Semiparametric Model Based Bootstrap

Guang Cheng

Department of Statistics, Purdue University

The 2nd IMS-APRM
July 3rd, 2012

Outline

Motivation I

- As a general inference procedure, the bootstrap has been applied to the semiparametric models in a wide variety of contexts, e.g. Biostatistics, Survival Analysis and Econometrics, for a long time.

Motivation I

- As a general inference procedure, the bootstrap has been applied to the semiparametric models in a wide variety of contexts, e.g. Biostatistics, Survival Analysis and Econometrics, for a long time.
- There are two types of bootstrap sampling: (I) Nonparametric Bootstrap; (II) Parametric Bootstrap (also known as model based bootstrap).

Motivation I

- As a general inference procedure, the bootstrap has been applied to the semiparametric models in a wide variety of contexts, e.g. Biostatistics, Survival Analysis and Econometrics, for a long time.
- There are two types of bootstrap sampling: (I) Nonparametric Bootstrap; (II) Parametric Bootstrap (also known as model based bootstrap).
- By taking the model structure into account, the model-based bootstrap usually has better small sample performances than those resampling based bootstrap such as the nonparametric bootstrap. Comparing with the inconsistency of nonparametric bootstrap, we expect that the model based bootstrap is valid in drawing inferences for non-root-n convergent nonparametric components in semiparametric models.

Motivation II

- Cheng and Huang (2010) and Cheng (2012) established the bootstrap distribution and moment consistency for the semiparametric M-estimation (Euclidean part). Their conclusions apply to a broad class of bootstrap methods with exchangeable bootstrap weights including the nonparametric bootstrap.

Motivation II

- Cheng and Huang (2010) and Cheng (2012) established the bootstrap distribution and moment consistency for the semiparametric M-estimation (Euclidean part). Their conclusions apply to a broad class of bootstrap methods with exchangeable bootstrap weights including the nonparametric bootstrap.
- As far as we are aware, no rigorous theoretical justifications on the model-based bootstrap are existent for the semiparametric models despite its superior empirical performance.

Motivation II

- Cheng and Huang (2010) and Cheng (2012) established the bootstrap distribution and moment consistency for the semiparametric M-estimation (Euclidean part). Their conclusions apply to a broad class of bootstrap methods with exchangeable bootstrap weights including the nonparametric bootstrap.
- As far as we are aware, no rigorous theoretical justifications on the model-based bootstrap are existent for the semiparametric models despite its superior empirical performance.
- Some empirical processes tool of independent interest, i.e., uniform L_s maximal inequality, has been developed for proving the theoretical validity of the model based bootstrap.

Intuition I

- We first unify the semiparametric estimation based on the original data and bootstrap data at the technical level. This unification provides theoretical insights that are helpful for understanding future theoretical development.

Intuition I

- We first unify the semiparametric estimation based on the original data and bootstrap data at the technical level. This unification provides theoretical insights that are helpful for understanding future theoretical development.
- In this talk, we assume that the observed data $\mathcal{X}_n \equiv (X_1, \ldots, X_n)$ are i.i.d. from the probability space $(\mathcal{X}, \mathcal{A}, P_X)$. Define $P_X f = \int f dP_X$ and $\mathbb{P}_n f = \frac{1}{n} \sum_{i=1}^{n} f(X_i)$ for any measurable function f.

Intuition I

- We first unify the semiparametric estimation based on the original data and bootstrap data at the technical level. This unification provides theoretical insights that are helpful for understanding future theoretical development.
- In this talk, we assume that the observed data $\mathcal{X}_n \equiv (X_1, \ldots, X_n)$ are i.i.d. from the probability space $(\mathcal{X}, \mathcal{A}, P_X)$. Define $P_X f = \int f dP_X$ and $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$ for any measurable function f.
- Denote $\ell(\theta, \eta)$ as the semiparametric log-likelihood and $(\theta, \eta) \in \Theta \times \mathcal{H}$ as the parameter space.

