

# KEYPOINT SUMMARY I

## I. RANDOM VARIABLES

Random variable: a random variable  $X$  is a function from a sample space  $S$  into  $\mathbb{R}$  with certain probabilities.

### A. Discrete Random Variables

1. Discrete random variable:  $\mathbb{P}(X = x_i) = p_i$
2. Bivariate random variable:  $\mathbb{P}(X = x_i, Y = y_j) = p_{ij}$
3. Marginal probability:

$$\mathbb{P}(X = x_i) = \sum_j \mathbb{P}(X = x_i, Y = y_j)$$

4. Conditional probability:

$$\mathbb{P}(X = x_i | Y = y_j) = \frac{\mathbb{P}(X = x_i, Y = y_j)}{\mathbb{P}(Y = y_j)}$$

5. Independence:

$$\mathbb{P}(X = x_i, Y = y_j) = \mathbb{P}(X = x_i)\mathbb{P}(Y = y_j)$$

### B. Continuous Random Variables

1. Continuous random variable:

$$\mathbb{P}(x_1 \leq X \leq x_2) = \int_{x_1}^{x_2} f(x)dx$$

where  $f(x)$  is called the density function.

2. Cumulative distribution function:

$$F(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^x f(t)dt$$

3. Bivariate continuous random variable:

$$\mathbb{P}(x_1 \leq X \leq x_2, y_1 \leq Y \leq y_2) = \int_{y_1}^{y_2} \int_{x_1}^{x_2} f(x, y)dx dy$$

where  $f(x, y)$  is called the joint density function.  
For more general case,

$$\mathbb{P}((X, Y) \in S) = \iint_S f(x, y)dx dy$$

If region  $S$  can be characterized by  $a \leq x \leq b$ ,  $h(x) \leq y \leq g(x)$ , we have

$$\iint_S f(x, y)dx dy = \int_a^b \left[ \int_{h(x)}^{g(x)} f(x, y)dy \right] dx$$

4. Marginal probability (marginal density):

$$\begin{aligned} \mathbb{P}(x_1 \leq X \leq x_2) &= \mathbb{P}(x_1 \leq X \leq x_2, -\infty \leq Y \leq +\infty) \\ &= \int_{-\infty}^{+\infty} f(x, y)dy \end{aligned}$$

5. Conditional density:

$$f(x, y|S) = \begin{cases} \frac{f(x, y)}{\mathbb{P}[(x, y) \in S]} & \text{for } (x, y) \in S \\ 0 & \text{otherwise} \end{cases}$$

If region  $S$  can be characterized by  $a \leq x \leq b$ ,  $h(x) \leq y \leq g(x)$ , the condition density of  $X$  is

$$f(x|S) = \begin{cases} \frac{\int_{h(x)}^{g(x)} f(x, y)dy}{\mathbb{P}[(x, y) \in S]} & \text{for } a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

Especially, the conditional density of  $X$  given  $y_1 \leq Y \leq y_2$  is defined by

$$f(x|y_1 \leq Y \leq y_2) = \frac{\int_{y_1}^{y_2} f(x, y)dy}{\int_{-\infty}^{+\infty} \left[ \int_{y_1}^{y_2} f(x, y)dy \right] dx}$$

The conditional density of  $X$  given  $Y = y_1$  is

$$f(x|Y = y_1) = \frac{f(x, y_1)}{f(y_1)}$$

6. Independence:  $f(x, y) = f(x)f(y)$
7. Random variable transformation:  $Y = \phi(X)$

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y < y) = \mathbb{P}(\phi(X) < y) \\ &= \mathbb{P}(X < \phi^{-1}(y)) \\ &= F_X(\phi^{-1}(y)) \\ f_Y(y) &= \frac{dF_Y(y)}{dy} \end{aligned}$$

If  $\phi$  is a strictly monotonic differentiable function,

$$f_Y(y) = f[\phi^{-1}(y)] \left| \frac{d\phi^{-1}(y)}{dy} \right|$$

## II. MOMENTS

### A. Expectation

1. Expected value (discrete):  $\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \mathbb{P}(x_i)$

2. Expected value (continuous):  $\mathbb{E}(X) = \int_{-\infty}^{\infty} xf(x)dx$
3. Median:  $m$  such that  $\mathbb{P}(X \leq m) = \frac{1}{2}$
4. Expectation of transformation of random variable:

$$\mathbb{E}[\phi(X)] = \sum_{i=1}^{\infty} \phi(x_i)\mathbb{P}(x_i)$$

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{\infty} \phi(x)f(x)dx$$

$$\mathbb{E}[\phi(X, Y)] = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \phi(x_i, y_j)\mathbb{P}(x_i, y_j)$$

$$\mathbb{E}[\phi(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \phi(x, y)f(x, y)dxdy$$

5. Properties of expectation:

- $\mathbb{E}(\alpha) = \alpha$
- $\mathbb{E}(\alpha X + \beta Y) = \alpha\mathbb{E}(X) + \beta\mathbb{E}(Y)$
- If  $X, Y$  are *independent*,  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$

## B. Conditional Expectation

1. Conditional expectation:

- $\mathbb{E}[\phi(X, Y)|X] = \sum_{j=1}^{\infty} \phi(X, y_j)\mathbb{P}(y_j|X)$
- $\mathbb{E}[\phi(X, Y)|X] = \int_{-\infty}^{\infty} \phi(X, y)f(y|X)dy$
- Note  $\mathbb{E}[\phi(X, Y)|X]$  is a function of  $X$ .

