STAT 824 sp 2025 Lec 10 slides

Bootstrap beyond the mean: Statistical functionals

Karl Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.


Percentile bootstrap
Unstudentized
Consistency
Bootstrap for regression
Bootstrap
Berry-Esseen
Statistical functionals
Studentized
Second-order correctness
Edgeworth expansion

probability measure (not necessarily a cdf)

Let $X_1, \dots, X_n \overset{iid}{\sim} F$.

D is space of prob. distributions

Define $\theta = T(F)$, where $T : D \to \mathbb{R}$

Such a T is a statistical functional.

E.g.

$$\mu = T(F) = \int x \, dF(x)$$

$$\sigma^2 = T(F) = \int \left(x - \int t \, dF(t) \right)^2 dF(x)$$

Quantile τ: $\quad q_\tau = T(F) = \inf\{ x : F(x) \geq \tau \}$, F is cdf

Plug-in estimator:

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

where $\hat{F}_n$ is the empirical dist. of $X_1, \dots, X_n$.

has cdf $\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leq x)$

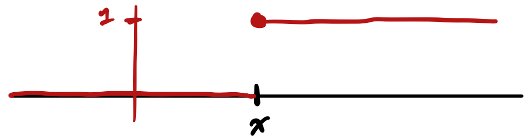
The prob. measure $\hat{F}_n$ is

$$\hat{F}_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i},$$

where δ_x (Dirac measure) is the dist. putting unit mass on the point x.

The cdf corresponding to δ_x

If $T \sim \delta_x$, then $P(T \le t) = \mathbb{1}(t \le x)$



<u>Find plug-in estimator for</u>

(1) $\mu = T(F) = \int x \, dF(x)$

$$\hat{\mu} = T(\hat{F}_n) = \int x \, d\hat{F}_n(x)$$

$$= \int x \, d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{X_i}\right)(x)$$

$$= \frac{1}{n}\sum_{i=1}^{n} \int x \, d\delta_{X_i}(x)$$

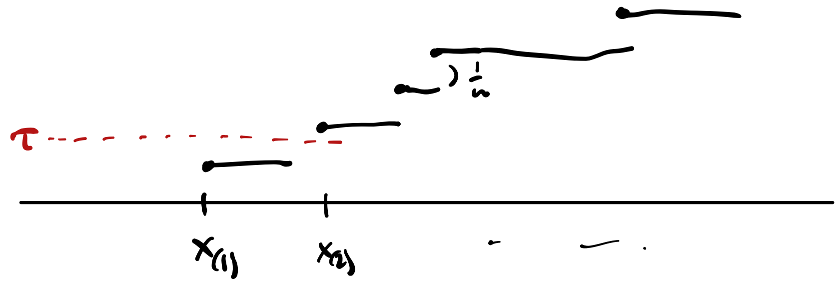
$$= \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$= \bar{X}_n$$

(2) $\hat{\sigma}^2 = T(\hat{F}_n) = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$

(3) $\hat{\xi}_\tau = T(\hat{F}_n) = \inf\{x : \hat{F}_n(x) \ge \tau\}$

$= X_{(\lceil \tau n \rceil)}$ ← order statistic


τ
1/n
$X_{(1)}$
$X_{(2)}$

Throughout, let $X_1, \dots, X_n$ be iid with distribution F

Statistical functional

A *statistical functional* is a function $T : \mathcal{D} \to \mathbb{R}$, where $\mathcal{D}$ is the space of probability distributions.

- Define a quantity $\theta \in \mathbb{R}$ of interest as $\theta = T(F)$.
- Consider *plug-in* estimator $\hat{\theta}_n = T(\hat{F}_n)$, where $\hat{F}_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$
- The notation δ_x represents the distribution placing unit mass on x.
- We will use F, $\hat{F}_n$ to denote distributions and their cdfs interchangeably.

We rely on the Glivenko-Cantelli theorem for consistency of plug-in estimators:

Theorem (Glivenko-Cantelli Theorem)

If $X_1, \ldots, X_n$ is a rs from a distribution with cdf F, then

$$P\left(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |\hat{F}_n(x) - F(x)| = 0\right) = 1$$

Suggests $\hat{\theta}_n = T(\hat{F}_n)$ should get close to $\theta = T(F)$.