Intuition II

In the real world, we have (under identifiability conditions)

$$
\begin{aligned}
\text{True value:} \quad (\theta_0, \eta_0) &= \arg \sup_{\theta \in \Theta, \eta \in \mathcal{H}} P_X \ell(\theta, \eta), \\
\text{MLE:} \quad (\widehat{\theta}, \widehat{\eta}) &= \arg \sup_{\theta \in \Theta, \eta \in \mathcal{H}} \mathbb{P}_n \ell(\theta, \eta),
\end{aligned}
$$

where $(X_1, \ldots, X_n) \overset{i.i.d.}{\sim} P_X$. Note that $P_{\theta_0, \eta_0} = P_X$.

Intuition III

- The bootstrap method mimics the real data-generating process by drawing the bootstrap data $\mathcal{X}_n^* \equiv (X_1^*, \ldots, X_n^*) \overset{i.i.d.}{\sim} \widehat{P}_n$, where $\widehat{P}_n$ is some distribution estimate for P_X based on the original data $\mathcal{X}_n$.

Intuition III

- The bootstrap method mimics the real data-generating process by drawing the bootstrap data $\mathcal{X}_n^* \equiv (X_1^*, \ldots, X_n^*) \overset{i.i.d.}{\sim} \widehat{P}_n$, where $\widehat{P}_n$ is some distribution estimate for P_X based on the original data $\mathcal{X}_n$.
- In the parallel bootstrap world, we have

$$\text{Bootstrap "true" value:} \quad (\theta_0^*, \eta_0^*) = \arg \sup_{\theta \in \Theta, \eta \in \mathcal{H}} \widehat{P}_n \ell(\theta, \eta),$$

$$\text{Bootstrap MLE:} \quad (\widehat{\theta}^*, \widehat{\eta}^*) = \arg \sup_{\theta \in \Theta, \eta \in \mathcal{H}} \widehat{\mathbb{P}}_n^* \ell(\theta, \eta),$$

where $\widehat{\mathbb{P}}_n^* f = \frac{1}{n} \sum_{i=1}^n f(X_i^*)$.

Intuition IV

- In the nonparametric bootstrap, we draw bootstrap samples from the c.d.f. $\mathbb{P}_n$. In this case, $\widehat{P}_n = \mathbb{P}_n$ and $(\theta_0^*, \eta_0^*) = (\widehat{\theta}, \widehat{\eta})$. However, the discrete $\mathbb{P}_n$ may fail to capture some properties of the underlying distribution P_X, e.g., its smoothness, that may be crucial for the problem under consideration.

Intuition IV

- In the nonparametric bootstrap, we draw bootstrap samples from the c.d.f. $\mathbb{P}_n$. In this case, $\widehat{P}_n = \mathbb{P}_n$ and $(\theta_0^*, \eta_0^*) = (\widehat{\theta}, \widehat{\eta})$. However, the discrete $\mathbb{P}_n$ may fail to capture some properties of the underlying distribution P_X, e.g., its smoothness, that may be crucial for the problem under consideration.
- Hence, in this talk, we consider sampling from some smooth estimated distribution $\widehat{P} \equiv P_{\widetilde{\theta},\widetilde{\eta}}$, where $P_{\theta,\eta}$ is the distribution of semiparametric models, and the initial estimate $(\widetilde{\theta}, \widetilde{\eta})$ (unnecessarily MLE) is computed based on $\mathcal{X}_n$. This is the so-called semiparametric model based bootstrap in which the model information is naturally incorporated into the resampling process. In this case, $\widehat{P}_n = \widehat{P}$ and $(\theta_0^*, \eta_0^*) = (\widetilde{\theta}, \widetilde{\eta})$.

