}}$$

where $(R_{n1}, \ldots, R_{nn})$ is a random permutation of the numbers $1, 2, \ldots, n$

3.7.14: Bootstrap without replacement

- In this case the assertion of the theorem can be phrased in terms of the empirical measures

$$\tilde{\mathbb{P}}_{k,n} = \frac{1}{k} \sum_{i=1}^{k} \delta_{X_{R_{ni}}}$$

  where $(R_{n1}, \ldots, R_{nn})$ is a random permutation of the numbers $1, 2, \ldots, n$

- If both $k \to \infty$ and $n - k \to \infty$ then the sequence

$$\sqrt{\frac{nk}{n-k}} \left(\tilde{\mathbb{P}}_{k,n} - \mathbb{P}_n \right)$$

  converges conditionally in distribution to a tight Brownian bridge

Lemma 3.7.15 For each n, let $(a_{n1}, \dots, a_{nn})$ and $(W_{n1}, \dots, W_{nn})$ be a vector of numbers and an exchangeable random vector such that

$$\bar{a}_n = 0; \quad \frac{1}{n}\sum_{i=1}^{n} a_{ni}^2 \to \sigma^2 > 0; \quad \lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} a_{ni}^2\{|a_{ni}| > M\} = 0;$$

$$\bar{W}_n = 0; \quad \frac{1}{n}\sum_{i=1}^{n} W_{ni}^2 \xrightarrow{P} \tau^2 > 0; \qquad \frac{1}{n}\max_{1\le i\le n} W_{ni}^2 \xrightarrow{P} 0.$$

Then the sequence $n^{-1/2}\sum_{i=1}^{n} a_{ni}W_{ni}$ converges weakly to a $N(0, \sigma^2\tau^2)$-distribution.

Proof of Lemma 3.7.15

- We are going to show that every subsequence of $\{n\}$ has a further subsequence along which

$$\sup_{h \in BL_1} \left| \mathrm{E}_R h \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{ni} W_{n,R_n i} \right) - \mathrm{E} h(\xi) \right| \to 0$$

  almost surely, for a standard normal variable ξ
- Then this bounded Lipschitz distance converges to zero in probability along the whole sequence $\{n\}$

Proof of Lemma 3.7.15

- We are going to show that every subsequence of $\{n\}$ has a further subsequence along which

$$\sup_{h \in BL_1} \left| \mathrm{E}_R h \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} a_{ni} W_{n,R_n i} \right) - \mathrm{E} h(\xi) \right| \to 0$$

  almost surely, for a standard normal variable ξ
- Then this bounded Lipschitz distance converges to zero in probability along the whole sequence $\{n\}$
- Take the expectation with respect to the W_{ni} to see that the sequence $n^{-1/2} \sum a_{ni} W_{n,R_{ni}}$ is unconditionally asymptotically normal as well
- By the exchangeability of W_{nj}, this sequence is equal in distribution to the sequence $n^{-1/2} \sum a_{ni} W_{ni}$

Proof of Lemma 3.7.15

- Without loss of generality, assume that both $n^{-1}\sum_{i=1}^n a_{ni}^2$ and $n^{-1}\sum_{i=1}^n W_{ni}^2$ are equal to 1 for every n

Proof of Lemma 3.7.15

- Without loss of generality, assume that both $n^{-1}\sum_{i=1}^{n} a_{ni}^2$ and $n^{-1}\sum_{i=1}^{n} W_{ni}^2$ are equal to 1 for every n
- We are going to use the rank central limit theorem (Proposition A.5.3)

$$\sum_i \sum_j \frac{a_{ni}^2 W_{nj}^2}{nn}\{|a_{ni}W_{nj}| > \varepsilon\sqrt{n}\}$$

$$\leq \frac{1}{n}\sum_i a_{ni}^2\{|a_{ni}| \max_j |W_{nj}| > \varepsilon\sqrt{n}\}$$