Examples of statistical functionals:

1. The mean: $\mu = \int x dF(x)$
2. The variance: $\sigma^2 = \int (x - \int t dF(t))^2 dF(x)$
3. The τth quantile: $\xi_\tau = \inf\{x : F(x) \geq \tau\}$
4. Smooth function of a mean $g(\mu) = g(\int x dF(x))$
5. The ξ-trimmed mean: $\mu_\xi = (1 - 2\xi)^{-1} \int_{F^{-1}(\xi)}^{F^{-1}(1-\xi)} x dF(x)$

Exercise: Write down the plug-in estimators $T(\hat{F}_n)$ for each of the above.

We want something like this:

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) \xrightarrow{D} N(0, \sigma_T^2)$$

Use an expansion (similar to Taylor expansion)

Something like this?

von Mises derivative

$$T(\hat{F}_n) = T(F) + T_F^{(1)}(\hat{F}_n - F) + R(\hat{F}_n - F)$$

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + R$$

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \underbrace{\sqrt{n}\, T_F^{(1)}(\hat{F}_n - F)}_{\xrightarrow{D} N(0,\sigma_T^2)\ ?} + \underbrace{\sqrt{n}\, R(\hat{F}_n - F)}_{\xrightarrow{P} 0\ ?}$$

We want central limit theorems for $\sqrt{n}(T(\hat{F}_n) - T(F))$.

von Mises expansion for statistical functionals

An expansion for stat. functionals called a *von Mises expansion* lets us write

$$\sqrt{n}(T(\hat{F}_n) - T(F)) = \sqrt{n}T_F^{(1)}(\hat{F}_n - F) + \sqrt{n}R(\hat{F}_n - F),$$

where (like a Taylor expansion but for functions $T : \mathcal{D} \to \mathbb{R}$)

- $T_F^{(1)}(\hat{F}_n - F)$ is a *von Mises derivative*.
- $R(\hat{F}_n - F)$ is a (hopefully small) remainder term.

Under some conditions (covered later) we have

$$\sqrt{n}T^{(1)}(\hat{F}_n - F) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_F(X_i) \xrightarrow{\text{D}} \text{Normal}(0, \sigma_T^2),$$

where φ_F is the *influence curve* of the functional T at F and $\sigma_T^2 = \text{Var}\,\varphi_F(X_1)$.

$T(F + \epsilon(\delta_x - F)) = T((1-\epsilon)F + \epsilon\delta_x)$

mixture of F and δ_x

How to find the influence curve

The influence curve is given by $\varphi_F(x) = \frac{d}{d\varepsilon} T(F + \varepsilon(\delta_x - F))\Big|_{\varepsilon=0}$.

Influence curves play an important role in *robust estimation*.

The IC expresses change in $T(F)$ due to perturbing F by adding a point mass at x.

Exercise: Find the influence curves φ_F for these functionals:

1. $\mu = \int x dF(x)$
2. $\sigma^2 = \int (x - \int t dF(t))^2 dF(x)$
3. $g(\mu) = g(\int x dF(x))$ for some function g

$$\varphi_F(x) = \frac{d}{d\varepsilon} T(F + \varepsilon(\delta_x - F))\Big|_{\varepsilon=0}.$$

(1) $\mu = T(F) = \int t \, dF(t)$

$$\varphi_F(x) = \frac{d}{d\epsilon} T\left(F + \epsilon(\delta_x - F)\right)\Big|_{\epsilon=0}$$

$$= \frac{d}{d\epsilon} \int t \, d\left(F + \epsilon(\delta_x - F)\right)(t)\Big|_{\epsilon=0}$$

$(1-\epsilon)F + \epsilon\delta_x$

$$= \frac{d}{d\epsilon}\Big[(1-\epsilon)\underbrace{\int t \, dF(t)}_{\mu} + \epsilon \underbrace{\int t \, d\delta_x(t)}_{x}\Big]\Big|_{\epsilon=0}$$

$$= \frac{d}{d\epsilon}\left[(1-\epsilon)\mu + \epsilon x\right]\Big|_{\epsilon=0}$$

$$= x - \mu.$$

von Mises expansion

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_F(X_i) + \sqrt{n} R(\hat{F}_n - F)$$

