南方科技大学 STA203《概率论基础》期末互助课堂知识点梳理笔记

互助课堂导生：师珑天 (Longtian Shi)

个人主页：https://statslt.github.io/

基于 2025 年秋张卓松副教授讲义 Lecture Notes 1-9

2025 年 12 月 19 日

目录

1 概率论公理体系与基本概念 (Lecture 1-4)

1.1 概率空间与基本公理

概率空间

概率空间由三元组 $(\Omega, \mathcal{F}, \mathbb{P})$ 构成：

- **样本空间** Ω：所有可能结果的集合
- **事件域** $\mathcal{F}$：Ω 的子集构成的 σ-代数，满足：
  1. $\Omega \in \mathcal{F}$
  2. 若 $A \in \mathcal{F}$，则 $A^c \in \mathcal{F}$
  3. 若 $A_1, A_2, \ldots \in \mathcal{F}$，则 $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$
- **概率测度** $\mathbb{P}$：映射 $\mathbb{P} : \mathcal{F} \to [0, 1]$，满足：
  1. 非负性：$\mathbb{P}(A) \geq 0$，$\forall A \in \mathcal{F}$
  2. 规范性：$\mathbb{P}(\Omega) = 1$
  3. 可列可加性：若 $A_1, A_2, \ldots$ 互不相容，则 $\mathbb{P}\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i)$

博雷尔域 $\mathcal{B}(\mathbb{R})$

$\mathcal{B}(\mathbb{R})$ 是包含所有区间 $(a, b]$ 的最小 σ-代数。包含：

- 单点集：$\{a\} = \bigcap_{n=1}^{\infty} (a - \frac{1}{n}, a]$
- 开区间：$(a, b) = (a, b] \setminus \{b\}$
- 闭区间：$[a, b] = (a, b] \cup \{a\}$

1.2 条件概率与独立性

条件概率

对于事件 A, B，且 $\mathbb{P}(B) > 0$，定义：

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

乘法公式

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\mathbb{P}(A_3|A_1 \cap A_2) \cdots \mathbb{P}(A_n|A_1 \cap \cdots \cap A_{n-1})$$

全概率公式

若 $\{B_i\}_{i=1}^{\infty}$ 是样本空间的划分（互不相容且 $\bigcup_i B_i = \Omega$），则：

$$\mathbb{P}(A) = \sum_{i=1}^{\infty} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

贝叶斯公式

$$\mathbb{P}(B_i|A) = \frac{\mathbb{P}(A|B_i)\mathbb{P}(B_i)}{\sum_{j=1}^{\infty} \mathbb{P}(A|B_j)\mathbb{P}(B_j)}$$

独立性

- 事件 A, B 独立：$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- 事件 $A_1, \ldots, A_n$ 相互独立：对任意子集 $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$，有

$$\mathbb{P}(A_{i_1} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1}) \cdots \mathbb{P}(A_{i_k})$$

互斥 vs 独立

关键区别:

- 互斥：$A \cap B = \emptyset$，$\mathbb{P}(A \cap B) = 0$
- 独立：$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$
- **除非 $\mathbb{P}(A) = 0$ 或 $\mathbb{P}(B) = 0$，否则互斥事件一定不独立！**

2 随机变量及其分布 (重点内容)

2.1 随机变量的形式化定义

随机变量

设 $(\Omega, \mathcal{F}, \mathbb{P})$ 为概率空间。函数 $X : \Omega \to \mathbb{R}$ 称为随机变量，如果对任意 $B \in \mathcal{B}(\mathbb{R})$，有：

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}$$

即原像可测。

2.2 分布函数及其性质

累积分布函数 CDF

$F_X(x) = \mathbb{P}(X \leq x) = \mathbb{P}(\{\omega : X(\omega) \leq x\})$，$x \in \mathbb{R}$

CDF 的性质

任意分布函数 F 满足：

1. 单调不减：$x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
2. 右连续：$\lim_{h \to 0^+} F(x+h) = F(x)$
3. $\lim_{x \to -\infty} F(x) = 0$，$\lim_{x \to \infty} F(x) = 1$