Proof of Lemma 3.7.15

- Without loss of generality, assume that both $n^{-1}\sum_{i=1}^n a_{ni}^2$ and $n^{-1}\sum_{i=1}^n W_{ni}^2$ are equal to 1 for every n
- We are going to use the rank central limit theorem (Proposition A.5.3)

$$\sum_i \sum_j \frac{a_{ni}^2 W_{nj}^2}{nn} \{|a_{ni} W_{nj}| > \varepsilon\sqrt{n}\}$$
$$\leq \frac{1}{n} \sum_i a_{ni}^2 \{|a_{ni}| \max_j |W_{nj}| > \varepsilon\sqrt{n}\}$$

- This expression converges to zero in probability as $n \to \infty$, for every $\varepsilon > 0$ since
  - $\max |W_{nj}|/\sqrt{n}$ converges to zero in probability array
  - $\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^n a_{ni}^2 \{|a_{ni}| > M\} = 0$

Proof of Lemma 3.7.15

- Combination with the assumption $1/n \max_{1\le i\le n} W_{ni}^2 \xrightarrow{P} 0$ shows that every subsequence of $\{n\}$ has a further subsequence along which, for every $\varepsilon > 0$,

$$\max_{1\le j\le n} \frac{1}{\sqrt{n}} |W_{nj}| \xrightarrow{as} 0 \quad \sum_i \sum_j \frac{a_{ni}^2 W_{nj}^2}{nn} \{|a_{ni} W_{ni}| > \varepsilon\sqrt{n}\} \xrightarrow{as} 0$$

Proof of Lemma 3.7.15

- Combination with the assumption $1/n \max_{1\leq i\leq n} W_{ni}^2 \xrightarrow{P} 0$ shows that every subsequence of $\{n\}$ has a further subsequence along which, for every $\varepsilon > 0$,

$$\max_{1\leq j\leq n} \frac{1}{\sqrt{n}}|W_{nj}| \xrightarrow{as} 0 \quad \sum_i \sum_j \frac{a_{ni}^2 W_{nj}^2}{nn}\{|a_{ni}W_{ni}| > \varepsilon\sqrt{n}\} \xrightarrow{as} 0$$

- For almost every realization of the W_{nj}, the conditions of rank central limit theorem (Proposition A.5.3) are satisfied along the subsequence.

Proof of Lemma 3.7.15

- Combination with the assumption $1/n \max_{1\leq i\leq n} W_{ni}^2 \xrightarrow{P} 0$ shows that every subsequence of $\{n\}$ has a further subsequence along which, for every $\varepsilon > 0$,

$$\max_{1\leq j\leq n} \frac{1}{\sqrt{n}}|W_{nj}| \xrightarrow{as} 0 \quad \sum_i \sum_j \frac{a_{ni}^2 W_{nj}^2}{nn}\{|a_{ni}W_{ni}| > \varepsilon\sqrt{n}\} \xrightarrow{as} 0$$

- For almost every realization of the W_{nj}, the conditions of rank central limit theorem (Proposition A.5.3) are satisfied along the subsequence.
- Conditionally on the W_{nj}, the subsequence of rank statistics $n^{-1/2}\sum_{i=1}^n a_{ni}W_{n,R_ni}$ is asymptotically standard normally distributed.

Lemma 3.7.16: multiplier Glivenko-Cantelli theorem. Let $\mathcal{F}$ be a Glivenko-Cantelli class of measurable functions. For each n, let $(W_{n1}, \ldots, W_{nn})$ be an exchangeable nonnegative random vector independent of $X_1, X_2, \ldots$ such that $\sum_{i=1}^{n} W_{ni} = 1$ and $\max_{1 \leq i \leq n} |W_{ni}|$ converges to zero in probability. Then, for every $\varepsilon > 0$, as $n \to \infty$,

$$P_W\left(\left\|\sum_{i=1}^{n} W_{ni}(\delta_{X_i} - P)\right\|_{\mathcal{F}}^{*} > \varepsilon\right) \xrightarrow{as*} 0.$$