For $\mu = T(F) = \int x \, dF(x)$ we have

$$\sqrt{n}\left(\bar{X}_n - \mu\right) = \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)}_{} + \underbrace{\sqrt{n} R(\hat{F}_n - F)}_{=0}$$

$$\xrightarrow{D} N(0, \sigma^2)$$

$$\sigma_T^2 = \mathrm{Var}\, \varphi_F(X_1) = \mathrm{Var}(X_1 - \mu) = \sigma^2$$

(3)

$$\theta = T(F) = g\left(\int x\,dF(x)\right)$$

$$\hat{\theta} = T(\hat{F}_n) = g\left(\underbrace{\int x\,d\hat{F}_n(x)}_{\bar{X}_n}\right) = g(\bar{X}_n)$$

Then

$$\varphi_F(x) = \frac{d}{d\epsilon} T\left(F + \epsilon(\delta_x - F)\right)\Big|_{\epsilon=0}$$

$$= \frac{d}{d\epsilon}\, g\left(\int t\, d\left(F + \epsilon(\delta_x - F)\right)(t)\right)\Big|_{\epsilon=0}$$

$$= \frac{d}{d\epsilon}\, g\left((1-\epsilon)\mu + \epsilon x\right)\Big|_{\epsilon=0}$$

$$= g'\left((1-\epsilon)\mu + \epsilon x\right)(x-\mu)\Big|_{\epsilon=0}$$

$$= g'(\mu)(x-\mu)$$

$$\sqrt{n}\left(T(\hat{F}_n) - T(F)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \varphi_F(X_i) + \sqrt{n}\,R(\hat{F}_n - F)$$

$$\sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} g'(\mu)(X_i - \mu) + \sqrt{n}\,R(\hat{F}_n - F)$$

$$= \underbrace{\sqrt{n}(\bar{X}_n - \mu)\,g'(\mu)}_{\xrightarrow{D} N(0,\,[g'(\mu)]^2\sigma^2)} + \sqrt{n}\,R(\hat{F}_n - F)$$

$$\sigma_T^2 = \mathrm{Var}\,\varphi_F(X_1) = \mathrm{Var}\left(g'(\mu)(X_1 - \mu)\right) = [g'(\mu)]^2\sigma^2$$

$$\sqrt{n}\, R(\hat{F}_n - F) = \sqrt{n}\left(g(\bar{X}_n) - g(\mu)\right) - \sqrt{n}\,(\bar{X}_n - \mu)\, g'(\mu)$$

Exercise: Write down the expansion

$$\sqrt{n}(T(\hat{F}_n) - T(F)) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi_F(X_i) + \sqrt{n}R(\hat{F}_n - F),$$

and give the asymptotic distribution of $\sqrt{n}(T(\hat{F}_n) - T(F))$ for:

1. $\mu = \int x dF(x)$
2. $p_A = \int_A dF(x)$
3. $g(\mu) = g(\int x dF(x))$

von Mises expansion for the τth quantile

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} F$ with continuous density f and consider the $\tau$the quantile

$$\xi_\tau = T(F) = \inf\{x : F(x) \geq \tau\} = F^{-1}(\tau)$$

The influence function is

$\sqrt{n}(\hat{\xi}_\tau - \xi_\tau) \xrightarrow{D} N\left(0, \frac{\tau(1-\tau)}{f^2(\xi_\tau)}\right)$

$$\varphi_F(x) = \frac{\tau - \mathbf{1}(x \leq \xi_\tau)}{f(\xi_\tau)}, \qquad \text{provided } f(\xi_\tau) > 0.$$

$\sigma_\tau^2 = \operatorname{Var} \varphi_F(X_1) = \frac{1}{f^2(\xi_\tau)} \operatorname{Var}(\tau - \mathbb{1}(X_1 \leq \xi_\tau)) = \frac{\tau(1-\tau)}{f^2(\xi_\tau)}$

Exercise:

1. Give the von Mises expansion of $\sqrt{n}(\hat{\xi}_\tau - \xi_\tau)$.
2. Make a conjecture about the asymptotic distribution of $\sqrt{n}(\hat{\xi}_\tau - \xi_\tau)$.

We have $\sqrt{n}R(\hat{F}_n - F) \to 0$ in probability as $n \to \infty$, by Ghosh (1971) [2].

