More Results on Bootstrapping Empirical Processes

Yangjianchen Xu

Department of Biostatistics
University of North Carolina at Chapel Hill

08/26/2021

Overview

- Continuous Mapping Results
- The Bootstrap for Glivenko-Cantelli Classes
- A Simple Z-Estimator Master Theorem

Continuous Mapping Results

We now assume a more general set-up, where

- $\hat{X}_n$ is a bootstrapped process in a Banach space $(\mathbb{D}, \|\cdot\|)$
- and is composed of the sample data $\mathcal{X}_n \equiv (X_1, \ldots, X_n)$
- and a random weight vector $M_n \in \mathbb{R}^n$ independent of $\mathcal{X}_n$.

We do not require that $X_1, \ldots, X_n$ be i.i.d.

In this section, we obtain two continuous mapping results.

- The first result, Proposition 1 (10.7 in Kosorok's book), is a simple continuous mapping results for the very special case of Lipschitz continuous maps.
- It is applicable to both the in-probability or outer-almost sure versions of bootstrap consistency.
- The second result, Theorem 1 (10.8 in Kosorok's book), is a considerably deeper result for general continuous maps applied to bootstraps which are consistent in probability.

Result for Lipschitz continuous maps

Proposition 1

Let $\mathbb{D}$ and $\mathbb{E}$ be Banach spaces, X a tight random variable on $\mathbb{D}$, and $g : \mathbb{D} \mapsto \mathbb{E}$ Lipschitz continuous. We have the following:

1. *If $\hat{X}_n \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} X$, then $g(\hat{X}_n) \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} g(X)$.*
2. *If $\hat{X}_n \underset{M}{\overset{\mathrm{as*}}{\rightsquigarrow}} X$, then $g(\hat{X}_n) \underset{M}{\overset{\mathrm{as*}}{\rightsquigarrow}} g(X)$.*

Recall that we use the notation $\hat{X}_n \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} X$ to mean that $\sup_{h \in BL_1} \left| \mathrm{E}_M h(\hat{X}_n) - \mathrm{E} h(X) \right| \overset{\mathrm{P}}{\to} 0$ and $\mathrm{E}_M h(\hat{X}_n)^* - \mathrm{E}_M h(\hat{X}_n)_* \overset{\mathrm{P}}{\to} 0$, for all $h \in BL_1$, where the subscript M in the expectations indicates conditional expectation over the weights M given the remaining data, and where $h(\hat{X}_n)^*$ and $h(\hat{X}_n)_*$ denote measurable majorants and minorants with respect to the joint data (including the weights M).

Proof of Proposition 1

Proof. Let $c_0 < \infty$ be the Lipschitz constant for g, and, without loss of generality, assume $c_0 \geq 1$. Note that for any $h \in BL_1(\mathbb{E})$, the map $x \mapsto h(g(x))$ is an element of $c_0 BL_1(\mathbb{D})$. Thus

$$\begin{aligned}\sup_{h \in BL_1(\mathrm{E})} \left| \mathrm{E}_M h(g(\hat{X}_n)) - \mathrm{E}h(g(X)) \right| &\leq \sup_{h \in c_0 BL_1(\mathbb{D})} \left| \mathrm{E}_M h(\hat{X}_n) - \mathrm{E}h(X) \right| \\ &= c_0 \sup_{h \in BL_1(\mathbb{D})} \left| \mathrm{E}_M h(\hat{X}_n) - \mathrm{E}h(X) \right|\end{aligned}$$

and the desired result follows by the respective definitions of $\underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}}$ and $\underset{M}{\overset{\mathrm{as}*}{\rightsquigarrow}}$.

Result for general continuous maps

Theorem 1

Let $g : \mathbb{D} \mapsto \mathbb{E}$ be continuous at all points in $\mathbb{D}_0 \subset \mathbb{D}$, where $\mathbb{D}$ and $\mathbb{E}$ are Banach spaces and $\mathbb{D}_0$ is closed. Assume that $M_n \mapsto h(\hat{X}_n)$ is measurable for every $h \in C_b(\mathbb{D})$ outer almost surely. Then if $\hat{X}_n \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} X$ in $\mathbb{D}$, where X is tight and $\mathrm{P}^* (X \in \mathbb{D}_0) = 1, g(\hat{X}_n) \underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}} g(X)$.