反之，满足这三条性质的函数一定是某个随机变量的分布函数。

2.3 离散型随机变量

概率质量函数 PMF

若 X 取值于可数集 $S = \{x_1, x_2, \ldots\}$，则其 PMF 为：

$$p(x_i) = \mathbb{P}(X = x_i) \geq 0, \quad \sum_i p(x_i) = 1$$

约定：本笔记中采用的几何分布支持为 $k = 1, 2, \ldots$（首次成功前的试验次数），

$$\mathbb{P}(X = k) = (1-p)^{k-1} p, \quad 0 < p < 1.$$

（若采用支持 $k = 0, 1, \ldots$，需相应改为 $(1-p)^k p$ 并调整期望等式。）

表 1: 常见离散分布及其性质

分布	**PMF** $p(k)$	**支撑集/参数**	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
伯努利 Bernoulli(p)	$p^k(1-p)^{1-k}$	$k \in \{0,1\}$	p	$p(1-p)$
二项 Binomial(n,p)	$\binom{n}{k}p^k(1-p)^{n-k}$	$k=0,\ldots,n$	np	$np(1-p)$
几何 Geometric(p)	$(1-p)^{k-1}p$	$k=1,2,\ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
泊松 Poisson(λ)	$e^{-\lambda}\frac{\lambda^k}{k!}$	$k=0,1,\ldots$	λ	λ

几何分布的无记忆性

若 $X \sim \mathrm{Geometric}(p)$，则对任意 $m,n \geq 1$：

$$\mathbb{P}(X > m+n | X > m) = \mathbb{P}(X > n)$$

这是唯一具有离散无记忆性的分布。

2.4 连续型随机变量

概率密度函数 PDF

若存在非负可积函数 $f(x)$，使得对任意 $B \in \mathcal{B}(\mathbb{R})$：

$$\mathbb{P}(X \in B) = \int_B f(x)dx$$

则称 X 为连续型随机变量，$f(x)$ 为其 PDF。

PDF 的性质

1. $f(x) \geq 0$ 对几乎所有 x 成立
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $\mathbb{P}(a < X \leq b) = \int_a^b f(x)dx$
4. 在 f 的连续点处，$F'(x) = f(x)$

指数分布的无记忆性

若 $X \sim \mathrm{Exp}(\lambda)$，则对任意 $s,t > 0$：

$$\mathbb{P}(X > s+t | X > t) = \mathbb{P}(X > s) = e^{-\lambda s}$$

这是唯一具有连续无记忆性的分布。

表 2: 常见连续分布及其性质

分布	**PDF** $f(x)$	**支持集/参数**	$\mathbb{E}[X]$	$\mathrm{Var}(X)$
均匀 Uniform(a,b)	$\frac{1}{b-a}$	$x\in[a,b], a<b$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
指数 Exp(λ)	$\lambda e^{-\lambda x}$	$x\ge 0, \lambda>0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
正态 Normal(μ,σ^2)	$\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x\in\mathbb{R}, \mu\in\mathbb{R}, \sigma>0$	μ	σ^2

2.5 随机变量函数的分布

设 X 为随机变量，$Y=g(X)$。求 Y 的分布有以下方法：

1. **CDF 法（通用）**：

$$F_Y(y)=\mathbb{P}(Y\le y)=\mathbb{P}(g(X)\le y)=\int_{\{x:g(x)\le y\}} f_X(x)dx$$

然后对 y 求导得 $f_Y(y)$。

2. **变换公式法（g 严格单调）**：若 g 严格单调可微，反函数 $h=g^{-1}$，则

$$f_Y(y)=f_X(h(y))\cdot|h'(y)|$$

若 g 在某区间内不是单调的，应把满足 $g(x)=y$ 的所有根 x_i 加和：

$$f_Y(y)=\sum_{x_i:g(x_i)=y} f_X(x_i)\cdot\frac{1}{|g'(x_i)|}.$$

3. **特殊变换公式**：

- $Y=X^2$：$f_Y(y)=\frac{1}{2\sqrt{y}}[f_X(\sqrt{y})+f_X(-\sqrt{y})], \quad y>0$
- $Y=|X|$：$f_Y(y)=f_X(y)+f_X(-y), \quad y>0$

