

博士资格考试复习资料

概率论与数理统计博士资格考试复习资料

Probability Foundations and Limits · Stochastic Processes · Mathematical
Statistics

本资料面向博士资格考试复习。在读者具备基础微积分、线性代数和概率知识的前提下，可独立用于系统复习、证明训练和历年真题准备。本书覆盖六期正式试卷、63 道大题、115 个显式子问和 128 个最细作答单元，并同时标明 2026 秋及以前考纲与 2027 春起新版考纲。线性模型作为新版模块完整纳入。

2026 年 7 月

使用说明

全书依赖顺序为：测度与积分、条件期望、收敛与极限定理、鞅和随机过程、充分性与估计、检验与 Bayesian、渐近统计、线性模型。Delta 方法只在第 4 章证明，后续直接引用；条件期望只在第 3 章建立，鞅与 Rao–Blackwell 不重复定义。

考试范围提醒

GOE、GFF、树渗流等背景只展开到完成正式题所需深度。六份本地答案候选继续标为 un-verified_candidate，不覆盖独立核验解答，也不称为官方答案。

目录

使用说明	ii
第一部分 Probability Foundations and Limits	1
第一章 概率空间、随机变量与分布	2
1.1 为什么从概率空间开始	2
1.2 由集合运算读懂随机事件	3
1.3 随机变量与分布	3
1.4 随机向量、联合分布与独立性	4
1.5 变量变换：先找原像，再求分布	5
1.6 分位数与概率积分变换	5
1.7 测度连续性与概率的可列结构	6
1.8 分布函数完整刻画分布	7
1.9 多维变量变换与 Jacobian	7
1.10 概率积分变换与分位数构造	8
第二章 积分、期望与概率不等式	9
2.1 从简单函数到期望	9
2.2 四个基本极限定理	9
2.3 尾积分公式与矩	11
2.4 四类常用不等式	11
2.5 生成函数：何时能用，能推出什么	12
2.6 特征函数：唯一性、反演与 Lévy 连续性	12
2.7 单调收敛、Fatou 与支配收敛的依赖链	14
2.8 尾积分公式与矩的层蛋糕表示	15
2.9 Jensen、条件 Jensen 与等号	15
2.10 Chernoff 方法的优化步骤	16
2.11 生成函数的唯一性与解析边界	16
第三章 条件期望、条件分布与条件独立	17
3.1 为什么条件期望仍是随机变量	17
3.2 必须熟练的运算规则	17
3.3 L^2 中的正交投影	18
3.4 按随机变量条件化	18
3.5 Gaussian 条件分布	20
3.6 条件独立与图式分解	20
3.7 停止信息与未来信息的雏形	20
3.8 条件期望的存在、唯一与测试函数刻画	21
3.9 塔式法则与“提出已知量”	22
3.10 L^2 投影与方差分解	22
3.11 Gaussian 条件分布的 Schur 补	22
3.12 条件独立的因子分解	23
第四章 收敛模式、一致可积与 Delta 方法	24
4.1 四种收敛及其关系	24
4.2 Portmanteau、连续映射与 Slutsky	25
4.3 一致可积：把概率收敛升级为均值收敛	25
4.4 Delta 方法	26
4.5 随机级数中的几乎处处与 L^1 收敛	27
4.6 收敛关系与反例体系	29
4.7 子列原理	29
4.8 Portmanteau 的开闭集版本	30
4.9 一致可积与 Vitali 定理	30

4.10	Delta 方法的余项条件	30
第五章	Borel–Cantelli、稀有事件与随机级数	32
5.1	$\limsup A_n$ 就是无穷多次发生	32
5.2	第一 Borel–Cantelli 引理	32
5.3	第二 Borel–Cantelli 引理	33
5.4	两两独立版本	33
5.5	从尾概率判断单点极值	34
5.6	随机级数：大跳、中心部分与方差	35
5.7	零一律的边界	35
5.8	第二 Borel–Cantelli 中独立性不可删除	36
5.9	两两独立版本的二阶矩证明	37
5.10	极值的上、下界模板	37
5.11	Kolmogorov 三系列定理：完整判别与证明	38
第六章	最大不等式与强大数律	41
6.1	为什么只控制终点还不够	41
6.2	方差可和的随机级数	42
6.3	有限方差强大数律	42
6.4	仅有一阶矩时：截断到底解决什么	43
6.5	当普通均值不收敛：区块振荡法	44
6.6	Kolmogorov 最大不等式的停时分割	47
6.7	方差可和级数与 Kronecker 引理	47
6.8	仅一阶矩下截断法的完整主线	47
6.9	重尾下普通均值为什么失败	48
第七章	中心极限定理与三角阵列	50
7.1	从大数律到中心极限定理	50
7.2	三角阵列与 Lindeberg 条件	51
7.3	Lyapunov 条件	52
7.4	方差尺度不是机械相加：先拆稀有大跳	53
7.5	CLT 之后的变换	54
7.6	经典 CLT 的特征函数证明	55
7.7	Lindeberg 条件为何排除单个大项	56
7.8	Lyapunov 条件与 Lindeberg	56
7.9	Berry–Esseen 与正态近似误差	56
7.10	Cramér–Wold 与多维 CLT	57
第二部分	Stochastic Processes	58
第八章	鞅、停时与可选停止	59
8.1	滤过、适应性与停时	59
8.2	鞅、构造与不等式	59
8.3	可选停止与 Wald 等式	60
8.4	鞅的三种构造	61
8.5	Doob 最大不等式	61
8.6	可选停止的安全版本	62
8.7	赌徒破产与指数鞅	62
8.8	Wald 等式与条件	62
第九章	鞅收敛与一致可积	64
9.1	上穿不等式与几乎处处收敛	64
9.2	L^p 、一致可积与闭鞅	64
9.3	上穿不等式的买低卖高证明	65
9.4	非负超鞅收敛	66
9.5	闭鞅与一致可积	66
9.6	L^p 有界鞅	66
9.7	Doob 分解	67
第十章	Markov 链、Poisson 过程与首达递推	68
10.1	齐次 Markov 性、条件独立与半群	68
10.2	停时、强 Markov 性与首达边值问题	69
10.3	常返、暂留与沟通类	71
10.4	平稳分布的构造、唯一性与详细平衡	72
10.5	周期、时刻分布与时间平均	74
10.6	Poisson 过程：指数等待时间、条件到达与分裂	74

10.7 连续时间 Markov 链与生成元入口	75
第十一章 Brownian motion、反射原理与指数鞅	77
11.1 Brownian motion、反射原理与二次变差	77
11.2 Brownian motion 的有限维分布与协方差	78
11.3 反射原理与最大值分布	78
11.4 指数鞅与漂移命中	79
11.5 二次变差与非经典可微性	79
11.6 Brownian bridge 与条件高斯	80
第十二章 Galton–Watson 分枝过程与树上临界	81
12.1 Galton–Watson 分枝过程	81
12.2 综合题与易错点	81
12.3 生成函数迭代与灭绝概率	82
12.4 归一化人口鞅	83
12.5 总后代数与随机游走编码	83
12.6 树渗流的 boundary 身份	83
第三部分 Mathematical Statistics	85
第十三章 次序统计量、充分性、完备性与 Basu	86
13.1 统计模型、统计量与充分性	86
13.2 完备性、辅助统计量与 Basu	86
13.3 因子分解定理的双向证明	87
13.4 极小充分统计量的似然比判据	88
13.5 一参数指数族与完备性	88
13.6 Rao–Blackwell、Lehmann–Scheffé 与 Basu	88
13.7 次序统计量的联合密度	89
第十四章 参数估计、Fisher 信息与风险	90
14.1 矩估计、MLE 与存在性	90
14.2 偏差、MSE 与 Rao–Blackwell	90
14.3 Fisher 信息与 Cramér–Rao 下界	91
14.4 矩估计、MLE 与参数空间边界	92
14.5 偏差、方差、MSE 与风险	92
14.6 Fisher 信息与 score 恒等式	92
14.7 Cramér–Rao 下界的完整证明	93
14.8 置信区间与枢轴量	93
14.9 正则 MLE 的渐近效率	94
第十五章 假设检验、似然比与置信区间	95
15.1 置信区间、枢轴量与正态模型	95
15.2 检验、错误概率、功效与 p 值	96
15.3 Neyman–Pearson 引理	96
15.4 单调似然比、UMP 与检验反演	96
15.5 似然比、Wald 与 score	97
15.6 卡方检验与拟合优度	97
15.7 Neyman–Pearson 引理的交换证明	98
15.8 单调似然比与 Karlin–Rubin	98
15.9 检验反演与置信集	99
15.10 似然比、Wald 与 score 的局部关系	99
15.11 p 值、尺寸与功效	99
15.12 卡方拟合优度与参数估计扣自由度	100
15.13 精确离散置信区间与最短性边界	100
第十六章 Bayesian 推断、共轭与决策	102
16.1 Bayesian 后验、决策与共轭	102
16.2 后验分布与后验预测	103
16.3 共轭族的三个完整例子	104
16.4 Bayes 决策与常见损失	104
16.5 可信区间与置信区间	105
16.6 非信息先验与 proper posterior	105
16.7 Bayes factor 与先验敏感性	106
第十七章 渐近统计：一致性、正态性与非正则极限	107
17.1 一致性与 MLE 渐近理论	107

17.2 M-estimation 的一致性模板	108
17.3 score 展开与渐近正态	108
17.4 样本分位数与非正则速率	108
17.5 典型题型与易错点	108
17.6 估计量一致性的三条路线	109
17.7 M-estimator 的渐近线性化	109
17.8 渐近等价、Slutsky 与 studentization	110
17.9 样本分位数的 Bahadur 线性化	110
17.10 Bootstrap 的目标与失效	111
17.11 局部备择与检验功效	111
第十八章 线性模型：投影、可估性与 t/F 推断	112
18.1 考纲版本与模型设定	112
18.2 最小二乘、正规方程与投影	112
18.3 Gauss–Markov 定理与 BLUE	113
18.4 正态误差下的分布理论	113
18.5 一般线性假设、F 检验与置信椭圆	113
18.6 ANOVA、回归与预测	114
18.7 秩亏、可估函数与广义逆	114
18.8 投影矩阵与最小二乘几何	115
18.9 Gauss–Markov 定理的矩阵证明	115
18.10 正态模型下的独立分解	116
18.11 一般线性假设与 F 检验	116
18.12 ANOVA 与平方和分解	117
18.13 秩亏、可估函数与广义逆	117
18.14 异方差与广义最小二乘	117
18.15 预测区间与均值响应区间	118
18.16 同时推断、拟合不足与诊断边界	118
第四部分 概率论与数理统计历年真题与完整解答	120
第十九章 2023 年秋季博士资格考试	121
19.1 第 1 题：二维高斯、边缘与极坐标 (原卷 p.1, 10 分)	121
19.2 第 2 题：Laplace 增量的指数鞅 (原卷 p.1, 10 分)	122
19.3 第 3 题：随机重抛硬币的 Markov 链 (原卷 p.1, 10 分)	122
19.4 第 4 题：三骰等待时间最优策略 (原卷 p.2, 10 分)	123
19.5 第 5 题：正则树渗流 (原卷 p.2, 15 分)	123
19.6 第 6 题：高斯随机矩阵 CLT (原卷 pp.2–3, 15 分)	124
19.7 第 7 题：正态模型的最强/UMP 检验 (原卷 p.3, 15 分)	124
19.8 第 8 题：Cauchy 极小充分与双样本 MLE (原卷 p.3, 15 分)	124
第二十章 2024 年春季博士资格考试	126
20.1 第 1 题：无限方差与非标准 CLT (原卷 p.1, 10 分)	126
20.2 第 3 题：乘积收敛与反例 (原卷 p.1, 10 分)	126
20.3 第 4 题：偏置随机游动首次 (原卷 pp.1–2, 15 分)	127
20.4 第 5 题：Markov 链的递推表示 (原卷 p.2, 10 分)	127
20.5 第 6 题：Brownian 首次时 (原卷 p.2, 10 分)	128
20.6 第 7 题：高斯均值估计的 minimax 性 (原卷 p.2, 15 分)	128
20.7 第 8 题：平移均匀分布的相合与渐近 (原卷 p.2, 10 分)	128
20.8 第 9 题：均值方差同参数的 MLE (原卷 p.2, 10 分)	129
20.9 第 10 题：Pareto 均值的 UMP 检验 (原卷 p.2, 10 分)	129
20.10 第 11 题：约束正态均值的 UMVU (原卷 p.2, 15 分)	129
第二十一章 2024 年秋季博士资格考试	131
21.1 第 1 题：均匀和的四阶矩与 Irwin–Hall 分布 (原卷 p.1, 10 分)	131
21.2 第 3 题：次指数尾、矩与 mgf 等价 (原卷 pp.1–2, 10 分)	132
21.3 第 4 题：正则树与 b 叉树随机游动 (原卷 p.2, 10 分)	132
21.4 第 6 题：Brownian 存活条件与 Brownian bridge/GFF (原卷 pp.2–3, 15 分)	133
21.5 第 7 题：只记录符号后的推断 (原卷 p.3, 10 分)	133
21.6 第 8 题：线性模型 BLUE 与信息矩阵 (原卷 p.3, 10 分)	134
21.7 第 9 题：两个估计量的渐近比较 (原卷 pp.3–4, 10 分)	134
21.8 第 10 题：受限正态模型的 UMVU 与 admissibility (原卷 p.4, 15 分)	135
21.9 第 11 题：参数依赖支集的矩法、MLE 与相合 (原卷 p.4, 15 分)	135
第二十二章 2025 年春季概率论与数理统计博士资格考试	136
22.1 第 1 题：样本极值的联合分布与条件期望 (原卷 p.1, 10 分)	136

22.2 第 2 题: 高斯向量的线性表示 (原卷 p.1, 10 分)	137
22.3 第 3 题: 随机游走的矩母函数与退出时生成函数 (原卷 p.1, 15 分)	137
22.4 第 4 题: 非一致有界增量鞅与停时 (原卷 p.1, 10 分)	138
22.5 第 5 题: Brownian 对称退出与符号独立性 (原卷 pp.1-2, 10 分)	139
22.6 第 6 题: 条件二阶矩刻画正态分布 (原卷 p.2, 15 分)	139
22.7 第 7 题: 均匀分布端点的广义 Bayes 估计 (原卷 p.2, 10 分)	140
22.8 第 8 题: Gamma 尺度族的检验与 MLE 渐近分布 (原卷 p.2, 10 分)	140
22.9 第 9 题: 样本相关系数与最小充分性 (原卷 p.2, 15 分)	141
22.10 第 10 题: 绝对值矩估计与可容许性 (原卷 pp.2-3, 10 分)	142
22.11 第 11 题: 带位置参数指数分布的非正则 MLE (原卷 p.3, 15 分)	142
第二十三章 2025 年秋季概率论与数理统计博士资格考试	144
23.1 第 1 题: 随机幂的矩 (原卷 p.1, 10 分)	144
23.2 第 2 题: 三角区域上的条件分布 (原卷 p.1, 10 分)	144
23.3 第 3 题: 鞅的严格凸等号情形 (原卷 pp.1-2, 10 分)	145
23.4 第 5 题: 指数 Brownian 鞅的积分 (原卷 p.2, 10 分)	146
23.5 第 6 题: 二阶 GOE 的特征值与特征向量 (原卷 p.2, 15 分)	146
23.6 第 7 题: 充分统计量上的相合估计 (原卷 p.2, 10 分)	147
23.7 第 8 题: 指数分布尾概率的 UMVU (原卷 p.3, 15 分)	147
23.8 第 9 题: Bernoulli 泛函的 UMVU 与可容许性 (原卷 p.3, 15 分)	148
23.9 第 10 题: 对数正态先验下的端点 MAP (原卷 p.3, 10 分)	148
23.10 第 11 题: 离散均匀族的 UMP 检验 (原卷 p.3, 10 分)	149
第二十四章 2026 年春季概率论与数理统计博士资格考试	150
24.1 第 1 题: 高斯测度的指数倾斜 (原卷 p.1, 10 分)	150
24.2 第 3 题: 随机连接面条的环数 (原卷 p.2, 10 分)	150
24.3 第 4 题: Brownian 指数积分的 Laplace 原理 (原卷 p.2, 10 分)	151
24.4 第 5 题: Galton-Watson 过程基础 (原卷 p.2, 10 分)	152
24.5 第 6 题: 分枝树继承性质与渗流临界值 (原卷 pp.2-3, 15 分)	152
24.6 第 7 题: 变换指数族的 MLR 与 UMP (原卷 p.3, 10 分)	153
24.7 第 8 题: 离散分布的样本分位数相合性 (原卷 p.3, 15 分)	153
24.8 第 9 题: 两点先验的 Bayes 与 minimax 规则 (原卷 p.3, 10 分)	154
24.9 第 10 题: Poisson 平方均值参数的 UMVU 与 MLE (原卷 p.4, 10 分)	154
24.10 第 11 题: 偏斜正态模型 (原卷 p.4, 15 分)	155

