Bootstrap Limit Theory

Jon A. Wellner

University of Washington, Seattle

Research Day Talk, University of Washington

September 27, 2010

Outline

- 1: Efron's nonparametric bootstrap.
- 2: Bootstrapping the mean: limit theory.
- 3: Bootstrapping the empirical process: limit theory.
- 4: Exchangeably weighted bootstrap:
  limit theory and a problem.
- 5: Conjectures, approaches, and connections.

1. Efron's nonparametric bootstrap

.

- $X_1, \ldots, X_n$ i.i.d. P on $(S, \mathcal{S})$.
- Empirical measure $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$.
- $T(P)$ is some functional of P.
- Natural estimator of $T(P)$: "plug-in" estimator $T(\mathbb{P}_n) \equiv T_n$.

1. Efron's nonparametric bootstrap

Consider estimation of:

- A: $b_n(P) \equiv n\{E_P(T_n) - T(P)\}$.
- B: $n\sigma_n^2(P) \equiv nVar_P(T_n)$.
- C: $\kappa_{3,n}(P) \equiv E_P[T_n - E_P(T_n)]^3/\sigma_n^3(P)$.
- D: $H_n(x, P) \equiv Pr_P(\sqrt{n}(T_n - T(P)) \le x)$.
- E: $K_n(x, P) \equiv Pr_P(\sqrt{n}\|\mathbb{P}_n - P\|_{Kol} \le x)$.
- F: $L_n(x, P) \equiv Pr_P(\sqrt{n}\|\mathbb{P}_n - P\|_{\mathcal{F}} \le x)$ where $\mathcal{F}$ is a class of functions for which the central limit theorem holds uniformly over $\mathcal{F}$ (i.e. a *Donsker class*).

1. Efron's nonparametric bootstrap

(ideal) nonparametric bootstrap estimates: estimate P by $\mathbb{P}_n$. This yields the following nonparametric bootstrap estimates in examples A - F: with $\hat{X}_1, \ldots, \hat{X}_n$ i.i.d. $\mathbb{P}_n$, $\hat{T}_n \equiv T(\hat{\mathbb{P}}_n)$,

- A': $b_n(\mathbb{P}_n) \equiv n\{E_{\mathbb{P}_n}(\hat{T}_n) - T(\mathbb{P}_n)\}$.
- B': $n\sigma_n^2(\mathbb{P}_n) \equiv nVar_{\mathbb{P}_n}(\hat{T}_n)$.
- C': $\kappa_{3,n}(\mathbb{P}_n) \equiv E_{\mathbb{P}_n}[\hat{T}_n - E_{\mathbb{P}_n}(\hat{T}_n^*)]^3/\sigma_n^3(\mathbb{P}_n)$.
- D': $H_n(x, \mathbb{P}_n) \equiv Pr_{\mathbb{P}_n}(\sqrt{n}(\hat{T}_n - T(\mathbb{P}_n)) \le x)$.
- E': $K_n(x, \mathbb{P}_n) \equiv Pr_{\mathbb{P}_n}(\sqrt{n}\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{Kol} \le x)$. $(S = R)$
- F': $L_n(x, \mathbb{P}_n) \equiv Pr_{\mathbb{P}_n}(\sqrt{n}\|\hat{\mathbb{P}}_n - \mathbb{P}_n\|_{\mathcal{F}} \le x)$, ($S$ general) where $\mathcal{F}$ is a class of real-valued functions for which the central limit theorem holds uniformly over $\mathcal{F}$ (i.e. a *Donsker class*).

1. Efron's nonparametric bootstrap

For estimation of $T(P)$ by $T(\mathbb{P}_n)$ suppose that

$$\sqrt{n}(T(\mathbb{P}_n) - T(P)) \rightarrow_d N(0, V^2(P)).$$

Goal: show that the bootstrap estimator satisfies

$$\sqrt{n}(T(\widehat{\mathbb{P}}_n) - T(\mathbb{P}_n)) \rightarrow_d N(0, V^2(P)) \qquad \text{in probability or a.s.}$$

conditionally on $X_1, \ldots, X_n$ (i.e. conditionally on $\mathbb{P}_n$).