Proof. As in the proof of the implication (ii) $\Rightarrow$ (i) of Theorem 10.4 in Kosorok's book, we can argue that $\hat{X}_n \rightsquigarrow X$ unconditionally, and thus $g(\hat{X}_n) \rightsquigarrow g(X)$ unconditionally by the standard continuous mapping theorem. Moreover, we can replace $\mathbb{E}$ with its closed linear span so that the restriction of g to $\mathbb{D}_0$ has an extension $\tilde{g} : \mathbb{D} \mapsto \mathbb{E}$ which is continuous on all of $\mathbb{D}$ by Dugundji's extension theorem (Theorem 2 below).

Proof of Theorem 1, continued

Theorem 2 (Dugundji's extension theorem)

Let X be an arbitrary metric space, A a closed subset of X, L a locally convex linear space (which includes Banach vector spaces), and $f : A \mapsto L$ a continuous map. Then there exists a continuous extension of $f, F : X \mapsto L$. Moreover, $F(X)$ lies in the closed linear span of the convex hull of $f(A)$.

Thus $(g(\hat{X}_n), \tilde{g}(\hat{X}_n)) \rightsquigarrow (g(X), \tilde{g}(X))$, and hence $g(\hat{X}_n) - \tilde{g}(\hat{X}_n) \xrightarrow{\mathrm{P}} 0$. Therefore we can assume without loss of generality that g is continuous on all of $\mathbb{D}$. We can also assume without loss of generality that $\mathbb{D}_0$ is a separable Banach space since X is tight. Hence $\mathbb{E}_0 \equiv g(\mathbb{D}_0)$ is also a separable Banach space.

Proof of Theorem 1, continued

Fix $\epsilon > 0$. There now exists a compact $K \subset \mathbb{E}_0$ such that $\mathrm{P}(g(X) \notin K) < \epsilon$. By Theorem 10.10 in Kosorok's book, we know there exists an integer $k < \infty$, elements $z_1, \ldots, z_k \in C[0,1]$, continuous functions $f_1, \ldots, f_k : \mathbb{E} \mapsto \mathbb{R}$, and a Lipschitz continuous function $J : \overline{\text{lin}}\,(z_1, \ldots, z_k) \mapsto \mathbb{E}$, such that the map $x \mapsto T_\epsilon(x) \equiv J\left(\sum_{j=1}^k z_j f_j(x)\right)$ has domain $\mathbb{E}$ and range $\subset \mathbb{E}$ and satisfies $\sup_{x \in K} \|T_\epsilon(x) - x\| < \epsilon$. Let $BL_1 \equiv BL_1(\mathbb{E})$. We now have

$$\begin{aligned}
&\sup_{h \in BL_1} \left|\mathrm{E}_M h(g(\hat{X}_n)) - \mathrm{E}h(g(X))\,\right| \\
&\quad \leq \sup_{h \in BL_1} \left|\mathrm{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathrm{E}h\left(T_\epsilon g(X)\right)\right| \\
&\qquad + \mathrm{E}_M \left\{\left\|T_\epsilon g(\hat{X}_n) - g(\hat{X}_n)\right\| \wedge 2\right\} + \mathrm{E}\left\{\|T_\epsilon g(X) - g(X)\| \wedge 2\right\}
\end{aligned}$$

Proof of Theorem 1, continued

However, the outer expectation of the second term on the right converges to the third term, as $n \to \infty$, by the usual continuous mapping theorem. Thus, provided

$$\sup_{h \in BL_1} \left| \mathrm{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathrm{E}\, h(T_\epsilon g(X)) \right| \xrightarrow{\mathrm{P}} 0 \tag{1}$$

we have that

$$\begin{aligned} \lim_{n \to \infty} \sup \mathrm{E}^* \left\{ \sup_{h \in BL_1} \left| \mathrm{E}_M h(g(\hat{X}_n)) - \mathrm{E} h(g(X)) \right| \right\} & \\ \leq 2\mathrm{E} \left\{ \| T_\epsilon g(X) - g(X) \| \wedge 2 \right\} & \qquad (2) \\ \leq 2\mathrm{E} \left\| \{ T_\epsilon g(X) - g(X) \} 1\{ g(X) \in K \} \right\| + 4\mathrm{P}(g(X) \notin K) & \\ < 6\epsilon & \end{aligned}$$