第一部分

Probability Foundations and Limits

第一章 概率空间、随机变量与分布

学习目标

本章建立全书共同语言：从样本空间上的事件与概率测度出发，严格说明随机变量、分布、独立性和随机向量。学完后应能把“抛硬币”“取最大值”“变量变换”等叙述翻译为可计算的测度问题，并能识别离散与连续公式背后的同一结构。

1.1 为什么从概率空间开始

概率题常以随机变量开头，但随机变量本身不是“会随机变化的数”，而是把基本结果映到数轴的可测函数。三元组

$$(\Omega, \mathcal{F}, \mathbb{P})$$

称为概率空间： Ω 是所有基本结果的集合， $\mathcal{F}$ 是允许赋予概率的事件族， $\mathbb{P}$ 是总质量为 1 的测度。若只写 $\mathbb{P}(A)$ 而不说明 A 是否为事件，则极限、条件概率和积分中的可测性问题都会被掩盖。

定义 1.1 (概率空间). 集合族 $\mathcal{F} \subseteq 2^\Omega$ 若满足

$$\Omega \in \mathcal{F}, \quad A \in \mathcal{F} \implies A^c \in \mathcal{F}, \quad A_n \in \mathcal{F} \implies \bigcup_{n \geq 1} A_n \in \mathcal{F},$$

则称为 σ -代数。若 $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$ 满足 $\mathbb{P}(\Omega) = 1$ ，且对两两不交的 A_n 有

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) = \sum_{n \geq 1} \mathbb{P}(A_n),$$

则 $(\Omega, \mathcal{F}, \mathbb{P})$ 是概率空间。

命题 1.1 (测度的连续性). 若 $A_n \uparrow A$ ，则 $\mathbb{P}(A_n) \uparrow \mathbb{P}(A)$ ；若 $A_n \downarrow A$ ，则 $\mathbb{P}(A_n) \downarrow \mathbb{P}(A)$ 。

证明. 令 $B_1 = A_1$ ， $B_n = A_n \setminus A_{n-1}$ 。这些集合两两不交且 $A = \bigcup_n B_n$ ，于是

$$\mathbb{P}(A_n) = \sum_{k \leq n} \mathbb{P}(B_k) \uparrow \sum_{k \geq 1} \mathbb{P}(B_k) = \mathbb{P}(A).$$

递减情形对补集使用递增结论即可；这里用到了 $\mathbb{P}(\Omega) < \infty$ 。

□

易错点

有限可加不能替代可列可加。概率极限中出现的 $\limsup A_n$ 、 $\liminf A_n$ 都含可列并交；若只掌握有限并的公式，Borel–Cantelli 引理和几乎处处收敛便无从成立。

1.2 由集合运算读懂随机事件

对任意事件列 (A_n) ，定义

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n, \quad \liminf_{n \rightarrow \infty} A_n = \bigcup_{m \geq 1} \bigcap_{n \geq m} A_n.$$

前者表示“无穷多个 A_n 发生”，也记作 $\{A_n \text{ i.o.}\}$ ；后者表示“从某一项起所有 A_n 都发生”。逐点看指标函数可得

$$\mathbf{1}_{\liminf A_n} \leq \liminf_n \mathbf{1}_{A_n} \leq \limsup_n \mathbf{1}_{A_n} \leq \mathbf{1}_{\limsup A_n}.$$

这些集合恒等式将在第 五 章直接转化为尾事件判断。

例题 1.1 (至少一次与首次发生). 若 A_n 表示第 n 次试验成功，则

$$\{\text{至少成功一次}\} = \bigcup_{n \geq 1} A_n, \quad \{T = n\} = A_n \cap \bigcap_{k < n} A_k^c,$$

其中 $T = \inf\{n : A_n \text{ 发生}\}$ 。若 A_n 独立且 $\mathbb{P}(A_n) = p$ ，则 $\mathbb{P}(T = n) = (1 - p)^{n-1}p$ 。

1.3 随机变量与分布

定义 1.2 (随机变量与分布). 函数 $X : \Omega \rightarrow \mathbb{R}$ 若满足

$$\{X \leq x\} = X^{-1}((-\infty, x]) \in \mathcal{F}, \quad x \in \mathbb{R},$$

则称为实值随机变量。其分布是实数轴上的概率测度

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

分布函数为 $F_X(x) = \mathbb{P}(X \leq x)$ 。

只检查半直线足够，是因为这些半直线生成 Borel σ -代数。分布函数恰好满足：单调不减、右连续、 $F(-\infty) = 0$ 、 $F(+\infty) = 1$ ；反过来，每个满足这些条件的函数都唯一对应一个概率分布。

命题 1.2 (由分布函数计算区间概率). 对 $a < b$,

$$\mathbb{P}(a < X \leq b) = F_X(b) - F_X(a), \quad \mathbb{P}(X = a) = F_X(a) - F_X(a-).$$

因此 F_X 在 a 连续当且仅当 $\mathbb{P}(X = a) = 0$ 。

证明. 事件 $\{X \leq b\}$ 是 $\{X \leq a\}$ 与 $\{a < X \leq b\}$ 的不交并。第二式令 $x \uparrow a$ 并用测度连续性即可。 $\square$

例题 1.2 (混合分布). 设 $\mathbb{P}(X = 0) = 1/3$, 且在事件 $\{X \neq 0\}$ 上 X 服从参数为 1 的指数分布。则

$$F_X(x) = \begin{cases} 0, & x < 0, \\ \frac{1}{3}, & x = 0, \\ 1 - \frac{2}{3}e^{-x}, & x > 0. \end{cases}$$

它既有原子又有密度部分。把“有分布函数”等同于“有密度”会漏掉原子质量。

1.3.1 分布相同不等于随机变量相等

记 $X \stackrel{d}{=} Y$ 表示 $\mu_X = \mu_Y$ 。这只比较边缘分布, 不说明 X, Y 在同一概率空间上的逐点关系。例如 $U \sim \text{Unif}(0, 1)$ 时, $U \stackrel{d}{=} 1 - U$, 但二者通常不相等。独立性更不能由边缘分布决定。

易错点

“同分布”是推前测度相同; “几乎处处相等”是 $\mathbb{P}(X = Y) = 1$ 。后者推出前者, 反向不成立。以后截断变量 $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}$ 一般也不与 X_n 同分布。

1.4 随机向量、联合分布与独立性

随机向量 $X = (X_1, \dots, X_d)$ 是从 Ω 到 $\mathbb{R}^d$ 的可测映射。联合分布函数

$$F_X(x_1, \dots, x_d) = \mathbb{P}(X_1 \leq x_1, \dots, X_d \leq x_d)$$

决定全部联合概率; 各边缘分布只由令其余坐标趋于 $+\infty$ 得到。

定义 1.3 (独立). σ -代数 $\mathcal{G}_i$ 相互独立, 若任取有限个不同指标及 $A_i \in \mathcal{G}_i$ 均有

$$\mathbb{P}\left(\bigcap_i A_i\right) = \prod_i \mathbb{P}(A_i).$$

随机变量 X_i 相互独立, 若 $\sigma(X_i)$ 相互独立。对两个变量, 等价地, 所有 Borel 集 A, B 满足

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

命题 1.3 (独立变量的函数仍独立). 若 $X_1, \dots, X_k$ 相互独立, 且 g_i 为 Borel 可测函数, 则 $g_1(X_1), \dots, g_k(X_k)$ 相互独立。

证明. $\{g_i(X_i) \in B_i\} = \{X_i \in g_i^{-1}(B_i)\} \in \sigma(X_i)$, 代入独立 σ -代数的定义即可。□

反例 1.1 (两两独立不推出相互独立). 令 U, V 为独立 Rademacher 变量, 取 $W = UV$ 。任意两个都独立, 但 $UVW = 1$ 恒成立, 故三者不相互独立。资格考试中若要把多个概率相乘, 必须确认所需的是相互独立还是仅两两独立。

1.5 变量变换：先找原像，再求分布

设 $Y = g(X)$ 。最稳妥的起点不是背 Jacobian，而是

$$\mathbb{P}(Y \in B) = \mathbb{P}(X \in g^{-1}(B)).$$

若 g 严格递增， $F_Y(y) = F_X(g^{-1}(y))$ ；若 g 严格递减， $F_Y(y) = 1 - F_X(g^{-1}(y))$ 。当 X 有密度且 g 分段一一可微时，

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

例题 1.3 (平方变换必须累加所有原像). 若 $X \sim N(0, 1)$, $Y = X^2$, 则 $y > 0$ 有两个原像 $\pm\sqrt{y}$,

$$f_Y(y) = \frac{\phi(\sqrt{y}) + \phi(-\sqrt{y})}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}.$$

漏掉负根会使密度积分只有 $1/2$ 。

定理 1.4 (二维变换公式). 设 (X, Y) 有联合密度 $f_{X,Y}$, 映射 $T(x, y) = (u, v)$ 在区域内一一、连续可微, 逆映射为 $(x(u, v), y(u, v))$, 且 Jacobian 非零, 则

$$f_{U,V}(u, v) = f_{X,Y}(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$$

在像区域内成立, 区域外为零。

证明. 对任意有界可测函数 φ , 普通多元换元公式给出

$$\begin{aligned} \mathbb{E}\varphi(U, V) &= \iint \varphi(T(x, y)) f_{X,Y}(x, y) dx dy \\ &= \iint \varphi(u, v) f_{X,Y}(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv. \end{aligned}$$

由密度刻画即得结论。若映射不是一一的，应分支求和。 $\square$

例题 1.4 (和与比例). 若 X, Y 独立且均服从 $\text{Exp}(1)$, 令 $S = X + Y$, $U = X/(X + Y)$ 。逆变换为 $x = su, y = s(1 - u)$, Jacobian 为 s , 像区域为 $s > 0, 0 < u < 1$ 。故

$$f_{S,U}(s, u) = e^{-s} s \mathbf{1}_{\{s>0, 0<u<1\}},$$

因而 $S \sim \Gamma(2, 1)$ 、 $U \sim \text{Unif}(0, 1)$, 且二者独立。这里“密度因子可分离”只有在支撑集也是直积时才推出独立。

1.6 分位数与概率积分变换

定义 1.4 (广义逆). 分布函数 F 的广义逆定义为

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}, \quad 0 < u < 1.$$

定理 1.5 (概率积分变换). 若 $U \sim \text{Unif}(0, 1)$, 则 $F^{-1}(U)$ 的分布函数为 F . 若 F 连续, 则 $F(X) \sim \text{Unif}(0, 1)$; 若 F 有跳跃, 后一结论一般不成立。

证明. 广义逆满足 $F^{-1}(u) \leq x$ 当且仅当 $u \leq F(x)$, 故

$$\mathbb{P}(F^{-1}(U) \leq x) = \mathbb{P}(U \leq F(x)) = F(x).$$

连续情形可由广义逆关系或分位数逼近得到第二个结论。□

答题方法

处理分布与变换题时依次检查:

1. 写清随机对象及其支撑;
2. 用事件原像先求分布函数, 或写出完整逆变换;
3. 多对一映射对所有原像求和;
4. 变换后重新写支撑集, 不能只变公式;
5. 用总积分为 1、边缘分布或极端情形作校验。

本节结论

概率空间提供可列运算, 随机变量把样本空间推到可计算的状态空间, 分布记录了单个变量的全部概率信息, 但不记录变量之间的耦合。联合分布与独立性是多变量计算的核心; 变量变换的本质始终是求事件原像。

自测

1. 构造三个两两独立但不相互独立的 Bernoulli 变量, 并指出哪一个乘积公式失败。
2. 设 X 的分布函数在 0 处跳跃 $1/4$, 求 $\mathbb{P}(X = 0)$ 。
3. 若 X, Y 独立标准正态, 求 $(X + Y, X - Y)$ 的联合分布并判断独立性。
4. 说明为何对含原子的分布, $F(X)$ 未必均匀。

自测核对要点 第 1 题可取 $U, V, U \oplus V$; 三者同为 Bernoulli(1/2) 且两两独立, 但第三个由前两个确定。第 2 题答案为 $1/4$ 。第 3 题所得向量为中心 Gaussian, 两坐标协方差为 0、方差均为 2, 故独立。第 4 题中 $F(X)$ 在跳点会产生原子, 不可能具有连续均匀分布。

1.7 测度连续性与概率的可列结构

若事件 $A_n \uparrow A$, 可写成不交并

$$A = A_1 \dot{\cup} \bigcup_{n \geq 2} (A_n \setminus A_{n-1}),$$

由可列可加性得到

$$\mathbb{P}(A_n) \uparrow \mathbb{P}(A).$$

若 $A_n \downarrow A$, 因为概率测度有限,

$$\mathbb{P}(A_n) = \mathbb{P}(A_1) - \mathbb{P}(A_1 \setminus A_n) \downarrow \mathbb{P}(A_1) - \mathbb{P}(A_1 \setminus A) = \mathbb{P}(A).$$

第二个结论对一般无限测度需要 $\mu(A_1) < \infty$. 概率空间自动满足这一有限性。

Boole 不等式由不交化得到

$$\mathbb{P}\left(\bigcup_{n \geq 1} A_n\right) \leq \sum_{n \geq 1} \mathbb{P}(A_n).$$

它不要求独立。Borel–Cantelli 第一引理正是在尾并上使用它。

1.8 分布函数完整刻画分布

随机变量 X 的分布函数

$$F_X(x) = \mathbb{P}(X \leq x)$$

单调不减、右连续, 满足

$$\lim_{x \rightarrow -\infty} F_X(x) = 0, \quad \lim_{x \rightarrow \infty} F_X(x) = 1.$$

反过来, 每个满足这些性质的函数都唯一决定一个 Borel 概率测度。跳跃大小

$$F_X(x) - F_X(x-) = \mathbb{P}(X = x)$$

给原子概率; 连续不代表有密度, Cantor 分布是连续但奇异的标准反例。

1.9 多维变量变换与 Jacobian

若 $Y = T(X)$, $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$ 在有关区域一一可微且逆 Jacobian 非奇异, 则

$$f_Y(y) = f_X(T^{-1}y) |\det DT^{-1}(y)|.$$

若映射多对一, 要对所有逆像分支求和。

例题 1.5 (比值与和). 设 X, Y 独立指数分布, 参数同为 1. 令

$$S = X + Y, \quad U = \frac{X}{X + Y}.$$

逆变换 $X = US, Y = (1 - U)S$, 区域 $s > 0, 0 < u < 1$, Jacobian 绝对值为 s . 于是

$$f_{S,U}(s, u) = e^{-s}s.$$

它分解为

$$(se^{-s}) \cdot 1,$$

故 $S \sim \Gamma(2, 1)$, $U \sim \text{Unif}(0, 1)$, 且二者独立。

1.10 概率积分变换与分位数构造

若 F 连续, 则 $F(X) \sim \text{Unif}(0, 1)$ 。证明: 对 $0 < u < 1$, 利用广义逆

$$F^{-1}(u) = \inf\{x : F(x) \geq u\}$$

处理平坦区, 可得

$$\mathbb{P}(F(X) \leq u) = u.$$

若 F 有跳跃, $F(X)$ 一般不均匀。反向构造始终成立: 若 $U \sim \text{Unif}(0, 1)$, 则

$$X = F^{-1}(U)$$

具有分布函数 F 。这是逆变换采样的数学基础。

易错点

“同分布”只说明推前测度相同。它不推出两个随机变量定义在同一概率空间, 不推出逐点相等, 也不推出相互独立。耦合题必须另行指定联合分布。

自测

1. 证明递减事件列的概率连续性, 并指出有限测度条件。
2. 分布函数的跳跃如何对应原子?
3. 多对一变量变换为什么要对逆像分支求和?
4. $F(X)$ 在 F 不连续时为何未必均匀?

