

概率论知识大纲

本文整理大学数学系需要掌握的概率论主线。重点不是孤立罗列概念，而是说明概念之间的依赖关系，以及每一层工具解决什么问题。

目录

1. 总体逻辑	3
2. 概率空间：事件和概率测度	3
2.1. 基本对象	3
2.2. 关键公式	3
2.3. 逻辑关系	4
3. 随机变量和分布	4
3.1. 随机变量	4
3.2. 离散和连续情形	4
3.3. 随机向量	5
3.4. 逻辑关系	5
4. 常见分布	5
4.1. 离散分布	5
4.2. 连续分布	6
4.3. 统计中常见分布	6
4.4. 逻辑关系	6
5. 期望、方差和积分观点	6
5.1. 期望	6
5.2. 方差和协方差	7
5.3. 常用不等式	7
5.4. 逻辑关系	7
6. 独立性和条件概率	8
6.1. 条件概率	8
6.2. 独立性	8
6.3. 条件期望初步	8
6.4. 逻辑关系	8
7. 生成函数、特征函数和变换方法	8
8. 收敛概念	9
9. 大数定律和中心极限定理	10
9.1. 大数定律	10
9.2. 中心极限定理	10
9.3. 逻辑关系	10
10. 条件期望、鞅和停时	11
10.1. 条件期望	11
10.2. 鞅	11
10.3. 逻辑关系	11
11. 随机过程	11
11.1. Markov 链	11
11.2. Poisson 过程	12
11.3. Brownian motion	12
11.4. 逻辑关系	12

12. 与数理统计的接口	12
12.1. 样本和似然	12
12.2. 估计量性质	13
12.3. 假设检验	13
12.4. 逻辑关系	13
13. 更高阶方向	13
14. 推荐学习顺序	14
15. 资源	14
16. 问题	14

1. 总体逻辑

概率论的主线可以这样组织：

1. 样本空间和事件
2. 概率测度
3. 随机变量
4. 分布
5. 期望和积分
6. 独立性和条件概率
7. 收敛和极限定理
8. 随机过程
9. 统计推断和应用

其中最核心的抽象是：

- 事件是集合，概率是集合上的测度。
- 随机变量是从样本空间到数值空间的 measurable 函数。
- 分布把随机变量的问题转化为数轴、欧氏空间或者一般空间上的测度问题。
- 期望是对随机变量做积分，不只是“平均值”。
- 极限定理解释大量随机现象为什么会稳定，或者为什么会近似正态。

2. 概率空间：事件和概率测度

2.1. 基本对象

概率论首先需要有一个可以谈论“事件”的空间。一个概率空间写作：

$$(\Omega, \mathcal{F}, P) \tag{1}$$

其中：

- Ω 是样本空间，表示所有可能结果。
- $\mathcal{F}$ 是 Ω 上的 sigma 代数，表示允许讨论概率的事件族。
- P 是概率测度。

概率测度满足三条公理：

$$P(A) \geq 0 \tag{2}$$

$$P(\Omega) = 1 \tag{3}$$

$$P(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) \quad \text{当} \quad A_i \cap A_j = \emptyset, i \neq j \tag{4}$$

2.2. 关键公式

补事件：

$$P(A^c) = 1 - P(A) \tag{5}$$

容斥公式的二事件版本：

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \tag{6}$$

连续性：

$$A_n \text{ increases to } A \Rightarrow P(A_n) \rightarrow P(A) \quad (7)$$

$$A_n \text{ decreases to } A \Rightarrow P(A_n) \rightarrow P(A) \quad (8)$$

2.3. 逻辑关系

- 没有 sigma 代数，就无法严谨讨论哪些集合可以赋予概率。
- 没有可数可加性，就无法处理极限事件，例如“无限次试验中至少发生一次”。
- 后面的随机变量、分布、条件期望，本质都建立在概率空间上。