均匀分布的平方

设 $X\sim U(0,1)$，求 $Y=X^2$ 的分布。

解：对于 $0\le y\le 1$，

$$F_Y(y)=\mathbb{P}(X^2\le y)=\mathbb{P}(X\le\sqrt{y})=\sqrt{y}$$

求导得：

$$f_Y(y)=\frac{1}{2\sqrt{y}}, \quad 0<y<1$$

这是 Beta 分布的特例：$Y\sim\mathrm{Beta}(\frac{1}{2},1)$。

2.6 期望、方差与高阶矩

期望

- 离散型：$\mathbb{E}[X] = \sum_x x \cdot p(x)$（要求绝对收敛）
- 连续型：$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$（要求绝对收敛）

期望的性质

1. 线性性：$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y]$
2. 单调性：若 $X \leq Y$ a.s.，则 $\mathbb{E}[X] \leq \mathbb{E}[Y]$
3. 函数期望：$\mathbb{E}[g(X)] = \begin{cases} \sum_x g(x)p(x) & \text{离散} \\ \int_{-\infty}^{\infty} g(x)f(x)dx & \text{连续} \end{cases}$

方差与矩

- 方差：$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
- k 阶原点矩：$\mathbb{E}[X^k]$
- k 阶中心矩：$\mathbb{E}[(X - \mathbb{E}[X])^k]$

方差的性质

1. $\mathrm{Var}(X) \geq 0$
2. $\mathrm{Var}(aX + b) = a^2\mathrm{Var}(X)$
3. 若 X, Y 独立，则 $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
4. 更一般地：$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) + 2\sum_{i<j} \mathrm{Cov}(X_i, X_j)$

3 多维随机变量与联合分布 (Lecture 6)

3.1 联合分布与边缘分布

联合 CDF

$F_{X,Y}(x,y) = \mathbb{P}(X \le x, Y \le y)$，必须满足：

1. **矩形不等式 (2-increasing)**：对任意 $x_1 < x_2, y_1 < y_2$，有

$$F(x_2,y_2) - F(x_1,y_2) - F(x_2,y_1) + F(x_1,y_1) \ge 0$$

（该式等于 $\mathbb{P}(x_1 < X \le x_2, y_1 < Y \le y_2)$）

2. 对每个变量单调不减，且关于每个变量右连续
3. $\lim_{x\to-\infty} F(x,y) = 0$，$\lim_{y\to-\infty} F(x,y) = 0$
4. $\lim_{x,y\to\infty} F(x,y) = 1$

联合密度

若存在 $f(x,y) \ge 0$，使得对任意 $D \subseteq \mathbb{R}^2$：

$$\mathbb{P}((X,Y) \in D) = \iint_D f(x,y)dxdy$$

则 $f(x,y)$ 为联合密度函数。

边缘密度

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)dy, \quad f_Y(y) = \int_{-\infty}^{\infty} f(x,y)dx$$

3.2 独立性

随机变量的独立性

X, Y 独立 $\iff$ 对任意 $A, B \in \mathcal{B}(\mathbb{R})$:

$$\mathbb{P}(X \in A, Y \in B) = \mathbb{P}(X \in A)\mathbb{P}(Y \in B)$$

等价条件:

- $F_{X,Y}(x,y) = F_X(x)F_Y(y)$
- $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ (连续情形)
- $p_{X,Y}(x,y) = p_X(x)p_Y(y)$ (离散情形)

独立性的判别法

若联合密度可分解为 $f(x,y) = g(x)h(y)$,且定义域为矩形区域(即 x, y 取值范围独立),则 X 与 Y 独立。否则通常不独立。

非独立的特殊情况

设 (X,Y) 在单位圆内均匀分布: $f(x,y) = \begin{cases} \frac{1}{\pi}, & x^2+y^2 \leq 1 \\ 0, & \text{否则} \end{cases}$ 虽然密度函数看似" 可分解" (为常数),但定义域 $x^2+y^2 \leq 1$ 耦合了 x 和 y,故 X 与 Y 不独立。

