Boostrap:

Goal: Estimate the statistical functional

$$\theta = T(F_x)$$

$X_1, \dots, X_n \sim \text{iid } F_X$ with estimate

$$\hat{\theta} = g(X)$$

Q: How accurate is this $\hat{\theta}$?

$$\widehat{\mathrm{Var}_F}(\hat{\theta}) = \mathrm{Var}_{\hat{F}}(\hat{\theta})$$

$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le x)$$

$$I(X_i \le x) \sim \mathrm{Bern}(F(x))$$

$$\sum_{i=1}^{n} I(X_i \le x) \sim \mathrm{Binom}(n, F(x))$$

$$\mathbb{E}[\hat{F}(x)] = F(x)$$

$$\operatorname{Var}(\hat{F}(x)) = \frac{F(x)(1-F(x))}{n}$$

<u>Ex</u>: $\theta = \mu = \int x \, dF$

$$\hat{\theta} = g(X) = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\operatorname{Var}_F(\hat{\theta}) = \frac{1}{n}\underbrace{\operatorname{Var}_F(X_1)}_{\sigma^2} = \frac{\sigma^2}{n}$$

$$\sigma^2 = \int (x-\mu)^2 \, dF$$

$$\operatorname{Var}_{\hat{F}}(\hat{\theta}) = \frac{\hat{\sigma}^2}{n} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\theta})^2$$

<u>Rmk</u>: Not always possible to use pluggin.

Sample $X^* = (x_1^*, \dots, x_n^*)$ $x_i^* \stackrel{iid}{\sim} \hat{F}$

Compute $\hat{\theta}^* = g(x^*)$

Repeate B times.

$$\mathrm{Var}_{\hat{F}}(\hat{\theta}) \simeq \frac{1}{B} \sum_{i=1}^{B} \left(\hat{\theta}^{*(b)} - \bar{\hat{\theta}}^* \right)^2$$

$$\doteq \widehat{\mathrm{Var}}_{Boot}(\hat{\theta})$$

<u>Q</u>: How do we sample from $\hat{F}$ directly?

<u>A</u>: Sample with replacement.

Inverse CDF method same as sampling with replacement.

$$\mathrm{Var}_F[\hat{\theta}] \underbrace{\underbrace{\simeq}_{O(1/\sqrt{n})}}_{\text{I don't control.}} \mathrm{Var}_{\hat{F}}(\hat{\theta}) \underbrace{\underbrace{\overset{\text{Boot}}{\simeq}}_{O(1/\sqrt{B})}}_{\text{I control}} \widehat{\mathrm{Var}}_{\text{Boot}}(\hat{\theta})$$

<u>Rmk:</u> Usually use these to build empirical confidence intervals.

$100(1-\alpha)\%$ CI

$$\left(\hat{\theta}^*_{(B\alpha/2)}, \hat{\theta}^*_{(B(1-\alpha/2))} \right)$$