Proof of Theorem 1, continued

Now note that for each $h \in BL_1, h(J(\sum_{j=1}^{k} z_j a_j)) = \tilde{h}(a_1, \ldots, a_k)$ for all $(a_1, \ldots, a_k) \in \mathbb{R}^k$ and some $\tilde{h} \in c_0 BL_1\left(\mathbb{R}^k\right)$, where $1 \leq c_0 < \infty$ (this follows since J is Lipschitz continuous and $\|\sum_{j=1}^{k} z_j a_j\| \leq \max_{1\leq j\leq k} |a_j| \times \sum_{j=1}^{k} \|z_j\|$). Hence

$$\begin{aligned} &\sup_{h\in BL_1} \left|\mathrm{E}_M h(T_\epsilon g(\hat{X}_n)) - \mathrm{E}h(T_\epsilon g(X))\right| \\ &\leq \sup_{h\in c_0 BL_1(\mathbb{R}^k)} \left|\mathrm{E}_M h(u(\hat{X}_n)) - \mathrm{E}h(u(X))\right| \\ &= c_0 \sup_{h\in BL_1(\mathbb{R}^k)} \left|\mathrm{E}_M h(u(\hat{X}_n)) - \mathrm{E}h(u(X))\right| \end{aligned} \tag{3}$$

where $x \mapsto u(x) \equiv (f_1(g(x)), \ldots, f_k(g(x)))$.

Proof of Theorem 1, continued

Fix any $v : \mathbb{R}^k \mapsto [0, 1]$ which is Lipschitz continuous (the Lipschitz constant may be > 1). Then, since $\hat{X}_n \rightsquigarrow X$ unconditionally,

$$\mathrm{E}^* \left\{ \mathrm{E}_M v(u(\hat{X}_n))^* - \mathrm{E}_M v(u(\hat{X}_n))_* \right\} \leq \mathrm{E}^* \left\{ v(u(\hat{X}_n))^* - v(u(\hat{X}_n))_* \right\} \to 0,$$

where sub- and super-script $*$ denote measurable majorants and minorants, respectively, with respect to the joint probability space of $(\mathcal{X}_n, M_n)$. Thus,

$$\left| \mathrm{E}_M v(u(\hat{X}_n)) - \mathrm{E}_M v(u(\hat{X}_n))^* \right| \xrightarrow{\mathrm{P}} 0. \qquad (4)$$

Note that we are using at this point the outer almost sure measurability of $M_n \mapsto v(u(\hat{X}_n))$ to ensure that $E_M v(u(\hat{X}_n))$ is well defined, even if the resulting random expectation is not itself measurable.

Proof of Theorem 1, continued

Now, for every subsequence n', there exists a further subsequence n'' such that $\hat{X}_{n''} \overset{\text{as}*}{\underset{M}{\rightsquigarrow}} X$. This means that for this subsequence, the set B of data subsequences $\{\mathcal{X}_{n''} : n \geq 1\}$ for which $\mathrm{E}_M v(u(\hat{X}_{n''})) - \mathrm{E}v(u(X)) \to 0$ has inner probability 1. Combining this with (4) and Proposition 7.22 in Kosorok's book, we obtain that $\mathrm{E}_M v(u(\hat{X}_n)) - \mathrm{E}v(u(X)) \xrightarrow{\mathrm{P}} 0$. Since v was an arbitrary real, Lipschitz continuous function on $\mathbb{R}^k$, we now have by Part (i) of Lemma 3 below followed by Lemma 4 below, that

$$\sup_{h \in BL_1(\mathbb{R}^k)} \left| \mathrm{E}_M h(u(\hat{X}_n)) - \mathrm{E}h(u(X)) \right| \xrightarrow{\mathrm{P}} 0$$

Combining this with (3), we obtain that (1) is satisfied. The desired result now follows from (2), since $\epsilon > 0$ was arbitrary.

Proof of Theorem 1, continued

Lemma 3

Let X_n *and* X *be random variables in* $\mathbb{R}^k$ *for all* $n \geq 1$. *Define* $\mathcal{S} \subset [\mathbb{R} \cup \{-\infty, \infty\}]^k$ *to be the set of all continuity points of* $t \mapsto F(t) \equiv \mathrm{P}(X \leq t)$ *and* H *to be the set of all Lipschitz continuous functions* $h : \mathbb{R}^k \mapsto [0, 1]$ *(the Lipschitz constants may be* > 1*). Then, provided the expectations are well defined, we have:*

