Bootstrap → Wasserman 整理

$X \sim F(\cdot)$

① interest: $T_n = g(X_1, \dots, X_n)$

Natural Question: ① Confidence interval

② se → $V_F(T_n)$

Simple Version

→ a) if T_n is simple, like $T_n = n^{-1} \sum_{i=1}^{n} X_i$ → $\mathbb{E}[X] = \int x \, dF(x)$ estimate

then ① $\dfrac{T_n - \mathbb{E}[X]}{\frac{S_n}{\sqrt{n}}} \rightsquigarrow N(0,1) \Longrightarrow$ CI

b) Complex Version

② $V_F(T_n) = \dfrac{V_F(X)}{n}$ $V_F(X) \longleftrightarrow V_{\hat{F}_n}(X)$ plug-in estimator

Theory guarantee ↓ $\hat{F}_n \approx F$

Framework (Plug-in Estimator)

1、总体 $\xrightarrow{\text{interest}}$ $T(F)$

2. 估计 $\xrightarrow{X_1, \dots, X_n}$ $T(\hat{F}_n)$ (F unknown)

plug-in estimator → very good → nice result for linear functional T → asymptotic Gaussian

also hadamard differentiable T

it is also $T(\hat{F}_n) = g(X_1, \dots, X_n) = T_n$

Question: How to find $\hat{se}$?

1. Theorem gives one calculation of $\hat{se}$!

2. Bootstrap → can also give CI

gives Plug-In Estimate CI

Bootstrap → ① estimate $V_F(T_n) \longleftrightarrow$ se

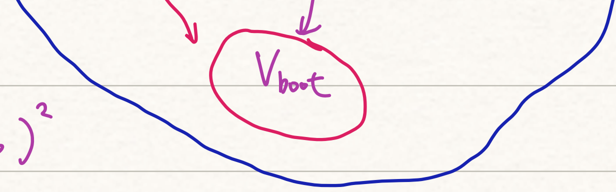
→ ② estimate CI for T_n

① se → our interest: $V_F(T_n) \xleftrightarrow{O(\frac{1}{\sqrt{n}})} V_{\hat{F}_n}(T_n)$

approximation

CLT $O(\frac{1}{\sqrt{B}})$

⇒ same technique can be used on CDF of T_n ⇒ have the similar convergence analysis



$\boxed{\text{SLLN:}}$ $V_{boot} = \frac{1}{B}\sum_{b=1}^{B}\left(T_{n,b}^{*} - \frac{1}{B}\sum_{b=1}^{B}T_{n,b}^{*}\right)^2$

$$= \frac{1}{B}\sum_{b=1}^{B}T_{n,b}^{*2} - \left(\frac{1}{B}\sum_{b=1}^{B}T_{n,b}^{*}\right)^2 \xrightarrow{a.s.} \mathbb{E}_{\hat{F}_n}(T_n^2) - \mathbb{E}_{\hat{F}_n}(T_n)^2$$

$$= \mathbb{V}_{\hat{F}_n}(T_n)$$

$\boxed{\text{CLT}}$ $V_{boot} \xrightarrow{O(\frac{1}{\sqrt{B}})} \mathbb{V}_{\hat{F}_n}(T_n) \longrightarrow$ here, $(X_1, \dots, X_n)$ is fixed

$\left\{\begin{array}{l} \hat{\theta}_n = T(\hat{P}_n) \to g(X_1, \dots, X_n) \\ \theta = T(P) \longrightarrow X \sim P \end{array}\right.$

Basis

② $\boxed{\text{CI}}$ $\longrightarrow$ Focus on $\boxed{F_n(t) = \mathbb{P}_P(\sqrt{n}(\hat{\theta}_n - \theta) \le t)}$

$\left\{\begin{array}{l} \hat{\theta}_n = g(X_1, \dots, X_n) \\ X \sim P(\cdot) \end{array}\right.$

$\left\{\begin{array}{l} \theta_n^* = T(\hat{P}_n^*) \to g(X_1, \dots, X_n) \\ \hat{\theta}_n = T(\hat{P}_n) \ \boxed{X_i \sim \hat{P}_n} \end{array}\right.$

$F_n \xrightarrow{O(\frac{1}{\sqrt{n}})} L$

$L \xrightarrow{O_p(\frac{1}{\sqrt{n}})} \hat{L}$

$\hat{F}_n \xrightarrow{O(\frac{1}{\sqrt{n}})} \hat{L}$

$\hat{F}_n(t) = \mathbb{P}_{\hat{P}_n}(\sqrt{n}(\hat{\theta}_n - \theta) \le t)$ $\boxed{\hat{P}_n \text{ is fixed}}$

$$= \mathbb{P}_{\hat{P}_n}(\sqrt{n}(\theta_n^* - \hat{\theta}_n) \le t)$$

$$= \mathbb{E}_{X \sim \hat{P}_n}[\mathbb{1}\{\sqrt{n}(\theta_n^*(X) - \hat{\theta}_n) \le t\}]$$

$O(\frac{1}{\sqrt{B}})$

generate quantile for basis $\longleftarrow$ $\boxed{\bar{F}_n(t) = \frac{1}{B}\sum_{b=1}^{B}\mathbb{1}\{\sqrt{n}(\theta_{n,b}^* - \hat{\theta}_n) \le t\}}$

$\Rightarrow$ $\boxed{\text{CI}: \left[\hat{\theta}_n - \frac{t_{1-\frac{\alpha}{2}}}{\sqrt{n}}, \ \hat{\theta}_n - \frac{t_{\frac{\alpha}{2}}}{\sqrt{n}}\right]}$

Update 11/17

① Naive case

$X_1, \dots, X_n \rightsquigarrow$ estimate $\mathbb{E}[X]$

$$T(X_1, \dots, X_n) = \bar{X} = n^{-1} \sum_{i=1}^{n} X_i \xrightarrow{d} N(\mu, \frac{\sigma^2}{n})$$

$$\xrightarrow{d} N(\mu, \frac{\hat{\sigma}^2}{n})$$

Standard error of estimator (slutsky)

$\propto$ population variance $\rightarrow$ $\boxed{\text{estimated by plug-in}}$

② Real Case

$X_1, \dots, X_n \rightsquigarrow$ median $[X]$

$$\boxed{T(X_1, \dots, X_n) = \text{median}(X_1, \dots, X_n)}$$

a general approach to achieve the estimator

$\Downarrow$

plug-in estimator

1) Suppose from plug-in principle,

we know that $T(X_1, \dots, X_n) \xrightarrow{d} G(\text{median}[X], \sigma_T^2)$

Question: How to find σ_T^2 ? $\rightarrow$ construct CI

2) if we do not know the Asymptotic Behaviour,

then how to achieve the Confidence Interval?