3.3 条件分布

条件密度

给定 $Y = y$ 下 X 的条件密度:

$$f_{X|Y}(x|y) = \frac{f(x,y)}{f_Y(y)}, \quad \text{若} f_Y(y) > 0$$

条件期望

$$\mathbb{E}[X|Y=y] = \int_{-\infty}^{\infty} x f_{X|Y}(x|y)dx$$

$\mathbb{E}[X|Y]$ 是一个随机变量,其取值为 $\mathbb{E}[X|Y=y]$ 当 $Y = y$。

全期望公式（Tower Property）

$$\mathbb{E}[\mathbb{E}[X|Y]] = \mathbb{E}[X]$$

即 $\int \mathbb{E}[X|Y=y]f_Y(y)dy = \mathbb{E}[X]$

全方差公式

$$\mathrm{Var}(X) = \mathbb{E}[\mathrm{Var}(X|Y)] + \mathrm{Var}(\mathbb{E}[X|Y])$$

3.4 协方差与相关系数

协方差

$$\mathrm{Cov}(X,Y) = \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

相关系数

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

满足 $-1 \le \rho_{XY} \le 1$

相关系数的性质

1. $|\rho_{XY}| = 1 \iff$ 存在 a, b 使得 $Y = aX + b$ a.s.
2. 若 X, Y 独立，则 $\rho_{XY} = 0$（独立 $\Rightarrow$ 不相关）
3. 反之不成立：不相关 $\nRightarrow$ 独立

经典反例：不相关但不独立

设 $X \sim N(0,1)$，$Y = X^2$。计算：

$$\begin{gathered}
\mathbb{E}[X] = 0, \quad \mathbb{E}[Y] = \mathbb{E}[X^2] = 1 \\
\mathbb{E}[XY] = \mathbb{E}[X^3] = 0 \quad \text{(奇函数，对称性)} \\
\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0
\end{gathered}$$

故 $\rho_{XY} = 0$，但不独立：Y 完全由 X 决定。

3.5 随机向量的变换

设 (X, Y) 有联合密度 $f_{XY}(x, y)$，变换 $(U, V) = (g_1(X, Y), g_2(X, Y))$ 可逆，雅可比行列式 $J = \frac{\partial(x,y)}{\partial(u,v)} \neq 0$，则：

$$f_{UV}(u, v) = f_{XY}(x(u, v), y(u, v))|J|$$

和与差的分布

设 $U = X + Y$，$V = X - Y$，则：

$$f_{UV}(u, v) = \frac{1}{2} f_{XY}\left(\frac{u+v}{2}, \frac{u-v}{2}\right)$$

特别地，若 X, Y 独立，则 U 的边际密度为卷积：

$$f_U(u) = \int_{-\infty}^{\infty} f_X\left(\frac{u+v}{2}\right) f_Y\left(\frac{u-v}{2}\right) \frac{1}{2} dv = \int_{-\infty}^{\infty} f_X(z) f_Y(u-z) dz$$

3.6 和的分布与卷积

卷积公式

若 X, Y 独立连续，$Z = X + Y$，则：

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx = (f_X * f_Y)(z)$$

表 3: 独立随机变量和的可加性

X 的分布	Y 的分布	$X + Y$ 的分布
Binomial(n, p)	Binomial(m, p)	Binomial$(n + m, p)$
Poisson(λ_1)	Poisson(λ_2)	Poisson$(\lambda_1 + \lambda_2)$
Normal(μ_1, σ_1^2)	Normal(μ_2, σ_2^2)	Normal$(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$
Gamma(α_1, λ)	Gamma(α_2, λ)	Gamma$(\alpha_1 + \alpha_2, \lambda)$
Exp(λ)	Exp(λ)	Gamma$(2, \lambda)$
Chi-squared(n)	Chi-squared(m)	Chi-squared$(n + m)$

3.7 顺序统计量

设 $X_1, \ldots, X_n$ i.i.d.，分布函数 F，密度 f。

- 第 k 个顺序统计量 $X_{(k)}$ 的密度：

$$f_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} [1 - F(x)]^{n-k} f(x)$$

- 最小值 $X_{(1)}$ 的分布：$F_{\min}(x) = 1 - [1 - F(x)]^n$
- 最大值 $X_{(n)}$ 的分布：$F_{\max}(x) = [F(x)]^n$
- 极差 $R = X_{(n)} - X_{(1)}$ 的分布：

$$f_R(r) = n(n-1)\int_{-\infty}^{\infty}[F(x+r) - F(x)]^{n-2} f(x)f(x+r)dx, \quad r > 0$$