需要掌握：

- 用集合语言表示事件。
- 证明事件之间的包含、互斥、独立。
- 使用概率公理推导基本公式。
- 理解 Borel 集合为什么自然出现在实值随机变量中。

3. 随机变量和分布

3.1. 随机变量

随机变量不是“会随机变化的数”，而是定义在概率空间上的可测函数：

$$X : \Omega \rightarrow \mathbb{R} \quad (9)$$

可测性保证下面这种事件属于 $\mathcal{F}$ ：

$$\{\omega \in \Omega : X(\omega) \leq x\} \in \mathcal{F} \quad (10)$$

分布函数定义为：

$$F_{X(x)} = P(X \leq x) \quad (11)$$

更一般地， X 的分布是 $\mathbb{R}$ 上的概率测度：

$$\mu_{X(B)} = P(X \in B), \quad B \in \mathcal{B}(\mathbb{R}) \quad (12)$$

3.2. 离散和连续情形

离散随机变量：

$$P(X \in A) = \sum_{x \in A} P(X = x) \quad (13)$$

连续随机变量：

$$P(a \leq X \leq b) = \int_a^b f_{X(x)} dx \quad (14)$$

其中 f_X 是密度函数，并满足：

$$f_{X(x)} \geq 0, \quad \int_{-\infty}^{\infty} f_{X(x)} dx = 1 \quad (15)$$

分布函数和密度的关系：

$$F_{X(x)} = \int_{-\infty}^x f_{X(t)} dt, \quad f_{X(x)} = F'_{X(x)}(x) \quad (16)$$

3.3. 随机向量

随机向量的联合分布：

$$F_{X,Y}(x,y) = P(X \leq x, Y \leq y) \quad (17)$$

边缘分布：

$$f_{X(x)} = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dy \quad (18)$$

$$f_{Y(y)} = \int_{-\infty}^{\infty} f_{X,Y}(x,y) dx \quad (19)$$

3.4. 逻辑关系

- 随机变量把抽象的随机结果变成可以计算的数值。
- 分布把 $P(X \in A)$ 从样本空间搬到数值空间。
- 联合分布描述多个随机变量的整体关系，边缘分布只是其中一部分信息。

需要掌握：

- 从分布函数判断随机变量类型。
- 在离散、连续、混合情形下计算概率。
- 从联合分布求边缘分布。
- 对随机变量做函数变换，例如 $Y = g(X)$ 。
- 理解“同分布”和“相等”不是一回事。

4. 常见分布

常见分布不是背公式，而是理解它们对应的随机机制。

4.1. 离散分布

Bernoulli 分布：

$$X \sim \text{Bernoulli}(p), \quad P(X=1)=p, \quad P(X=0)=1-p \quad (20)$$

Binomial 分布：

$$X \sim \text{Bin}(n,p), \quad P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (21)$$

Geometric 分布：

$$X \sim \text{Geom}(p), \quad P(X=k) = (1-p)^{k-1} p, \quad k=1,2,\dots \quad (22)$$

Poisson 分布：

$$X \sim \text{Poisson}(\lambda), \quad P(X=k) = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k=0,1,2,\dots \quad (23)$$

Poisson 分布是二项分布的稀有事件极限：

$$\text{Bin}\left(n, \frac{\lambda}{n}\right) \rightarrow \text{Poisson}(\lambda) \quad (24)$$

4.2. 连续分布

Uniform 分布:

$$X \sim \text{Unif}(a, b), \quad f(x) = \frac{1}{b-a} 1_{a \leq x \leq b} \quad (25)$$

Exponential 分布:

$$X \sim \text{Exp}(\lambda), \quad f(x) = \lambda e^{-\lambda x} 1_{x \geq 0} \quad (26)$$

无记忆性:

$$P(X > s+t \mid X > s) = P(X > t) \quad (27)$$

Gamma 分布:

$$X \sim \text{Gamma}(\alpha, \lambda), \quad f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0 \quad (28)$$