Table of Contents

Consistency of bootstrap for Hadamard differentiable functionals

If T is Hadamard differentiable and $\sigma_T^2 = \text{Var}\,\varphi_F(X_1) < \infty$, then

$$\sup_{x\in\mathbb{R}} \left| P_*\left(\sqrt{n}(T(\hat{F}_n^*) - T(\hat{F}_n)) \leq x\right) - P\left(\sqrt{n}(T(\hat{F}_n) - T(F)) \leq x\right)\right| \to 0$$

in probability as $n \to \infty$.

In the above $\hat{F}_n^* = \frac{1}{n}\sum_{i=1}^n \delta_{X_i^*}$, where $X_1^*, \dots, X_n^* | X_1, \dots, X_n \overset{\text{ind}}{\sim} \hat{F}_n$.

Many interesting statistical functionals are Hadamard differentiable (defined later).

Exercise: Given B sorted Monte Carlo reps $T^{*(1)}, \dots, T^{*(B)}$ of $T(\hat{F}_n^*)$, justify

$$\left(2 \cdot T(\hat{F}_n) - T^{*(\lceil(1-\alpha/2)B\rceil)}, 2 \cdot T(\hat{F}_n) - T^{*(\lceil(\alpha/2)B\rceil)}\right)$$

as an asymptotic $(1-\alpha) \times 100\%$ confidence interval for $T(F)$.

Exercise: Let F be the Gamma(α, β) and construct 95% bootstrap CIs for

$$T(F) = \frac{\int (x-\mu)^3 dF(x)}{(\int (x-\mu)^2 dF(x))^{3/2}}, \text{ where } \mu = \int x dF(x).$$

Run simulations with $n = 30$ and $n = 100$ and assess coverage. Note $T(F) = \frac{2}{\sqrt{\alpha}}$.

$$\mu_\xi = T(F) = \frac{1}{1-2\xi}\int_{F^{-1}(\xi)}^{F^{-1}(1-\xi)} x\, dF(x) \qquad \hat{\mu}_\xi = \frac{1}{n(1-2\xi)}\sum_{i=\lceil \xi n\rceil}^{\lceil (1-\xi)n\rceil} X_{(i)}$$

Exercise: Coverage of 95% bootstrap CIs for the ξ-trimmed mean when $X_1, \dots, X_n$ are independent realizations of the random variable

$$X = D(G - ab) + (1 - D)\upsilon|T|,$$

where D, G, and T are independent random variables such that

- $D \sim \text{Bernoulli}(\delta)$
- $G \sim \text{Gamma}(a, b)$
- $T \sim t_2$ has the t distribution with degrees of freedom 2

Coverage over 500 datasets under $a = 2$, $b = 3$, $\delta = 0.8$, $\xi = 0.10$, $B = 500$:

n	20	40	80	160	320
coverage	0.89	0.89	0.94	0.94	0.92

Exercise: Let $F = \text{Uniform}(0, \theta)$. Then $\theta = T(F) = \inf\{x : F(x) \geq 1\}$.

1. Find the asymptotic distribution of $\sqrt{n}(T(\hat{F}_n) - T(F))$.
2. Find the asymptotic distribution of $n(T(\hat{F}_n) - T(F))$.
3. Consider behavior as $n \to \infty$ of the quantity

$$\sup_{x \in \mathbb{R}} \left| P_*(n(T(\hat{F}_n^*) - T(\hat{F}_n)) \leq x) - P(n(T(\hat{F}_n) - T(F)) \leq x) \right|.$$

Does the bootstrap work in this setting?

Table of Contents

von Mises derivative. See the book of Luisa Fernholz [1].

The *von Mises derivative of T at F in the direction* G is defined as

$$T_F^{(1)}(G - F) = \frac{d}{d\varepsilon} T(F + \varepsilon(G - F))\Big|_{\varepsilon=0},$$

provided there exists a function φ_F, not depending on G, such that

$$T_F^{(1)}(G - F) = \int \varphi_F(x) d(G - F)(x),$$

with $\int \varphi_F(x) dF(x) = 0$; in this case $\varphi_F(x) = T_F^{(1)}(\delta_x - F)$.

The function φ_F is called the *influence curve* of the functional T at F.