1. *If* $\mathrm{E}\left[h\left(X_n\right) \mid \mathcal{Y}_n\right] \overset{\mathrm{P}}{\to} \mathrm{E}h(X)$ *for all* $h \in H$, *then* $\sup_{t \in A} \left| \mathrm{P}\left(X_n \leq t \mid \mathcal{Y}_n\right) - F(t) \right| \overset{\mathrm{P}}{\to} 0$ *for all closed* $A \subset \mathcal{S}$;
2. *If* $\mathrm{E}\left[h\left(X_n\right) \mid \mathcal{Y}_n\right] \overset{\mathrm{as}*}{\to} \mathrm{E}h(X)$ *for all* $h \in H$, *then* $\sup_{t \in A} \left| \mathrm{P}\left(X_n \leq t \mid \mathcal{Y}_n\right) - F(t) \right| \overset{\mathrm{as}*}{\to} 0$ *for all closed* $A \subset \mathcal{S}$.

Proof of Theorem 1, continued

Lemma 4

Let $\{F_n\}$ *and* F *be distribution functions on* $\mathbb{R}^k$*, and let* $\mathcal{S} \subset [\mathbb{R} \cup \{-\infty, \infty\}]^k$ *be the set of all continuity points of* F*. Then the following are equivalent:*

1. $\sup_{t \in A} |F_n(t) - F(t)| \to 0$ *for all closed* $A \subset \mathcal{S}$.
2. $\sup_{h \in BL_1(\mathbb{R}^k)} \left| \int_{\mathbb{R}^k} h \, (dF_n - dF) \right| \to 0$.

The Bootstrap for Glivenko-Cantelli Classes

We now present several results for the bootstrap applied to Glivenko-Cantelli classes. The primary use of these results is to assist verification of consistency of bootstrapped estimators.

- The first theorem (Theorem 5) consists of various multiplier bootstrap results, and it is followed by a corollary (Corollary 6) which applies to certain weighted bootstrap results.
- The final theorem of this section (Theorem 10.15) gives corresponding results for the multinomial bootstrap.

Theorem 5

Let $\mathcal{F}$ be a class of measurable functions, and let $\xi_1, \ldots, \xi_n$ be i.i.d. nonconstant random variables with $0 < \mathrm{E}|\xi| < \infty$ and independent of the sample data $X_1, \ldots, X_n$. Let $\mathbb{W}_n \equiv n^{-1} \sum_{i=1}^{n} \xi_i \left(\delta_{X_i} - P\right)$ and $\tilde{\mathbb{W}}_n \equiv n^{-1} \sum_{i=1}^{n} \left(\xi_i - \bar{\xi}\right) \delta_{X_i}$, where $\bar{\xi} \equiv n^{-1} \sum_{i=1}^{n} \xi_i$. Then the following are equivalent:

- (i) *$\mathcal{F}$ is strong Glivenko-Cantelli;*
- (ii) $\|\mathbb{W}_n\|_{\mathcal{F}} \overset{as*}{\to} 0$;
- (iii) $\mathrm{E}_\xi \|\mathbb{W}_n\|_{\mathcal{F}} \overset{as*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$;
- (iv) *For every* $\eta > 0, \mathrm{P}\left(\|\mathbb{W}_n\|_{\mathcal{F}} > \eta \mid \mathcal{X}_n\right) \overset{\text{as*}}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$ *where* $\mathcal{X}_n \equiv (X_1, \ldots, X_n)$;
- (v) *For every* $\eta > 0, \mathrm{P}\left(\|\mathbb{W}_n\|_{\mathcal{F}}^* > \eta \mid \mathcal{X}_n\right) \overset{\text{as*}}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$, *for some version of* $\|\mathbb{W}_n\|_{\mathcal{F}}^*$, *where the superscript* $*$ *denotes a measurable majorant with respect to* $(\xi_1, \ldots, \xi_n, X_1, \ldots, X_n)$ *jointly;*

Theorem 5

(vi) $\|\tilde{\mathbb{W}}_n\|_{\mathcal{F}} \overset{\text{as*}}{\to} 0$;

(vii) $\mathrm{E}_\xi\|\tilde{\mathbb{W}}_n\|_{\mathcal{F}} \overset{\text{as*}}{\to} 0$ and $P^*\|f - Pf\|_{\mathcal{F}} < \infty$;

(viii) For every $\eta > 0, \mathrm{P}(\|\tilde{\mathbb{W}}_n\|_{\mathcal{F}} > \eta \mid \mathcal{X}_n) \overset{\text{as*}}{\to} 0$ and $P^*\|f - Pf\|_{\mathcal{F}} < \infty$;

(ix) For every $\eta > 0, \mathrm{P}(\|\tilde{\mathbb{W}}_n\|^*_{\mathcal{F}} > \eta \mid \mathcal{X}_n) \overset{\text{as*}}{\to} 0$ and $P^*\|f - Pf\|_{\mathcal{F}} < \infty$ for some version of $\|\tilde{\mathbb{W}}_n\|^*_{\mathcal{F}}$.

