

# STAT 4100 Lecture Note

Week Ten (Nov 14, 16, & 18, 2022)

Zhiyang Zhou (zhiyang.zhou@umanitoba.ca, zhiyanggeezhou.github.io)

2022/Nov/19 01:30:20

---

## Important inequalities

### Markov's inequality (CB Lemma 3.8.3 & HMC Thm 1.10.2)

- If  $\Pr(X \geq 0) = 1$  and  $EX^k$  exists, then, for all  $r, k > 0$ ,

$$\Pr(X \geq r) \leq EX^k / r^k.$$

### Chebyshev's inequality (CB Thm 3.6.1 & Example 3.6.2 & HMC Thm 1.10.3)

- A corollary of Markov's inequality
- Let  $X \sim (\mu_X, \sigma_X^2)$ . Then, for each  $r > 0$ ,

$$\Pr\{|X - \mu_X| \geq r\sigma_X\} = \Pr\{(X - \mu_X)^2 / \sigma_X^2 \geq r^2\} \leq r^{-2}.$$

### Cauchy-Schwarz inequality (CB Thm 4.7.3)

- $X$  and  $Y$  are both r.v.s. Then  $|E(XY)| \leq E|XY| \leq \sqrt{EX^2}\sqrt{EY^2}$ .
  - Because

$$\frac{X^2}{EX^2} + \frac{Y^2}{EY^2} \geq \frac{2|XY|}{\sqrt{EX^2}\sqrt{EY^2}}$$

## Convexity and concavity

- Convex set: for any two points in the set, the whole line segment that joins them is also in the set
- Let  $\mathcal{D}$  be a convex set. Then real-valued function  $f : \mathcal{D} \rightarrow \mathbb{R}$  is convex  $\iff f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$  for all  $x_1, x_2 \in \mathcal{D}$  and all  $\lambda \in [0, 1]$ .
  - If  $f$  is twice-differentiable, then  $f$  is convex (on  $\mathcal{D}$ )  $\iff f''(x) \geq 0$  for each  $x \in \mathcal{D}$ .
- $f$  is concave  $\iff -f$  is convex.

### Example Lec17.1

- Check the convexity/concavity of following functions.
  - a.  $f(x) = \exp(x)$ ,  $x \in \mathbb{R}$ .
  - b.  $f(x) = \ln x$ ,  $x \in \mathbb{R}^+$ .
  - c.  $f(x) = x^2$ ,  $x \in \mathbb{R}$ .
  - d.  $f(x) = x^{-1}$ ,  $x \in \mathbb{R} \setminus \{0\}$ .
  - e.  $f(x) = x^{-2}$ ,  $x \in \mathbb{R} \setminus \{0\}$ .

## Jensen's inequality (CB Thm 4.7.7 & HMC Thm 1.10.5)

- If  $f$  is convex on  $(a, b)$  and  $EX \in (a, b)$ , then

$$E\{f(X)\} \geq f(EX).$$

### Example Lec17.2

- Let  $X$  be a positive random variable, i.e.,  $\Pr(X > 0) = 1$ . Argue that
  - a.  $E(-\ln X) \geq \ln(1/EX)$ ;
  - b.  $EX^3 \geq (EX)^3$ .

## Convergence of random variables

### Definitions

- Convergence in probability (CB Def 5.5.1), say  $X_n \xrightarrow{P} X$ : for each  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr(|X_n - X| < \varepsilon) = 1, \quad \text{or equivalently,} \quad \lim_{n \rightarrow \infty} \Pr(|X_n - X| > \varepsilon) = 0.$$

- Almost sure convergence (CB Def 5.5.6), say  $X_n \xrightarrow{\text{a.s.}} X$ :

$$\Pr\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1$$

or equivalently, for each  $\varepsilon > 0$ ,

$$\Pr\left(\lim_{n \rightarrow \infty} |X_n - X| < \varepsilon\right) = 1$$

- Convergence in distribution (CB Def 5.5.10), say  $X_n \xrightarrow{d} X$ : for each  $x$  with  $\Pr(X = x) = 0$ ,

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x);$$

or equivalently (and optionally), for each  $t \in \mathbb{R}$ ,

$$\lim_{n \rightarrow \infty} E\{\exp(itX_n)\} = E\{\exp(itX)\};$$

or equivalently (and optionally), for each  $f$  continuous and bounded within  $\text{supp}(X)$ ,

$$\lim_{n \rightarrow \infty} E\{f(X_n)\} = E\{f(X)\}.$$

- For the third equivalent statement, the boundedness of  $f$  is essential. Hence the convergence in distribution doesn't imply the convergence of moments; see CB Example 10.1.10.

### Example Lec18.1 (optional)

- Assume that  $X(s) = 0$  for all  $s \in [0, 1]$  and

$$X_n(s) = \begin{cases} 1, & s \in [\frac{n}{2^{\lfloor \log_2 n \rfloor}} - 1, \frac{n+1}{2^{\lfloor \log_2 n \rfloor}} - 1] \\ 0, & \text{elsewhere.} \end{cases}$$

Then the convergence of  $X_n$  to  $X$  is in probability but NOT almost surely.

### CB Example 5.5.11

- (Limiting distribution of the maximum of uniforms) if iid  $X_1, \dots, X_n$  follow  $\text{Unif}(0, 1)$ , then  $n(1 - X_{(n)}) \xrightarrow{d} \text{exponential}(1)$ .

