STA355 Lecture 4

Recall: empirical cdf and functional estimation

- For linear functionals of the form

$$T(F) = \int r(x)\, dF(x) = \mathbb{E}[r(X)] \quad \text{(where } X \sim F\text{)}$$

then $T(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n} r(X_i)$

- $T(\hat{F}_n) \xrightarrow[p]{a.s.} T(F)$ by LLN

- $$\frac{\left(T(\hat{F}_n) - T(F)\right)}{se\left(T(\hat{F}_n)\right)} \to N(0,1) \text{ by CLT.}$$

- $\left(se\left(T(\hat{F}_n)\right) = \frac{1}{\sqrt{n}}\sqrt{var(r(X))} \right)$.

- Some linear functionals cannot be written in this form.

General linear functionals: any T s.t.

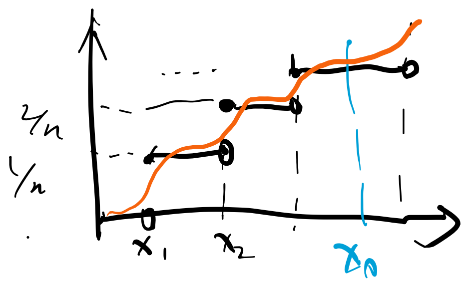
$$T(aF_1 + bF_2) = a\,T(F_1) + b\,T(F_2)$$

$\forall\, a, b \in \mathbb{R}$, F_1, F_2 cdfs.

e.g. $T(F) = F'(x_0)$.

$$T(\hat{F}_n) = 0 \quad \text{w.p. } 1.$$

$\hat{F}_n$:



Solution: (to be discussed)

smoothing.

Nonlinear functionals that can be estimated well.

e.g. Lipschitz functional T

$$|T(F_1) - T(F_2)| \le L \cdot \|F_1 - F_2\|_\infty.$$

$$\|F_1 - F_2\|_\infty := \sup_{x \in \mathbb{R}} |F_1(x) - F_2(x)|.$$

Consequence of cdf results:

- $T(\hat{F}_n) \xrightarrow{P} T(F)$.
- (By DKW). $\mathbb{P}\left(|T(\hat{F}_n) - T(F)| \ge \varepsilon\right) \le 2\exp\left(-\frac{2n\varepsilon^2}{L^2}\right)$.

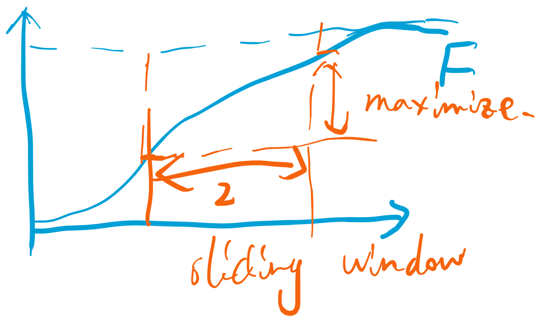
Equivalently w.p. $1-\alpha$,

$$|T(\hat{F}_n) - T(F)| \le L \cdot \sqrt{\frac{\log(2/\alpha)}{2n}}.$$

Concrete instance:

$$T(F) := \max_{x \in \mathbb{R}} \mathbb{P}\left(X \in (x-1, x+1]\right)$$

$$= \max_{x \in \mathbb{R}} \left(F(x+1) - F(x-1)\right).$$



$$\left|T(F_1) - T(F_2)\right| = \left| \max_{x} \left(F_1(x+1) - F_1(x-1)\right) - \max_{x}\left(F_2(x+1) - F_2(x-1)\right) \right|$$

Let x_1, x_2 be maximizers under F_1 and F_2 respectively.

$$\left(F_1(x_1+1) - F_1(x_1-1)\right) - \left(F_2(x_2+1) - F_2(x_2-1)\right)$$

$$\leq \left(F_1(x_1+1) - F_1(x_1-1)\right) - \left(F_2(x_1+1) - F_2(x_1-1)\right)$$

(Because x_2 is the maximizer).

$$\leq 2 \cdot \sup_{x \in \mathbb{R}} \left|F_1(x) - F_2(x)\right|.$$

Similarly,

$$\left(F_1(x_1+1) - F_1(x_1-1)\right) - \left(F_2(x_2+1) - F_2(x_2-1)\right) \geq -2 \sup_{x \in \mathbb{R}} \left|F_1(x) - F_2(x)\right|.$$

So T is 2-Lipschitz.

Bootstrap.

Motivation: confidence interval construction.

$$\frac{\sqrt{n}\,(\hat{T}_n - T(F))}{\sigma} \xrightarrow{d} N(0,1). \quad \text{in many cases.}$$

Need ~~to~~ know σ to construct CIs.

- Easy to estimate for empirical estimators (sample variance)

$$\hat{\sigma}_n^2 = \frac{1}{n}\sum_{i=1}^{n}\left(r(X_i) - \hat{T}_n\right)^2$$

for estimating $T(F) := \mathbb{E}[r(X)]$, $\hat{T}_n := \frac{1}{n}\sum_{i=1}^{n} r(X_i)$.