The distinctions between (iv) and (v) and between (viii) and (ix) are not as trivial as they appear. This is because the measurable majorants involved are computed with respect to $(\xi_1, \ldots, \xi_n, X_1, \ldots, X_1)$ jointly, and thus the differences between $\|\mathbb{W}_n\|_{\mathcal{F}}$ and $\|\mathbb{W}_n\|^*_{\mathcal{F}}$ or between $\|\tilde{W}_n\|_{\mathcal{F}}$ and $\|\tilde{W}_n\|_{\mathcal{F}}$ may be nontrivial.

Results for certain weighted bootstrap

Corollary 6

Let $\mathcal{F}$ be a class of measurable functions, and let $\xi_1, \ldots, \xi_n$ be i.i.d. nonconstant, nonnegative random variables with $0 < \mathrm{E}\xi < \infty$ and independent of $X_1, \ldots, X_n$. Let $\tilde{\mathbb{P}}_n \equiv n^{-1} \sum_{i=1}^{n} \left(\xi_i / \bar{\xi}\right) \delta_{X_i}$ where we set $\tilde{\mathbb{P}}_n = 0$ when $\bar{\xi} = 0$. Then the following are equivalent:

(i) *$\mathcal{F}$ is strong Glivenko-Cantelli;*

(ii) $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{\mathrm{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(iii) $\mathrm{E}_\xi \|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{as*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(iv) *For every* $\eta > 0, \mathrm{P}(\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} > \eta \mid \mathcal{X}_n) \overset{\mathrm{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(v) *For every* $\eta > 0, \mathrm{P}(\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}}^* > \eta \mid \mathcal{X}_n) \overset{as*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*, for some version of* $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}}^*$*;*

If in addition $P(\xi = 0) = 0$, then the requirement that $P^*\|f - Pf\|_{\mathcal{F}} < \infty$ in (ii) may be dropped.

Proof of Corollary 6

Proof. Since the processes $\mathbb{P}_n - P$ and $\tilde{\mathbb{P}}_n - \mathbb{P}_n$ do not change when the class $\mathcal{F}$ is replaced with $\dot{\mathcal{F}} \equiv \{f - Pf : f \in \mathcal{F}\}$, we can assume $\|P\|_{\mathcal{F}} = 0$ without loss of generality. Let the envelope of $\dot{\mathcal{F}}$ be denoted $\dot{F} \equiv \|f\|^*_{\dot{\mathcal{F}}}$. Since multiplying the ξ_i by a constant does not change $\xi_i/\bar{\xi}$, we can also assume $\mathrm{E}\xi = 1$ without loss of generality. The fact that the conditional expressions in (iii) and (iv) are well defined can be argued is in the proof of Theorem 5, and we do not repeat the details here.

Proof of Corollary 6, continued

(i)⇒(ii): Since

$$\tilde{\mathbb{P}}_n - \mathbb{P}_n - \tilde{\mathbb{W}}_n = \left(\frac{1}{\bar{\xi}} - 1\right) 1\{\bar{\xi} > 0\}\tilde{\mathbb{W}}_n - 1\{\bar{\xi} = 0\}\mathbb{P}_n, \tag{5}$$

(ii) follows by Theorem 5 and the fact that $\bar{\xi} \overset{\text{as*}}{\to} 1$.
(ii)⇒(i): Note that

$$\tilde{\mathbb{P}}_n - \mathbb{P}_n - \tilde{\mathbb{W}}_n = -(\bar{\xi} - 1)1\{\bar{\xi} > 0\}(\tilde{\mathbb{P}}_n - \mathbb{P}_n) - 1\{\bar{\xi} = 0\}\mathbb{P}_n, \tag{6}$$

The first term on the right $\overset{\text{as*}}{\to}$ 0 by (ii), while the second term on the right is bounded in absolute value by
$1\{\bar{\xi} = 0\} \|\mathbb{P}_n\|_{\dot{\mathcal{F}}} \leq 1\{\bar{\xi} = 0\}\mathbb{P}_n\dot{F} \overset{\text{as*}}{\to} 0$, by the moment condition.