4 多元正态分布 (Lecture 8)

4.1 二元正态分布

$(X, Y) \sim N(\mu_X, \mu_Y; \sigma_X^2, \sigma_Y^2; \rho)$，联合密度：

$$f(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_X)^2}{\sigma_X^2} - 2\rho\frac{(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} + \frac{(y-\mu_Y)^2}{\sigma_Y^2}\right]\right\}$$

二元正态的性质

1. 边际分布：$X \sim N(\mu_X, \sigma_X^2)$，$Y \sim N(\mu_Y, \sigma_Y^2)$
2. 条件分布：$Y|X = x \sim N\left(\mu_Y + \rho\frac{\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1-\rho^2)\right)$
3. 独立性：X, Y 独立 $\iff \rho = 0$
4. 线性组合：$aX + bY \sim N(a\mu_X + b\mu_Y, a^2\sigma_X^2 + 2ab\rho\sigma_X\sigma_Y + b^2\sigma_Y^2)$

4.2 多元正态分布的定义

多元正态分布

随机向量 $\boldsymbol{X} = (X_1, \ldots, X_p)^T$ 服从 p 维正态分布 $N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，若其联合密度为：

$$f(\boldsymbol{x}) = (2\pi)^{-p/2}|\boldsymbol{\Sigma}|^{-1/2}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

其中 $\boldsymbol{\mu} \in \mathbb{R}^p$ 为均值向量，$\boldsymbol{\Sigma} \in \mathbb{R}^{p\times p}$ 为对称正定协方差矩阵。

矩母函数定义

更一般地，$\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ 当且仅当其矩母函数为：

$$M_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}[\exp(\boldsymbol{t}^T\boldsymbol{X})] = \exp\left(\boldsymbol{t}^T\boldsymbol{\mu} + \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t}\right), \quad \boldsymbol{t} \in \mathbb{R}^p$$

此定义允许 $\boldsymbol{\Sigma}$ 半正定（奇异正态）。

多元正态的性质

1. **线性变换**：若 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{A} \in \mathbb{R}^{q\times p}$，$\boldsymbol{b} \in \mathbb{R}^q$，则
$$\boldsymbol{AX} + \boldsymbol{b} \sim N_q(\boldsymbol{A\mu} + \boldsymbol{b}, \boldsymbol{A\Sigma A}^T)$$
2. **边际分布**：任意子向量仍服从多元正态
3. **条件分布**：给定部分变量，其余变量仍服从多元正态
4. **独立性**：对于多元正态，不相关 $\iff$ 独立。即：
$$\mathrm{Cov}(\boldsymbol{X}_1, \boldsymbol{X}_2) = \boldsymbol{0} \iff \boldsymbol{X}_1 \text{ 与 } \boldsymbol{X}_2 \text{ 独立}$$
5. **二次型**：若 $\boldsymbol{X} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$，$\boldsymbol{\Sigma} > 0$，则
$$(\boldsymbol{X} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X} - \boldsymbol{\mu}) \sim \chi^2(p)$$

条件分布公式

设 $\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}\right)$，则

$$\boldsymbol{X}_1 | \boldsymbol{X}_2 = \boldsymbol{x}_2 \sim N\left(\boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\boldsymbol{x}_2 - \boldsymbol{\mu}_2), \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21}\right)$$

5 期望的进一步性质与矩母函数 (Lecture 7)

5.1 协方差矩阵与相关系数矩阵

对于随机向量 $\boldsymbol{X} = (X_1, \ldots, X_p)^T$：

- 协方差矩阵：$\boldsymbol{\Sigma} = (\sigma_{ij})$，其中 $\sigma_{ij} = \mathrm{Cov}(X_i, X_j)$
- 相关系数矩阵：$\boldsymbol{R} = (\rho_{ij})$，其中 $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$

协方差的双线性性

$$\mathrm{Cov}\left(\sum_{i=1}^n a_i X_i, \sum_{j=1}^m b_j Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m a_i b_j \mathrm{Cov}(X_i, Y_j)$$
$$\mathrm{Var}\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i^2 \mathrm{Var}(X_i) + 2\sum_{i<j} a_i a_j \mathrm{Cov}(X_i, X_j)$$