Proof of Lemma 3.7.16

- By a modification of the multiplier inequality given by Lemma 3.7.7 ($L_1 \to L_r$ with $r < 1$), with $Z_i = \delta_{X_i} - P$

$$
\begin{aligned}
E_W \left\| \sum_{i=1}^{n} W_{ni} Z_i \right\|_{\mathcal{F}}^{*r} \leq & (n_0 - 1) \mathrm{E} \max_{1 \leq i \leq n} W_{ni}^r \frac{1}{n} \sum_{j=1}^{n} ||Z_j||_{\mathcal{F}}^{*r} \\
& + (n \mathrm{E} W_{n1})^r \max_{n_0 \leq k \leq n} \mathrm{E}_R \left\| \frac{1}{k} \sum_{j=1}^{k} Z_{R_j} \right\|_{\mathcal{F}}^{*r}
\end{aligned}
$$

Proof of Lemma 3.7.16

- By a modification of the multiplier inequality given by Lemma 3.7.7 ($L_1 \to L_r$ with $r < 1$), with $Z_i = \delta_{X_i} - P$

$$E_W \left\|\sum_{i=1}^n W_{ni} Z_i\right\|_{\mathcal{F}}^{*r} \leq (n_0 - 1)\mathrm{E} \max_{1 \leq i \leq n} W_{ni}^r \frac{1}{n} \sum_{j=1}^n ||Z_j||_{\mathcal{F}}^{*r} + (n\mathrm{E}W_{n1})^r \max_{n_0 \leq k \leq n} \mathrm{E}_R \left\|\frac{1}{k}\sum_{j=1}^k Z_{R_j}\right\|_{\mathcal{F}}^{*r}$$

- The average $n^{-1}\sum_{i=1}^n ||Z_i||_{\mathcal{F}}^{*r}$ is bounded by $\mathbb{P}_n F^{*r} + P^* F^r$, which converges almost surely to $2P^* F^r$

Proof of Lemma 3.7.16

- By a modification of the multiplier inequality given by Lemma 3.7.7 ($L_1 \to L_r$ with $r < 1$), with $Z_i = \delta_{X_i} - P$

$$E_W \left\|\sum_{i=1}^{n} W_{ni} Z_i\right\|_{\mathcal{F}}^{*r} \leq (n_0 - 1)\mathrm{E} \max_{1\leq i\leq n} W_{ni}^r \frac{1}{n}\sum_{j=1}^{n} ||Z_j||_{\mathcal{F}}^{*r} + (n\mathrm{E}W_{n1})^r \max_{n_0\leq k\leq n} \mathrm{E}_R \left\|\frac{1}{k}\sum_{j=1}^{k} Z_{R_j}\right\|_{\mathcal{F}}^{*r}$$

- The average $n^{-1}\sum_{i=1}^{n} ||Z_i||_{\mathcal{F}}^{*r}$ is bounded by $\mathbb{P}_n F^{*r} + P^* F^r$, which converges almost surely to $2P^* F^r$
- Since $\max_{1\leq i\leq n} |W_{ni}| \xrightarrow{P} 0$ and W_{ni} take their values in $[0,1]$, the sequence $\mathrm{E}\max_{1\leq i\leq n} W_{ni}^r$ converges to zero by the dominated convergence theorem

Proof of Lemma 3.7.16

- By a modification of the multiplier inequality given by Lemma 3.7.7 ($L_1 \to L_r$ with $r < 1$), with $Z_i = \delta_{X_i} - P$

$$E_W \left\| \sum_{i=1}^{n} W_{ni} Z_i \right\|_{\mathcal{F}}^{*r} \leq (n_0 - 1) \mathrm{E} \max_{1 \leq i \leq n} W_{ni}^r \frac{1}{n} \sum_{j=1}^{n} ||Z_j||_{\mathcal{F}}^{*r} + (n \mathrm{E} W_{n1})^r \max_{n_0 \leq k \leq n} \mathrm{E}_R \left\| \frac{1}{k} \sum_{j=1}^{k} Z_{R_j} \right\|_{\mathcal{F}}^{*r}$$

- The average $n^{-1} \sum_{i=1}^{n} ||Z_i||_{\mathcal{F}}^{*r}$ is bounded by $\mathbb{P}_n F^{*r} + P^* F^r$, which converges almost surely to $2P^* F^r$
- Since $\max_{1 \leq i \leq n} |W_{ni}| \xrightarrow{P} 0$ and W_{ni} take their values in $[0, 1]$, the sequence $\mathrm{E} \max_{1 \leq i \leq n} W_{ni}^r$ converges to zero by the dominated convergence theorem
- The first term on the right converges almost surely to zero for every fixed n_0