2. Bootstrapping the mean: limit theory

- $T(P) = E_P(X) = \int x dP(x) \equiv \mu(P)$ when $S = R$.
- If $E_P(X^2) < \infty$, then by the classical (Lindeberg) CLT

$$\sqrt{n}(T(\mathbb{P}_n) - T(P)) = \sqrt{n}(\overline{X}_n - \mu(P)) \rightarrow_d N(0, Var_P(X)).$$

The corresponding statement for the bootstrap is:

Theorem. (Bickel-Freedman, 1981)
If $E_P X^2 < \infty$, then for a.e. sequence $X_1, X_2, \ldots$,

$$\sqrt{n}(T(\widehat{\mathbb{P}}_n) - T(\mathbb{P}_n)) = \sqrt{n}(\overline{\widehat{X}}_n - \overline{X}_n) \ \rightarrow_d N(0, Var_P(X)).$$

Conclusion: The bootstrap "works" for estimation of the mean if $E_P(X^2) < \infty$.

3. Bootstrapping the empirical process: limit theory

Empirical process notation and theory:

- $X_1, \ldots, X_n$ i.i.d. P on $(S, \mathcal{S})$
- $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ = empirical measure.
- $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P)$ = empirical process.
- $\mathcal{F}$ = a class of real-valued (measurable) functions defined on S
- $P(f) = \int f dP = E_P f(X)$,
  $\mathbb{P}_n(f) = \int f d\mathbb{P}_n = n^{-1} \sum_{i=1}^n f(X_i)$,
  $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f))$.
- If $\mathcal{F} \subset L_1(P)$, then $\mathbb{P}_n(f) \rightarrow_{a.s.} P(f)$ for each $f \in \mathcal{F}$.
- If $\mathcal{F} \subset L_2(P)$, then $\mathbb{G}_n(f) = \sqrt{n}(\mathbb{P}_n(f) - P(f)) \rightarrow_d \mathbb{G}_P(f) \sim N(0, Var_P(f(X)))$, and

$$\mathbb{G}_n \stackrel{f.d.}{\rightarrow} \mathbb{G}_P.$$

3. Bootstrapping the empirical process: limit theory

Some terminology:

- $\mathcal{F}$ is "$P-$Glivenko-Cantelli", or $\mathcal{F} \in GC(P)$, if

$$\|\mathbb{P}_n - P\|_{\mathcal{F}} \rightarrow_{a.s.} 0.$$

- $\mathcal{F}$ is "$P-$Donsker", or $\mathcal{F} \in CLT(P)$, if

$$\mathbb{G}_n \Rightarrow \mathbb{G}_P \qquad \text{in} \quad \ell^\infty(\mathcal{F});$$

  where $\mathbb{G}_P$ is a $P-$Brownian bridge process indexed by $\mathcal{F}$ i.e.

$$E^* H(\mathbb{G}_n) \rightarrow EH(\mathbb{G}_P)$$

  for all bounded and continuous functions $H : \ell(\mathcal{F}) \rightarrow R$.

3. Bootstrapping the empirical process: limit theory

Bootstrapping $\mathbb{P}_n$ and $\mathbb{G}_n$:

- $\hat{X}_1, \ldots, \hat{X}_n$ i.i.d. $\mathbb{P}_n$ (viewing $\mathbb{P}_n$ as fixed).
- $\widehat{\mathbb{P}}_n = n^{-1} \sum_{i=1}^{n} \delta_{\hat{X}_i}$ = bootstrap empirical measure.
- $\widehat{\mathbb{G}}_n = \sqrt{n}(\widehat{\mathbb{P}}_n - \mathbb{P}_n)$ =bootstrap empirical process.
- Question: does $\widehat{\mathbb{G}}_n$ mimic $\mathbb{G}_n$ asymptotically?