Exercise: Find $T_F^{(1)}(G - F)$ and $T_F^{(1)}(\delta_x - F)$ for $T(F) = \int x dF(x)$.

von Mises expansion for M estimators

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} F$, and for some function $\psi : \mathbb{R}^2 \to \mathbb{R}$ consider

$$\begin{aligned} \theta_0 &= T(F) = \text{ value of } t \text{ which solves } \int \psi(x,t)dF(x) = 0 \\ \hat{\theta}_n &= T(F) = \text{ value of } t \text{ which solves } \int \psi(x,t)d\hat{F}_n(x) = 0 \end{aligned}$$

The von Mises derivative is

$$T_F^{(1)}(G - F) = -\frac{\lambda_G(T(F))}{\lambda_F'(T(F))},$$

where $\lambda_F(t) = \int \psi(x,t)dF(x)$ and $\lambda_G(t) = \int \psi(x,t)dG(x)$, $t \in \mathbb{R}$.

Exercise:

1. Give the von Mises expansion of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.
2. Make a conjecture about the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$.
3. Discuss connection to maximum likelihood estimators (ψ as score function).

von Mises expansion for L estimators

Let $X_1, \dots, X_n \overset{\text{ind}}{\sim} F$, and for some function $J : (0,1) \to \mathbb{R}$ consider

$$\theta_0 = T(F) = \int_0^1 J(u)F^{-1}(u)du$$
$$\hat{\theta}_n = T(F) = \int_0^1 J(u)\hat{F}_n^{-1}(u)du$$

The von Mises derivative is

$$T_F^{(1)}(G - F) = \int_{-\infty}^{\infty} J(F(y))[F(y) - G(y)]dy.$$

Exercise:

1. Find $u_1, \dots, u_n$ such that $\hat{\theta}_n = \sum_{i=1}^n u_i X_{(i)}$, with $X_{(1)} < \dots < X_{(n)}$.
2. Identify the function J that gives the α-trimmed mean μ_α.

See handwritten notes for von Mises expansion of $\sqrt{n}(\hat{\mu}_\alpha - \mu_\alpha)$.

Central limit theorem for Hadamard differentiable functionals. See [3].

If T is a Hadamard differentiable functional then

1. $\sqrt{n}(T(\hat{F}_n) - T(F)) \to \text{Normal}(0, \vartheta)$ in distribution as $n \to \infty$, with

$$\vartheta = \int [T_F^{(1)}(\delta_x - F)]^2 dF(x).$$

2. $\sqrt{n}(T(\hat{F}_n) - T(F))/\hat{\vartheta}^{1/2} \to \text{Normal}(0, 1)$ in distribution as $n \to \infty$, with

$$\hat{\vartheta} = \int [T_{\hat{F}_n}^{(1)}(\delta_x - \hat{F}_n)]^2 d\hat{F}_n(x).$$

Result (ii) validates $T(\hat{F}_n) \pm z_{\alpha/2}\sqrt{\hat{\vartheta}/n}$ as an asymp. $(1-\alpha)100\%$ CI for $T(F)$.

Exercise: Find $\hat{\vartheta}$ for

1. $T(F) = \int x dF(x)$
2. $T(F) = g(\int x dF(x))$.

Let $\mathcal{D}$ be the space of linear combinations of probability distributions.

Hadamard differentiability

A functional $T : \mathcal{D} \to \mathbb{R}$ is *Hadamard differentiable* at $F \in \mathcal{D}$ in the direction $G \in \mathcal{D}$ if there exists a linear function $T_F^{(1)} : \mathcal{D} \to \mathbb{R}$ such that

$$\lim_{n \to \infty} \left| \frac{T(F + \varepsilon_n(G_n - F)) - T(F)}{\varepsilon_n} - T_F^{(1)}(G - F) \right| = 0,$$

for all sequences $G_n \in \mathcal{D}$ such that $\|G_n - G\|_\infty \to 0$ and $\varepsilon_n \downarrow 0$ as $n \to \infty$

Luisa Fernholz [1] gives conditions under which

1. M-estimators
2. L-estimators
3. R-estimators (rank based estimators)

satisfy Hadamard differentiability.

Quantiles do not, but asymptotic Normality of $\sqrt{n}(\hat{\xi}_\tau - \xi_\tau)$ can still be shown.

Luisa Turrin Fernholz.
Von Mises calculus for statistical functionals, volume 19.
Springer Science & Business Media, 2012.

Jayanta K Ghosh.
A new proof of the bahadur representation of quantiles and an application.
The Annals of Mathematical Statistics, pages 1957–1961, 1971.

Larry Wasserman.
All of nonparametric statistics.
Springer Science & Business Media, 2006.