STA732
Statistical Inference

Lecture 14: Basics in large sample theory

Yuansi Chen

Spring 2023

Duke University

https://www2.stat.duke.edu/courses/Spring23/sta732.01/

Recap from Lecture 13

Two more strategies to find minimax estimator

1. Least favorable prior sequence
2. Submodel restriction

Minimax estimators are not always admissible, but unique minimax estimator is.

Motivation from finite-sample to large-sample theory

- So far, everything has been finite-sample. We often rely on the simple form of the model P (e.g. exponential family) to do exact calculation to prove unbiasedness, Bayes, minimax etc.
- Exact calculation may be intractable when models are more complicated.
- We may use approximations
- Approximations by Gaussian are often possible as $n \to \infty$. But the asymptotic theory is only interesting for "reasonably" large sample size.

In large-sample i.i.d. case, MLE is approximately optimal!

Let $X_i \overset{\text{i.i.d.}}{\sim} p_\theta$ for $i = 1, \dots, n$ for a generic "regular" p_θ. Then MLE is approximately optimal:

- consistent
- asympotically normal
- asympotically efficient
- locally asympotically minimax (won't cover)

Similar optimality results hold for tests and confidence intervals based on likelihood

Goal of Lecture 14

1. Convergence of random variables
    - Almost sure convergence
    - Convergence in probability
    - Convergence in distribution
2. Calculus for functions of converging random variables
    - Continuous mapping theorem
    - Slutsky's theorem
    - Delta method

Chap. 8.1 - 8.2 of Keener or Chap. 1.8, 6.1 of Lehmann and Casella

Convergence of random variables

Convergence in probability

Let $Y_1, Y_2, \ldots \in \mathbb{R}^d$ be a sequence of random variables.

Def. convergence in probability, 8.1 Keener

The sequence $Y_1, Y_2, \ldots$ converges in probability to a random variable Y (denoted as $Y_n \xrightarrow{p} Y$) if for every $\epsilon > 0$

$$\mathbb{P}\left(|Y_n - Y| > \epsilon\right) \to 0$$

Remark:

In most cases in this course, we only need convergence in probability to a constant c

$$Y_n \xrightarrow{p} c.$$

Example: show convergence in probability via Chebyshev

Thm. Chebyshev's inequality. 8.2 in Keener

For any random variable X and any constant $a > 0$,

$$\mathbb{P}(|X| \geq a) \leq \frac{\mathbb{E}[X^2]}{a^2}.$$

Prop. 8.3 in Keener

If $\mathbb{E}\left(Y_n - Y\right)^2 \to 0$ as $n \to \infty$, then $Y_n \xrightarrow{p} Y$.

Convergence in distribution

Def. convergence in distribution, 8.9 Keener

The sequence $Y_1, Y_2, \dots$ converges in distribution to a random variable Y (noted $Y_n \Rightarrow Y$ or $Y_n \xrightarrow{d} Y$) if

$$\mathbb{E}f(Y_n) \to \mathbb{E}f(Y)$$

for all bounded continuous function f.

Alternative definition, 8.7 Keener

Let $F_n(y) = \mathbb{P}(Y_n \leq y), F(y) = \mathbb{P}(Y \leq y)$, then $Y_n \Rightarrow X$ iff $F_n(x) \to F(x)$ for all x such that F is continuous at x.

Example

When convergence in distribution, cumulative distribution function does not have to converge everywhere.

Let $Y_n \sim \delta_{\frac{1}{n}}$, $Y \sim \delta_0$, then $Y_n \Rightarrow Y$ (according to the first definition). For the cumulative distribution function

$$F_n(y) = \mathbf{1}_{\frac{1}{n} \leq x} \to \mathbf{1}_{0 \leq x}$$

except at $x = 0$.

Convergence in probability vs. convergence in distribution (1)

Convergence in probability implies convergence in distribution

Prop.

If $Y_n \overset{p}{\to} Y$, then $Y_n \overset{d}{\to} Y$

proof: show the cdf converges at continuous points

Useful lemma to prove the above Prop.

$$\mathbb{P}(Y \leq a) \leq \mathbb{P}(X \leq a + \epsilon) + \mathbb{P}(|Y - X| > \epsilon)$$

Convergence in probability vs. convergence in distribution (2)

Convergence in distribution to a constant implies convergence in probability

Prop.

$Y_n \xrightarrow{p} c$ iff $Y_n \xrightarrow{d} \delta_c$

Consider $f_\epsilon(x) = \min\left\{1, \frac{|x-c|}{\epsilon}\right\}$

Consistency

Def. consistency

Given a sequence of statistical models $\mathscr{P}_n = \{P_{n,\theta}, \theta \in \Omega\}$ with data $X_n \sim P_{n,\theta}$, we say $\delta_n(X_n)$ is consistent for $g(\theta)$ if

$$\delta_n(X_n) \xrightarrow{p} g(\theta),$$

that is,

$$\mathbb{P}_\theta(|\delta_n(X_n) - g(\theta)| > \epsilon) \to 0$$

We may omit the index n on $P_{n,\theta}$ if it is clear

Sufficient condition for consistency

Let $R(\theta, \delta_n) = \mathbb{E}_\theta \left(\delta_n - g(\theta)\right)^2$ is the mean squared error (or risk under squared error loss), by Chebyshev's inequality, if $R(\theta, \delta_n) \to 0$ as $n \to \infty$ for any θ, then δ_n is consistent.

Limit theorems

Let $X_1, X_2, \ldots$ be i.i.d. random variables. Let $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.