Proof of Corollary 6, continued

(ii)⇒(iii): The method of proof will be to use the expansion (5) to show that $\mathrm{E}_\xi \|\tilde{\mathbb{P}}_n - \mathbb{P}_n - \tilde{W}_n\|_{\mathcal{F}} \overset{\text{as*}}{\to} 0$. Then (iii) will follow by Theorem 5 and the established equivalence between (ii) and (i). Along this vein, we have by symmetry followed by an application of Theorem 9.29 in Kosorok's book that

$$\begin{aligned}
\mathrm{E}_\xi \left\{ \left| \frac{1}{\bar{\xi}} - 1 \right| 1\{\bar{\xi} > 0\} \|\tilde{\mathbb{W}}_n\|_{\dot{\mathcal{F}}} \right\} &\le \frac{1}{n} \sum_{i=1}^{n} \dot{F}(X_i)\, \mathrm{E}_\xi \left\{ \xi_i \left| \frac{1}{\bar{\xi}} - 1 \right| 1\{\bar{\xi} > 0\} \right] \\
&= \mathbb{P}_n \dot{F} \mathrm{E}_\xi \{ |1 - \bar{\xi}| 1\{\bar{\xi} > 0\} \} \\
&\overset{\text{as*}}{\to} 0
\end{aligned}$$

Since also $\mathrm{E}_\xi[1\{\bar{\xi} = 0\}]\, \|\mathbb{P}_n\|_{\dot{\mathcal{F}}} \overset{\text{as*}}{\to} 0$, the desired conclusion follows.
(iii)⇒(iv): This is obvious.

Proof of Corollary 6, continued

(iv)$\Rightarrow$(i): Consider again expansion (6). The moment conditions easily give us, conditional on $X_1, X_2, \ldots$, that $1\{\bar{\xi} = 0\} \|\mathbb{P}_n\|_{\dot{\mathcal{F}}} \leq 1\{\bar{\xi} = 0\}\mathbb{P}_n \dot{F} \xrightarrow{\text{P}} 0$ for almost all sequences $X_1, X_2, \ldots$ By (iv), we also obtain that $|\bar{\xi} - 1| 1\{\bar{\xi} > 0\} \|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \xrightarrow{\text{P}} 0$ for almost all sequences $X_1, X_2, \ldots$. Thus Assertion (viii) of Theorem 5 follows, and we obtain (i).

If $P(\xi = 0) = 0$, then $1\{\bar{\xi} = 0\}\mathbb{P}_n = 0$ almost surely, and we no longer need the moment condition $P\dot{F} < \infty$ in the proofs of (ii) $\Rightarrow$ (i) and (ii) $\Rightarrow$(iii), and thus the moment condition in (ii) can be dropped in this setting.

Proof of Corollary 6, continued

(ii)$\Rightarrow$(v): Assertion (ii) implies that there exists a measurable set B of infinite sequences $(\xi_1, X_1), (\xi_2, X_2), \ldots$ with $\mathrm{P}(B) = 1$ such that $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} \to 0$ on B for some version of $\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}}$. Let $E_{\xi,\infty}$ be the expectation taken over the infinite sequence $\xi_1, \xi_2, \ldots$ holding the infinite sequence $X_1, X_2, \ldots$ fixed. By the bounded convergence theorem, we have for any $\eta > 0$ and almost all sequences $X_1, X_2, \ldots$

$$\begin{aligned}
\limsup_{n\to\infty} \mathrm{P}(\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta \mid \mathcal{X}_n) &= \limsup_{n\to\infty} \mathrm{E}_{\xi,\infty} 1\left\{\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta\right\} 1\{B\} \\
&= \mathrm{E}_{\xi,\infty} \limsup_{n\to\infty} 1\left\{\|\tilde{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta\right\} 1\{B\} \\
&= 0.
\end{aligned}$$

Thus (v) follows.

(v)$\Rightarrow$(iv): This is obvious.