2. Law of iterated expectations:

- $\mathbb{E}[\phi(X, Y)] = \mathbb{E}[\mathbb{E}[\phi(X, Y)|X]]$
- $\mathbb{E}[\phi(X)|X] = \phi(X)$
- $\mathbb{E}[\phi(X)Y|X] = \phi(X)\mathbb{E}[Y|X]$

## C. Variance and Covariance

1. Variance:

- $\text{Var}(X) = \mathbb{E}[X - \mathbb{E}(X)]^2$
- $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$
- $\text{Var}(\alpha X + \beta) = \alpha^2 \text{Var}(X)$
- $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$
- $X, Y$  independent,  $\text{Var}(X_1 + \dots + X_n) = \sum_{i=1}^n \text{Var}(X_i)$

2. Covariance:

- $\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))]$
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$
- $X, Y$  independent  $\not\Rightarrow \text{Cov}(X, Y) = 0$

3. Correlation:

- $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}$
- $\text{Corr}(\alpha X, \beta Y) = \text{Corr}(X, Y)$
- $-1 \leq \text{Corr}(X, Y) \leq 1$

## D. Random Variable Predictor

The best *predictor* of  $Y$  based on  $X$  is the predicting function  $\phi(X)$  that minimizes  $\mathbb{E}[Y - \phi(X)]^2$ .

1. The best predictor of  $Y$  based on  $X$  is  $\mathbb{E}[Y|X]$ .

$$\begin{aligned} \mathbb{E}[Y - \phi(X)]^2 &= \mathbb{E}[Y - \mathbb{E}(Y|X) + \mathbb{E}(Y|X) - \phi(X)]^2 \\ &= \mathbb{E}[Y - \mathbb{E}(Y|X)]^2 + \mathbb{E}[\mathbb{E}(Y|X) - \phi(X)]^2 \\ &\quad + \underbrace{2\mathbb{E}[(Y - \mathbb{E}(Y|X))(\mathbb{E}(Y|X) - \phi(X))]}_{=0} \\ &\geq \mathbb{E}[Y - \mathbb{E}(Y|X)]^2 \end{aligned}$$

2. The best linear predictor  $\phi(X) = \alpha + \beta X$  is given by

$$\beta^* = \frac{\text{Cov}(X, Y)}{\text{Var}(X)}, \alpha^* = \mathbb{E}(Y) - \beta^* \mathbb{E}(X)$$

Minimize  $\mathbb{E}(Y - \alpha - \beta X)^2$ , the FOC gives  $\alpha^*, \beta^*$ .

## III. NORMAL RANDOM VARIABLES

### A. Normal Distribution

1. Normal density function:

$$f(x) = \frac{1}{\sqrt{1\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right]$$

Gaussian integral:  $\int_{-\infty}^{+\infty} e^{-x^2} dx = \sqrt{\pi}$ .

2.  $\mathbb{E}(X) = \mu, \text{Var}(X) = \sigma^2$

Note:  $\int x e^{-\frac{1}{2}x^2} dx = -e^{-\frac{1}{2}x^2}$

3. If  $X \sim N(\mu, \sigma^2)$ ,  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$

4. If  $X \sim N(\mu, \sigma^2)$ , and  $Y = \alpha + \beta X$ , then we have  $Y \sim N(\alpha + \beta\mu, \beta^2\sigma^2)$ .

5. If  $X_1, X_2$  are *independent* and  $X_1 \sim N(\mu_1, \sigma_1^2)$ ,  $X_2 \sim N(\mu_2, \sigma_2^2)$ , then we have  $X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$ . If  $X_1, X_2$  are not independent,  $X_1 + X_2$  is *not* necessarily normal.

## B. Bivariate Normal Distribution

1. Multivariate normal density function:

$$f(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right]$$

where  $\Sigma$  is the variance-covariance matrix.

