Lecture 27

Review for the Midterm Test

Chapter 1: Probability and Distributions
Chapter 2: Sampling Distribution
Chapter 3: Point Estimation
Chapter 4: CI Estimation
Appendix C

(i) Midterm Test 2025

- **Time and Date: 7:50am – 9:50am, May 17 (Saturday), 120 minutes**

- **Venue: The Third Teaching Building, Room 107–108 (Class I, 125 students), Room 108 (Class II, 25 students)**

- **Range: Chapters 1–4, Appendix C**

- **Assessment: Midterm test (25%)**

(ii) Distribution of Questions in the Midterm Test 2025

- There are **five** questions in the midterm test with a total of **100** marks.
- **Q1** has **20** sub-questions with **2** marks per sub-question, ranging from Chapters 1 to 4 & Appendix C; Directly giving your answers.
- **Q2–Q3** are in Chapter 2;
- **Q4** is in Chapter 3 and **Q5** is in Chapter 4.

(iii) The Policy of Closed Book Midterm Test

- Please bring one calculator and check the battery.
- Please bring two pens/pencils in case one is not available.
- You can prepare anything on one side of an A4 paper and bring it with you to the test venue.
- You are not allowed to bring any other material (including iPhone/iPad) to the test venue.

0) Important concepts/formulae

0.1 The definitions of quantile and median.

0.2 Given conditional expectation/variance to find the expectation and variance:

$$E(X) = E\{E(X|Y)\},$$

$$\mathrm{Var}(X) = E\{\mathrm{Var}(X|Y)\} + \mathrm{Var}\{E(X|Y)\}.$$

0.3 What are the corresponding supports of the exponential and gamma distributions?

0.4 What is the advantage of (1.53) over (1.46) to define the multivariate normal distribution?

1) Given two conditional densities, to find the marginal densities

$$f_X(x) f_{(Y|X)}(y|x) = f_Y(y) f_{(X|Y)}(x|y)$$

1. **Continuous case**

$$f_X(x) \propto \frac{f_{(X|Y)}(x|y_0)}{f_{(Y|X)}(y_0|x)},$$

$$f_X(x) = \left\{ \int \frac{f_{(Y|X)}(y|x)}{f_{(X|Y)}(x|y)} \mathrm{d}y \right\}^{-1}.$$

$$\Pr(X = x_i)\Pr(Y = y_j|X = x_i)$$
$$= \Pr(Y = y_j)\Pr(X = x_i|Y = y_j)$$

2. Discrete case

$$p_i = \Pr(X = x_i) \propto \frac{\Pr(X = x_i|Y = y_{j0})}{\Pr(Y = y_{j0}|X = x_i)} \hat{=} q_i,$$
$$p_i = \frac{q_i}{\sum_{i'} q_{i'}},$$

$$\Pr(X = x_i) = \left\{ \sum_j \frac{\Pr(Y = y_j|X = x_i)}{\Pr(X = x_i|Y = y_j)} \right\}^{-1}.$$

2) Six Methods to Find the Distribution of the Function of Random Variables (§2.1)

3. **Cumulative distribution function technique (§2.1.1)**

4. **Transformation technique (§2.1.2)**
(a) Monotone transformation

$$g(y) = f(x) \times \left| \mathrm{d}x / \mathrm{d}y \right| . \qquad (2.1)$$

(b) Piecewise monotone transformation

$$g(y) = \sum_{i=1}^{n} f(h_i^{-1}(y)) \times \left|\frac{\mathrm{d}h_i^{-1}(y)}{\mathrm{d}y}\right|. \qquad (2.2)$$

e.g., $X \sim N(0,1)$, then $Y = X^2 \sim \chi^2(1)$.

(c) Bivariate transformation

$$g(y_1, y_2) = f(x_1, x_2) \times \left|\frac{\partial(x_1, x_2)}{\partial(y_1, y_2)}\right|. \qquad (2.3)$$

– Examples 2.8 and 2.9 (p.65–68)

(d) Multivariate transformation

$$g(y_1, \ldots, y_n) = f(x_1, \ldots, x_n) \times \left| \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \right| .$$

5. Moment generating function technique (§2.1.3)

5.1 Expectation technique (T2.5)

5.2 Mixture technique (T3.1)

5.3 SR technique (T3.2)

3) Order Statistics (§2.4, p.81)

6. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F(x)$**. The cdf and pdf of** $X_{(1)} = \min\{X_1, \ldots, X_n\}$

$$\begin{aligned} G_1(x) &= 1 - [1 - F(x)]^n, \\ g_1(x) &= n[1 - F(x)]^{n-1} f(x). \end{aligned}$$

7. Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F(x)$**. The cdf and pdf of** $X_{(n)} = \max\{X_1, \ldots, X_n\}$

$$\begin{aligned} G_n(x) &= [F(x)]^n, \\ g_n(x) &= n[F(x)]^{n-1} f(x). \end{aligned}$$

If $\{X_i\}_{i=1}^n \overset{\text{ind}}{\sim} F_i(\cdot)$**, what about?**

8. The cdf and pdf of $X_{(r)}$

$$G_r(x) = \sum_{i=r}^{n} \binom{n}{i} F^i(x)[1 - F(x)]^{n-i}, \quad (2.21)$$

$$g_r(x) = \frac{n!}{(r-1)!(n-r)!} f(x) F^{r-1}(x)[1 - F(x)]^{n-r}. \quad (2.23)$$

9. The joint pdf of $X_{(1)}, \ldots, X_{(n)}$ is

$$g_{X_{(1)},\ldots,X_{(n)}}\left(x_{(1)}, \ldots, x_{(n)}\right) = n! f_X\left(x_{(1)}\right) \cdots f_X\left(x_{(n)}\right). \tag{2.27}$$

where $x_{(1)} \leqslant \cdots \leqslant x_{(n)}$ and $f_X(\cdot)$ is the density function of the population random variable X, i.e., $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f_X(x)$.

4) Central Limit Theorem (§2.5.5, p.94)

10. **Theorem 2.9** **Let $\{X_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. r.v.'s with common mean μ and common variance $\sigma^2 > 0$. Let**

$$\bar{X}_n = \sum_{i=1}^{n} X_i/n \quad \textbf{and} \quad Y_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma,$$

then $Y_n \xrightarrow{\mathrm{L}} Z$, where $Z \sim N(0,1)$. ||

5) Point Estimation (Chapter 3)

11. Joint density and likelihood function (p.103)

$$L(\boldsymbol{\theta}) = f(\boldsymbol{x};\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i;\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \boldsymbol{\Theta}.$$

– **Example 3.3 (p.108):**

$$f(\boldsymbol{x};\theta) = \begin{cases} \theta^{-n}, & \textbf{if } 0 < x_i \leqslant \theta\textbf{, } i = 1, \ldots, n\textbf{,} \\ 0, & \textbf{elsewhere.} \end{cases}$$

$$L(\theta) = \begin{cases} \theta^{-n}, & \textbf{if } \theta \geqslant x_{(n)} \hat{=} \max\{x_1, \ldots, x_n\}\textbf{,} \\ \\ 0, & \textbf{elsewhere.} \end{cases}$$

12. MLE and mle (p.105)

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L(\boldsymbol{\theta}) = \arg \max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \ell(\boldsymbol{\theta}),$$

then $\boldsymbol{\theta}^* = \boldsymbol{g}(x_1, \ldots, x_n)$ **is called the** *maximum likelihood estimate* **(mle) of** $\boldsymbol{\theta}$ **and** $\hat{\boldsymbol{\theta}} = \boldsymbol{g}(X_1, \ldots, X_n)$ **is called the MLE of** $\boldsymbol{\theta}$**.**

13. Unrestricted MLE

Let

$$\frac{\mathrm{d}\ell(\theta)}{\mathrm{d}\theta} = 0,$$

we can obtain the unrestricted MLE.

- **Example 3.1: Bernoulli(θ)**
- **Example 3.2:** $N(\mu, \sigma^2)$

14. How to find MLEs of parameters in other distributions, e.g.

- $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ **Poisson**(λ),
- $X_i \overset{\text{ind}}{\sim}$ **Binomial**(n_i, θ),
- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} U[\theta_1, \theta_2]$,
- $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exponential}(\beta)$,
- $X_i \overset{\text{ind}}{\sim} \text{Gamma}(n_i, \beta)$.