- Hard in general. e.g. $\hat{\theta}_n$ obtained from some optimization procedure.

Idea: by G-C or DKW. we know $\hat{F}_n \approx F$.

$$se_F(\hat{T}_n) = \sqrt{\mathbb{E}\left[|\hat{T}_n - \mathbb{E}(\hat{T}_n)|^2\right]}$$

$\Downarrow$ replace with expectations under empirical distribution.

$$se_{\hat{F}_n}(\hat{T}_n).$$

Step I.

Real world:

$$F \Rightarrow \underbrace{X_1, X_2, \cdots X_n}_{\text{random}} \overset{iid}{\sim} F \Rightarrow T_n = g(X_1, X_2, \cdots X_n).$$

"Bootstrap world":

$$\hat{F}_n \Rightarrow \underbrace{X_1^*, X_2^*, \cdots, X_n^*}_{\text{Additional randomization.}} \overset{iid}{\sim} \hat{F}_n \Rightarrow T_n^* = g(X_1^*, X_2^*, \cdots X_n^*)$$

Conditionally on data $X_1, \cdots, X_n$, T_n^* is a r.v.

$$\hat{v}^2_{\text{bootstrap, ideal}} = \operatorname{var}(T_n^* \mid X_1, X_2, \cdots X_n).$$

How to compute the idealized bootstrap variance?

Step II. simulation

Framework of Monte Carlo simulation:

Suppose we can sample $Y_1, Y_2, \cdots Y_B \overset{iid}{\sim} G$.

$$\frac{1}{B}\sum_{i=1}^{B} f(Y_i) \xrightarrow{P} \mathbb{E}_G[f(Y)].$$

All happening in your computer

Repeat the "bootstrap world" procedure B times,

and compute $T_{n,1}^*, T_{n,2}^*, \cdots, T_{n,B}^*$.

$\overset{iid}{\sim}$ Conditional distr'l of T_n^* given $X_1, X_2, \cdots X_n$.

$$\hat{V}^2_{bootstrap} = \frac{1}{B}\sum_{b=1}^{B}\left(T^*_{n,b} - \frac{1}{B}\sum_{i=1}^{B} T^*_{n,i}\right)^2.$$

Hope to have:

$$\mathrm{var}_F(\hat{T}_n) \; \overset{?}{\approx} \; \hat{V}^2_{bootstrap,\,idealized} \; \approx \; \hat{V}^2_{bootstrap}$$

$\hat{V}^2_{bootstrap,\,idealized} = \mathrm{var}_{\hat{F}}(\hat{T}_n)$

may and may not be close

err → 0 as $B \to +\infty$.

(easy to get non-asymptotic bounds).

Depends on property of $\hat{T}_n$, T, etc.

Bootstrap confidence intervals.

- "Normal interval".
  - Compute $\hat{T}_n$ on $X_1 \cdots X_n$
  - Compute $\hat{V}_{bootstrap}$ by simulation

$$C_n = \left[\hat{T}_n - \hat{V}_{bootstrap}\, z_{\alpha/2},\ \hat{T}_n + \hat{V}_{bootstrap}\, z_{\alpha/2}\right].$$

where z_β is the $(1-\beta)$ quantile of $N(0,1)$.

works when $\widehat{T}_n$ is close to normal

& $\widehat{F}_n$ is close to F.

- "Pivotal interval".

$$\theta = T(F), \qquad \widehat{\theta}_n = T(\widehat{F}_n).$$

In CI construction, we usually study the asymptotic distribution of $R_n = \widehat{\theta}_n - \theta$.

Let $\widehat{\theta}_n^*$ be bootstrap replication of $\widehat{\theta}_n$

Following the principle of bootstrap,

$$\boxed{\widehat{\theta}_n^* \mid X_1 \cdots X_n \overset{d}{\approx} \widehat{\theta}_n.}$$

CI construction:

- Define $H(r) = \mathbb{P}_F(R_n < r)$
- $C_n := \left[\widehat{\theta}_n - H^{-1}(1-\frac{\alpha}{2}), \widehat{\theta}_n - H^{-1}(\frac{\alpha}{2})\right]$.

(not neccessarily normal cdf).

$$\mathbb{P}(C_n \ni \theta) = 1-\alpha.$$

$$\hat{H}_{boot,ideal}(r) := \mathbb{P}\left(\underset{\substack{\| \\ T(\text{bootstrap})}}{\hat{\theta}_n^*} - \underset{\substack{\| \\ T(\hat{F}_n)}}{\hat{\theta}_n} \le r \,\middle|\, X_1, X_2, \dots, X_n\right).$$

Bootstrap world: $T(\text{bootstrap}) \quad T(\hat{F}_n)$

Real world: $T(\hat{F}_n) - T(F) = R_n$

Simulation:

$$\hat{H}_{boot}(r) = \frac{1}{B} \sum_{b=1}^{B} \mathbb{1}\{\hat{\theta}_{n,b}^* - \hat{\theta}_n \le r\}.$$

$$C_n = \left[\hat{\theta}_n - \hat{H}^{-1}(1-\alpha/2),\ \hat{\theta}_n - \hat{H}^{-1}(\alpha/2)\right].$$

Rmk: doesn't require asymptotic normality.