2. If  $(X, Y)$  are bivariate normally distributed

$$(X, Y) \sim N \left( \begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right)$$

Then we have

$$X \sim N(\mu_X, \sigma_X^2)$$

$$Y \sim N(\mu_Y, \sigma_Y^2)$$

$$X|Y \sim N \left( \mu_X + \rho \frac{\sigma_X}{\sigma_Y} (Y - \mu_Y), \sigma_X^2 (1 - \rho^2) \right)$$

$$Y|X \sim N \left( \mu_Y + \rho \frac{\sigma_Y}{\sigma_X} (X - \mu_X), \sigma_Y^2 (1 - \rho^2) \right)$$

$$\alpha X + \beta Y \sim N(\alpha\mu_X + \beta\mu_Y, \alpha^2\sigma_X^2 + \beta^2\sigma_Y^2 + 2\alpha\beta\rho\sigma_X\sigma_Y)$$

## IV. LARGE SAMPLE THEORY

### A. Modes of Convergence

1. Convergence in probability:  $X_n \xrightarrow{p} X$  if

$$\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}(|X_n - X| < \epsilon) = 1$$

The random variable  $Y = X_n - X \rightarrow 0$  as  $n \rightarrow \infty$ .

2. Convergence in distribution:  $X_n \xrightarrow{d} X$  if

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

where  $F_n$  and  $F$  are the cumulative distribution functions of random variables  $X_n$  and  $X$  respectively.

3.  $X_n \xrightarrow{p} X \not\Rightarrow X_n \xrightarrow{d} X$

4. Continuous mapping theorem:

Let  $g(\cdot)$  be a *continuous* function. Then

$$\cdot X_n \xrightarrow{p} X \implies g(X_n) \xrightarrow{p} g(X)$$

$$\cdot X_n \xrightarrow{d} X \implies g(X_n) \xrightarrow{d} g(X)$$

5. Slutsky Theorem: If  $X_n \xrightarrow{d} X$  and  $Y_n \xrightarrow{p} \alpha$ , then

$$\cdot X_n + Y_n \xrightarrow{d} X + \alpha$$

$$\cdot X_n Y_n \xrightarrow{d} \alpha X$$

$$\cdot X_n / Y_n \xrightarrow{d} X / \alpha \quad (\alpha \neq 0)$$

## B. Law of Large Numbers

1. Chebyshev's inequality: suppose  $g(\cdot)$  is a *nonnegative* continuous function, then we have

$$\mathbb{P}[g(X) \geq \epsilon] \leq \frac{\mathbb{E}[g(X)]}{\epsilon}, \forall \epsilon > 0$$

2. Law of Large Numbers: Let  $\{X_i\}$  be *i.i.d.* random variables with  $\mathbb{E}(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2 < \infty$ . Define

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

Then  $\bar{X}_n \xrightarrow{p} \mu$  as  $n \rightarrow \infty$ .

$$\begin{aligned} \mathbb{P}(|\bar{X}_n - \mu| > \epsilon) &= \mathbb{P}(|\bar{X}_n - \mu|^2 > \epsilon^2) \\ &\leq \frac{\mathbb{E}(\bar{X}_n - \mu)^2}{\epsilon^2} \\ &= \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 \end{aligned}$$

## C. Moment Generating Functions

1. The moment generating function of  $X$ :

$$M_X(t) = \mathbb{E}(e^{Xt}) = 1 + \mathbb{E}(X)t + \mathbb{E}(X^2)\frac{t^2}{2!} + \dots$$

2. Given  $M_X(t)$ ,  $\mathbb{E}(X^n) = \left. \frac{d^n M_X(t)}{dt^n} \right|_{t=0}$

3. If  $M_X(t) = M_Y(t)$ , then  $F_X(x) = F_Y(x)$ .

4. If  $X, Y$  independent,  $M_{X+Y}(t) = M_X(t)M_Y(t)$

5. If  $\lim_{n \rightarrow \infty} M_{X_n}(t) = M_X(t)$ , then  $X_n \xrightarrow{d} X$ .

6. If  $X \sim N(\mu, \sigma^2)$ ,  $M_X(t) = \exp(\mu + \frac{1}{2}\sigma^2 t^2)$ .

## D. Central Limit Theorem

Let  $\{X_i\}$  be *i.i.d.* random variables with  $\mathbb{E}(X_i) = \mu$  and  $\text{Var}(X_i) = \sigma^2 < \infty$ . Define  $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ . Then

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

*Proof.* Let  $Y = (\bar{X}_n - \mu)/\sigma$ .

$$(i) \ M_{\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}}(t) \stackrel{\text{iid}}{=} M_Y \left( n^{-\frac{1}{2}} t \right)^n$$

$$(ii) \ M_Y \left( n^{-\frac{1}{2}} t \right)^n \stackrel{\text{Taylor}}{=} 1 + \frac{t^2}{2n} + o(n^{-1})$$

$$(iii) \ M_Y \left( n^{-\frac{1}{2}} t \right)^n = \left[ \left( 1 + \frac{t^2}{2n} + o(n^{-1}) \right)^{\frac{2n}{t^2}} \right]^{\frac{t^2}{2}} \rightarrow e^{\frac{t^2}{2}}$$