Normal 分布:

$$X \sim N(\mu, \sigma^2), \quad f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (29)$$

4.3. 统计中常见分布

若 $Z_i \sim N(0, 1)$ 且相互独立, 则:

$$\sum_{i=1}^n Z_i^2 \sim \text{chi}_n^2 \quad (30)$$

若 $Z \sim N(0, 1)$, $U \sim \text{chi}_n^2$ 且独立, 则:

$$T = \frac{Z}{\sqrt{\frac{U}{n}}} \sim t_n \quad (31)$$

这些分布是正态样本推断、置信区间和假设检验的基础。

4.4. 逻辑关系

- Bernoulli $\rightarrow$ Binomial $\rightarrow$ Poisson 是从单次试验到计数极限。
- Exponential $\rightarrow$ Gamma 是从一次等待到多次等待。
- Normal $\rightarrow$ Chi-square/t/F 是数理统计的重要基础。
- Multinomial $\rightarrow$ Dirichlet 常用于贝叶斯统计。

5. 期望、方差和积分观点

5.1. 期望

离散情形:

$$E[X] = \sum_x xP(X=x) \quad (32)$$

连续情形:

$$E[X] = \int_{-\infty}^{\infty} x f_{X(x)} dx \quad (33)$$

一般情形中，期望是 Lebesgue 积分：

$$E[X] = \int_{\Omega} X dP \quad (34)$$

函数的期望可以直接对分布积分：

$$E[g(X)] = \int_{\mathbb{R}} g(x) d\mu_{X(x)} \quad (35)$$

连续情形写成：

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_{X(x)} dx \quad (36)$$

5.2. 方差和协方差

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \quad (37)$$

$$\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y] \quad (38)$$

$$\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X) \text{Var}(Y)}} \quad (39)$$

5.3. 常用不等式

Markov 不等式：

$$P(X \geq a) \leq \frac{E[X]}{a} \quad (X \geq 0, a > 0) \quad (40)$$

Chebyshev 不等式：

$$P(|X - E[X]| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2} \quad (41)$$

Jensen 不等式：

$$\varphi(E[X]) \leq E[\varphi(X)] \quad \text{当 } \varphi \text{ 为凸函数} \quad (42)$$

Cauchy-Schwarz 不等式：

$$|E[XY]| \leq \sqrt{E[X^2]E[Y^2]} \quad (43)$$

5.4. 逻辑关系

- 期望是对随机变量积分。
- 方差是二阶矩给出的波动尺度。
- 协方差描述两个随机变量的线性关联。
- 不等式把难以精确计算的概率转化为可控上界。

6. 独立性和条件概率

6.1. 条件概率

条件概率定义：

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad P(B) > 0 \quad (44)$$

乘法公式：

$$P(A \cap B) = P(A | B)P(B) \quad (45)$$

全概率公式：

$$P(A) = \sum_i P(A | B_i)P(B_i) \quad (46)$$

Bayes 公式：

$$P(B_j | A) = \frac{P(A | B_j)P(B_j)}{\sum_i P(A | B_i)P(B_i)} \quad (47)$$

6.2. 独立性

事件独立：

$$P(A \cap B) = P(A)P(B) \quad (48)$$

随机变量独立：

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B) \quad (49)$$

有密度时，独立性等价于：

$$f_{X,Y}(x, y) = f_{X(x)}f_{Y(y)} \quad (50)$$

6.3. 条件期望初步

条件期望可以理解为给定信息后的最佳预测：

$$E[X | Y] = g(Y) \quad (51)$$

并满足：

$$E[X] = E[E[X | Y]] \quad (52)$$

如果 X 和 Y 独立，则：

$$E[X | Y] = E[X] \quad (53)$$

6.4. 逻辑关系

- 条件概率描述信息更新之后概率如何变化。
- 独立性表示一个对象不提供关于另一个对象的信息。
- 条件期望是给定信息后的最佳预测，也是鞅论的基础。

7. 生成函数、特征函数和变换方法

这些工具的作用是把分布问题变成函数问题。

概率母函数:

$$G_{X(s)} = E[s^X] = \sum_{k=0}^{\infty} P(X = k) s^k \quad (54)$$

矩母函数:

$$M_{X(t)} = E[e^{tX}] \quad (55)$$

特征函数:

$$\varphi_{X(t)} = E[e^{itX}] \quad (56)$$

Laplace 变换:

$$L_{X(t)} = E[e^{-tX}] \quad (57)$$

独立和的特征函数:

$$X \text{ 与 } Y \text{ 独立} \Rightarrow \varphi_{X+Y}(t) = \varphi_{X(t)} \varphi_{Y(t)} \quad (58)$$

卷积公式:

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f_{X(x)} f_{Y(z-x)} dx \quad (59)$$

逻辑关系:

- 独立随机变量之和对应分布卷积。
- 卷积在变换之后变成乘法。
- 特征函数总是存在，比矩母函数更一般。
- 极限定理常通过特征函数证明。

8. 收敛概念

概率论里有多种收敛，它们强弱不同。

几乎处处收敛:

$$P\left(\lim_{n \rightarrow \infty} X_n = X\right) = 1 \quad (60)$$

依概率收敛:

$$\forall \varepsilon > 0, \quad P(|X_n - X| > \varepsilon) \rightarrow 0 \quad (61)$$

$$L^p \quad (62)$$

收敛:

$$E[|X_n - X|^p] \rightarrow 0 \quad (63)$$

依分布收敛:

$$F_{X_n}(x) \rightarrow F_{X(x)} \quad \text{在 } F_X \text{ 的连续点} \quad (64)$$

常见蕴含关系:

$$X_n \rightarrow X \text{ in } L^p \Rightarrow X_n \rightarrow X \text{ in probability} \Rightarrow X_n \rightarrow X \text{ in distribution} \quad (65)$$

几乎处处收敛也推出依概率收敛：

$$X_n \rightarrow X \text{ a.s.} \Rightarrow X_n \rightarrow X \text{ in probability} \quad (66)$$

Borel-Cantelli 引理：

$$\sum_{n=1}^{\infty} P(A_n) < \infty \Rightarrow P(A_n \text{ i.o.}) = 0 \quad (67)$$

如果 A_n 相互独立，并且 $\sum_n P(A_n) = \infty$ ，则：

$$P(A_n \text{ i.o.}) = 1 \quad (68)$$

一致可积用于把收敛和期望交换，是理解条件期望和极限定理的重要技术条件。

9. 大数定律和中心极限定理

9.1. 大数定律

设 $X_1, X_2, \dots$ 独立同分布， $E[X_1] = \mu$ ，样本均值为：

$$|(X)_n = \frac{1}{n} \sum_{i=1}^n X_i \quad (69)$$

弱大数定律：

$$|(X)_n \rightarrow \mu \text{ in probability} \quad (70)$$

强大数定律：

$$|(X)_n \rightarrow \mu \text{ a.s.} \quad (71)$$

直观含义：大量独立重复实验的平均值会稳定到期望。

9.2. 中心极限定理

若 X_i 独立同分布， $E[X_i] = \mu$ ， $\text{Var}(X_i) = \sigma^2 < \infty$ ，则：

$$\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \rightarrow N(0, 1) \text{ in distribution} \quad (72)$$

等价地：

$$\sqrt{n} \frac{|(X)_n - \mu}{\sigma} \rightarrow N(0, 1) \quad (73)$$

Delta 方法：

$$\sqrt{n}(T_n - \theta) \rightarrow N(0, \sigma^2) \Rightarrow \sqrt{n}(g(T_n) - g(\theta)) \rightarrow N(0, (g'(\theta))^2 \sigma^2) \quad (74)$$