Intuition V

- In view of the above analysis, we refer the distribution consistency as "the distribution of $\sqrt{n}(\widehat{\theta}^* - \widetilde{\theta})$ asymptotically imitates that of $\sqrt{n}(\widehat{\theta} - \theta_0)$ conditional on $\mathcal{X}_n$", and refer the variance consistency as "$nE_{\mathcal{X}^*|\mathcal{X}_n}(\widehat{\theta}^* - \widetilde{\theta})^{\otimes 2}$ consistently estimates the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta_0)$".

Intuition V

- In view of the above analysis, we refer the distribution consistency as "the distribution of $\sqrt{n}(\widehat{\theta}^* - \widetilde{\theta})$ asymptotically imitates that of $\sqrt{n}(\widehat{\theta} - \theta_0)$ conditional on $\mathcal{X}_n$", and refer the variance consistency as "$nE_{\mathcal{X}^*|\mathcal{X}_n}(\widehat{\theta}^* - \widetilde{\theta})^{\otimes 2}$ consistently estimates the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta_0)$".
- The above consistency results trivially hold if $(\widetilde{\theta}, \widetilde{\eta}) = (\theta_0, \eta_0)$. Therefore, we expect the model based bootstrap to be valid if (i) $P_{\theta,\eta}$ satisfies some smoothness conditions w.r.t. (θ, η); (ii) $(\widetilde{\theta}, \widetilde{\eta})$ converges to (θ_0, η_0) in some sense such that the estimated empirical processes $\widehat{\mathbb{G}}_n^* \equiv \sqrt{n}(\widehat{\mathbb{P}}_n^* - \widehat{P})$ (built upon $\widehat{P}$) perturbs around $\mathbb{G}_n \equiv \sqrt{n}(\mathbb{P}_n - P_X)$ (built upon P_X) and has proper asymptotic continuity modulus.

Semiparametric Models

- Random Variable $X \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$

Semiparametric Models

- Random Variable $X \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
    - θ: Euclidean parameter of interest;

Semiparametric Models

- Random Variable $X \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.

Semiparametric Models

- Random Variable $X \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.
- Recall that $\ell(\theta, \eta)$ represents the log-likelihood. Define the efficient score function as $\widetilde{\ell}(\theta, \eta)$.

Semiparametric Models

- Random Variable $X \sim \left\{ P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H} \right\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.
- Recall that $\ell(\theta, \eta)$ represents the log-likelihood. Define the efficient score function as $\widetilde{\ell}(\theta, \eta)$.
- Example I: The Cox regression model with survival data

Semiparametric Models

- Random Variable $X \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.
- Recall that $\ell(\theta, \eta)$ represents the log-likelihood. Define the efficient score function as $\widetilde{\ell}(\theta, \eta)$.
- Example I: The Cox regression model with survival data
  - θ: regression parameter;
  - η: cumulative hazard function.

Semiparametric Models

- Random Variable $X \sim \left\{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\right\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.
- Recall that $\ell(\theta, \eta)$ represents the log-likelihood. Define the efficient score function as $\widetilde{\ell}(\theta, \eta)$.
- Example I: The Cox regression model with survival data
  - θ: regression parameter;
  - η: cumulative hazard function.
- Example II: The partly linear model: $Y = W'\theta + f(T) + \epsilon$

Semiparametric Models

- Random Variable $X \sim \{P_{\theta,\eta} : \theta \in \Theta \subset \mathbb{R}^k, \eta \in \mathcal{H}\}$
  - θ: Euclidean parameter of interest;
  - η: a *possibly* infinite dimensional *nuisance* parameter, e.g. some function.
- Recall that $\ell(\theta, \eta)$ represents the log-likelihood. Define the efficient score function as $\widetilde{\ell}(\theta, \eta)$.
- Example I: The Cox regression model with survival data
  - θ: regression parameter;
  - η: cumulative hazard function.
- Example II: The partly linear model: $Y = W'\theta + f(T) + \epsilon$
  - θ: linear regression parameter;
  - f: nonlinear smooth function.

