

SDS 384 11: Theoretical Statistics

Lecture 18: Bootstrap

Purnamrita Sarkar
Department of Statistics and Data Science
The University of Texas at Austin

The bootstrap

- Data $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} P$
- Some estimator $\hat{\theta}$ of parameter of interest θ .
- Want:

$$P\left(\hat{\theta} - \kappa_{\alpha}\hat{\sigma} \leq \theta \leq \hat{\theta} + \kappa_{1-\alpha}\hat{\sigma}\right) \geq 1 - 2\alpha,$$

where $\kappa_{\alpha}, \kappa_{1-\alpha}$ are the quantiles of $(\hat{\theta} - \theta)/\hat{\sigma}$

- The distribution of $(\hat{\theta} - \theta)/\hat{\sigma}$ depends on P .
- Often this distribution is normal, but with unknown parameters.

Bootstrap: plug in principle

True model	Bootstrapped model
------------	--------------------

$\hat{\theta}$	$\hat{\theta}^*$
----------------	------------------

$\hat{\sigma}$	σ
----------------	----------

$\frac{\hat{\theta}-\theta}{\hat{\sigma}}$	$\frac{\hat{\theta}^*-\hat{\theta}}{\hat{\sigma}^*}$
--	--

Empirical bootstrap

How do you estimate P ?

Empirical Bootstrap $\hat{P} = \hat{P}_n$

Generate m samples $(X_1^*, \dots, X_n^*)^{(j)}$, $j = 1 : m$.

Each giving a $(\hat{\theta}^*, \hat{\sigma}^*)$ pair.

Compute the κ_α quantile

of the distribution of $\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*}$

Parametric bootstrap $\hat{P} = P_{\hat{\theta}}$

Consistency

- We want, as $n \rightarrow \infty$,

$$\sup_x \left| P \left(\frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq x \right) - \hat{P}_n \left(\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \leq x \right) \right| \xrightarrow{P} 0$$

- We assume:

$$P \left(\frac{\hat{\theta} - \theta}{\hat{\sigma}} \leq x \right) \rightarrow F(x)$$

- F is continuous, and so it is enough to show that:

$$\hat{P}_n \left(\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}^*} \leq x \right) \xrightarrow{P} F(x).$$

Theorem

If X_1, \dots, X_n are iid with mean μ and variance σ^2 , then conditioned on the data, for almost every sequence X_1^n

$$\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \xrightarrow{d} N(0, \sigma^2)$$

- First note that $E[\bar{X}_n^*|X_1^n] = E[X_1^*|X_1^n] = \bar{X}_n$
- Now note that $\text{var}(\bar{X}_n^*|X_1^n) = E[(X_1 - \bar{X}_n)^2|X_1^n] = \hat{\sigma}^2$
- Can we use CLT?
- The \hat{P}_n is changing for each n .
- Need to check the Lindeberg condition.

Lindeberg-feller CLT for triangular arrays

X_{11}

X_{21}, X_{22}

X_{21}, X_{22}, X_{23}

...

Theorem

For each n let $(X_{ni})_{i=1}^n$ be independent random variables with mean zero and variance σ_{ni}^2 . Let $Z_n = \sum_{i=1}^n X_{ni}$ and $B_n^2 = \text{var}(Z_n)$. Then

$Z_n/B_n \xrightarrow{d} N(0, 1)$, as long as the **Lindeberg condition** holds.

The Lindeberg condition

Definition (Lindeberg condition)

For every $\epsilon > 0$,

$$\frac{1}{B_n^2} \sum_{j=1}^n E[X_{nj}^2 1(|X_{nj}| \geq \epsilon B_n)] \rightarrow 0 \text{ as } n \rightarrow \infty \quad (1)$$

Converse: If $\frac{\sigma_{nj}^2}{B_n^2} \rightarrow 0$ as $n \rightarrow \infty$, i.e. no one variance plays a significant role in the limit, and if $Z_n/B_n \xrightarrow{d} N(0,1)$, then the Lindeberg condition holds.

Necessary and Sufficient: If $\frac{\sigma_{nj}^2}{B_n^2} \rightarrow 0$, the the Lindeberg condition is necessary and sufficient to show the CLT.

Does the Lindeberg condition hold?