Slutsky 定理的常用形式：

$$X_n \rightarrow X \text{ in distribution, } Y_n \rightarrow c \text{ in probability} \Rightarrow X_n + Y_n \rightarrow X + c \quad (75)$$

$$X_n Y_n \rightarrow cX \text{ in distribution} \quad (76)$$

9.3. 逻辑关系

- 大数定律说明样本平均会稳定到期望。

- 中心极限定理说明样本平均的波动在适当缩放后趋向正态。
- Slutsky 定理和 Delta 方法把极限定理扩展到函数和估计量。

10. 条件期望、鞅和停时

10.1. 条件期望

设 $\mathcal{G} \subset \mathcal{F}$ 。条件期望 $E[X | \mathcal{G}]$ 是 $\mathcal{G}$ -可测随机变量，并满足：

$$\int_G E[X | \mathcal{G}] dP = \int_G X dP, \quad \forall G \in \mathcal{G} \quad (77)$$

塔性质：

$$E[E[X | \mathcal{G}]] = E[X] \quad (78)$$

若 $\mathcal{H} \subset \mathcal{G}$ ，则：

$$E[E[X | \mathcal{G}] | \mathcal{H}] = E[X | \mathcal{H}] \quad (79)$$

10.2. 鞅

给定滤过 $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \dots$ ，过程 (X_n) 是鞅，如果：

$$X_n \text{ is } \mathcal{F}_n\text{-measurable} \quad (80)$$

$$E[X_n] < \infty \quad (81)$$

$$E[X_{n+1} | \mathcal{F}_n] = X_n \quad (82)$$

上鞅和下鞅分别把最后一个等号改成 $\leq$ 和 $\geq$ 。

停时：

$$\tau \text{ 是停时} \Leftrightarrow \{\tau \leq n\} \in \mathcal{F}_n \quad (83)$$

Optional stopping theorem 的典型结论是，在适当有界性或可积性条件下：

$$E[X_\tau] = E[X_0] \quad (84)$$

10.3. 逻辑关系

- 条件期望是“给定信息后的平均”。
- 鞅描述公平游戏或者无漂移过程。
- 停时描述依赖历史信息的随机时间。
- 鞅工具可以处理随机过程中的最大值、收敛和停止问题。

11. 随机过程

随机过程是按时间索引的一族随机变量：

$$\{X_t : t \in T\} \quad (85)$$

11.1. Markov 链

Markov 性：

$$P(X_{n+1} = j | X_n = i, X_{n-1}, \dots, X_0) = P(X_{n+1} = j | X_n = i) \quad (86)$$

转移矩阵:

$$P = (p_{ij}), \quad p_{ij} = P(X_{n+1} = j \mid X_n = i) \quad (87)$$

$$n \quad (88)$$

步转移概率:

$$P^n = P^n \quad (89)$$

平稳分布:

$$\pi P = \pi, \quad \sum_i \pi_i = 1 \quad (90)$$

11.2. Poisson 过程

计数过程 $(N_t)_{t \geq 0}$ 是强度为 λ 的 Poisson 过程时:

$$N_t \sim \text{Poisson}(\lambda t) \quad (91)$$

独立增量:

$$N_t - N_s \text{ 与过去独立 } (t > s) \quad (92)$$

平稳增量:

$$N_t - N_s \sim \text{Poisson}(\lambda(t - s)) \quad (93)$$

11.3. Brownian motion

标准 Brownian motion $(B_t)_{t \geq 0}$ 满足:

$$B_0 = 0 \quad (94)$$

$$B_t - B_s \sim N(0, t - s), \quad t > s \quad (95)$$

并且具有独立增量和连续路径。

11.4. 逻辑关系

- 随机变量描述一个随机对象，随机过程描述随时间演化的随机对象。
- Markov 性表示未来只依赖当前状态。
- Poisson 过程是连续时间计数模型。
- Brownian motion 是连续时间极限模型，也是随机分析的入口。