第二章 积分、期望与概率不等式

学习目标

本章把期望理解为 Lebesgue 积分，而不是只会“密度乘函数”。重点掌握非负函数逼近、三大收敛定理、Tonelli–Fubini 的使用条件、尾积分公式和常用矩不等式。这些工具将直接支撑条件期望、强大数律和鞅收敛。

2.1 从简单函数到期望

非负简单随机变量可写成 $S = \sum_{k=1}^m a_k \mathbf{1}_{A_k}$ ，其中 $a_k \geq 0$ 、 A_k 两两不交。定义

$$\mathbb{E}S = \sum_{k=1}^m a_k \mathbb{P}(A_k).$$

对非负随机变量 X ，取简单函数 $0 \leq S_n \uparrow X$ ，定义

$$\mathbb{E}X = \sup\{\mathbb{E}S : 0 \leq S \leq X, S \text{ 为简单函数}\}.$$

这个值允许为 $+\infty$ 。一般实值 X 写作 $X = X^+ - X^-$ ；当 $\mathbb{E}X^+ + \mathbb{E}X^- < \infty$ 时称 X 可积，并定义 $\mathbb{E}X = \mathbb{E}X^+ - \mathbb{E}X^-$ 。若正负两部分都为无穷，期望没有定义，不能写成“ $\infty - \infty = 0$ ”。

命题 2.1 (只由分布决定的积分). 若 g 为 Borel 可测且 $g(X)$ 非负或可积，则

$$\mathbb{E}g(X) = \int_{\mathbb{R}} g(x) \mu_X(\mathrm{d}x).$$

特别地，若 μ_X 有密度 f_X ，则右端为 $\int g(x)f_X(x) \mathrm{d}x$ ；离散情形则为求和。

证明. 对指标函数 $g = \mathbf{1}_B$ ，等式就是分布的定义；由线性推广到简单函数，再以单调收敛推广到非负函数，最后分解正负部分。□

2.2 四个基本极限定理

定理 2.2 (单调收敛定理). 若 $0 \leq X_n \uparrow X$ 几乎处处，则

$$\mathbb{E}X_n \uparrow \mathbb{E}X.$$

证明. 单调性给出 $\lim_n \mathbb{E}X_n \leq \mathbb{E}X$ 。反向对任意非负简单函数 $S \leq X$ 和 $0 < c < 1$ ，令 $A_n =$

$\{X_n \geq cS\}$ 。逐点有 $A_n \uparrow \{S > 0\}$, 故

$$\liminf_n \mathbb{E}X_n \geq \lim_n \mathbb{E}(cS\mathbf{1}_{A_n}) = c\mathbb{E}S.$$

依次对所有 $S \leq X$ 取上确界并令 $c \uparrow 1$ 。 □

定理 2.3 (Fatou 引理). 若 $X_n \geq 0$, 则

$$\mathbb{E}(\liminf_n X_n) \leq \liminf_n \mathbb{E}X_n.$$

证明. 令 $Y_m = \inf_{n \geq m} X_n$, 则 $Y_m \uparrow \liminf_n X_n$ 。由单调收敛,

$$\mathbb{E} \liminf_n X_n = \lim_m \mathbb{E}Y_m \leq \lim_m \inf_{n \geq m} \mathbb{E}X_n.$$

□

定理 2.4 (控制收敛定理). 若 $X_n \rightarrow X$ 几乎处处, 且存在可积 Y 使 $|X_n| \leq Y$, 则 X 可积且

$$\mathbb{E}|X_n - X| \rightarrow 0, \quad \mathbb{E}X_n \rightarrow \mathbb{E}X.$$

证明. 极限保持 $|X| \leq Y$ 。对非负变量 $2Y - |X_n - X|$ 用 Fatou 引理:

$$2\mathbb{E}Y \leq \liminf_n (2\mathbb{E}Y - \mathbb{E}|X_n - X|),$$

从而 $\limsup_n \mathbb{E}|X_n - X| \leq 0$ 。最后 $|\mathbb{E}X_n - \mathbb{E}X| \leq \mathbb{E}|X_n - X|$ 。 □

定理 2.5 (Tonelli–Fubini). 设 f 是乘积空间上的可测函数。

1. 若 $f \geq 0$, 则两个迭代积分与乘积积分相等, 值允许为 $+\infty$;
2. 若 $\iint |f| < \infty$, 则 f 的两个迭代积分几乎处处存在且有限, 三者相等。

易错点

Tonelli 允许先交换非负积分再判断是否有限; Fubini 对有正有负的函数要求绝对可积。条件收敛的二重级数随求和顺序改变, 是“随意交换期望与求和”最典型的反例。

例题 2.1 (期望与无穷和交换). 若 $X_n \geq 0$, 则 Tonelli 给出

$$\mathbb{E} \sum_{n \geq 1} X_n = \sum_{n \geq 1} \mathbb{E}X_n,$$

无须预先知道任何一边有限。若右端有限, 则左端有限, 从而 $\sum_n X_n < \infty$ 几乎处处。这一简单结论常用于证明随机级数绝对收敛。

2.3 尾积分公式与矩

定理 2.6 (尾积分公式). 若 $X \geq 0$, 则对 $p > 0$,

$$\mathbb{E}X^p = p \int_0^\infty t^{p-1} \mathbb{P}(X > t) dt.$$

证明. 逐点恒等式

$$X^p = \int_0^\infty p t^{p-1} \mathbf{1}_{\{X > t\}} dt$$

由普通微积分成立; 因被积函数非负, Tonelli 允许交换积分与期望。□

例题 2.2 (幂尾与矩的临界阶). 若存在常数 $c_1, c_2 > 0$ 使 $c_1 t^{-\alpha} \leq \mathbb{P}(X > t) \leq c_2 t^{-\alpha}$ 对充分大 t 成立, 则 $\mathbb{E}X^p < \infty$ 当且仅当 $p < \alpha$ 。资格考试中出现幂尾时, 先比较 $p - 1 - \alpha$ 与 -1 , 比直接求密度更快。

2.4 四类常用不等式

定理 2.7 (Markov、Chebyshev 与 Jensen). 1. 若 $X \geq 0$, 则 $\mathbb{P}(X \geq a) \leq \mathbb{E}X/a$;

2. 若 $X \in L^2$, 则 $\mathbb{P}(|X - \mathbb{E}X| \geq t) \leq \text{Var}(X)/t^2$;

3. 若 φ 为凸函数且相关期望存在, 则 $\varphi(\mathbb{E}X) \leq \mathbb{E}\varphi(X)$ 。

证明. 第一式由 $X \geq a \mathbf{1}_{\{X \geq a\}}$; 第二式对 $(X - \mathbb{E}X)^2$ 使用第一式。对 Jensen, 先设 φ 在 $m = \mathbb{E}X$ 处有支撑直线 $\varphi(x) \geq \varphi(m) + s(x - m)$, 两边取期望。一般凸函数可用次梯度或支撑线逼近。□

定理 2.8 (Hölder 与 Minkowski). 若 $p, q > 1$ 且 $1/p + 1/q = 1$, 则

$$\mathbb{E}|XY| \leq \|X\|_p \|Y\|_q.$$

对 $p \geq 1$,

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p.$$

证明. 先把 X, Y 除以各自范数, 再用 Young 不等式 $ab \leq a^p/p + b^q/q$, 得到 Hölder。若 $\|X + Y\|_p > 0$, 则

$$\mathbb{E}|X + Y|^p \leq \mathbb{E}(|X| |X + Y|^{p-1}) + \mathbb{E}(|Y| |X + Y|^{p-1}).$$

分别使用 Hölder, 再除以 $\|X + Y\|_p^{p-1}$, 即得 Minkowski。□

推论 2.9 (Cauchy-Schwarz 与 L^p 嵌入).

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

在概率空间上若 $1 \leq p \leq q$, 则 $\|X\|_p \leq \|X\|_q$ 。

2.5 生成函数：何时能用，能推出什么

矩母函数 $M_X(t) = \mathbb{E}e^{tX}$ 若在 0 的邻域内有限，则唯一决定分布，并可在适当条件下逐次求导得到矩。特征函数

$$\phi_X(t) = \mathbb{E}e^{itX}$$

对所有实 t 都存在，且同样唯一决定分布。独立和满足

$$\phi_{X+Y}(t) = \phi_X(t)\phi_Y(t).$$

第七章将用它识别正态极限。

命题 2.10 (Chernoff 方法). 若 $M_X(t) < \infty$ 对某个 $t > 0$ 成立, 则

$$\mathbb{P}(X \geq x) \leq \inf_{t>0} e^{-tx} M_X(t),$$

其中下确界只在矩母函数有限的 t 上取。

证明. 对非负变量 e^{tX} 用 Markov 不等式:

$$\mathbb{P}(X \geq x) = \mathbb{P}(e^{tX} \geq e^{tx}) \leq e^{-tx} \mathbb{E}e^{tX},$$

再优化 t 。 □

例题 2.3 (Bernoulli 和的一个指数界). 若 B_i 独立且 $\mathbb{P}(B_i = 1) = p$, $S_n = \sum_i B_i$, 则对 $a > p$,

$$\mathbb{P}(S_n \geq na) \leq \inf_{t>0} \exp\{n[\log(1-p+pe^t) - ta]\}.$$

最优 t 给出二元相对熵 $D(a||p)$, 故概率至多 $e^{-nD(a||p)}$ 。在第六章的区块计数中, 只需这一指数衰减, 不必追求最优常数。

反例 2.1 (所有矩存在仍未必由矩唯一决定). 对数正态分布具有任意阶有限矩, 但存在不同分布与它有同样的全部矩。因此“矩相同”只有在矩确定性条件下才推出同分布; 资格考试更安全的识别工具是特征函数唯一性。

2.6 特征函数：唯一性、反演与 Lévy 连续性

特征函数 $\phi_X(t) = \mathbb{E}e^{itX}$ 总存在, 满足 $\phi_X(0) = 1$ 、 $\phi_X(-t) = \overline{\phi_X(t)}$ 及一致连续性。若 X, Y 独立, 则 $\phi_{X+Y} = \phi_X\phi_Y$; 若 $a, b \in \mathbb{R}$, 则 $\phi_{aX+b}(t) = e^{itb}\phi_X(at)$ 。这些公式都直接来自期望与独立乘积分解, 不能在不独立时使用。

定理 2.11 (反演与唯一性). 在分布函数 F 的连续点 $a < b$, 有

$$\mathbb{P}(a < X < b) + \frac{1}{2}\mathbb{P}(X = a) + \frac{1}{2}\mathbb{P}(X = b) = \lim_{T \rightarrow \infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-ita} - e^{-itb}}{it} \phi_X(t) dt.$$

因此特征函数唯一决定分布。若 $\phi_X \in L^1(\mathbb{R})$, 则 X 有连续密度

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_X(t) dt.$$

证明. 先把区间指标与 Dirichlet 核的截断 Fourier 表示相卷积。对固定 $x \neq a, b$, 截断积分趋于 $\mathbf{1}_{(a,b)}(x)$; 在端点趋于 $1/2$ 。核的对称截断和分部积分给一致可积的尾控制, 于是可对 X 取期望, 得到第一式。若两个分布有同一特征函数, 第一式给出它们在所有连续端点区间上的概率相同, 故分布相同。

当 $\phi_X \in L^1$ 时, Fourier 逆变换定义的 f 连续。对任意光滑紧支撑函数 g , Fubini 给

$$\int g(x) f(x) dx = \frac{1}{2\pi} \int \hat{g}(-t) \phi_X(t) dt = \mathbb{E}g(X).$$

以光滑函数逼近区间指标便知 f 是 X 的密度。 □

命题 2.12 (矩与零点展开). 若 $\mathbb{E}|X|^k < \infty$, 则 ϕ_X 可微到 k 阶且

$$\phi_X^{(j)}(t) = \mathbb{E}[(iX)^j e^{itX}], \quad \phi_X^{(j)}(0) = i^j \mathbb{E}X^j, \quad 0 \leq j \leq k.$$

特别地, 若 $\mathbb{E}X = 0, \mathbb{E}X^2 = \sigma^2 < \infty$,

$$\phi_X(t) = 1 - \frac{\sigma^2 t^2}{2} + o(t^2).$$

反向推断矩存在需要额外条件; 仅凭 ϕ 在零点形式上可微不能无条件交换导数与期望。

证明. 对差商使用 $|(e^{ihX} - 1)/h| \leq |X|$, 再逐阶用控制收敛。二阶 Peano 展开应用于 ϕ 即得末式。 □

定理 2.13 (Lévy 连续性定理). $X_n \Rightarrow X$ 当且仅当 $\phi_{X_n}(t) \rightarrow \phi_X(t)$ 对每个 t 成立。更一般地, 若 $\phi_{X_n}(t) \rightarrow \phi(t)$ 且 ϕ 在 0 连续, 则 ϕ 是某个概率分布的特征函数, 并且 X_n 依分布收敛到该分布。

证明. 正向由有界连续函数 $x \mapsto e^{itx}$ 的弱收敛刻画立即得到。反向的关键是零点连续性给紧性: 选小 $u > 0$ 使 $1 - \Re\phi(t)$ 在 $[-u, u]$ 上小; 利用

$$\frac{1}{2u} \int_{-u}^u (1 - \cos(tx)) dt$$

在 $|x|$ 大时有统一正下界, 可把 $\sup_{n \geq n_0} \mathbb{P}(|X_n| > R)$ 压小, 有限个早期 n 再单独处理。由紧性取任意弱收敛子列, 其极限特征函数由正向部分必为 ϕ ; 定理 2.11 的唯一性说明所有子列极限相同, 故全列收敛。 □

例题 2.4 (正态和的识别). 若独立 $X_j \sim N(0, \sigma_j^2)$, 则

$$\phi_{\sum_j X_j}(t) = \prod_j e^{-\sigma_j^2 t^2 / 2} = e^{-(\sum_j \sigma_j^2) t^2 / 2},$$