Law of large numbers (LLN)

If $\mathbb{E}\,|X_i| < \infty, \mathbb{E}X_i = \mu$, then $\bar{X}_n \xrightarrow{p} \mu$

actually $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$ also works

Central limit theorem (CLT)

If $\mathbb{E}X_n = \mu$, $\mathrm{Var}(X_n) = \Sigma$, then

$$\sqrt{n}(\bar{X}_n - \mu) \Rightarrow \mathcal{N}(0, \Sigma).$$

see Keener 8.12 for proofs and discussions

Calculus for functions of converging random variables

Continuous mapping theorem

Thm. 8.5 8.11 in Keener

Let g be a continuous function. $Y_1, Y_2, \ldots$ random variables.

- If $Y_n \Rightarrow Y$, then $g(Y_n) \Rightarrow g(Y)$
- If $Y_n \xrightarrow{p} c$, then $g(Y_n) \xrightarrow{p} g(c)$

Slutsky's theorem

Thm. Slutsky's theorem, 8.13 in Keener
Assume $X_n \Rightarrow X, Y_n \xrightarrow{p} c$, then

- $X_n + Y_n \Rightarrow X + c$
- $X_n \cdot Y_n \Rightarrow cX$
- $X_n / Y_n \Rightarrow X/c$ if $c \neq 0$

proof: show that $X_n \Rightarrow X, Y_n \xrightarrow{p} c$ implies $(X_n, Y_n) \Rightarrow (X, c)$, then apply cts mapping

Delta method

Delta method, 8.14 in Keener

Suppose

- $\sqrt{n}(X_n - \mu) \Rightarrow \mathcal{N}(0, \sigma^2)$
- $f(\cdot)$ is differentiable at μ

Then

$$\sqrt{n}\left(f(X_n) - f(\mu)\right) \Rightarrow \mathcal{N}(0, [f'(\mu)]^2\sigma^2)$$

Informal statement:

$X_n \approx \mathcal{N}(\mu, \frac{\sigma^2}{n})$ implies $f(X_n) \approx \mathcal{N}(f(\mu), [f'(\mu)]^2 \frac{\sigma^2}{n})$

Small o and big O notation (1)

Def. scale of magnitude for constants, 8.20 Keener

Let a_n and $b_n, n \geq 1$ be constants. Then

- $a_n = o(b_n)$ as $n \to \infty$ means that $a_n/b_n \to 0$ as $n \to \infty$
- $a_n = O(b_n)$ as $n \to \infty$ means that $\limsup_{n\to\infty} |a_n/b_n| < \infty$

Small o and big O notation (2)

Def. scale of magnitude for random variables, 8.21 Keener

Let $X_n, Y_n, n \geq 1$ be random variables, let $b_n, n \geq 1$ be constants. Then

- $X_n = o_p(b_n)$ as $n \to \infty$ means that $X_n/b_n \xrightarrow{p} 0$ as $n \to \infty$
- $X_n = O_p(1)$ as $n \to \infty$ means that

$$\sup_n \mathbb{P}(|X_n| > K) \to 0$$

  as $K \to \infty$
- $X_n = O_p(b_n)$ means that $X_n/b_n = O_p(1)$ as $n \to \infty$.

Prop. 8.24 in Keener

If $X_n = O_p(a_n)$ with $a_n \to 0$, and if $f(\epsilon) = o(\epsilon^\alpha)$ as $\epsilon \to 0$ with $\alpha > 0$, then

$$f(X_n) = o_p(a_n^\alpha)$$

proof omitted, go through the definitions

Proof of the Delta method

By the first assumption, $X_n = \mu + O_p(1/\sqrt{n})$ By Taylor expansion, we have

$$f(\mu + \epsilon) = f(\mu) + \epsilon f'(\mu) + o(\epsilon)$$

as $\epsilon \to 0$. Applying Prop. 8.24, we have

$$f(X_n) = f(\mu) + f'(\mu)(X_n - \mu) + o_p(1/\sqrt{n}).$$

Rearraging the terms, we obtain

$$\sqrt{n}\left(f(\bar{X}_n) - f(\mu)\right) = \sqrt{n}\left(\bar{X}_n - \mu\right) f'(\mu) + o_p(1) \Rightarrow \mathcal{N}(0, [f'(\mu)]^2\sigma^2)$$

Picture for the Delta method

Example

Suppose $X_1, \ldots X_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2), Y_1, \ldots, Y_n \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\nu, \tau^2), X_i, Y_i$ are independent.

- Find the asymptotic distribution of $\left(\bar{X} + \bar{Y}\right)^2$

Extensions of the Delta method

- Multivariate Delta method
- Higher-order Taylor expansion if first derivative is 0:

$$f(X_n) = f(\mu) + f'(\mu)(X_n - \mu) + \frac{f''(\mu)}{2}\left(X_n - \mu\right)^2 + o_p(1/n).$$

If $f'(\mu) = 0$, then

$$\begin{aligned} n\left(f(X_n) - f(\mu)\right) &= \frac{f''(\mu)}{2}\left(\sqrt{n}(X_n - \mu)\right)^2 + o_p(1) \\ &\Rightarrow \frac{f''(\mu)\sigma^2}{2}\chi_1^2 \end{aligned}$$

Summary

- Convergence in probability vs. convergence in distribution
- Basic strategy to find asympotic distribution:
    - Apply central limit theorem to some X_n
    - Use continuous mapping/Slutsky/Delta method for $f(X_n)$

What is next?

Derive asymptotic distribution of the MLE

Thank you