

# STA732

## Statistical Inference

Lecture 14: Basics in large sample theory

---

Yuansi Chen

Spring 2022

Duke University

<https://www2.stat.duke.edu/courses/Spring22/sta732.01/>



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  $n \rightarrow \infty$ . But the asymptotic theory is only interesting for “reasonably” large sample size.

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

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

- consistent
- asymptotically normal
- asymptotically efficient
- locally asymptotically minimax

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

**Def. convergence in probability, 8.1 Keener**

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

$$\mathbb{P}(|Y_n - Y| > \epsilon) \rightarrow 0$$

**Remark:**

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

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

**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  $\mathbb{E}(Y_n - Y)^2 \rightarrow 0$  as  $n \rightarrow \infty$ , then  $Y_n \xrightarrow{p} Y$ .



### 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

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

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  $F_n(x) \rightarrow F(x)$  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

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

except at  $x = 0$ .

## Convergence in probability vs. convergence in distribution (1)

Convergence in probability implies convergence in distribution

**Prop.**

If  $Y_n \xrightarrow{p} Y$ , then  $Y_n \xrightarrow{d} Y$

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 distribution to a constant implies convergence in probability

**Prop.**

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

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

## Def. consistency

Given a sequence of statistical models  $\mathcal{P}_n = \{P_{n,\theta}, \theta \in \Omega\}$  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,

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

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

## Sufficient condition for consistency

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

## Limit theorems

Let  $X_1, X_2, \dots$  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  $\mathbb{E}X_n = \mu$ ,  $\text{Var}(X_n) = \Sigma$ , 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

---



### Thm. 8.5 8.11 in Keener

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

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

## 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, 8.14 in Keener

Suppose

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

Then

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

Informal statement:

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

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

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

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

as  $k \rightarrow \infty$

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

**Prop. 8.24 in Keener**

If  $X_n = O_p(a_n)$  with  $a_n \rightarrow 0$ , and if  $f(\epsilon) = o(\epsilon^\alpha)$  as  $\epsilon \rightarrow 0$  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  $\epsilon \rightarrow 0$ . Applying Prop. 8.24, we have

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

Rearranging the terms, we obtain

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

## Picture for the Delta method



## Example

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

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

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

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

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

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

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

Derive asymptotic distribution of the MLE

Thank you for attending  
See you on Monday in Old Chem  
025