故和仍为正态。若没有独立性, 乘积步骤失效, 单知道各边缘正态不足以推出和的上述方差公式。

答题方法

特征函数题按“存在性—独立乘积—零点展开—逐点极限—检查极限在零点连续—Lévy 定理”作答。需要由特征函数反推分布时写明使用唯一性或反演条件；需要求密度时额外检查 $\phi \in L^1$ ，不能把形式 Fourier 逆变换当作自动合法。

答题方法

面对“极限与期望能否交换”，先按下面次序寻找工具：

1. 非负且单调：单调收敛；
2. 非负但不单调：Fatou 给单侧不等式；
3. 有共同可积控制：控制收敛；
4. 只有尾部可控：先证明一致可积，再用第 四 章的 Vitali 判别；
5. 涉及无穷和或二重积分：先判断 Tonelli 的非负性或 Fubini 的绝对可积性。

本节结论

Lebesgue 积分把离散、连续和混合分布统一起来。单调收敛与 Fatou 处理非负量，控制收敛处理共同上界，Tonelli–Fubini 控制交换顺序；尾积分和矩不等式则把概率尾界与积分矩相互转换。

自测

1. 构造 $X_n \rightarrow 0$ 几乎处处但 $\mathbb{E}X_n = 1$ 的例子，并指出控制收敛缺少什么。
2. 证明若 $\sum_n \mathbb{E}|X_n| < \infty$ ，则 $\sum_n |X_n| < \infty$ 几乎处处。
3. 若 $\mathbb{P}(X > t) \leq e^{-t}$ ，证明 $\mathbb{E}X^p \leq \Gamma(p+1)$ 。
4. 说明为什么矩母函数在 0 处有限而非在邻域有限，不足以逐项求导。

自测核对要点 第 1 题可取 $X_n = n\mathbf{1}_{(0,1/n)}$ ；缺少共同可积控制，也不一致可积。第 2 题对非负级数用 Tonelli。第 3 题代入尾积分公式。第 4 题需要在 0 附近有指数可积控制，才能把导数与期望交换。

2.7 单调收敛、Fatou 与支配收敛的依赖链

若 $0 \leq X_n \uparrow X$ ，单调收敛定理给

$$\mathbb{E}X_n \uparrow \mathbb{E}X$$

，允许右侧为无穷。对任意非负 X_n ，令

$$Y_m = \inf_{n \geq m} X_n.$$

$Y_m \uparrow \liminf X_n$, 且 $Y_m \leq X_n$ 对 $n \geq m$, 故

$$\mathbb{E} \liminf X_n = \lim_m \mathbb{E} Y_m \leq \liminf_n \mathbb{E} X_n.$$

这就是 Fatou 引理。

若 $X_n \rightarrow X$ a.s. 且 $|X_n| \leq Y \in L^1$, 对 $Y \pm X_n \geq 0$ 使用 Fatou, 得到

$$\mathbb{E}(Y + X) \leq \liminf \mathbb{E}(Y + X_n), \quad \mathbb{E}(Y - X) \leq \liminf \mathbb{E}(Y - X_n),$$

合并推出 $\mathbb{E} X_n \rightarrow \mathbb{E} X$ 。进一步对 $|X_n - X| \leq 2Y$ 再用支配收敛, 得 L^1 收敛。

反例 2.2.

$$X_n = n \mathbf{1}_{(0, 1/n)}$$

在 $(0, 1)$ 上 a.s. 收敛到零, 但 $\mathbb{E} X_n = 1$ 。没有共同可积支配, 也不一致可积, 故不能交换极限与期望。

2.8 尾积分公式与矩的层蛋糕表示

对非负随机变量,

$$X = \int_0^\infty \mathbf{1}\{X > t\} dt.$$

Tonelli 定理给

$$\mathbb{E} X = \int_0^\infty \mathbb{P}(X > t) dt.$$

更一般地, 对 $p > 0$,

$$X^p = \int_0^\infty p t^{p-1} \mathbf{1}\{X > t\} dt,$$

所以

$$\mathbb{E} X^p = p \int_0^\infty t^{p-1} \mathbb{P}(X > t) dt.$$

这把矩条件转化为尾概率可积性, 是截断与极值题的常用入口。

2.9 Jensen、条件 Jensen 与等号

若 φ 凸且有关期望存在,

$$\varphi(\mathbb{E} X) \leq \mathbb{E} \varphi(X).$$

条件版本为

$$\varphi(\mathbb{E}[X | \mathcal{G}]) \leq \mathbb{E}[\varphi(X) | \mathcal{G}] \quad a.s.$$

若 φ 严格凸, 等号要求 X 在相应条件下几乎处处为常数。对 $\varphi(x) = e^x$ 得指数矩控制; 对 $\varphi(x) = x^2$ 得

$$(\mathbb{E}[X | \mathcal{G}])^2 \leq \mathbb{E}[X^2 | \mathcal{G}].$$

2.10 Chernoff 方法的优化步骤

对任意 $t > 0$, Markov 不等式给

$$\mathbb{P}(X \geq a) = \mathbb{P}(e^{tX} \geq e^{ta}) \leq e^{-ta} \mathbb{E}e^{tX}.$$

所以

$$\mathbb{P}(X \geq a) \leq \inf_{t>0} \exp\{\log M_X(t) - ta\}.$$

关键不只是“套 MGF”，而是检查 MGF 在所选 t 有限并优化。若 $S_n \sim \text{Bin}(n, p)$, 对 $a = nq$, $q > p$, 优化给

$$\begin{aligned} \mathbb{P}(S_n \geq nq) &\leq \exp\{-nD(q\|p)\}, \\ D(q\|p) &= q \log \frac{q}{p} + (1-q) \log \frac{1-q}{1-p}. \end{aligned}$$

2.11 生成函数的唯一性与解析边界

概率生成函数

$$G_X(s) = \mathbb{E}s^X$$

适用于非负整数值变量, $|s| \leq 1$ 自动有限。独立和满足乘法

$$G_{X+Y} = G_X G_Y.$$

在 $s = 1-$ 的导数给阶乘矩:

$$G'_X(1-) = \mathbb{E}X, \quad G''_X(1-) = \mathbb{E}[X(X-1)]$$

, 允许无穷。MGF 在零邻域有限时唯一决定分布; 只在单点有限不够, 重尾分布可能除 $t = 0$ 外正侧 MGF 全为无穷。

答题方法

交换极限与期望前依次检查: 非负单调、共同可积支配、一致可积。尾概率题先尝试层蛋糕; 指数尾题写出 Chernoff 的参数优化, 而不是只给未优化的 Markov 界。

自测

1. 从单调收敛证明 Fatou, 再由 Fatou 推出支配收敛。
2. 推导 p 阶矩的尾积分公式。
3. 条件 Jensen 的等号在严格凸时意味着什么?
4. Chernoff 界中为什么必须检查 MGF 的定义域?

第三章 条件期望、条件分布与条件独立

学习目标

本章把条件期望从“条件密度公式”提升为关于信息的随机变量。需要掌握条件期望的测度论定义、几乎处处唯一性、塔式性质、提出已知量和 L^2 投影解释，并能在离散、连续及 Gaussian 情形中完成计算。

3.1 为什么条件期望仍是随机变量

给定子 σ -代数 $\mathcal{G} \subseteq \mathcal{F}$ ，它代表已经知道的信息。 $\mathbb{E}(X | \mathcal{G})$ 必须依赖这份信息而变化，所以不是通常意义下的常数。其定义只要求两件事：对 $\mathcal{G}$ 可测；在每个 $G \in \mathcal{G}$ 上保留原变量的平均质量。

定义 3.1 (条件期望). 若 $X \in L^1$ ，随机变量 Y 称为 X 关于 $\mathcal{G}$ 的条件期望，若

$$Y \text{ 是 } \mathcal{G}\text{-可测的, } \int_G Y \, d\mathbb{P} = \int_G X \, d\mathbb{P}, \quad G \in \mathcal{G}.$$

记 $Y = \mathbb{E}(X | \mathcal{G})$ 。两个版本只要求几乎处处相等。

定理 3.1 (存在唯一性). 对每个 $X \in L^1$ ， $\mathbb{E}(X | \mathcal{G})$ 存在且几乎处处唯一。

证明. 先设 $X \geq 0$ ，在 $(\Omega, \mathcal{G})$ 上定义有限测度 $\nu(G) = \int_G X \, d\mathbb{P}$ 。由于 $\nu \ll \mathbb{P}|_{\mathcal{G}}$ ，Radon-Nikodym 定理给出 $\mathcal{G}$ -可测 $Y \geq 0$ 使 $\nu(G) = \int_G Y \, d\mathbb{P}$ 。一般 X 对 X^+, X^- 分别处理。

若 Y, Z 都满足定义，令 $G = \{Y > Z\} \in \mathcal{G}$ ，则 $\int_G (Y - Z) \, d\mathbb{P} = 0$ ，故 $Y \leq Z$ 几乎处处；交换二者得到唯一性。□

3.2 必须熟练的运算规则

定理 3.2 (条件期望规则). 设下列变量均满足所需可积性。

1. 线性: $\mathbb{E}(aX + bY | \mathcal{G}) = a\mathbb{E}(X | \mathcal{G}) + b\mathbb{E}(Y | \mathcal{G})$;
2. 单调性: $X \leq Y$ 推出 $\mathbb{E}(X | \mathcal{G}) \leq \mathbb{E}(Y | \mathcal{G})$;
3. 提出已知量: 若 Z 为 $\mathcal{G}$ -可测且 $ZX \in L^1$ ，则

$$\mathbb{E}(ZX | \mathcal{G}) = Z\mathbb{E}(X | \mathcal{G});$$

4. 塔式性质: 若 $\mathcal{H} \subseteq \mathcal{G}$ ，则

$$\mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}] = \mathbb{E}(X | \mathcal{H});$$

5. 若 X 与 $\mathcal{G}$ 独立, 则 $\mathbb{E}(X | \mathcal{G}) = \mathbb{E}X$;

6. 条件 Jensen: 凸 φ 满足

$$\varphi(\mathbb{E}[X | \mathcal{G}]) \leq \mathbb{E}[\varphi(X) | \mathcal{G}].$$

证明. 线性与单调性由定义和唯一性得到. 提出已知量先对 $Z = \mathbf{1}_A, A \in \mathcal{G}$ 验证, 再推广到简单函数、非负变量, 最后截断到一般可积情形. 塔式性质中, 内层条件期望已经是 $\mathcal{G}$ -可测; 对每个 $H \in \mathcal{H}$,

$$\int_H \mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}] d\mathbb{P} = \int_H \mathbb{E}(X | \mathcal{G}) d\mathbb{P} = \int_H X d\mathbb{P}.$$

独立情形验证常数 $\mathbb{E}X$ 满足定义. 条件 Jensen 可先对支撑直线证明, 再作可测逼近. □

易错点

塔式性质要求 σ -代数嵌套. 对任意两个信息族 $\mathcal{G}, \mathcal{H}$, 一般不能把 $\mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}]$ 简化为某个“较小者”。

3.3 L^2 中的正交投影

若 $X \in L^2$, 则 $\mathbb{E}(X | \mathcal{G})$ 是闭子空间

$$L^2(\mathcal{G}) = \{Z \in L^2 : Z \text{ 为 } \mathcal{G}\text{-可测}\}$$

上的正交投影。

定理 3.3 (最小均方误差). 若 $X \in L^2, Y = \mathbb{E}(X | \mathcal{G})$, 则对任意 $Z \in L^2(\mathcal{G})$,

$$\mathbb{E}(X - Y)Z = 0, \quad \mathbb{E}(X - Z)^2 = \mathbb{E}(X - Y)^2 + \mathbb{E}(Y - Z)^2.$$

因此 Y 是所有基于 $\mathcal{G}$ 的平方损失预测器中唯一的最优者。

证明. 先对有界 $\mathcal{G}$ -可测 Z 用提出已知量:

$$\mathbb{E}[(X - Y)Z] = \mathbb{E}\{\mathbb{E}[(X - Y)Z | \mathcal{G}]\} = 0.$$

对一般 $Z \in L^2$ 截断并用 Cauchy-Schwarz 取极限. 展开 $X - Z = (X - Y) + (Y - Z)$, 交叉项为零, 即得 Pythagoras 恒等式. □

推论 3.4 (全方差公式). 若 $X \in L^2$, 则

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | \mathcal{G})] + \text{Var}(\mathbb{E}[X | \mathcal{G}]).$$

3.4 按随机变量条件化

记 $\mathbb{E}(X | Y) = \mathbb{E}(X | \sigma(Y))$. Doob-Dynkin 引理说明存在 Borel 函数 g 使 $\mathbb{E}(X | Y) = g(Y)$ 几乎处处. 计算任务是确定这个 g .

3.4.1 离散情形

若 $\mathbb{P}(Y = y) > 0$, 则

$$g(y) = \mathbb{E}(X | Y = y) = \sum_x x \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)}.$$

右边只在边缘概率为正的点有约束；零概率点如何定义不影响条件期望版本。

定理 3.5 (全概率公式与 Bayes 公式). 若 $B_1, B_2, \dots$ 两两不交、 $\bigcup_k B_k = \Omega$, 且 $\mathbb{P}(B_k) > 0$, 则对任意事件 A ,

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A | B_k) \mathbb{P}(B_k).$$

若再有 $\mathbb{P}(A) > 0$, 则

$$\mathbb{P}(B_j | A) = \frac{\mathbb{P}(A | B_j) \mathbb{P}(B_j)}{\sum_k \mathbb{P}(A | B_k) \mathbb{P}(B_k)}.$$

证明. 由 $A = \bigcup_k (A \cap B_k)$ 的不交并和概率的可列可加性,

$$\mathbb{P}(A) = \sum_k \mathbb{P}(A \cap B_k) = \sum_k \mathbb{P}(A | B_k) \mathbb{P}(B_k).$$

再把条件概率定义 $\mathbb{P}(B_j | A) = \mathbb{P}(A \cap B_j) / \mathbb{P}(A)$ 的分子写成 $\mathbb{P}(A | B_j) \mathbb{P}(B_j)$, 分母使用刚证得的全概率公式, 即得 Bayes 公式. 条件 $\mathbb{P}(A) > 0$ 不能省略, 因为它是后验条件概率的分母. $\square$

例题 3.1 (诊断试验中的基率). 疾病先验概率为 0.01, 灵敏度为 0.95, 假阳性率为 0.05. 阳性后的患病概率为

$$\frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \approx 0.161.$$

灵敏度很高不等于阳性后的患病概率很高；分母中的基率是 Bayes 题最常漏掉的量。

3.4.2 有联合密度的情形

若 (X, Y) 有联合密度, 且 $f_Y(y) > 0$, 条件密度为

$$f_{X|Y}(x | y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \mathbb{E}(X | Y = y) = \int x f_{X|Y}(x | y) dx.$$

这是一种正则条件分布的版本, 并不是事件 $\{Y = y\}$ 上的初等比值, 因为连续变量通常满足 $\mathbb{P}(Y = y) = 0$.