Estimation Consistency Theorem (Theorem I)

Suppose that (θ_0, η_0) is well-separated and $(\widetilde{\theta}, \widetilde{\eta})$ is consistent. Define

$$\begin{aligned} \mathcal{B}_\delta &= \{(\theta,\eta) : d((\theta,\eta),(\theta_0,\eta_0)) \leq \delta\}, \\ \mathcal{P}_\delta &= \{P_{\theta,\eta} : (\theta,\eta) \in \mathcal{B}_\delta\}. \end{aligned}$$

If the class $\mathcal{L} \equiv \{\ell(X;\theta,\eta) : (\theta,\eta) \in \Theta \times \mathcal{H}\}$ is $\mathcal{P}_\delta$-Uniform Glivenko-Cantelli for some $\delta > 0$ and

$$\lim_{\delta \downarrow 0} \sup_{(\theta,\eta) \in \mathcal{B}_\delta} \|(dP_{\theta,\eta}/dP_X) - 1\|_2 = 0, \tag{1}$$

then $(\widehat{\theta}^*, \widehat{\eta}^*)$ is conditionally consistency given $\mathcal{X}_n$, i.e.,

$$P_{X^*|\mathcal{X}_n}(d((\widehat{\theta}^*, \widehat{\eta}^*),(\theta_0,\eta_0)) > \epsilon) \xrightarrow{P_X} 0 \quad \text{for every } \epsilon > 0.$$

Distribution Consistency Conditions

Define $\mathcal{R}_\delta = \left\{ \widetilde{\ell}(\theta, \eta) - \widetilde{\ell}(\theta_0, \eta_0) : (\theta, \eta) \in \mathcal{B}_\delta \right\}$.

Distribution Consistency Conditions

Define $\mathcal{R}_\delta = \left\{ \widetilde{\ell}(\theta, \eta) - \widetilde{\ell}(\theta_0, \eta_0) : (\theta, \eta) \in \mathcal{B}_\delta \right\}$.

- Regularity conditions (those in guaranteeing the asymptotic normality of MLE $\widehat{\theta}$);

Distribution Consistency Conditions

Define $\mathcal{R}_\delta = \left\{ \widetilde{\ell}(\theta, \eta) - \widetilde{\ell}(\theta_0, \eta_0) : (\theta, \eta) \in \mathcal{B}_\delta \right\}$.

- Regularity conditions (those in guaranteeing the asymptotic normality of MLE $\widehat{\theta}$);
- Assume that $\mathcal{T}_\delta \equiv \{\widetilde{\ell}(\theta, \eta) : (\theta, \eta) \in \mathcal{B}_\delta\}$ is $\mathcal{P}_{\delta'}$-Uniform Donsker and $\sup_{f \in \mathcal{T}_\delta} |(\widehat{P} - P_X)(f^2)| \xrightarrow{P_X} 0$ for some $\delta, \delta' > 0$;

Distribution Consistency Conditions

Define $\mathcal{R}_\delta = \left\{ \widetilde{\ell}(\theta, \eta) - \widetilde{\ell}(\theta_0, \eta_0) : (\theta, \eta) \in \mathcal{B}_\delta \right\}$.

- Regularity conditions (those in guaranteeing the asymptotic normality of MLE $\widehat{\theta}$);
- Assume that $\mathcal{T}_\delta \equiv \{\widetilde{\ell}(\theta, \eta) : (\theta, \eta) \in \mathcal{B}_\delta\}$ is $\mathcal{P}_{\delta'}$-Uniform Donsker and $\sup_{f \in \mathcal{T}_\delta} |(\widehat{P} - P_X)(f^2)| \xrightarrow{P_X} 0$ for some $\delta, \delta' > 0$;
- $\sup_{r \in \mathcal{R}_{\delta_n}} \|r\|_2 = O(\delta_n)$ for any sequence $\delta_n \to 0$;

Distribution Consistency Conditions

Define $\mathcal{R}_\delta = \left\{\widetilde{\ell}(\theta,\eta) - \widetilde{\ell}(\theta_0,\eta_0) : (\theta,\eta) \in \mathcal{B}_\delta\right\}$.