Results for the multinomial bootstrap

Theorem 7

Let $\mathcal{F}$ be a class of measurable functions, and let the multinomial vectors W_n in $\hat{\mathbb{P}}_n$ be independent of the data. Then the following are equivalent:

(i) *$\mathcal{F}$ is strong Glivenko-Cantelli;*

(ii) $\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{\mathrm{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(iii) $\mathrm{E}_W\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \overset{as*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(iv) *For every* $\eta > 0, \mathrm{P}(\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} > \eta \mid \mathcal{X}_n) \overset{\mathrm{as}*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*;*

(v) *For every* $\eta > 0, \mathrm{P}(\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}} > \eta \mid \mathcal{X}_n) \overset{as*}{\to} 0$ *and* $P^*\|f - Pf\|_{\mathcal{F}} < \infty$*, for some version of* $\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|^*_{\mathcal{F}}$*;*

A Simple Z-Estimator Master Theorem

Consider Z-estimation based on the estimating equation

$$\theta \mapsto \Psi_n(\theta) \equiv \mathbb{P}_n \psi_\theta$$

where $\theta \in \Theta \subset \mathbb{R}^p$ and $x \mapsto \psi_\theta(x)$ is a measurable p-vector valued function for each θ. Define the map

$$\theta \mapsto \Psi(\theta) \equiv P\psi_\theta,$$

and assume $\theta_0 \in \Theta$ satisfies $\Psi(\theta_0) = 0$.

A Simple Z-Estimator Master Theorem

Let $\hat{\theta}_n$ be an approximate zero of Ψ_n, and let $\hat{\theta}_n^\circ$ be an approximate zero of the bootstrapped estimating equation

$$\theta \mapsto \Psi_n^\circ(\theta) \equiv \mathbb{P}_n^\circ \psi_\theta$$

where $\mathbb{P}_n^\circ$ is either

- $\tilde{\mathbb{P}}_n$ of Corollary 6, with $\xi_1, \ldots, \xi_n$ satisfying the conditions specified in the first paragraph of Section 10.1.3 (the multiplier bootstrap),
- or $\hat{\mathbb{P}}_n$ of Theorem 7 (the multinomial bootstrap).

A Simple Z-Estimator Master Theorem

The goal is to determine reasonably general conditions under which

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z$$

where Z is mean zero normally distributed, and

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{\mathrm{P}}{\rightsquigarrow}} k_0 Z$$

Here, we use $\underset{\circ}{\overset{\mathrm{P}}{\rightsquigarrow}}$ to denote either $\underset{\mathbb{P}}{\overset{\mathrm{P}}{\rightsquigarrow}}$ or $\underset{M}{\overset{\mathrm{P}}{\rightsquigarrow}}$ depending on which bootstrap is being used, and

- $k_0 = \tau/\mu$ for the multiplier bootstrap,
- while $k_0 = 1$ for the multinomial bootstrap.

A Simple Z-Estimator Master Theorem

Theorem 8

Let $\Theta \subset \mathbb{R}^p$ *be open, and assume* $\theta_0 \in \Theta$ *satisfies* $\Psi(\theta_0) = 0$. *Also assume the following:*

(A) *For any sequence* $\{\theta_n\} \in \Theta, \Psi(\theta_n) \to 0$ *implies* $\|\theta_n - \theta_0\| \to 0$*;*

(B) *The class* $\{\psi_\theta : \theta \in \Theta\}$ *is strong Glivenko-Cantelli;*

(C) *For some* $\eta > 0$*, the class* $\mathcal{F} \equiv \{\psi_\theta : \theta \in \Theta, \|\theta - \theta_0\| \leq \eta\}$ *is Donsker and* $P\|\psi_\theta - \psi_{\theta_0}\|^2 \to 0$ *as* $\|\theta - \theta_0\| \to 0$

(D) $P\|\psi_{\theta_0}\|^2 < \infty$ *and* $\Psi(\theta)$ *is differentiable at* θ_0 *with nonsingular derivative matrix* V_{θ_0}*;*

(E) $\Psi_n(\hat{\theta}_n) = o_P\left(n^{-1/2}\right)$ *and* $\Psi_n^\circ(\hat{\theta}_n^\circ) = o_P\left(n^{-1/2}\right)$.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z \sim N(0, V_{\theta_0}^{-1} P\left[\psi_{\theta_0}\psi_{\theta_0}^T\right]\left(V_{\theta_0}^{-1}\right)^T)$$

and $\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) \underset{\circ}{\overset{\mathrm{P}}{\rightsquigarrow}} k_0 Z$.