例题 3.2 (Poisson 分拆). 设 $X \sim \text{Pois}(\lambda)$ 、 $Y \sim \text{Pois}(\mu)$ 独立. 对 $S = X + Y$,

$$\mathbb{P}(X = k | S = n) = \binom{n}{k} \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\frac{\mu}{\lambda + \mu} \right)^{n-k}.$$

因此

$$\mathbb{E}(X | S) = \frac{\lambda}{\lambda + \mu} S.$$

这个例子同时展示了：先找条件分布可以计算全部条件矩；若只求条件均值，也可猜线性形式后按定义验证。

3.5 Gaussian 条件分布

定理 3.6 (分块 Gaussian 条件公式). 若

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right),$$

且 Σ_{YY} 可逆, 则

$$X | Y = y \sim N(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}).$$

证明. 令

$$Z = X - \mu_X - \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y).$$

直接计算得 $\text{Cov}(Z, Y) = 0$ 。 (Z, Y) 仍联合 Gaussian, 所以零协方差推出独立。于是给定 $Y = y$ 时, Z 的分布不变; 其协方差正是 Schur 补 $\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$, 平移即得结论。 $\square$

易错点

“不相关推出独立”只对联合 Gaussian 等特殊族成立。一般情况下 $X \sim \text{Unif}(-1, 1)$ 与 X^2 满足 $\text{Cov}(X, X^2) = 0$, 却显然不独立。

3.6 条件独立与图式分解

定义 3.2 (条件独立). 给定 $\mathcal{G}$, 随机变量 X, Y 条件独立, 若对任意有界 Borel 函数 f, g ,

$$\mathbb{E}[f(X)g(Y) | \mathcal{G}] = \mathbb{E}[f(X) | \mathcal{G}]\mathbb{E}[g(Y) | \mathcal{G}] \quad \text{几乎处处.}$$

若存在条件密度, 等价地

$$f_{X,Y|Z}(x, y | z) = f_{X|Z}(x | z)f_{Y|Z}(y | z)$$

对几乎所有 z 成立。条件独立与无条件独立没有包含关系: 共同原因 Z 会使 X, Y 边缘相关但给定 Z 后独立; 另一方面, 给定碰撞变量又可能制造依赖。

反例 3.1 (条件化制造依赖). 掷两枚独立公平硬币, 令 $X, Y \in \{0, 1\}$ 表示正面, $S = X + Y$ 。 X, Y 无条件独立, 但给定 $S = 1$ 后必有 $Y = 1 - X$, 所以不再独立。

3.7 停止信息与未来信息的雏形

设 $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$ 。如果 X_{n+1} 与 $\mathcal{F}_n$ 独立且可积, 则

$$\mathbb{E}(X_{n+1} | \mathcal{F}_n) = \mathbb{E}X_{n+1}.$$

若 $S_n = \sum_{k=1}^n (X_k - \mathbb{E}X_k)$, 则

$$\mathbb{E}(S_{n+1} | \mathcal{F}_n) = S_n.$$

这正是第 8 章将使用的鞅结构。此处的重点是：条件期望把“已知过去后，下一步平均增量为零”写成可迭代的方程。

答题方法

条件期望题的标准检查顺序：

1. 明确条件对象对应的 σ -代数；
2. 先判断是否可用独立性、对称性或“提出已知量”；
3. 离散或有密度时求条件分布；Gaussian 时优先作线性残差；
4. 用塔式性质核验 $\mathbb{E}[\mathbb{E}(X | \mathcal{G})] = \mathbb{E}X$ ；
5. 若求最优平方预测，直接调用正交投影，而不是重新做点态微分。

本节结论

条件期望由可测性和积分保持刻画；Radon–Nikodym 定理保证存在，几乎处处唯一解释了版本差异。塔式性质组织信息层级， L^2 投影解释最优预测，条件分布则提供具体计算接口。

自测

1. 证明 $\mathbb{E}[1_A | \mathcal{G}]$ 可解释为给定信息后的条件概率。
2. 若 X, Y 独立且可积，证明 $\mathbb{E}(X | X + Y)$ 一般不等于 $\mathbb{E}X$ 。
3. 对二元标准正态且相关系数为 ρ 的 (X, Y) ，求 $X | Y = y$ 。
4. 给出“无条件独立但给定 Z 后不独立”的例子。

自测核对要点 第 1 题直接核验定义。第 2 题可取独立 Poisson，条件均值随总和变化。第 3 题为 $N(\rho y, 1 - \rho^2)$ 。第 4 题可用本章两枚硬币并令 $Z = X + Y$ 的反例。

3.8 条件期望的存在、唯一与测试函数刻画

对 $X \in L^1$ 和子 σ -代数 $\mathcal{G}$ ，条件期望 $Z = \mathbb{E}[X | \mathcal{G}]$ 满足：

1. Z 对 $\mathcal{G}$ 可测；
2. 对每个 $G \in \mathcal{G}$,

$$\int_G Z \, d\mathbb{P} = \int_G X \, d\mathbb{P}.$$

Radon–Nikodym 定理给存在性；若 Z_1, Z_2 都满足，取 $G = \{Z_1 > Z_2\} \in \mathcal{G}$ ，则 $\int_G (Z_1 - Z_2) = 0$ ，非负性迫使该差在 G 上 a.s. 为零，同理反向，故 a.s. 唯一。

等价地，对任意有界 $\mathcal{G}$ -可测 H ，

$$\mathbb{E}[HX] = \mathbb{E}[H \mathbb{E}(X | \mathcal{G})].$$

从指示函数扩展到简单函数，再用有界收敛即可。

3.9 塔式法则与“提出已知量”

若 $\mathcal{H} \subseteq \mathcal{G}$ ，则

$$\mathbb{E}[\mathbb{E}(X | \mathcal{G}) | \mathcal{H}] = \mathbb{E}(X | \mathcal{H}).$$

证明只需核验右侧对 $\mathcal{H}$ 可测，并对每个 $H \in \mathcal{H}$ 积分相同。

若 Y 为 $\mathcal{G}$ -可测且 $YX \in L^1$ ，在适当可积条件下

$$\mathbb{E}[YX | \mathcal{G}] = Y\mathbb{E}[X | \mathcal{G}].$$

先对有界简单 Y 证明，再截断。不能把依赖未来信息、仅与 X 独立但不 $\mathcal{G}$ -可测的变量随意提出。

3.10 L^2 投影与方差分解

当 $X \in L^2$ ， $\mathbb{E}[X | \mathcal{G}]$ 是闭子空间 $L^2(\mathcal{G})$ 上的正交投影。对任意 $Y \in L^2(\mathcal{G})$ ，

$$\mathbb{E}[(X - \mathbb{E}[X | \mathcal{G}])Y] = 0.$$

因此

$$\mathbb{E}(X - Y)^2 = \mathbb{E}(X - \mathbb{E}[X | \mathcal{G}])^2 + \mathbb{E}(\mathbb{E}[X | \mathcal{G}] - Y)^2.$$

条件均值是所有 $\mathcal{G}$ -可测平方损失预测器中唯一的 a.s. 最优解。

取 $Y = \mathbb{E}X$ 得总方差公式

$$\text{Var}(X) = \mathbb{E}[\text{Var}(X | \mathcal{G})] + \text{Var}(\mathbb{E}[X | \mathcal{G}]).$$

3.11 Gaussian 条件分布的 Schur 补

若

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N \left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix} \right)$$

且 Σ_{YY} 可逆，则

$$X | Y = y \sim N(\mu_X + \Sigma_{XY}\Sigma_{YY}^{-1}(y - \mu_Y), \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}).$$

可通过把

$$R = X - \mu_X - \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mu_Y)$$

与 Y 的协方差算为零；联合高斯下不相关推出独立。条件协方差是 Schur 补，与观测值 y 无关。

3.12 条件独立的因子分解

$X \perp Y \mid Z$ 表示对合适可测集

$$\mathbb{P}(X \in A, Y \in B \mid Z) = \mathbb{P}(X \in A \mid Z)\mathbb{P}(Y \in B \mid Z) \quad a.s.$$

在有密度时等价于

$$p(x, y \mid z) = p(x \mid z)p(y \mid z)$$

几乎处处。条件独立不推出无条件独立；例如令 $Z \sim \text{Bernoulli}(1/2)$ ，给定 Z 后取 $X = Y = Z$ ，则给定 Z 时两者退化而条件独立，但无条件完全相关。

易错点

“给定后变独立”与“独立所以条件分布不变”是不同命题。写图模型分解时要标明条件集合，并确认 collider 条件化可能反而引入依赖。

自测

1. 用测试函数刻画证明塔式法则。
2. 条件期望为什么是平方损失最优预测器？
3. 推导 Gaussian 条件均值，并解释 Schur 补为何半正定。
4. 构造条件独立但无条件不独立的例子。

第四章 收敛模式、一致可积与 Delta 方法

学习目标

本章建立极限定理的语法：区分几乎处处、依概率、 L^p 与依分布收敛；掌握连续映射、Slutsky、Portmanteau 与一致可积；能够判断“分布收敛是否能推出矩收敛”，并完整使用 Delta 方法。章末两道正式真题只在本章保留完整解答。

4.1 四种收敛及其关系

定义 4.1 (常用收敛模式). 设 X_n, X 定义在同一概率空间上。

1. $X_n \rightarrow X$ 几乎处处: $\mathbb{P}(\lim_n X_n = X) = 1$;
2. $X_n \rightarrow X$ 依概率: 对每个 $\varepsilon > 0$, $\mathbb{P}(|X_n - X| > \varepsilon) \rightarrow 0$;
3. $X_n \rightarrow X$ 于 L^p : $\mathbb{E}|X_n - X|^p \rightarrow 0$;
4. $X_n \Rightarrow X$ 依分布: 在 F_X 的每个连续点 x , $F_{X_n}(x) \rightarrow F_X(x)$ 。

基本蕴含关系为

$$L^p \implies \text{依概率} \implies \text{依分布}, \quad \text{几乎处处} \implies \text{依概率}.$$

几乎处处与 L^p 一般互不蕴含；依分布最弱，而且 X_n, X 甚至不必定义在同一概率空间。

关键蕴含的证明. L^p 到依概率由 Markov 不等式：

$$\mathbb{P}(|X_n - X| > \varepsilon) \leq \varepsilon^{-p} \mathbb{E}|X_n - X|^p.$$

几乎处处到依概率可对 $\mathbf{1}_{\{|X_n - X| > \varepsilon\}}$ 使用控制收敛。依概率到依分布可对 F_X 的连续点使用夹逼：

$$\mathbb{P}(X_n \leq x) \leq \mathbb{P}(X \leq x + \varepsilon) + \mathbb{P}(|X_n - X| > \varepsilon),$$

并写出相反方向，再令 $n \rightarrow \infty, \varepsilon \downarrow 0$ 。

□

反例 4.1 (逆向不成立). 在 $(0, 1)$ 上取 $X_n = n^{1/p} \mathbf{1}_{(0, 1/n)}$ ，则 $X_n \rightarrow 0$ 几乎处处，但 $\mathbb{E}|X_n|^p = 1$ 。另一方面，按“打字机序列”依次扫描二进区间，可构造 $X_n \rightarrow 0$ 依概率但不几乎处处收敛的序列。若 $X_n \stackrel{d}{=} X$ ，则恒有 $X_n \Rightarrow X$ ，但在同一空间上可令 $X_n = -X$ ，不必依概率收敛。

4.2 Portmanteau、连续映射与 Slutsky

定理 4.1 (Portmanteau 定理的常用形式). 下列条件等价于 $X_n \Rightarrow X$:

1. 对每个有界连续函数 f , $\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X)$;
2. 对每个闭集 C , $\limsup_n \mathbb{P}(X_n \in C) \leq \mathbb{P}(X \in C)$;
3. 对每个开集 G , $\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G)$;
4. 对每个满足 $\mathbb{P}(X \in \partial A) = 0$ 的 Borel 集 A , $\mathbb{P}(X_n \in A) \rightarrow \mathbb{P}(X \in A)$.

定理 4.2 (连续映射与 Slutsky). 若 $X_n \Rightarrow X$, 且 g 在 X 几乎必然落入的点上连续, 则 $g(X_n) \Rightarrow g(X)$. 若另有 $Y_n \rightarrow c$ 依概率, 则

$$X_n + Y_n \Rightarrow X + c, \quad X_n Y_n \Rightarrow cX, \quad X_n / Y_n \Rightarrow X/c \quad (c \neq 0).$$

证明. 连续映射可由 Portmanteau 的有界连续函数形式证明: $f \circ g$ 在 g 连续的区域上连续, 再用几乎必然连续版本的推广. Slutsky 的核心是 $(X_n, Y_n) \Rightarrow (X, c)$; 可用紧性与子列论证, 再对加法、乘法、除法使用连续映射. $\square$

易错点

分母连续映射必须避开 0; 函数在不连续点上的概率必须为 0. 例如 $X_n \Rightarrow 0$ 并不自动允许对 $g(x) = \mathbf{1}_{\{x=0\}}$ 使用连续映射。

4.3 一致可积: 把概率收敛升级为均值收敛

定义 4.2 (一致可积). 随机变量族 $\mathcal{X}$ 称为一致可积, 若

$$\lim_{K \rightarrow \infty} \sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbf{1}_{\{|X| > K\}}] = 0.$$

一致可积同时排除“少量但极大的质量逃向无穷”。它比 $\sup_n \mathbb{E}|X_n| < \infty$ 强; 后者只控制平均大小, 不控制尾质量统一消失。

定理 4.3 (Vitali 收敛判别). 若 $X_n \rightarrow X$ 依概率, 则

$$X_n \rightarrow X \text{ 于 } L^1 \iff \{X_n : n \geq 1\} \text{ 一致可积}.$$

证明. 先证充分性。依概率收敛蕴含每个子列都有进一步子列几乎处处收敛。对该进一步子列, 在集合 $\{|X_n| \leq K, |X| \leq K\}$ 上用有界性与依概率收敛得到截断差的期望趋零; 两侧尾部由一致可积和 Fatou 推出的 $X \in L^1$ 统一控制。故每个子列都有进一步子列在 L^1 收敛, 只能推出原序列在 L^1 收敛。

再证必要性。若 $X_n \rightarrow X$ 于 L^1 , 则 $X \in L^1$, 且逐点有安全估计

$$|X_n| \mathbf{1}_{\{|X_n| > K\}} \leq 2|X_n - X| + 2|X| \mathbf{1}_{\{|X| > K/2\}}.$$

确实, 在 $\{|X_n| > K, |X| \leq K/2\}$ 上, $|X_n - X| \geq |X_n| - |X| > |X_n|/2$, 故第一项足以控制左端; 在 $\{|X_n| > K, |X| > K/2\}$ 上, 用 $|X_n| \leq |X_n - X| + |X|$ 即得。取期望后, 对任意 $\varepsilon > 0$, 先选 N 使 $n \geq N$ 时 $2\mathbb{E}|X_n - X| < \varepsilon/2$, 再选 K 使 $2\mathbb{E}[|X|\mathbf{1}_{\{|X|>K/2\}}] < \varepsilon/2$ 。于是所有 $n \geq N$ 的尾期望统一小于 ε 。剩余有限个 $X_1, \dots, X_{N-1} \in L^1$ 的尾期望分别趋零, 增大 K 即同时控制, 故 $\{X_n : n \geq 1\}$ 一致可积。 $\square$

定理 4.4 (de la Vallée-Poussin 判别). 若存在凸递增函数 $\Phi : [0, \infty) \rightarrow [0, \infty)$, 满足 $\Phi(t)/t \rightarrow \infty$ 且

$$\sup_n \mathbb{E}\Phi(|X_n|) < \infty,$$

则 $\{X_n\}$ 一致可积。特别地, 若某个 $\delta > 0$ 使 $\sup_n \mathbb{E}|X_n|^{1+\delta} < \infty$, 则一致可积。