15. MLE with equality constraints

– **Example 3.6: Multinomial distribution**

16. MLE with inequality constraints

– **Example 3.7: Normal distribution with constraints $a \leqslant \mu \leqslant b$.**

17. Moment estimator
(§3.2): Replace MLE by Moment estimator in Item 14.

18. Bayesian estimator
(§3.3)

19. Efficiency (§3.4.2):

– **19.1 How to calculate the Fisher information (Theorem 3.4, p.132): If** $E[S(\theta)] = 0$, **then**

$$I_n(\theta) = nI(\theta),$$

where

$$\begin{aligned} I(\theta) &= E\left[\left(\frac{\mathrm{d}\log f(X;\theta)}{\mathrm{d}\theta}\right)^2\right] \\ &= E\left[-\frac{\mathrm{d}^2\log f(X;\theta)}{\mathrm{d}\theta^2}\right]. \end{aligned}$$

– 19.2 How to calculate the CR lower bound of $T(X_1, \ldots, X_n) = T(\mathbf{x})$, which is an unbiased estimator of $\tau(\theta)$ being an arbitrary function of θ (Theorem 3.3, p.130):

$$\frac{\{\tau'(\theta)\}^2}{I_n(\theta)} \leqslant \mathrm{Var}(T(\mathbf{x})). \tag{3.20}$$

– 19.3 In particular, when $\tau(\theta) = \theta$ and $T(\mathbf{x}) = \hat{\theta}$, then the CR lower bound of $\hat{\theta}$ is

$$\frac{1}{I_n(\theta)} \leqslant \mathrm{Var}(\hat{\theta}). \tag{3.22}$$

– 19.4 How to calculate the efficiency of an unbiased estimator $\hat{\theta}$ for θ (p. 136):

$$\mathbf{Eff}_{\hat{\theta}}(\theta) = \frac{1/I_n(\theta)}{\mathrm{Var}(\hat{\theta})}. \tag{3.26}$$

20. Sufficiency (§3.4.3):

– **20.1 The definition of a sufficient statistic.**

– **20.2 Use Theorem 3.5 (Factorization Theorem, p.141) to find a sufficient statistic $T(\mathbf{x})$ for θ:**

$$f(x_1, \ldots, x_n; \theta) = f(\boldsymbol{x}; \theta) = g(T(\boldsymbol{x}); \theta) \times h(\boldsymbol{x}), \tag{3.27}$$

– 20.3 Jointly sufficient statistics. **For example, let** $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Gamma}(\alpha, \beta)$**, the joint density is**

$$f(x; \alpha, \beta) = \prod_{i=1}^{n} \frac{\beta^{\alpha}}{\Gamma(\alpha)} x_i^{\alpha-1} e^{-\beta x_i},$$

then

$$\textstyle\prod_{i=1}^{n} X_i \quad \textbf{and} \quad \sum_{i=1}^{n} X_i$$

are jointly sufficient statistics of (α, β)**. So the distribution of**

$$(X_1, \ldots, X_n) | (\textstyle\prod_{i=1}^{n} X_i = t_1, \ \sum_{i=1}^{n} X_i = t_2)$$

does not depends on (α, β)**.**

– **20.4** **Let** $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} N(\mu, 1)$**, find the joint distribution of**

$$(X_1, \ldots, X_n) | (\sum_{i=1}^{n} X_i = t)$$

– **20.5** **Let** $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \textbf{Bernoulli}(\theta)$**, find the joint distribution of**

$$(X_1, \ldots, X_n) | (\sum_{i=1}^{n} X_i = t)$$

– **20.6** **Let** $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} \textbf{Poisson}(\lambda)$**, find the joint distribution of**

$$(X_1, \ldots, X_n) | (\sum_{i=1}^{n} X_i = t)$$

21. Data reduction

– Let $\{X_i\}_{i=1}^n \overset{\text{iid}}{\sim} f(x;\theta)$. To estimate the θ, first we need to find a sufficient statistic $T(X_1,\ldots,X_n) = T(\mathbf{x}) = T$ for θ.

– Then the MLE, moment estimator, Bayesian estimator of θ are functions of T, say $g_1(T), g_2(T), g_3(T)$.

– A pivotal quantity is a function of both T and θ, i.e., $g_4(T,\theta)$.

– Finally, the lower limit and upper limit of the CI of θ are also functions of T, say $[g_5(T), g_6(T)]$.

22. Completeness (§3.4.4):

– **22.1 How to prove that a statistic $T(X_1, \ldots, X_n)$ is complete for θ (Definition 3.9, p.147):**

The statistic T is said to be *complete* if for any $h(t)$,

$$E[h(T)] = 0 \quad \textbf{for all } \theta \in \Theta$$

implies that $h(T) = 0$ with probability 1.

– 22.2 How to find the unique UMVUE for θ (Theorem 3.7, Lehmann-Scheffé Theorem, p.149):

Step 1: To prove that $T(\mathbf{x})$ is sufficient for θ;
Step 2: To prove that $T(\mathbf{x})$ is complete for θ;
Step 3: To find a function of T, say, $g(T)$, which is an unbiased estimator of $\tau(\theta)$.