Proof of Lemma 3.7.16

- $n\mathrm{E}W_{n1}$ in the second term equals 1 by exchangeability and $\sum_{i=1}^{n} W_{ni} = 1$

Proof of Lemma 3.7.16

- $n\mathrm{E}W_{n1}$ in the second term equals 1 by exchangeability and $\sum_{i=1}^{n} W_{ni} = 1$
- By the triangle inequality, the second term

$$
\begin{aligned}
(n\mathrm{E}W_{n1})^r \max_{n_0 \le k \le n} \mathrm{E}_R \left\| \frac{1}{k} \sum_{j=1}^{k} Z_{R_j} \right\|_{\mathcal{F}}^{*r} \le & 2 \max_{n_0 - 1 \le k \le n} \mathrm{E}_R \left\| \frac{1}{k} \sum_{j=1}^{k} Z_{R_j} \right\|_{\mathcal{F}}^{*r} \\
= & 2 \max_{n_0 - 1 \le k \le n} \mathrm{E}(U_k^r | \mathcal{S}_n) \\
\le & 2\mathrm{E} \left(\max_{n_0 - 1 \le k} U_k^r | \mathcal{S}_n \right)
\end{aligned}
$$

where $U_k = ||k^{-1} \sum_{j=1}^{k} Z_j||_{\mathcal{F}}^{*}$, and $\mathcal{S}_n$ is the σ-field generated by all functions $f : \mathcal{X}^{\infty} \to R$ that are symmetric in their first n coordinates

Proof of Lemma 3.7.16

- The sequence $\mathcal{S}_n$ decreases to the symmetric σ-field $\mathcal{S}$, which consists of sets of probability 0 or 1 by the Hewitt-Savage zero–one law

- As $n \to \infty$,

$$\mathrm{E}\left(\max_{n_0-1\leq k} U_k^r|\mathcal{S}_n\right) \xrightarrow{as} \mathrm{E}\left(\max_{n_0-1\leq k} U_k^r|\mathcal{S}\right) = \mathrm{E}\left(\max_{n_0-1\leq k} U_k^r\right)$$

Proof of Lemma 3.7.16

- Since $\mathcal{F}$ is Glivenko-Cantelli, the sequence $U_k = ||k^{-1} \sum_{j=1}^{k} Z_j||_{\mathcal{F}}^{*}$ converges almost surely to zero

Proof of Lemma 3.7.16

- Since $\mathcal{F}$ is Glivenko-Cantelli, the sequence $U_k = ||k^{-1} \sum_{j=1}^{k} Z_j||_{\mathcal{F}}^{*}$ converges almost surely to zero
- $\mathrm{E}\left(\max_k U_k^r\right) < \infty$ by Problem 2.3.6

Proof of Lemma 3.7.16

- Since $\mathcal{F}$ is Glivenko-Cantelli, the sequence $U_k = ||k^{-1} \sum_{j=1}^{k} Z_j||_{\mathcal{F}}^{*}$ converges almost surely to zero

- $\mathrm{E}\left(\max_k U_k^r\right) < \infty$ by Problem 2.3.6

- $\mathrm{E}\left(\max_{n_0-1 \le k} U_k^r | \mathcal{S}_n\right) \xrightarrow{as} \mathrm{E}\left(\max_{n_0-1 \le k} U_k^r\right)$
  The right side converges to zero as $n_0 \to \infty$

- The left side converges to zero almost surely when $n \to \infty$ followed by $n_0 \to \infty$

Proof of Lemma 3.7.16

- To conclude, $E_W \left|\left|\sum_{i=1}^n W_{ni} Z_i\right|\right|_{\mathcal{F}}^{*r} \xrightarrow{a.s.} 0$
- Apply Markov's inequality to show

$$P_W\left(\left\|\sum_{i=1}^n W_{ni}(\delta_{X_i} - P)\right\|_{\mathcal{F}}^* > \varepsilon\right) \xrightarrow{as*} 0.$$