证明. 对任意 K ,

$$|X_n|\mathbf{1}_{\{|X_n|>K\}} \leq \sup_{t>K} \frac{t}{\Phi(t)} \Phi(|X_n|).$$

取期望、对 n 取上确界, 再令 $K \rightarrow \infty$ 。 $\square$

4.4 Delta 方法

定理 4.5 (一阶 Delta 方法). 若 $r_n(X_n - \theta) \Rightarrow Z$, $r_n \rightarrow \infty$, 且 g 在 θ 处可微, 则

$$r_n[g(X_n) - g(\theta)] \Rightarrow g'(\theta)Z.$$

证明. 依分布收敛蕴含 $r_n(X_n - \theta) = O_{\mathbb{P}}(1)$, 故 $X_n - \theta = O_{\mathbb{P}}(r_n^{-1}) \rightarrow 0$ 依概率。可微性给

$$g(X_n) - g(\theta) = g'(\theta)(X_n - \theta) + (X_n - \theta)\eta(X_n - \theta),$$

其中 $\eta(t) \rightarrow 0$ 。因此

$$r_n[g(X_n) - g(\theta)] = g'(\theta)r_n(X_n - \theta) + r_n(X_n - \theta)\eta(X_n - \theta).$$

第二项是 $O_{\mathbb{P}}(1)o_{\mathbb{P}}(1) = o_{\mathbb{P}}(1)$, 由 Slutsky 得结论。 $\square$

注 4.1 (导数为零时). 若 $g'(\theta) = 0$ 且 $g''(\theta) \neq 0$, 通常要改用二阶展开:

$$r_n^2[g(X_n) - g(\theta)] \Rightarrow \frac{1}{2}g''(\theta)Z^2.$$

直接套一阶 Delta 只得到退化极限 0, 并未找到正确尺度。

正式真题: 2024 年秋季概率统计 Q2

设

$$\sqrt{n}(Y_n - a) \Rightarrow N(0, \sigma^2).$$

1. 若 g 在 a 处可微且 $g'(a) \neq 0$, 证明

$$\sqrt{n}[g(Y_n) - g(a)] \Rightarrow N(0, \sigma^2[g'(a)]^2).$$

2. 若 $Z_n \sim \text{Bin}(n, p)$, $0 < p < 1$, 求

$$\sqrt{n} \left[\log \left(\frac{Z_n}{n} \right) - \log p \right]$$

的极限分布。

完整解答 (独立核验稿)

1. 这是 [定理 4.5](#) 取 $r_n = \sqrt{n}$ 、 $\theta = a$ 的直接结果。乘以常数 $g'(a)$ 后, 正态方差变为 $\sigma^2[g'(a)]^2$ 。

2. 令 $\hat{p}_n = Z_n/n$ 。由 Bernoulli 中心极限定理,

$$\sqrt{n}(\hat{p}_n - p) \Rightarrow N(0, p(1-p)).$$

在 $p > 0$ 处取 $g(x) = \log x$, $g'(p) = 1/p$, 故

$$\sqrt{n}[\log \hat{p}_n - \log p] \Rightarrow N\left(0, \frac{1-p}{p}\right).$$

严格说, 当 $Z_n = 0$ 时对数未定义。可在该事件上任意定义统计量, 因为 $\mathbb{P}(Z_n = 0) = (1-p)^n \rightarrow 0$; 这种在概率趋零事件上的修改不改变依分布极限。

4.5 随机级数中的几乎处处与 L^1 收敛

下一题展示常见的两步结构: 先用方差可和得到几乎处处收敛, 再用高阶矩控制建立一致可积, 从而升级到 L^1 。

正式真题: 2024 年秋季概率统计 Q5

设 X_k 独立同分布且服从参数为 1 的指数分布, 令

$$M_n = \sum_{k=1}^n \frac{X_k}{k}.$$

1. 证明 $M_n - \log n$ 几乎处处收敛;
2. 对 $0 < p < 1$, 证明 e^{pM_n}/n^p 在 L^1 中收敛。

完整解答 (独立核验稿)

1. 写成

$$M_n - \log n = \sum_{k=1}^n \frac{X_k - 1}{k} + \left(\sum_{k=1}^n \frac{1}{k} - \log n \right).$$

第二项收敛到 Euler 常数。第一项是独立中心随机变量级数，且

$$\sum_{k \geq 1} \text{Var} \left(\frac{X_k - 1}{k} \right) = \sum_{k \geq 1} \frac{1}{k^2} < \infty.$$

由 Kolmogorov 收敛判别（第 六 章证明所需最大不等式）该级数几乎处处收敛，故结论成立。

2. 记 $W_n = e^{pM_n}/n^p$ 。第一问给出 $W_n \rightarrow W$ 几乎处处，且 $0 < W < \infty$ 。只需证明 $\{W_n\}$ 一致可积。选 $q > 1$ 使 $pq < 1$ 。指数分布满足 $\mathbb{E}e^{tX} = 1/(1-t)$ ($t < 1$)，故

$$\mathbb{E}W_n^q = n^{-pq} \prod_{k=1}^n \mathbb{E}e^{pqX_k/k} = n^{-pq} \prod_{k=1}^n \left(1 - \frac{pq}{k}\right)^{-1}.$$

有限个满足 $k \leq pq$ 的项需要单独看：由于 $pq < 1$ ，所有 $k \geq 1$ 都处于矩母函数的定义域。利用

$$\prod_{k=1}^n \left(1 - \frac{a}{k}\right)^{-1} = \frac{\Gamma(1-a)\Gamma(n+1)}{\Gamma(n+1-a)} \sim \Gamma(1-a)n^a, \quad 0 < a < 1,$$

取 $a = pq$ 得 $\sup_n \mathbb{E}W_n^q < \infty$ 。由 定理 4.4, $\{W_n\}$ 一致可积；再由几乎处处收敛和 定理 4.3, $W_n \rightarrow W$ 于 L^1 。

易错点

上题不能只写“ W_n 几乎处处收敛，所以 L^1 收敛”。真正补上缺口的是 $q > 1$ 阶矩一致有界，从而一致可积。另一个常见错误是随意选 q ：必须同时满足 $q > 1$ 与 $pq < 1$ ，这正是题设 $p < 1$ 的用途。

答题方法

极限分布题先回答三个问题：

1. 已知的是哪一种收敛，目标又是哪一种？
2. 映射是否在极限变量的支撑上连续，分母或对数是否在小概率事件上失效？
3. 若目标涉及期望或 L^1 ，是否已有控制收敛上界，或需要证明一致可积？

Delta 方法还要检查导数是否为零，并把余项写成 $O_{\mathbb{P}}(1)o_{\mathbb{P}}(1)$ ，不能只写“由 Taylor 展开显然”。

本节结论

依分布收敛只控制有界连续检验函数，不能自动控制无界矩；一致可积正是把概率收敛升级为 L^1 收敛的尾部条件。连续映射与 Slutsky 负责传递极限，Delta 方法则把局部可微性转化为统计量的正确尺度与极限分布。

自测

1. 给出 $\sup_n \mathbb{E}|X_n| < \infty$ 但 $\{X_n\}$ 不一致可积的例子。
2. 若 $X_n \Rightarrow X$ 且 $\{|X_n|^p\}$ 一致可积, 证明 $\mathbb{E}|X_n|^p \rightarrow \mathbb{E}|X|^p$ 。
3. 若 $\sqrt{n}(\hat{\theta}_n - \theta) \Rightarrow N(0, \tau^2)$, 且 $g'(\theta) = 0, g''(\theta) \neq 0$, 写出二阶 Delta 极限。
4. 在 2024 秋 Q5 中说明为什么 $p = 1$ 时上述 L^q 论证失效。

自测核对要点 第 1 题可取 $X_n = n\mathbf{1}_{(0,1/n)}$ 。第 2 题对截断函数先用 Portmanteau, 再以一致可积控制尾部。第 3 题极限为 $\frac{1}{2}g''(\theta)\tau^2\chi_1^2$ 。第 4 题不存在同时满足 $q > 1$ 与 $pq < 1$ 的 q , 而且所需指数矩在首项处已到临界。

4.6 收敛关系与反例体系

在概率空间上,

$$X_n \xrightarrow{\text{a.s.}} X \implies X_n \xrightarrow{\mathbb{P}} X \implies X_n \xrightarrow{d} X.$$

若 X 为常数, 依分布收敛反过来推出依概率收敛。对 $p > 0$,

$$X_n \rightarrow X \text{ in } L^p \implies X_n \xrightarrow{\mathbb{P}} X$$

由 Markov 不等式得到。

反向一般不成立:

- $X_n = \mathbf{1}_{A_n}$, 若 $\mathbb{P}(A_n) \rightarrow 0$ 但事件安排使其无穷多次发生, 则依概率趋零却不 a.s.;
- $X_n = n\mathbf{1}_{(0,1/n)}$ a.s. 趋零却不 L^1 ;
- 独立同分布的 X_n 与 X_1 同分布, 故 $X_n \xrightarrow{d} X_1$, 但通常不依概率趋于 X_1 。

最后一例提醒: 依分布收敛只比较边际分布, 不要求随机变量同一实现接近。

4.7 子列原理

命题 4.6. $X_n \xrightarrow{\mathbb{P}} X$ 当且仅当每个子列都含有一个进一步子列 a.s. 收敛到 X 。

证明. 若依概率收敛, 从任意子列选 n_k 使

$$\mathbb{P}(|X_{n_k} - X| > 2^{-k}) \leq 2^{-k}.$$

Borel–Cantelli 给相应事件仅有限次发生, 故 $X_{n_k} \rightarrow X$ a.s.。

反之, 若不依概率收敛, 则存在 $\varepsilon, \delta > 0$ 和子列使

$$\mathbb{P}(|X_{n_k} - X| > \varepsilon) \geq \delta$$

对所有 k 成立。该子列的任何进一步 a.s. 收敛子列都会依概率收敛, 矛盾。 □

4.8 Portmanteau 的开闭集版本

$X_n \xrightarrow{d} X$ 等价于对所有有界连续 f ,

$$\mathbb{E}f(X_n) \rightarrow \mathbb{E}f(X).$$

也等价于

$$\limsup_n \mathbb{P}(X_n \in F) \leq \mathbb{P}(X \in F) \quad \text{对闭集 } F,$$

$$\liminf_n \mathbb{P}(X_n \in G) \geq \mathbb{P}(X \in G) \quad \text{对开集 } G.$$

对连续点 x 得

$$F_{X_n}(x) \rightarrow F_X(x).$$

若边界有正概率, 集合概率未必收敛; Portmanteau 只给单边界。

4.9 一致可积与 Vitali 定理

族 $\mathcal{X} \subset L^1$ 一致可积若

$$\lim_{K \rightarrow \infty} \sup_{X \in \mathcal{X}} \mathbb{E}[|X| \mathbf{1}_{\{|X| > K\}}] = 0.$$

L^p 有界 (某 $p > 1$) 蕴含 UI, 因为

$$|X| \mathbf{1}_{\{|X| > K\}} \leq \frac{|X|^p}{K^{p-1}}.$$

Vitali 定理给

$$X_n \xrightarrow{\mathbb{P}} X \quad \text{且} \quad \{X_n\} \text{ UI} \iff X_n \rightarrow X \text{ in } L^1.$$

必要性证明使用安全估计

$$|X_n| \mathbf{1}_{\{|X_n| > K\}} \leq 2|X_n - X| + 2|X| \mathbf{1}_{\{|X| > K/2\}},$$

而不是不成立的逐点尾部替换。先让 n 大控制第一项, 再对有限个早期 n 逐个控尾。

4.10 Delta 方法的余项条件

若

$$r_n(X_n - \theta) \xrightarrow{d} Z, \quad r_n \rightarrow \infty,$$

且 g 在 θ 可微, 则

$$r_n[g(X_n) - g(\theta)] = g'(\theta)r_n(X_n - \theta) + o_{\mathbb{P}}(1) \xrightarrow{d} g'(\theta)Z.$$

这里 Taylor 余项成为 $o_{\mathbb{P}}(1)$, 因为 $X_n \xrightarrow{\mathbb{P}} \theta$ 。若 $g'(\theta) = 0$, 一阶 Delta 只给退化极限; 若二阶可微且尺度合适,

$$r_n^2[g(X_n) - g(\theta)] \xrightarrow{d} \frac{1}{2}g''(\theta)Z^2.$$

答题方法

极限定理题先写收敛模式，再检查映射连续性、矩的 UI 与 Taylor 余项的随机阶。不要因边缘分布收敛就交换期望，也不要因一阶导为零就宣布“没有极限”。

自测

1. 给出依概率不推出 a.s. 的事件列例子。
2. 证明依概率收敛的子列原理。
3. Portmanteau 为什么只在分布函数连续点给收敛？
4. 用 L^p 有界证明一致可积。
5. 一阶 Delta 退化时，二阶尺度如何改变？

第五章 Borel–Cantelli、稀有事件与随机级数

学习目标

本章训练从数列极限中识别“事件发生无穷多次”。必须能够完整证明两条 Borel–Cantelli 引理，准确指出第二条所需独立性，掌握两两独立版本，并把尾概率、极值、稀有大跳和随机级数转化为可求和性问题。

5.1 $\limsup A_n$ 就是无穷多次发生

回顾

$$\limsup_{n \rightarrow \infty} A_n = \bigcap_{m \geq 1} \bigcup_{n \geq m} A_n = \{\omega : \omega \in A_n \text{ 对无穷多个 } n \text{ 成立}\}.$$

因此

$$\{X_n > a_n \text{ i.o.}\} = \left\{ \limsup_{n \rightarrow \infty} \frac{X_n}{a_n} > 1 \right\}$$

通常只是一侧包含，而不是自动相等；要得到精确极限，需分别对 $(1 + \varepsilon)a_n$ 与 $(1 - \varepsilon)a_n$ 建立上、下界，再令 ε 沿可数序列趋零。

5.2 第一 Borel–Cantelli 引理

定理 5.1 (第一 Borel–Cantelli 引理). 若

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty,$$

则 $\mathbb{P}(A_n \text{ i.o.}) = 0$ 。不需要独立性。

证明. 对每个 m ，并集上界给出

$$\mathbb{P}\left(\bigcup_{n \geq m} A_n\right) \leq \sum_{n \geq m} \mathbb{P}(A_n) \rightarrow 0.$$

集合 $\bigcup_{n \geq m} A_n$ 随 m 递减到 $\limsup A_n$ ，由测度连续性，

$$\mathbb{P}(\limsup A_n) = \lim_{m \rightarrow \infty} \mathbb{P}\left(\bigcup_{n \geq m} A_n\right) = 0.$$

□

第一引理的等价口语是：若发生概率可求和，则几乎必然只有有限多个例外。它常把概率估计转化成最终一致的逐点结论。

例题 5.1 (Markov 加 Borel–Cantelli). 若对某个 $p > 0$,

$$\sum_n \frac{\mathbb{E}|X_n|^p}{a_n^p} < \infty,$$

则对每个 $\varepsilon > 0$,

$$\sum_n \mathbb{P}(|X_n| > \varepsilon a_n) \leq \varepsilon^{-p} \sum_n \frac{\mathbb{E}|X_n|^p}{a_n^p} < \infty.$$

对有理 $\varepsilon > 0$ 取可数交，得到 $X_n/a_n \rightarrow 0$ 几乎处处。

5.3 第二 Borel–Cantelli 引理

定理 5.2 (第二 Borel–Cantelli 引理). 若事件 A_n 相互独立，且

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) = \infty,$$

则 $\mathbb{P}(A_n \text{ i.o.}) = 1$ 。

证明. 固定 m 。由独立性和 $1 - x \leq e^{-x}$,

$$\mathbb{P}\left(\bigcap_{n=m}^N A_n^c\right) = \prod_{n=m}^N (1 - \mathbb{P}(A_n)) \leq \exp\left(-\sum_{n=m}^N \mathbb{P}(A_n)\right) \rightarrow 0.$$