5.2 条件期望的性质

条件期望的性质

设 $\mathbb{E}[|X|] < \infty$，则：

1. 线性性：$\mathbb{E}[aX + bY|Z] = a\mathbb{E}[X|Z] + b\mathbb{E}[Y|Z]$ a.s.
2. 独立性：若 X 与 Z 独立，则 $\mathbb{E}[X|Z] = \mathbb{E}[X]$ a.s.
3. 可测性：$\mathbb{E}[g(Z)X|Z] = g(Z)\mathbb{E}[X|Z]$ a.s.
4. 迭代期望：$\mathbb{E}[\mathbb{E}[X|Y,Z]|Z] = \mathbb{E}[X|Z]$ a.s.

最优预测

对于任意函数 g，有：

$$\mathbb{E}[(X - \mathbb{E}[X|Y])^2] \le \mathbb{E}[(X - g(Y))^2]$$

即条件期望是最小均方误差预测。

5.3 矩母函数 (MGF)

矩母函数

$M_X(t) = \mathbb{E}[e^{tX}]$，定义域为使得期望存在的 t 的集合。

MGF 的性质

1. 矩的生成：若 $M_X(t)$ 在 $t = 0$ 的邻域内存在，则 $\mathbb{E}[X^n] = M_X^{(n)}(0)$
2. 唯一性：MGF 唯一确定分布（在存在区间内）
3. 独立和的 MGF：若 X, Y 独立，则 $M_{X+Y}(t) = M_X(t)M_Y(t)$
4. 线性变换：$M_{aX+b}(t) = e^{bt}M_X(at)$

5.4 特征函数 (CF)

特征函数

$\varphi_X(t) = \mathbb{E}[e^{itX}]$，其中 $i = \sqrt{-1}$。特征函数总是存在。

表 4: 常见分布的矩母函数

分布	**矩母函数** $M_X(t)$
Bernoulli(p)	$1-p+pe^t$
Binomial(n,p)	$(1-p+pe^t)^n$
Poisson(λ)	$\exp(\lambda(e^t-1))$
Geometric(p)	$\frac{pe^t}{1-(1-p)e^t}$, $t<-\ln(1-p)$
Uniform(a,b)	$\frac{e^{bt}-e^{at}}{(b-a)t}$, $t\neq 0$
Exp(λ)	$\frac{\lambda}{\lambda-t}$, $t<\lambda$
Normal(μ,σ^2)	$\exp(\mu t+\frac{1}{2}\sigma^2t^2)$
Gamma(α,λ)	$\left(\frac{\lambda}{\lambda-t}\right)^\alpha$, $t<\lambda$
Chi-squared(n)	$(1-2t)^{-n/2}$, $t<\frac{1}{2}$

MGF 与 CF 的关系

- 若 MGF 存在，则 CF 可通过 $t\to it$ 获得
- CF 总是存在且唯一确定分布
- Lévy 连续性定理：$X_n \xrightarrow{d} X \iff \varphi_{X_n}(t)\to\varphi_X(t)$ 对所有 t

6 极限定理 (Lecture 9)

6.1 收敛模式

各种收敛模式

设 $\{X_n\}$ 为随机变量序列，X 为随机变量。

1. **几乎必然收敛 (a.s.)**：$X_n \xrightarrow{a.s.} X$ 若 $\mathbb{P}(\lim_{n\to\infty} X_n = X)=1$
2. **依概率收敛**：$X_n \xrightarrow{P} X$ 若 $\forall\varepsilon>0$，$\lim_{n\to\infty}\mathbb{P}(|X_n-X|>\varepsilon)=0$
3. **依分布收敛**：$X_n \xrightarrow{d} X$ 若 $\lim_{n\to\infty}F_{X_n}(x)=F_X(x)$ 在所有 F_X 的连续点成立
4. L^p **收敛**：$X_n \xrightarrow{L^p} X$ 若 $\lim_{n\to\infty}\mathbb{E}[|X_n-X|^p]=0$，$p\geq 1$

收敛模式之间的关系

$$\begin{tikzcd} a.s. \arrow[r] \arrow[rr, bend right] & P \arrow[r] & d \\ & L^p \arrow[u] & \end{tikzcd}$$