Computation: $\hat{\theta}^*_{(\beta)}$: the β quantile of $(\hat{\theta}^*_{n,1}, \hat{\theta}^*_{n,2}, \dots, \hat{\theta}^*_{n,B})$

$$\hat{H}_{bootstrap}\left(1-\frac{\alpha}{2}\right) = \hat{\theta}^*_{(1-\frac{\alpha}{2})} - \hat{\theta}_n$$

$$\hat{H}_{bootstrap}(\alpha/2) = \hat{\theta}^*_{(\alpha/2)} - \hat{\theta}_n$$

$$CI_n = \left[2\hat{\theta}_n - \hat{\theta}^*_{(1-\frac{\alpha}{2})},\ 2\hat{\theta}_n - \hat{\theta}^*_{(\alpha/2)}\right].$$

· "Percentile interval".

$$C_n = \left[\hat{\theta}^*_{(\alpha/2)}, \quad \hat{\theta}^*_{(1-\alpha/2)} \right],$$

Ideally, we want to use quantile of $\hat{\theta}_n^* \mid X_1, \dots X_n$

use simulation sample quantiles.

When does bootstrap work?

- We know $\hat{F}_n \approx F$
- $T(\hat{F}_n) \overset{?}{\approx} T(F)$.
- $T(\text{bootstrap}) \overset{d}{\underset{?}{\approx}} T(\hat{F}_n)$.

Theorem: If $T(F)$ is Hadamard differentiable functional of F then for pivotal/percentile CIs.

$$\mathbb{P}\left(T(F) \in C_n \right) \to 1-\alpha \quad \begin{pmatrix} n \to +\infty \\ B \to +\infty \end{pmatrix}.$$

Hadamard differentiable:

$$\frac{T(F + t_n h_n) - T(F)}{t_n} \longrightarrow T_F'(h)$$

if $t_n \to 0$ and $h_n \to h$.

(t_n scalars, h_n functions).

When does bootstrap fail?

e.g. $X_1, X_2, \cdots X_n \overset{iid}{\sim} \mathrm{Unif}([0,\theta])$.

$$T(F) := \max\{x : x \in \mathrm{supp}(F)\}. \quad (= F^{-1}(1))$$

$$\hat{\theta}_n = T(\hat{F}_n) = \max\{X_1, X_2, \cdots X_n\}.$$

$$\mathbb{P}(\hat{\theta}_n \le t) = \mathbb{P}(X_1 \le t)^n = \left(\frac{t}{\theta}\right)^n. \quad \text{"true distribution"}.$$

In bootstrap world:

$X_1^*, X_2^*, \cdots, X_n^* \overset{iid}{\sim} \mathrm{Unif}(\{X_1, X_2, \cdots X_n\})$.

$$\hat{\theta}_n^* = \max\{X_1^*, X_2^*, \cdots, X_n^*\}$$

$$\hat{\theta}_n^* \mid X_1, X_2, \cdots X_n \quad \text{v.s.} \quad \hat{\theta}_n \quad ?$$

Indeed, we have

$$\mathbb{P}(\hat{\theta}_n^* = \hat{\theta}_n \mid X_1, X_2, \cdots X_n)$$

$$= \mathbb{P}(X_{max} \text{ selected in bootstrap samples} \mid X_1 \cdots X_n)$$

$$= 1 - \left(1 - \frac{1}{n}\right)^n \longrightarrow 1 - 1/e \approx 0.6$$

Real world: $\mathbb{P}(\hat{\theta}_n \le t) \le (t/\theta)^n$.

Bootstrap world. $\hat{\theta}_n^*$ concentrates on a single value w.p. $1-1/e$.

$\frac{1}{e}$ confidence set from bootstrap. $C_n = \{\max_{1\le i\le n} X_i\}$.

$$\mathbb{P}(C_n \ni \theta) = 0.$$

"Parametric bootstrap".

If we have parametric model $(P_\theta : \theta \in \Theta)$.

Estimator $\hat{\theta}_n$ from data.

Want to estimate var/CI/....

Parametric bootstrap world:

- $X_1^*, X_2^*, \dots X_n^* \overset{iid}{\sim} P_{\hat{\theta}_n}$
- Compute $\hat{\theta}_n^*$ on bootstrap samples
- Construct var estimator/CI.

Rmk:

- May become more accurate/reliable when parametric assumption is true.
- Sensitive to assumption.

Jackknife: "leave-one-out".

$\hat{\theta}_n = T(\hat{F}_n)$.

$$\hat{\theta}_{n,-i} = T\left(\frac{1}{n-1}\sum_{j\neq i} \mathbb{1}\{\cdot \le X_j\}\right).$$

"empirical distribution" over

$\{X_j : j \neq i\}$.

$$b_{jack} = (n-1)\left(\frac{1}{n}\sum_{i=1}^{n} \hat{\theta}_{n,-i} - \hat{\theta}_n\right).$$