- Regularity conditions (those in guaranteeing the asymptotic normality of MLE $\widehat{\theta}$);
- Assume that $\mathcal{T}_\delta \equiv \{\widetilde{\ell}(\theta,\eta) : (\theta,\eta) \in \mathcal{B}_\delta\}$ is $\mathcal{P}_{\delta'}$-Uniform Donsker and $\sup_{f\in\mathcal{T}_\delta} |(\widehat{P} - P_X)(f^2)| \overset{P_X}{\to} 0$ for some $\delta, \delta' > 0$;
- $\sup_{r\in\mathcal{R}_{\delta_n}} \|r\|_2 = O(\delta_n)$ for any sequence $\delta_n \to 0$;
- Uniform L_1 maximal inequality over $\mathcal{P}_{\delta'}$:

$$\sup_{P\in\mathcal{P}_{\delta'}} \|\|\mathbb{G}_n(P)\|_{\mathcal{R}_{\delta'_n}}\|_{P,1} = O(\delta'_n).$$

Distribution/Variance Consistency Theorem (Theorem II):

Suppose the above conditions hold. If

$$d((\widetilde{\theta}, \widetilde{\eta}), (\theta_0, \eta_0)) = o_{P_X}(n^{-1/4}),$$
$$\|d\widehat{P}/dP_X - 1\|_2 = O_{P_X}(n^{-1/2}), \tag{2}$$

and some proper convergence rate on $\widehat{\eta}^*$, then we have

$$\|\widehat{\theta}^* - \theta_0\| = O_{P_{X^*}}(n^{-1/2})$$
$$\sup_{x \in \mathbb{R}^d} \left| P_{X^*|\mathcal{X}_n}(\sqrt{n}(\widehat{\theta}^* - \widetilde{\theta}) \leq x) - P_X(\sqrt{n}(\widehat{\theta} - \theta_0) \leq x) \right| \xrightarrow{P_X} 0.$$

If we further strengthen the above uniform L_1 maximal inequality condition to the uniform L_2 maximal inequality condition, we also have the bootstrap variance consistency.

Remark 1:

Basically, we need to verify the following three types of conditions:

Remark 1:

Basically, we need to verify the following three types of conditions:

- Uniform Donsker and Uniform Glivenko-Cantelli Conditions: check some envelop conditions and uniform (bracketing) entropy conditions;

Remark 1:

Basically, we need to verify the following three types of conditions:

- Uniform Donsker and Uniform Glivenko-Cantelli Conditions: check some envelop conditions and uniform (bracketing) entropy conditions;
- Uniform L_s maximal inequality Conditions: some new empirical process results have been developed for this purpose;

Remark 1:

Basically, we need to verify the following three types of conditions:

- Uniform Donsker and Uniform Glivenko-Cantelli Conditions: check some envelop conditions and uniform (bracketing) entropy conditions;
- Uniform L_s maximal inequality Conditions: some new empirical process results have been developed for this purpose;
- Condition $\|dP_{\widetilde{\theta},\widetilde{\eta}}/dP_X - 1\|_2 = O_{P_X}(n^{-1/2})$: it is not easy to verify it when $\widetilde{\eta}$ converges at slower than root-n rate. However, we can remove this condition if $\widehat{\theta}^*$ is known to be root-n consistent.

Bootstrap Confidence Set

In the model-based bootstrap, we show that the bootstrap estimate $\widehat{\theta}^*$ centers around $\widetilde{\theta}$ rather than $\widehat{\theta}$ as in the case of nonparametric bootstrap. This is consistent with our intuitive analysis in the introduction, but will lead to somewhat different *percentile* bootstrap confidence set.