具体而言：

- $X_n \xrightarrow{a.s.} X \Rightarrow X_n \xrightarrow{P} X$
- $X_n \xrightarrow{L^p} X \Rightarrow X_n \xrightarrow{P} X$
- $X_n \xrightarrow{P} X \Rightarrow X_n \xrightarrow{d} X$
- 若 $X_n \xrightarrow{P} c$（常数），则 $X_n \xrightarrow{d} c$
- 若 $X_n \xrightarrow{d} c$（常数），则 $X_n \xrightarrow{P} c$

各种收敛的反例

1. **依概率收敛但不几乎必然收敛**：设 $\{X_n\}$ 独立，$\mathbb{P}(X_n = 1) = 1/n$，$\mathbb{P}(X_n = 0) = 1 - 1/n$。则 $X_n \xrightarrow{P} 0$，但由 Borel-Cantelli 引理，$X_n = 1$ i.o.，故不 a.s. 收敛。
2. **依分布收敛但不依概率收敛**：设 $X \sim N(0,1)$，$X_n = (-1)^n X$。则 $X_n \sim N(0,1)$，故 $X_n \xrightarrow{d} X$，但 $|X_n - X| = |1 - (-1)^n||X|$，不依概率收敛。
3. **几乎必然收敛但不 L^p 收敛**：设 $X_n = n^{1/p} 1_{[0,1/n]}$，则 $X_n \xrightarrow{a.s.} 0$，但 $\mathbb{E}[|X_n|^p] = 1$，不 L^p 收敛。

6.2 重要不等式

马尔可夫不等式

若 $X \geq 0$，则对任意 $a > 0$：$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$

切比雪夫不等式

若 $\mathbb{E}[X] = \mu$，$\mathrm{Var}(X) = \sigma^2$，则对任意 $k > 0$：

$$\mathbb{P}(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$$

等价地，$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

柯西-施瓦茨不等式

$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$

琴生不等式

若 φ 是凸函数，则 $\varphi(\mathbb{E}[X]) \leq \mathbb{E}[\varphi(X)]$

6.3 大数定律

弱大数定律（WLLN）

设 $\{X_n\}$ i.i.d.，$\mathbb{E}[|X_1|] < \infty$，$\mu = \mathbb{E}[X_1]$，则：

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

强大数定律（SLLN）

在相同条件下，$\bar{X}_n \xrightarrow{a.s.} \mu$

SLLN 证明思路（在存在四阶矩时）. 令 $\mu = \mathbb{E}[X_1]$, 记中心化变量 $Y_i = X_i - \mu$（则 $\mathbb{E}[Y_i] = 0$）。若 $\mathbb{E}[X_1^4] < \infty$，则

$$\mathbb{E}\left[(\bar{X}_n - \mu)^4\right] = \frac{1}{n^4}\mathbb{E}\left[\left(\sum_{i=1}^{n} Y_i\right)^4\right] = \frac{1}{n^4}\left[n\mathbb{E}[Y_1^4] + 3n(n-1)\left(\mathbb{E}[Y_1^2]\right)^2\right] = O\left(\frac{1}{n^2}\right).$$

由 Markov 不等式,

$$\mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\mathbb{E}[(\bar{X}_n - \mu)^4]}{\varepsilon^4} = O\left(\frac{1}{n^2}\right).$$

因此 $\sum_n \mathbb{P}(|\bar{X}_n - \mu| > \varepsilon) < \infty$，由 Borel–Cantelli 引理得 $\bar{X}_n \xrightarrow{a.s.} \mu$。 □

6.4 中心极限定理

林德伯格-莱维 CLT

设 $\{X_n\}$ i.i.d.，$\mathbb{E}[X_1] = \mu$，$\mathrm{Var}(X_1) = \sigma^2 < \infty$，则：

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$$

棣莫弗-拉普拉斯定理

二项分布的正态近似：设 $S_n \sim \text{Binomial}(n,p)$，则：

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} N(0,1)$$

近似时需连续性修正：

$$\mathbb{P}(a \le S_n \le b) \approx \Phi\left(\frac{b+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{a-\frac{1}{2}-np}{\sqrt{np(1-p)}}\right)$$