3. Bootstrapping the empirical process: limit theory

Answer: Yes! Let $F(x) \equiv \sup_{f\in\mathcal{F}} |f(x) - Pf|$ =centered envelope function for $\mathcal{F}$.

Theorem 1. (Giné and Zinn, 1990)
The following are equivalent:
A. $\mathcal{F}$ is $P-$Donsker; i.e. $\mathbb{G}_n \Rightarrow \mathbb{G}_P$ in $\ell^\infty(\mathcal{F})$.
B. $\widehat{\mathbb{G}}_n \Rightarrow \mathbb{G}_P$ in $\ell^\infty(\mathcal{P})$ "in probability" wrt $\mathbb{P}_n$.

Theorem 2. (Giné and Zinn, 1990)
The following are equivalent:
A. $\mathcal{F}$ is $P-$Donsker and $PF^2 < \infty$.
B. $\widehat{\mathbb{G}}_n \Rightarrow \mathbb{G}_P$ in $\ell^\infty(\mathcal{P})$ "almost surely" wrt $\mathbb{P}_n$.

Note that

$$\widehat{\mathbb{P}}_n \stackrel{d}{=} n^{-1} \sum_{i=1}^{n} M_{n,i} \delta_{X_i}$$

where

$$M_n \equiv (M_{n,1}, \ldots, M_{n,n}) \sim \text{Multinomial}_n(n, (1/n, \ldots, 1/n)).$$

4. Exchangeably weighted bootstrapping

The representation $\widehat{\mathbb{P}}_n \stackrel{d}{=} n^{-1}\sum_{i=1}^n M_{n,i}\delta_{X_i}$ suggests that it might be interesting to replace the multinomial weights $\{M_{n,i} : 1 \le i \le n\}$ in Efron's bootstrap by other (exchangeable) weights $\{W_{n,i} : 1 \le i \le n\}$. Then

$$\widehat{\mathbb{P}}_n^W \equiv \widehat{\mathbb{P}}_n = n^{-1}\sum_{i=1}^n W_{n,i}\delta_{X_i}.$$

Examples:

4. Exchangeably weighted bootstrapping

- $W_{n,i} = Y_i/\overline{Y}_i$ with $Y_1, \ldots, Y_n$ i.i.d. non-negative rv's with
$$\|Y\|_{2,1} \equiv \int_0^\infty \sqrt{P(Y > t)}dt < \infty.$$
If $Y_i \sim$ Exponential(1), then $\widehat{\mathbb{P}}_n$ is called the "Bayesian bootstrap".

- $(W_{n,1}, \ldots, W_{n,n}) = \sqrt{n/k}(M_{n,1}, \ldots, M_{n,n})$ where
$$M_n \equiv (M_{n,1}, \ldots, M_{n,n}) \sim \text{Multinomial}_n(k, (1/n, \ldots, 1/n))$$

- $w_n = (w_{n,1}, \ldots, w_{n,n})$ and $W_n = (w_{n,R_1}, \ldots, w_{n,R_n})$ where $R = (R_1, \ldots, R_n)$ is a random permutation of $\{1, \ldots, n\}$. With $1 \le k \le n$,
$$w_n = \left(\frac{n}{\sqrt{k(n-k)}}, \ldots, \frac{n}{\sqrt{k(n-k)}}, 0, \ldots, 0 \right)$$
this gives random sampling without replacement from $\mathbb{P}_n$.

4. Exchangeably weighted bootstrapping

Conditions on the $W_{n,i}$'s:

W1 $\sup_{n\ge 1} \|W_{n,1} - \overline{W}_n\|_{2,1} < \infty$.

W2 $n^{-1/2} E \max_{1\le i\le n} |W_{n,i} - \overline{W}_n| \to 0$.

W3 $n^{-1}\sum_{i=1}^n (W_{n,i} - \overline{W}_n)^2 \to_p c^2 > 0$.