令 $N \rightarrow \infty$ 得 $\mathbb{P}(\bigcap_{n \geq m} A_n^c) = 0$ ，即从第 m 项以后至少发生一次的概率为 1。再对 m 取可数交，得到无穷多次发生的概率为 1。□

反例 5.1 (独立性不能删除). 取同一个事件 A 满足 $0 < \mathbb{P}(A) < 1$ ，并令 $A_n = A$ 。则 $\sum_n \mathbb{P}(A_n) = \infty$ ，但

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(A) < 1.$$

发散概率和只说明“期望发生次数无穷”，强依赖可把这些次数全部捆在同一结果上。

5.4 两两独立版本

定理 5.3 (两两独立的第二引理). 若 A_n 两两独立且 $\sum_n \mathbb{P}(A_n) = \infty$ ，则 $\mathbb{P}(A_n \text{ i.o.}) = 1$ 。

证明. 令 $S_N = \sum_{n=1}^N \mathbf{1}_{A_n}$ ， $\mu_N = \mathbb{E}S_N = \sum_{n \leq N} \mathbb{P}(A_n)$ 。两两独立使协方差为零，故

$$\text{Var}(S_N) = \sum_{n \leq N} \mathbb{P}(A_n)(1 - \mathbb{P}(A_n)) \leq \mu_N.$$

选择子列 N_k 使 $\mu_{N_k} \geq k^2$ 且首次越过 k^2 。Chebyshev 给

$$\mathbb{P}\left(S_{N_k} \leq \frac{1}{2}\mu_{N_k}\right) \leq \frac{4\text{Var}(S_{N_k})}{\mu_{N_k}^2} \leq \frac{4}{\mu_{N_k}} \leq \frac{4}{k^2}.$$

由第一引理，几乎处处从某个 k 起 $S_{N_k} > \mu_{N_k}/2 \rightarrow \infty$ ，所以总发生次数无穷。□

5.5 从尾概率判断单点极值

设 X_n 独立且同分布。要判断 $X_n > a_n$ 是否无穷多次发生，只需研究

$$\sum_n \mathbb{P}(X_1 > a_n).$$

收敛用第一引理，发散用第二引理。若目标是最大值 $M_n = \max_{k \leq n} X_k$ ，下界可由单点大值给出，上界还需把所有早期值纳入：一旦单点越界只有有限次，这些有限历史值虽然可能很大，却终将低于不断增长的阈值 a_n 。

正式真题：2024 年春季概率统计 Q2

设 $X_1, X_2, \dots$ 独立同分布，且服从参数为 1 的指数分布。令 $M_n = \max_{1 \leq k \leq n} X_k$ 。

1. 证明

$$\limsup_{n \rightarrow \infty} \frac{X_n}{\log n} = 1 \quad \text{几乎处处};$$

2. 证明

$$\frac{M_n}{\log n} \rightarrow 1 \quad \text{几乎处处}.$$

完整解答（独立核验稿） 固定 $0 < \varepsilon < 1$ 。

1. 上界事件 $A_n^+ = \{X_n > (1 + \varepsilon) \log n\}$ 满足

$$\sum_n \mathbb{P}(A_n^+) = \sum_n n^{-(1+\varepsilon)} < \infty.$$

由 [定理 5.1](#)，几乎处处只有有限个 A_n^+ 发生，所以 $\limsup X_n / \log n \leq 1 + \varepsilon$ 。

下界事件 $A_n^- = \{X_n > (1 - \varepsilon) \log n\}$ 相互独立，且

$$\sum_n \mathbb{P}(A_n^-) = \sum_n n^{-(1-\varepsilon)} = \infty.$$

由 [定理 5.2](#)， A_n^- 几乎处处无穷多次发生，因此 $\limsup X_n / \log n \geq 1 - \varepsilon$ 。对有理 $\varepsilon \downarrow 0$ 取可数交，得到结论。

2. 下界由第一问立即得到：无穷多个 k 满足 $X_k > (1 - \varepsilon) \log k$ ，故在这些 k 上 $M_k / \log k > 1 - \varepsilon$ 。若要得到对所有充分大 n 的下界，可直接算

$$\mathbb{P}(M_n \leq (1 - \varepsilon) \log n) = \left(1 - n^{-(1-\varepsilon)}\right)^n \leq e^{-n^\varepsilon}.$$

右端可求和，由第一引理，几乎处处最终 $M_n > (1 - \varepsilon) \log n$ 。

对上界，第一问的上界说明几乎处处存在随机 N ，使对所有 $k \geq N$, $X_k \leq (1 + \varepsilon) \log k \leq (1 + \varepsilon) \log n$ ($k \leq n$)。剩余有限个历史值 $C = \max_{k < N} X_k < \infty$ 。由于 $(1 + \varepsilon) \log n \rightarrow \infty$ ，最终也有 $C \leq (1 + \varepsilon) \log n$ 。故最终

$$M_n \leq (1 + \varepsilon) \log n.$$

再对有理 $\varepsilon \downarrow 0$ 取可数交，得到 $M_n / \log n \rightarrow 1$ 。

易错点

从“最终每个新样本不超过阈值”推到最大值上界时，不能忘记有限个早期样本。正确补句是：早期最大值是有限随机常数，而 $(1 + \varepsilon) \log n$ 持续增长，所以最终也超过这段有限历史。另须明确 $0 < \varepsilon < 1$ ，否则下界阈值或调和级数判别失去原意。

5.6 随机级数：大跳、中心部分与方差

对独立随机变量级数 $\sum_n X_n$ ，Kolmogorov 三系列定理的结构是：截断在某个常数 $c > 0$ 后，级数几乎处处收敛当且仅当

$$\sum_n \mathbb{P}(|X_n| > c) < \infty, \quad \sum_n \mathbb{E}[X_n \mathbf{1}_{\{|X_n| \leq c\}}] \text{ 收敛},$$

$$\sum_n \text{Var}(X_n \mathbf{1}_{\{|X_n| \leq c\}}) < \infty.$$

第一系列排除无穷多个大跳，第二系列控制确定性漂移，第三系列控制中心随机波动。完整定理、必要性与充分性证明见本章的节 5.11；所用最大不等式及方差可和收敛判别也在章 六独立证明。

例题 5.2 (稀有大跳识别). 若独立 X_n 满足

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{2n^2}, \quad \mathbb{P}(X_n = 0) = 1 - \frac{1}{n^2},$$

则 $\sum_n \mathbb{P}(X_n \neq 0) < \infty$ ，所以几乎处处只有有限个非零项， $\sum_n X_n$ 几乎处处是有限和。尽管 $\text{Var}(X_n) = 1$ 不可求和，大跳稀有性仍直接解决了问题。这说明“方差可和”是充分条件，不是必要条件。

5.7 零一律的边界

$\{A_n \text{ i.o.}\}$ 若每个 $A_n \in \sigma(X_n)$ 且 X_n 独立，通常属于尾 σ -代数

$$\mathcal{T} = \bigcap_{m \geq 1} \sigma(X_m, X_{m+1}, \dots).$$

Kolmogorov 零一律说明独立序列的尾事件概率只能是 0 或 1。它不能替代 Borel–Cantelli 来判断究竟是哪一个值，却解释了许多极限事件为何呈现零一现象。

答题方法

遇到极值、稀有事件或“大跳”题时：

1. 把目标写成 $\{A_n \text{ i.o.}\}$ 或“最终不再发生”；
2. 精确计算或比较 $\sum_n \mathbb{P}(A_n)$ ；
3. 收敛侧用第一引理，不要求独立；
4. 发散侧先核验相互独立或两两独立，再选第二引理；
5. 极限常数用 $(1 \pm \varepsilon)$ 两侧夹逼，并对有理 $\varepsilon \downarrow 0$ 取可数交；
6. 从单点转到运行最大值时，单独处理有限历史。

本节结论

Borel–Cantelli 把概率和转化为最终行为：可求和意味着只有有限个例外；在独立或两两独立条件下，发散意味着无穷多次发生。极值题由上下阈值夹逼，随机级数则要把稀有大跳、漂移和中心波动分开。

自测

1. 若 X_n 独立同分布且 $\mathbb{P}(X_1 > x) = x^{-\alpha}$ ($x \geq 1$)，求 $\limsup \log X_n / \log n$ 。
2. 证明若 $\sum_n \mathbb{P}(|X_n - X| > \varepsilon) < \infty$ 对每个 $\varepsilon > 0$ 成立，则 $X_n \rightarrow X$ 几乎处处。
3. 找出反例 5.1 中第二引理失败的具体位置。
4. 说明两两独立版本为何要抽取 $\mu_{N_k} \geq k^2$ 的子列。

自测核对要点 第 1 题答案为 $1/\alpha$ ，对阈值 $n^{1/\alpha \pm \varepsilon}$ 使用两条引理。第 2 题只需对 $\varepsilon = 1/m$ 取可数交。第 3 题是补事件交集概率不能写成乘积。第 4 题使 Chebyshev 上界可求和，从而把依概率增长升级为几乎处处增长。

5.8 第二 Borel–Cantelli 中独立性不可删除

若 $\sum_n \mathbb{P}(A_n) = \infty$ 且 A_n 相互独立，则对任意 N ，

$$\mathbb{P}\left(\bigcap_{n=N}^m A_n^c\right) = \prod_{n=N}^m (1 - \mathbb{P}(A_n)) \leq \exp\left\{-\sum_{n=N}^m \mathbb{P}(A_n)\right\} \rightarrow 0.$$

故

$$\mathbb{P}\left(\bigcup_{n \geq N} A_n\right) = 1$$

对每个 N 成立，再取可列交得 $\mathbb{P}(A_n \text{ i.o.}) = 1$ 。

若删除独立性，取同一事件 A 满足 $0 < \mathbb{P}(A) < 1$ ，令 $A_n = A$ 。则概率和发散，但

$$\mathbb{P}(A_n \text{ i.o.}) = \mathbb{P}(A) < 1.$$

所以“概率和发散”自身不足以推出几乎必然无穷多次发生。

5.9 两两独立版本的二阶矩证明

令

$$N_n = \sum_{k=1}^n \mathbf{1}_{A_k}, \quad s_n = \mathbb{E}N_n = \sum_{k=1}^n \mathbb{P}(A_k).$$

若事件两两独立，则

$$\text{Var}(N_n) = \sum_{k=1}^n \mathbb{P}(A_k)[1 - \mathbb{P}(A_k)] \leq s_n.$$

当 $s_n \rightarrow \infty$ ，取子列 n_j 使 $s_{n_j} \geq j^2$ 。Chebyshev 给

$$\mathbb{P}\left(\left|\frac{N_{n_j}}{s_{n_j}} - 1\right| > \frac{1}{2}\right) \leq \frac{4}{s_{n_j}} \leq \frac{4}{j^2}.$$

Borel–Cantelli 推出子列上 $N_{n_j} \rightarrow \infty$ a.s.。因 N_n 单调，整个序列也趋无穷，故事件无穷多次发生。这里两两独立足以消去协方差，不需要完全独立的乘积公式。

5.10 极值的上、下界模板

设 $M_n = \max_{1 \leq k \leq n} X_k$ ， X_k i.i.d.。上界常选阈值 u_n 并使用

$$\mathbb{P}(M_n > u_n) \leq n\mathbb{P}(X_1 > u_n).$$

为获得 a.s. 最终上界，可在几何子列 n_r 上使右侧可和，再填补子列间隙。

下界用

$$\mathbb{P}(M_n \leq u_n) = [1 - \mathbb{P}(X_1 > u_n)]^n \leq \exp\{-n\mathbb{P}(X_1 > u_n)\}.$$

若右侧在子列上可和，则最终 $M_{n_r} > u_{n_r}$ 。

例题 5.3 (指数极值). 若 $X_k \sim \text{Exp}(1)$ ，取 $0 < \varepsilon < 1$ 。单点上尾

$$\mathbb{P}(X_n > (1 + \varepsilon) \log n) = n^{-(1+\varepsilon)}$$

可和，故只有有限个 n 越界。那几个有限历史值固定，而 $(1 + \varepsilon) \log n \rightarrow \infty$ ，所以它们最终也低于不断增长的阈值，从而

$$M_n \leq (1 + \varepsilon) \log n$$

最终成立。下界用几何子列和 $\mathbb{P}(M_n \leq (1 - \varepsilon) \log n) \leq e^{-n^\varepsilon}$ ，得到

$$\frac{M_n}{\log n} \rightarrow 1 \quad \text{a.s.}$$

5.11 Kolmogorov 三系列定理：完整判别与证明

随机级数的困难来自三种互不相同的机制：偶尔出现的巨大跳跃、截断后的确定性漂移，以及中心化后仍残留的随机振荡。三系列定理把它们分别交给一个可检查的数值级数。

定理 5.4 (Kolmogorov 三系列定理). 设 $X_1, X_2, \dots$ 相互独立。固定 $c > 0$ ，并定义截断变量

$$Y_n^{(c)} = X_n \mathbf{1}_{\{|X_n| \leq c\}}.$$

则随机级数 $\sum_{n \geq 1} X_n$ 几乎处处收敛，当且仅当下列三个条件同时成立：

$$\sum_{n \geq 1} \mathbb{P}(|X_n| > c) < \infty, \quad (5.1)$$

$$\sum_{n \geq 1} \mathbb{E} Y_n^{(c)} \text{ 在 } \mathbb{R} \text{ 中收敛}, \quad (5.2)$$

$$\sum_{n \geq 1} \text{Var}(Y_n^{(c)}) < \infty. \quad (5.3)$$

“固定 c ” 表示阈值在观察样本前确定，不能依赖 n 或样本路径。条件对某一个 $c > 0$ 成立与对每一个 $c > 0$ 成立等价。

证明. 充分性. 记 $Y_n = Y_n^{(c)}$ 以及

$$Z_n = Y_n - \mathbb{E} Y_n.$$

$\{Z_n\}$ 仍相互独立、中心且平方可积，并由 (5.3) 满足 $\sum_n \text{Var}(Z_n) < \infty$ 。Kolmogorov 最大不等式给出方差可和的中心级数收敛判别，详见定理 6.1、6.2；因此 $\sum_n Z_n$ 几乎处处收敛。再加上 (5.2) 中收敛的确定性级数，得到

$$\sum_n Y_n = \sum_n Z_n + \sum_n \mathbb{E} Y_n \text{ 几乎处处收敛}.$$

另一方面，由 (5.1) 和第一 Borel–Cantelli 引理， $\{|X_n| > c\}$ 几乎处处只发生有限次。于是几乎处处存在随机 $N(\omega)$ ，使所有 $n \geq N(\omega)$ 都有 $X_n = Y_n$ 。因此 $\sum_n (X_n - Y_n)$ 因而只是有限和，故 $\sum_n X_n$ 几乎处处收敛。

必要性：第一系列与截断级数. 反设 $\sum_n X_n$ 几乎处处收敛，则 $X_n \rightarrow 0$ 几乎处处，所以 $\{|X_n| > c\}$ 几乎处处只发生有限次。事件 $\{|X_n| > c\}$ 相互独立；若其概率和发散，第二 Borel–Cantelli 引理会给出它们无穷多次发生的概率为 1，矛盾。因此 (5.1) 成立。同样因为 $X_n = Y_n$ 最终成立， $\sum_n Y_n$ 也几乎处处收敛。