- Without loss of generality, we define $\tau^*_{n\alpha}$ as

$$P_{X|\mathcal{X}_n}\left(\widehat{\theta}^* \leq \tau^*_{n\alpha}\right) = \alpha.$$

- Without loss of generality, we define $\tau^*_{n\alpha}$ as

$$P_{X|\mathcal{X}_n}\left(\widehat{\theta}^* \leq \tau^*_{n\alpha}\right) = \alpha.$$

- Based on the bootstrap distributional consistency Theorem, we can approximate the α-th quantile of the distribution of $(\widehat{\theta} - \theta_0)$ by $(\tau^*_{n\alpha} - \widetilde{\theta})$. Thus, we construct the $(1-\alpha)$ percentile-type bootstrap confidence set as

$$BC_p(\alpha) = \left[(\widehat{\theta} - \widetilde{\theta}) + \tau^*_{n(\alpha/2)}, (\widehat{\theta} - \widetilde{\theta}) + \tau^*_{n(1-\alpha/2)}\right].$$

- Without loss of generality, we define $\kappa^*_{n\alpha}$ as

$$P_{X|\mathcal{X}_n}\left((\widehat{\theta}^* - \widetilde{\theta}) \leq \kappa^*_{n\alpha}\right) = \alpha.$$

- Without loss of generality, we define $\kappa^*_{n\alpha}$ as

$$P_{X|\mathcal{X}_n}\left((\widehat{\theta}^* - \widetilde{\theta}) \leq \kappa^*_{n\alpha}\right) = \alpha.$$

- Similarly, we can approximate the α-th quantile of $(\widehat{\theta} - \theta_0)$ by $\kappa^*_{n\alpha}$. Thus we construct the $(1-\alpha)$ hybrid-type bootstrap confidence set as

$$BC_h(\alpha) = \left[\widehat{\theta} - \kappa^*_{n(1-\alpha/2)}, \widehat{\theta} - \kappa^*_{n(\alpha/2)}\right].$$

Remark 2:

- Both the *percentile* bootstrap confidence set $BC_p(\alpha)$ and *hybrid* bootstrap confidence set $BC_h(\alpha)$ can be computed easily through routine bootstrap sampling.

Remark 2:

- Both the *percentile* bootstrap confidence set $BC_p(\alpha)$ and *hybrid* bootstrap confidence set $BC_h(\alpha)$ can be computed easily through routine bootstrap sampling.

- We can avoid estimating the asymptotic variance of $\widehat{\theta}$ when using $BC_p(\alpha)$ and $BC_h(\alpha)$.

The distributional consistency Theorem together with the quantile convergence Theorem implies the consistency of percentile-type and hybrid-type bootstrap confidence sets.

Bootstrap Confidence Set Corollary (Corollary I): Under the conditions in Theorem II, we have

$$
\begin{aligned}
Pr(\theta_0 \in BC_p(\alpha)) &\longrightarrow 1 - \alpha, \\
Pr(\theta_0 \in BC_h(\alpha)) &\longrightarrow 1 - \alpha,
\end{aligned}
$$

as $n \to \infty$.

Remark 3:

- We want to point out that $BC'_p(\alpha) = [\tau^*_{n(\alpha/2)}, \tau^*_{n(1-\alpha/2)}]$ is also consistent if $\widetilde{\theta}$ and $\widehat{\theta}$ shares the same limit distribution, i.e., $\widetilde{\theta} - \widehat{\theta} = o_{P_X}(n^{-1/2})$.

Remark 3:

- We want to point out that $BC_p'(\alpha) = [\tau^*_{n(\alpha/2)}, \tau^*_{n(1-\alpha/2)}]$ is also consistent if $\widetilde{\theta}$ and $\widehat{\theta}$ shares the same limit distribution, i.e., $\widetilde{\theta} - \widehat{\theta} = o_{P_X}(n^{-1/2})$.
- Provided the consistent estimator for the asymptotic covariance, e.g., bootstrap variance estimate, is available, we can show that the t-type bootstrap confidence set is also consistent by considering the Slutsky's Theorem.

Thank you for your attention....

Assistant Professor Guang Cheng
Department of Statistics, Purdue University
chengg@purdue.edu