6.5 其他重要极限定理

Borel-Cantelli 引理

1. 若 $\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty$，则 $\mathbb{P}(A_n \text{ i.o.}) = 0$
2. 若 $\{A_n\}$ 独立且 $\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty$，则 $\mathbb{P}(A_n \text{ i.o.}) = 1$

Slutsky 定理

若 $X_n \xrightarrow{d} X$，$Y_n \xrightarrow{P} c$（常数），则：

1. $X_n + Y_n \xrightarrow{d} X + c$
2. $X_n Y_n \xrightarrow{d} cX$
3. 若 $c \neq 0$，则 $X_n/Y_n \xrightarrow{d} X/c$

连续映射定理

若 $X_n \xrightarrow{d} X$，且 g 连续，则 $g(X_n) \xrightarrow{d} g(X)$ 若 $X_n \xrightarrow{P} X$，则 $g(X_n) \xrightarrow{P} g(X)$

6.6 弱收敛的理论工具

分布函数的弱收敛

$F_n \xrightarrow{w} F$ 若 $\lim_{n\to\infty} F_n(x) = F(x)$ 在所有 F 的连续点成立。

Helly 选择定理与紧性

1. **Helly 选择定理**：任意分布函数序列 $\{F_n\}$ 必存在一个子列 $\{F_{n_k}\}$，使得 $F_{n_k}(x)$ 在每一点都收敛于某个右连续非减函数 $F(x)$（注：F 不一定是概率分布函数，可能 $F(\infty) < 1$）。

2. **Prokhorov 定理**：若分布函数序列 $\{F_n\}$ 是**紧的 (Tight)**，即对任意 $\epsilon > 0$，存在 $M > 0$ 使得：
$$\inf_n \left(F_n(M) - F_n(-M)\right) > 1 - \epsilon$$
则其必存在弱收敛于**概率分布函数**的子列。

Lévy 连续性定理（最常用）

设 $\{X_n\}$ 对应特征函数为 $\varphi_n(t)$。$F_n \xrightarrow{w} F \iff \varphi_n(t) \to \varphi(t)$ 对所有 t，其中 φ_n, φ 为对应的特征函数。即：

- 若 $X_n \xrightarrow{d} X$，则 $\varphi_n(t) \to \varphi_X(t)$ 对所有 t。
- 反之，若 $\varphi_n(t)$ 对所有 t 收敛于某个函数 $\varphi(t)$，**且 $\varphi(t)$ 在 $t = 0$ 处连续**，则 $\varphi(t)$ 必是某个随机变量 X 的特征函数，且 $X_n \xrightarrow{d} X$。

Portmanteau 定理

以下陈述等价：

1. $X_n \xrightarrow{d} X$
2. $\mathbb{E}[g(X_n)] \to \mathbb{E}[g(X)]$ 对所有有界连续函数 g
3. $\limsup \mathbb{P}(X_n \in C) \leq \mathbb{P}(X \in C)$ 对所有闭集 C
4. $\liminf \mathbb{P}(X_n \in O) \geq \mathbb{P}(X \in O)$ 对所有开集 O

期末备考重点与策略

1. **概念理解**：深入理解概率空间、随机变量、分布函数、独立性、条件期望、各种收敛模式等核心概念。

2. **计算能力**：
   - 熟练掌握离散/连续随机变量的期望、方差计算
   - 随机变量函数分布的求法（CDF 法、变换公式法）
   - 多维随机变量的边际分布、条件分布、独立性判断

- 协方差、相关系数的计算
- 矩母函数、特征函数的计算与性质应用

3. **定理应用**：
   - 大数定律与中心极限定理的应用
   - 全概率公式、贝叶斯公式、全期望公式的应用
   - 不等式（Markov、Chebyshev、Cauchy-Schwarz）的应用
   - 极限定理（Slutsky、连续映射）的应用

4. **证明技巧**：
   - 收敛模式的证明与反例构造
   - 利用特征函数证明极限定理
   - 利用 Borel-Cantelli 引理证明 a.s. 收敛

5. **高频考点**：
   - 多维随机变量函数的分布
   - 协方差矩阵与相关系数矩阵
   - 条件期望与全方差公式
   - 中心极限定理的近似计算
   - 各种收敛模式的辨析与反例

祝大家期末考试顺利！