Theorem. (Praestgaard & W, 1993)
A. If $\mathcal{F}$ is $\mathcal{P}-$Donsker and W1-W3 hold, then

$$\widehat{\mathbb{G}}_n^W \equiv \sqrt{n}(\widehat{\mathbb{P}}_n^W - \mathbb{P}_n) \Rightarrow c\mathbb{G}_P \quad \text{in } \ell^\infty(\mathcal{F}). \tag{1}$$

in probability.
B. If $\mathcal{F}$ is $\mathcal{P}-$Donsker, $PF^2 < \infty$, and W1-W3 hold, then (**??**) holds almost surely.

4. Exchangeably weighted bootstrapping

Problem:

- The theorems of Giné and Zinn (1990) are if and only if.
- Our theorem for $\widehat{\mathbb{P}}_n$ and $\widehat{\mathbb{G}}_n$ is in just one direction: $\mathcal{F} \in CLT(P)$ implies bootstrap convergence.
- Question: Do natural converse theorems hold in the case of $\widehat{\mathbb{P}}_n^W$? If $\widehat{\mathbb{G}}_n^W \Rightarrow c\mathbb{G}_P$ in $\ell^\infty(\mathcal{F})$ in probability with $c > 0$, can we conclude that $\mathcal{F}$ is $P-$Donsker?

Conjectures, approaches, and connections

Conjectured Theorem 1'. The following are equivalent:
A. $\mathcal{F}$ is $P-$Donsker.
B. Some $\{W_n\}$ satisfying W1-W3 with $c > 0$, satisfies $\widehat{\mathbb{G}}_n^W \Rightarrow c\mathbb{G}_P$ in $\ell^\infty(\mathcal{F})$ in probability (and then the same holds for all sequences $\{W_n\}$ satisfying W1-W3).

Conjectured Theorem 2'. The following are equivalent:
A. $\mathcal{F}$ is $P-$Donsker and $PF^2 < \infty$.
B. Some $\{W_n\}$ satisfying W1-W3 with $c > 0$ satisfies $\widehat{\mathbb{G}}_n^W \Rightarrow c\mathbb{G}_P$ in $\ell^\infty(\mathcal{F})$ almost surely (and then the same holds for all sequences $\{W_n\}$ satisfying W1-W3).

Conjectures, approaches, and connections

- Q1. If for some $\{W_n\}$ satisfying W1-W3 with $c > 0$,

$$\widehat{\mathbb{G}}_n^W \Rightarrow c\mathbb{G}_P \quad \text{in} \quad \ell^\infty(\mathcal{F}) \quad \text{in prob} \tag{2}$$

  does the same hold for Efron's (multinomial) bootstrap? That is, do we have

$$\widehat{\mathbb{G}}_n^M \equiv \widehat{\mathbb{G}}_n \Rightarrow \mathbb{G}_P \quad \text{in} \quad \ell^\infty(\mathcal{F}) \quad \text{in prob?}$$

  If so, then we're done!

- If (**??**) does the same hold with weights $\tilde{W}_{n,j} = Y_j/\overline{Y}_n$ for i.i.d. Y_j's? If so, then we're done! Via Hoeffding's trick and the unconditional multiplier CLT.

- If (**??**) does the same hold with weights $\tilde{W}_{n,j} = Y_j$ for i.i.d. Y_j's? If so, then we're done!

- Obstructions? Extreme cases?

Other research interests and projects

- Empirical process theory: tools for estimation theory.
- Shape constrained estimation:
  - ▷ Estimation of smooth functionals of the Grenander estimator and other shape-constrained estimators.
  - ▷ Rates of convergence for shape-constrained estimators in $\mathbb{R}$ and $\mathbb{R}^d$.
- Semiparametric models:
  - ▷ Two - phase sampling methods (with missing data by design).
  - ▷ Semiparametric theory for high-dimensional settings with many predictor variables.