- Check if $E[(X_i^*)^2 1(|X_i^*| \geq \epsilon \sqrt{n\hat{\sigma}}) | X_1^n] \rightarrow 0$

$$E[(X_i^*)^2 1(|X_i^*| \geq \epsilon \sqrt{n\hat{\sigma}}) | X_1^n] = \frac{1}{n\hat{\sigma}^2} \sum_i X_i^2 1(|X_i| \geq \epsilon \sqrt{n\hat{\sigma}})$$

$$\text{When } \epsilon \sqrt{n\hat{\sigma}} \geq M \quad \leq \frac{C}{n} \sum_i X_i^2 1(|X_i| \geq M)$$

$$\stackrel{\text{a.s.}}{\rightarrow} E[X_i^2 1(|X_i| \geq M)]$$

= Arbitrarily small for sufficiently large M

Delta method for bootstrap

Theorem

If we have

- $\hat{\theta} \xrightarrow{\text{a.s.}} \theta$
- $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} T$
- *Conditionally, almost surely, $\sqrt{n}(\hat{\theta}^* - \hat{\theta}) \xrightarrow{d} T$*
- *ϕ is continuously differentiable in the neighborhood of θ , then conditionally almost surely,*

$$\sqrt{n}(\phi(\hat{\theta}^*) - \phi(\hat{\theta})) \xrightarrow{d} \phi'_{\theta}(T)$$

- *The traditional delta method gives:*

$$\sqrt{n}(\phi(\hat{\theta}) - \phi(\theta)) \xrightarrow{d} \phi'_{\theta}(T)$$

When does bootstrap fail

Example

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} F = U([0, \theta])$. $X_{(1)}, \dots, X_{(n)}$ are the order statistics.

- The true limiting distribution

$$P\left(\frac{n(\theta - X_{(n)})}{\theta} > x\right) = P\left(X_{(n)} \leq \theta(1 - x/n)\right) = (1 - x/n)^n \rightarrow e^{-x}$$

- The bootstrapped limiting distribution

$$P\left(\frac{n(X_{(n)} - X_{(n)}^*)}{X_{(n)}} = 0\right) = P(X_{(n)}^* = X_{(n)}) = (1 - (1 - 1/n)^n) \rightarrow 1 - 1/e$$

When does bootstrap fail

Example

Bootstrap works for U statistics, as long as they are not degenerate.

- Rule of thumb: as long as there is normal convergence bootstrap works.
- Are there more robust methods?

Subsampling

- Draw without replacement Y_i which is a size b subsample of X_1^n
- Repeat for all $\binom{n}{b}$ subsets.
- Calculate $\hat{\theta}_{n,b,i} = \hat{\theta}(Y_i)$ and use the empirical distribution of $\tau_b(\hat{\theta}_{n,b,i} - \hat{\theta})$ to approximate the distribution of $\tau_n(\hat{\theta} - \theta)$
- So we want the following

$$J_n(x, P) = P\left(\tau_n(\hat{\theta} - \theta) \leq x\right)$$

- Which we approximate by:

$$L_{n,b}(x) = \frac{1}{\binom{n}{b}} \sum_i 1\left(\tau_b(\hat{\theta}_{n,b,i} - \hat{\theta}) \leq x\right)$$

Theorem

If $b \rightarrow \infty$, $b/n \rightarrow 0$, $\tau_b/\tau_n \rightarrow 0$ as long as $\tau_n(\hat{\theta} - \theta) \xrightarrow{d} T$ for some T ,

$$J_n(x, P) - L_{n,b}(x) \rightarrow 0$$

- Since $\hat{\theta}_{n,b,i}$ is just an estimator on a smaller sample from the true distribution,

$$1\left(\tau_b(\hat{\theta}_b - \theta) \leq x\right) \xrightarrow{d} J(x, P)$$

$$\tau_b(\hat{\theta}_b - \theta) = \tau_b(\hat{\theta}_b - \hat{\theta}_n) + \underbrace{\tau_b(\hat{\theta}_n - \theta)}_{O_p(\tau_b/\tau_n) = o_p(1)}$$

Finishing the proof

- Recall that $\tau_n(\hat{\theta} - \theta) \xrightarrow{d} J(\cdot, P)$

- So suffices to show that

$$U_{n,b}(x) := \frac{1}{\binom{n}{b}} \sum_i 1\left(\tau_b(\hat{\theta}_{n,b,i} - \theta) \leq x\right) \xrightarrow{P} J(x, P)$$

$$U_{n,b}(x) - J(x, P) = U_{n,b}(x) - E[U_{n,b}(x)] + \underbrace{E[U_{n,b}(x)] - J(x, P)}_0$$

- Now recall your HW problem:

$$P\left(U_{n,b} - E[U_{n,b}] \geq \epsilon\right) \leq \exp(-\lfloor \frac{n}{b} \rfloor \epsilon^2) \rightarrow 0$$

since $b/n \rightarrow 0$.