A Simple Z-Estimator Master Theorem

- Condition (A) is one of several possible identifiability conditions.
- Condition (B) is a sufficient condition, when combined with (A), to yield consistency of a zero of Ψ_n.
- Condition (C) is needed for asymptotic normality of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ and is also not hard to verify in practice.
- Condition (D) enables application of the delta method at the appropriate juncture in the proof.
- Condition (E) is a specification of the level of approximation permitted in obtaining the zeros of the estimating equations.

Proof of Theorem 8

By (B) and (E),

$$\|\Psi(\hat{\theta}_n)\| \leq \|\Psi_n(\hat{\theta}_n)\| + \sup_{\theta\in\Theta} \|\Psi_n(\theta) - \Psi(\theta)\| \leq o_P(1)$$

Thus $\hat{\theta}_n \overset{\mathrm{P}}{\to} \theta_0$ by the identifiability Condition (A). By Assertion (ii) of either Corollary 6 or Theorem 7 (depending on which bootstrap is used), Condition (B) implies $\sup_{\theta\in\Theta} \|\Psi_n^\circ(\theta) - \Psi(\theta)\| \overset{\text{as*}}{\to} 0$. Thus reapplication of Conditions (A) and (E) yield $\hat{\theta}_n^\circ \overset{\mathrm{P}}{\to} \theta_0$.

Proof of Theorem 8, continued

By (C) and the consistency of $\hat{\theta}_n$, we have $\mathbb{G}_n\psi_{\hat{\theta}_n} - \mathbb{G}_n\psi_{\theta_0} \xrightarrow{\text{P}} 0$. Since (E) now implies that $\mathbb{G}_n\psi_{\hat{\theta}_n} = \sqrt{n}P(\psi_{\theta_0} - \psi_{\hat{\theta}_n}) + o_P(1)$, we can use the parametric (Euclidean) delta method plus differentiability of Ψ to obtain

$$\sqrt{n}V_{\theta_0}(\theta_0 - \hat{\theta}_n) + \sqrt{n}o_P(\|\hat{\theta}_n - \theta_0\|) = \mathbb{G}_n\psi_{\theta_0} + o_P(1). \qquad (7)$$

Since V_{θ_0} is nonsingular, this yields that $\sqrt{n}\|\hat{\theta}_n - \theta_0\|\,(1 + o_P(1)) = O_P(1)$, and thus $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_P(1)$. Combining this with (7), we obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -V_{\theta_0}^{-1}\sqrt{n}\mathbb{P}_n\psi_{\theta_0} + o_P(1) \qquad (8)$$

and thus $\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow Z$ with the specified covariance.

Proof of Theorem 8, continued

The first part of Condition (C) and Theorem 2.6 imply that $\mathbb{G}_n^\circ \equiv k_0^{-1}\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n) \rightsquigarrow \mathbb{G}$ in $\ell^\infty(\mathcal{F})$ unconditionally, by arguments similar to those used in the (ii) $\Rightarrow$ (i) part of the proof of Theorem 10.4. Combining this with the second part of Condition (C), we obtain $k_0\mathbb{G}_n^\circ(\psi_{\hat{\theta}_n^\circ}) + \mathbb{G}_n(\psi_{\hat{\theta}_n^\circ}) - k_0\mathbb{G}_n^\circ(\psi_{\theta_0}) - \mathbb{G}_n(\psi_{\theta_0}) \xrightarrow{\mathrm{P}} 0$. Condition (E) now implies $\sqrt{n}P(\psi_{\theta_0} - \psi_{\hat{\theta}_n^\circ}) = \sqrt{n}\mathbb{P}_n^\circ\psi_{\theta_0} + o_P(1)$. Using similar arguments to those used in the previous proof, we obtain

$$\sqrt{n}(\hat{\theta}_n^\circ - \theta_0) = -V_{\theta_0}^{-1}\sqrt{n}\mathbb{P}_n^\circ\psi_{\theta_0} + o_P(1) \qquad (9)$$

Combining with (8), we have

$$\sqrt{n}(\hat{\theta}_n^\circ - \hat{\theta}_n) = -V_{\theta}^{-1}\sqrt{n}(\mathbb{P}_n^\circ - \mathbb{P}_n)\psi_{\theta_0} + o_P(1) \qquad (10)$$

The desired conditional bootstrap convergence now follows from Theorem 2.6, Part (ii) or Part (iii) (depending on which bootstrap is used).