Then $g(T)$ is the unique UMVUE for $\tau(\theta)$.

23. Limiting Properties of MLE (§3.5):

$$[nI(\theta)]^{1/2}(\hat{\theta}_n - \theta) \xrightarrow{\mathrm{L}} Z \sim N(0,1) \quad \textbf{as} \quad n \to \infty. \tag{3.34}$$

– **23.1 Let $g(\cdot)$ be a function and its first derivative $g'(\cdot)$ exist. Then, using the first-order Taylor expansion, we have**

$$g(\hat{\theta}_n) \approx g(\theta) + (\hat{\theta}_n - \theta)g'(\theta)$$

$$\sim N(g(\theta),\ [g'(\theta)]^2 \mathrm{Var}(\hat{\theta}_n)),$$

i.e.

$$\frac{\sqrt{nI(\theta)}[g(\hat{\theta}_n) - g(\theta)]}{g'(\theta)} \dot{\sim} N(0,1). \tag{3.35}$$

– **23.2** **Let** $X_1, \ldots, X_n \overset{\text{iid}}{\sim}$ **Bernoulli**(θ), **then the MLE of** θ **is** $\hat{\theta}_n = (1/n)\sum_{i=1}^{n} X_i$. **Let** $g(x) = \arcsin\sqrt{x}$, **then** $g'(x) = \frac{1}{2\sqrt{x(1-x)}}$. **Note** $\mathrm{Var}(\hat{\theta}_n) = \mathrm{Var}(\bar{X}) = \theta(1-\theta)/n$ **so that**

$$[g'(\theta)]^2 \mathrm{Var}(\hat{\theta}_n) = \frac{1}{4n}$$

is a constant. From (3.35), we have

$$\frac{\arcsin\sqrt{\bar{X}} - \arcsin\sqrt{\theta}}{1/\sqrt{4n}} \dot{\sim} N(0,1),$$

which results in a CI for θ.

6) CI Estimation (Chapter 4)

24. **Upper α-th quantile points:**

$$
\begin{aligned}
\alpha &= \Pr\{Z > z_\alpha\}, \quad Z \sim N(0,1), \\
\alpha &= \Pr\{t(n) > t(\alpha, n)\}, \\
\alpha &= \Pr\{\chi^2(n) > \chi^2(\alpha, n)\}, \\
\alpha &= \Pr\{F(n,m) > F(\alpha, n, m)\}.
\end{aligned}
$$

25. **Pivotal quantity (Definition 4.1, p.164)**

26. The CI of normal mean (§4.2):

– **26.1** **If** σ_0^2 **is known, use the pivotal quantity**

$$\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma_0} \sim N(0,1)$$

to construct a $100(1-\alpha)\%$ **CI of** μ **as follows:**

$$\left[\bar{X} - z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}},\ \bar{X} + z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\right] \tag{4.4}$$

– **26.2 If σ^2 is unknown, use the pivotal quantity**

$$T = \frac{\sqrt{n}(\bar{X} - \mu)}{S} \sim t(n-1)$$

to construct a $100(1-\alpha)\%$ **CI of** μ **as follows:**

$$\left[\bar{X} - t(\alpha/2, n-1)\frac{S}{\sqrt{n}},\ \bar{X} + t(\alpha/2, n-1)\frac{S}{\sqrt{n}}\right] \tag{4.6}$$

27. The CI of normal variance (§4.4):

– **27.1 If $\mu = \mu_0$ is known, use the pivotal quantity**

$$\sum_{i=1}^{n} \frac{(X_i - \mu_0)^2}{\sigma^2} \sim \chi^2(n)$$

to construct a $100(1-\alpha)\%$ CI of σ^2 as follows:

$$\left[\frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\chi^2(\alpha/2, n)}, \frac{\sum_{i=1}^{n}(X_i - \mu_0)^2}{\chi^2(1-\alpha/2, n)}\right], \tag{4.14}$$

– **27.2 If μ is unknown, use the pivotal quantity**

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

to construct a $100(1-\alpha)\%$ **CI of** σ^2 **as follows:**

$$\left[\frac{(n-1)S^2}{\chi^2(\alpha/2, n-1)}, \frac{(n-1)S^2}{\chi^2(1-\alpha/2, n-1)}\right], \quad (4.15)$$

28. Large-Sample Confidence Intervals (§4.6, three methods)

7) Appendix C

Please review C.1 and C.2

29. All questions in Assignments 1–4.

30. All questions in Tutorials.

End of Lecture 27