12. 与数理统计的接口

概率论提供模型，统计学研究如何从数据反推模型。

12.1. 样本和似然

设样本 $X_1, \dots, X_n$ 来自密度或概率质量函数 $f(x \mid \theta)$ 。似然函数为:

$$L(\theta) = \prod_{i=1}^n f(X_i \mid \theta) \quad (96)$$

对数似然:

$$\ell(\theta) = \sum_{i=1}^n \log f(X_i | \theta) \quad (97)$$

最大似然估计：

$$\hat{\theta} = \arg \max_{\theta} L(\theta) = \arg \max_{\theta} \ell(\theta) \quad (98)$$

12.2. 估计量性质

无偏性：

$$E[\hat{\theta}] = \theta \quad (99)$$

一致性：

$$\hat{\theta}_n \rightarrow \theta \text{ in probability} \quad (100)$$

渐近正态性：

$$\sqrt{n}(\hat{\theta}_n - \theta) \rightarrow N(0, I(\theta)^{-1}) \quad (101)$$

12.3. 假设检验

p-value 的形式可以写成：

$$\text{p-value} = P_{H_0}(T(X) \geq T(x_{\text{obs}})) \quad (102)$$

它表示在零假设成立时，得到当前这样极端或更极端统计量的概率。

12.4. 逻辑关系

- 概率论：已知分布，推导数据性质。
- 数理统计：已知数据，推断分布或者参数。
- 极限定理是统计推断中近似方法的理论基础。

13. 更高阶方向

完成前面的内容之后，可以继续深入：

- 测度论概率。
- 随机微积分。
- 随机微分方程。
- 信息论。
- 大偏差理论。
- 随机场。
- 随机矩阵。
- 高维概率。
- 随机算法。
- 概率图模型。
- 因果推断。

这些方向的依赖关系大致是：

- 测度论概率 $\rightarrow$ 条件期望、鞅、随机过程。
- Brownian motion $\rightarrow$ 随机微积分 $\rightarrow$ 随机微分方程。

- 极限定理 → 大偏差、高维概率、随机矩阵。
- 条件独立 → 概率图模型、因果推断。
- 熵和 KL 散度 → 信息论、统计学习、变分推断。

14. 推荐学习顺序

1. 先学概率空间、随机变量、分布、期望。
2. 然后学常见分布、条件概率、独立性。
3. 接着学不等式、收敛概念、大数定律、中心极限定理。
4. 再学条件期望、鞅、Markov 链、Poisson 过程、Brownian motion。
5. 最后把概率论和数理统计、随机过程、高维概率连接起来。

最小闭环：

1. 概率空间
2. 随机变量
3. 分布
4. 期望
5. 独立性
6. 条件概率
7. 大数定律
8. 中心极限定理

如果这个闭环清楚，概率论的大部分基础内容就能组织起来。

15. 资源

<https://news.ycombinator.com/item?id=49583440>

- <https://seeing-theory.brown.edu/>
- <https://mbmlbook.com/>
- <https://xcelab.net/rm/>

16. 问题

- 为什么说，概率论还有学派的划分？
- 中新极限定理如何推出来的？
- Point Estimation
- Confidence Interval
- Bootstrap : 几种分布
 - uniform
 - Exponential
 - normal
 - student t
 - Chi-square
 - fisher snedecor
- Bayes' Theorem
- Likelihood Function
 - 这里又有好几种

- Prior to Posterior
- Regression Analysis
 - Ordinary Least Squares
 - Correlation

已知 10 个晶体管中有 7 个正品及 3 个次品, 每次任意抽取一个进行测试, 测试后不放回, 直到把 3 个次品都找到为止, 则需要测试 7 次的概率为 $2/15$, 那么是对还是错 A. 对 B. 错