## Connections

- The chain of implications

$$\xrightarrow{\text{a.s.}} \Rightarrow \xrightarrow{p} \Rightarrow \xrightarrow{d}$$

- (Continuous mapping theorem, CMT)  $h(\cdot)$  is continuous and  $X_n \xrightarrow{\text{a.s./}p/d} X \Rightarrow h(X_n) \xrightarrow{\text{a.s./}p/d} h(X)$ .
- (CB Thm 5.5.13 and Exercise 5.41)  $X_n \xrightarrow{d} \text{constant } c \Rightarrow X_n \xrightarrow{p} c$ .

- $X_n \xrightarrow{\text{a.s./}p} X$  and  $Y_n \xrightarrow{\text{a.s./}p} Y \Rightarrow$ 
  - $aX_n + bY_n \xrightarrow{\text{a.s./}p} aX + bY$
  - $X_n Y_n \xrightarrow{\text{a.s./}p} XY$
- (Slutsky's theorem, CB Thm 5.5.17)  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} \text{constant } c \Rightarrow$ 
  - $aX_n + bY_n \xrightarrow{d} aX + bc$ 
    - \* The requirement that  $Y_n \xrightarrow{d} \text{constant } c$  is important. Otherwise, consider the following counterexample: assuming  $X_n, X, Y \sim \text{Unif}(-1, 1)$  and  $Y_n = -X_n$ , one may find that  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{d} Y$  but  $X_n + Y_n$  does not converge to  $X + Y$  in distribution.
  - $X_n Y_n \xrightarrow{d} cX$

## CB Exercise 5.43

- $\sqrt{n}(X_n - c) \xrightarrow{d} \mathcal{N}(0, \sigma^2) \Rightarrow X_n \xrightarrow{p} c$ .

## Laws of large numbers (LLN, CB Thm 5.5.2 & 5.5.9)

- If  $X_1, X_2, \dots$  are iid with finite mean  $\mu$ , then
  - (Weak law of large numbers, WLLN)  $\bar{X}_n \xrightarrow{p} \mu$ ;
    - \* Proof using Chebyshev's inequality (if assuming finite variance as well)
  - (Strong law of large numbers, SLLN)  $\bar{X}_n \xrightarrow{\text{a.s.}} \mu$ .

## Central limit theorem (CLT)

- (CB Thm 5.5.15) if  $X_1, \dots, X_n$  are iid with finite mean  $\mu$  and finite variance  $\sigma^2$ , then

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} \xrightarrow{d} \mathcal{N}(0, 1).$$

- A generic approximation to the distribution of  $\bar{X}_n$
- (CB Example 5.5.18) assuming conditions for CLT and that  $T_n \xrightarrow{d} \sigma > 0$ ,

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{T_n} \xrightarrow{d} \mathcal{N}(0, 1).$$

- by the CMT, CLT and Slutsky's theorem

## CB Exercise 5.35 (optional)

- Derive the Stirling's formula:  $n! \approx n^{n+1/2} \exp(-n) \sqrt{2\pi}$ .

## Asymptotic properties of MLE

### Consistency (or consistence, CB Sec 10.1.1)

- $T_n = T_n(X_1, \dots, X_n)$  is consistent for  $\theta$  iff  $T_n \xrightarrow{p} \theta$  as  $n \rightarrow \infty$ 
  - A sufficient condition for consistency:  $E(T_n | \theta) \rightarrow \theta$  and  $\text{var}(T_n | \theta) \rightarrow 0$  as  $n \rightarrow \infty$

### CB Example 5.5.3

- Suppose that iid  $X_1, \dots, X_n \sim \mathcal{N}(\mu, \sigma^2)$ . Prove that
  - $S_n^2 = (n-1)^{-1} \sum_i (X_i - \bar{X}_n)^2$  is consistent for  $\sigma^2$ ;
  - $\widehat{\sigma}_{\text{ML}}^2 = n^{-1} \sum_i (X_i - \bar{X}_n)^2$  is consistent for  $\sigma^2$  too.

### Consistency of MLE (univariate case, CB Thm 10.1.6)

- $\hat{\theta}_{\text{ML}} \xrightarrow{p} \theta$  under regularity conditions (CB Sec 10.6.2):
  - $X_1, \dots, X_n$  are iid;
  - If  $\theta_1 \neq \theta_2$ , then  $f_X(x | \theta_1) \neq f_X(x | \theta_2)$ ;
  - $f_X(x | \theta)$  has a common support for each  $\theta \in \Theta$  and is differentiable with respect to  $\theta$ ;
    - \* Violated by, e.g.,  $\text{Unif}(0, \theta)$ ;
  - true  $\theta$  is an interior point of parameter space  $\Theta$ .

### Example of inconsistent MLE

There are independent  $X_{i1}, X_{i2} \sim \mathcal{N}(\mu_i, \sigma^2)$ ,  $i = 1, \dots, n$ . Then  $\widehat{\sigma}_{\text{ML}}^2$  is NOT consistent for  $\sigma^2$ .

### Examples of consistent MLE even when the regularity conditions are violated

- iid  $X_1, \dots, X_n \sim \text{Ber}(1)$
- iid  $X_1, \dots, X_n \sim \text{Unif}(0, \theta)$

**Take-home exercises (NOT to be submitted; to be potentially covered in labs)**