必要性：用对称化推出方差可和. 在乘积概率空间上取与 $\{Y_n\}$ 独立且同分布的副本 $\{Y'_n\}$ ，令

$$U_n = Y_n - Y'_n.$$

两份截断级数都几乎处处收敛，故 $\sum_n U_n$ 几乎处处收敛。而 U_n 相互独立、关于 0 对称，且 $|U_n| \leq 2c$ 。固定 $0 < |t| \leq (2c)^{-1}$ 。对 $m \leq n$ ，独立性和对称性给

$$\mathbb{E} \exp\left(it \sum_{k=m}^n U_k\right) = \prod_{k=m}^n a_k, \quad a_k = \mathbb{E} \cos(tU_k) = |\mathbb{E} e^{itY_k}|^2.$$

因为 $|tU_k| \leq 1$, 有

$$\cos 1 \leq a_k \leq 1.$$

级数 $\sum U_n$ 几乎处处收敛, 所以其尾和在概率意义下趋于 0; $\exp(it \sum_{k=m}^n U_k)$ 一致有界, 故其期望随 $m, n \rightarrow \infty$ 趋于 1。于是

$$\prod_{k=m}^n a_k \rightarrow 1.$$

对 $|u| \leq 1$, 初等不等式

$$1 - \cos u \geq \kappa u^2, \quad \kappa = 1 - \cos 1 > 0$$

成立, 因而

$$1 - a_k = \mathbb{E}[1 - \cos(tU_k)] \geq \kappa t^2 \mathbb{E}U_k^2 = 2\kappa t^2 \text{Var}(Y_k).$$

同时 $0 < a_k \leq 1$ 蕴含

$$\prod_{k=m}^n a_k \leq \exp\left\{-\sum_{k=m}^n (1 - a_k)\right\}.$$

左端尾乘积趋于 1, 故右端指数中的非负尾和必须趋于 0。于是 $\sum_k (1 - a_k) < \infty$, 进而 $\sum_k \text{Var}(Y_k) < \infty$, 即(5.3)。

必要性：恢复截断均值系列。由刚得到的方差可和性和 定理 6.2,

$$\sum_n (Y_n - \mathbb{E}Y_n)$$

几乎处处收敛。它与已经收敛的 $\sum_n Y_n$ 相减, 说明确定性部分和

$$\sum_{k=1}^n \mathbb{E}Y_k = \sum_{k=1}^n Y_k - \sum_{k=1}^n (Y_k - \mathbb{E}Y_k)$$

存在有限极限, 所以(5.2)成立。

阈值为何不影响结论。若三个条件对某个 c 成立, 充分性已推出 $\sum X_n$ 几乎处处收敛; 再把必要性应用于任意另一个 $d > 0$, 即得三个条件对 d 也成立。因此阈值可以为证明方便而选取, 但必须是预先固定的正数。□

例题 5.4 (Rademacher 系列: 平方可和恰是判据). 令 ε_n 独立且 $\mathbb{P}(\varepsilon_n = 1) = \mathbb{P}(\varepsilon_n = -1) = 1/2$, 并令 $X_n = a_n \varepsilon_n$, 其中 a_n 为确定实数。则

$$\sum_n a_n \varepsilon_n \text{ 几乎处处收敛} \iff \sum_n a_n^2 < \infty.$$

若右侧成立, 则 $a_n \rightarrow 0$, 取任意固定 $c > 0$ 后大跳最终消失, 截断均值为 0, 截断方差最终等于 a_n^2 , 三系列定理给出收敛。反之, 级数收敛迫使 $|a_n| = |X_n| \rightarrow 0$; 固定 c 后 $Y_n = X_n$ 最终成立, 由必要性得到 $\sum_n a_n^2 < \infty$ 。

反例 5.2 (独立性不能删除). 令单个 Rademacher 变量 ε 满足 $\mathbb{P}(\varepsilon = \pm 1) = 1/2$, 并取

$$X_n = \frac{\varepsilon}{n}.$$

取 $c = 1$ 时没有大跳, $\mathbb{E}X_n = 0$, 且 $\sum_n \text{Var}(X_n) = \sum_n n^{-2} < \infty$, 三个数值条件全部成立; 但

$$\sum_n X_n = \varepsilon \sum_n \frac{1}{n}$$

发散。失败原因是所有 X_n 由同一个 ε 决定, 并不独立。

注 5.1 (与两系列定理及“方差可和”的关系). 若独立变量本来就一致有界, 则第一系列自动满足; 三系列定理退化为“均值系列收敛加方差系列可和”的 Kolmogorov 两系列判据。若还中心, 则只剩方差可和这一充分条件。反过来, 不能把 $\sum \text{Var}(X_n) < \infty$ 当作一般必要条件: 前面的稀有跳例中 $\sum X_n$ 几乎处处只是有限和, 但未截断方差并不可和。

易错点

常见误用有四类: 用未截断方差替代第三系列; 把第二系列误写成绝对收敛; 允许阈值随 n 或样本改变却仍直接套定理; 忽略独立性而只核对三个数值级数。必要性证明中真正困难的是从截断级数收敛推出方差可和, 必须使用对称化和特征函数尾乘积 (或一个等价的严格论证), 不能从 $Y_n \rightarrow 0$ 直接跳到 $\sum \text{Var}(Y_n) < \infty$ 。

答题方法

随机级数题按以下顺序作答:

1. 先核验独立性, 并选一个预先固定的截断阈值 $c > 0$;
2. 计算大跳概率和, 判断是否只有有限个截断差;
3. 计算截断均值级数, 允许条件收敛但必须给出收敛理由;
4. 计算截断方差级数, 必要时用比较判别;
5. 三项全过才宣布几乎处处收敛; 任一项失败就指出是哪一种机制失控;
6. 若变量已中心且最终一致有界, 直接识别为平方可和判据。

自测

1. 构造依赖事件反例说明 BC2 的独立性必要。
2. 两两独立版本中为何选择 $s_{n_j} \geq j^2$?
3. 指数极值上界中为什么有限个例外不会影响最终最大值?
4. 三系列定理的三项分别控制什么机制?
5. 用定理 5.4 判别 $\sum_n \varepsilon_n / \sqrt{n}$ 是否几乎处处收敛。
6. 在必要性证明中, 为什么要引入独立副本而不能直接对 Y_n 的特征函数取实部?

第六章 最大不等式与强大数律

学习目标

本章完成从“中心独立增量”到强大数律的证明主链：Kolmogorov 最大不等式、方差可和级数、有限方差强大数律，以及仅有一阶矩时的截断法。最后以 2025 年秋季真题展示常规均值尺度被稀有大跳破坏时，如何用精确区块构造证明正负振荡。

6.1 为什么只控制终点还不够

Chebyshev 不等式控制 S_n ，但强大数律需要同时控制一段时间内的所有部分和。如果只在稀疏子列上证明 $S_{n_k}/n_k \rightarrow 0$ ，还必须填补子列之间的空隙。Kolmogorov 最大不等式正是把终点方差转换成路径最大值概率。

定理 6.1 (Kolmogorov 最大不等式). 设 $X_1, \dots, X_n$ 独立, $\mathbb{E}X_k = 0, \mathbb{E}X_k^2 < \infty$, 令 $S_k = \sum_{j=1}^k X_j$. 则对任意 $\lambda > 0$,

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| \geq \lambda\right) \leq \frac{\text{Var}(S_n)}{\lambda^2} = \frac{\sum_{k=1}^n \text{Var}(X_k)}{\lambda^2}.$$

证明. 令

$$A_k = \{|S_1| < \lambda, \dots, |S_{k-1}| < \lambda, |S_k| \geq \lambda\}.$$

A_k 两两不交，其并集就是最大值越界事件。又因为 $A_k \in \sigma(X_1, \dots, X_k)$ ，而 $S_n - S_k$ 与该 σ -代数独立且均值为零，

$$\mathbb{E}[S_k(S_n - S_k)\mathbf{1}_{A_k}] = 0.$$

于是

$$\begin{aligned} \mathbb{E}S_n^2 &\geq \sum_{k=1}^n \mathbb{E}[S_n^2 \mathbf{1}_{A_k}] \\ &= \sum_{k=1}^n \mathbb{E}[(S_k + (S_n - S_k))^2 \mathbf{1}_{A_k}] \\ &\geq \sum_{k=1}^n \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] \geq \lambda^2 \sum_{k=1}^n \mathbb{P}(A_k). \end{aligned}$$

最后利用独立中心增量使 $\mathbb{E}S_n^2 = \sum_k \text{Var}(X_k)$ 。 □

易错点

交叉项为零不是因为 S_k 与 $S_n - S_k$ 在事件 A_k 上“仍独立”，而是 $\mathbf{1}_{A_k} S_k$ 对过去可测，未来增量独立于整个过去且均值为零。若停止事件偷用了未来信息，这一步会失败。

6.2 方差可和的随机级数

定理 6.2 (Kolmogorov 收敛判别). 若 X_n 独立、中心、平方可积, 且

$$\sum_{n=1}^{\infty} \text{Var}(X_n) < \infty,$$

则 $\sum_n X_n$ 几乎处处且在 L^2 中收敛。

证明. L^2 收敛很直接: 对 $m > n$,

$$\mathbb{E}|S_m - S_n|^2 = \sum_{k=n+1}^m \text{Var}(X_k) \rightarrow 0,$$

所以 (S_n) 在完备空间 L^2 中 Cauchy。

为证几乎处处收敛, 选 n_r 使 $\sum_{k>n_r} \text{Var}(X_k) \leq 2^{-3r}$ 。由 [定理 6.1](#),

$$\mathbb{P}\left(\max_{n_r < k \leq n_{r+1}} |S_k - S_{n_r}| > 2^{-r}\right) \leq 2^{2r} \sum_{k>n_r} \text{Var}(X_k) \leq 2^{-r}.$$

Borel–Cantelli 说明这些区块内最大振幅最终至多 2^{-r} 。同时 S_{n_r} 的 L^2 极限存在, 从中可抽取几乎处处收敛子列; 区块振幅趋零便把收敛填充到全部 S_n 。□

6.3 有限方差强大数律

定理 6.3 (有限方差 SLLN). 若 X_n 独立同分布, $\mathbb{E}X_1 = \mu$, $\text{Var}(X_1) = \sigma^2 < \infty$, 则

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow \mu \quad \text{几乎处处}.$$

证明. 中心化后可设 $\mu = 0$, 记 $S_n = \sum_{k \leq n} X_k$ 。对二进终点, [定理 6.1](#) 给

$$\mathbb{P}\left(\max_{1 \leq k \leq 2^r} |S_k| \geq \varepsilon 2^r\right) \leq \frac{2^r \sigma^2}{\varepsilon^2 2^{2r}} = \frac{\sigma^2}{\varepsilon^2 2^r}.$$

右端对 r 可求和, 所以几乎处处最终

$$\max_{k \leq 2^r} |S_k| < \varepsilon 2^r.$$

若 $2^{r-1} < n \leq 2^r$, 则

$$\frac{|S_n|}{n} \leq \frac{\max_{k \leq 2^r} |S_k|}{2^{r-1}} < 2\varepsilon.$$

对有理 $\varepsilon \downarrow 0$ 取可数交, 得到 $S_n/n \rightarrow 0$ 。□

注 6.1. 上述证明用了整段前缀最大值, 常数略松但逻辑完整。也可先证 $S_{2^r}/2^r \rightarrow 0$, 再对二进区块增量使用最大不等式; 两条路线本质相同。

6.4 仅有一阶矩时：截断到底解决什么

定理 6.4 (Kolmogorov 强大数律). 若 X_n 独立同分布且 $\mathbb{E}|X_1| < \infty$, 则

$$\frac{1}{n} \sum_{k=1}^n X_k \longrightarrow \mathbb{E}X_1 \quad \text{几乎处处.}$$

证明. 令

$$Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}.$$

首先,

$$\sum_{n \geq 1} \mathbb{P}(X_n \neq Y_n) = \sum_{n \geq 1} \mathbb{P}(|X_1| > n) \leq \mathbb{E}|X_1| < \infty.$$

由第一 Borel–Cantelli, 引入截断只改变有限多项, 故

$$\frac{1}{n} \sum_{k \leq n} (X_k - Y_k) \rightarrow 0 \quad \text{几乎处处.}$$

其次, 截断后的中心波动满足

$$\begin{aligned} \sum_{n \geq 1} \frac{\text{Var}(Y_n)}{n^2} &\leq \sum_{n \geq 1} \frac{\mathbb{E}[X_1^2 \mathbf{1}_{\{|X_1| \leq n\}}]}{n^2} \\ &= \mathbb{E} \left[X_1^2 \sum_{n \geq \lceil |X_1| \rceil} \frac{1}{n^2} \right] \leq C \mathbb{E}|X_1| < \infty. \end{aligned}$$

于是由 定理 6.2,

$$\sum_{n \geq 1} \frac{Y_n - \mathbb{E}Y_n}{n}$$

几乎处处收敛。Kronecker 引理给出

$$\frac{1}{n} \sum_{k=1}^n (Y_k - \mathbb{E}Y_k) \rightarrow 0 \quad \text{几乎处处.}$$

这里 Kronecker 引理可由分部求和看出: 若 $T_n = \sum_{k \leq n} a_k/k$ 收敛, 则

$$\frac{1}{n} \sum_{k \leq n} a_k = T_n - \frac{1}{n} \sum_{k=1}^{n-1} T_k \longrightarrow 0$$

(对收敛序列使用 Cesàro)。

最后必须正确处理均值: X_n 与 X_1 同分布, 而不是 Y_n 与 X_n 同分布; 实际是

$$\mathbb{E}Y_n = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \longrightarrow \mathbb{E}X_1$$

由控制收敛成立。Cesàro 平均遂给 $n^{-1} \sum_{k \leq n} \mathbb{E}Y_k \rightarrow \mathbb{E}X_1$ 。合并三项即得结论。 $\square$

易错点

截断变量 $Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}$ 的分布随 n 改变, 绝不能写 “ Y_n 与 X_n 同分布”。严格表述是: X_n 与 X_1 同分布, 并且 $\mathbb{E}Y_n = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \rightarrow \mathbb{E}X_1$ 。

6.5 当普通均值不收敛: 区块振荡法

下面的正式真题中, $\mathbb{E}X_n = 0$, 但方差随 n 增长, 偶发的 $4n$ 大跳与持续的 -1 背景竞争。正向振荡由单个大跳产生; 负向振荡必须构造一整块 “没有大跳且负值足够多” 的区间。

正式真题: 2025 年秋季概率统计 Q4

设 $X_1, X_2, \dots$ 相互独立, 且

$$\mathbb{P}(X_n = -1) = \frac{1}{2}, \quad \mathbb{P}(X_n = 4n) = \frac{1}{8n}, \quad \mathbb{P}(X_n = 0) = \frac{1}{2} - \frac{1}{8n}.$$

令 $S_n = \sum_{k=1}^n X_k$ 。证明

$$\liminf_{n \rightarrow \infty} \frac{S_n}{n} < 0 < \limsup_{n \rightarrow \infty} \frac{S_n}{n} \quad \text{几乎处处}.$$

完整解答 (独立核验稿) **正向振荡。** 令 $A_n = \{X_n = 4n\}$ 。这些事件独立, 且

$$\sum_n \mathbb{P}(A_n) = \sum_n \frac{1}{8n} = \infty.$$

由第二 Borel–Cantelli, A_n 几乎处处发生无穷多次。又因每个 $X_j \geq -1$, 在 A_n 上

$$S_n = S_{n-1} + 4n \geq -(n-1) + 4n = 3n + 1,$$

故 $\limsup S_n/n \geq 3 > 0$ 。

负向振荡: 精确区块。 取整数 $N_2 \geq 2$, 并从 $r \geq 2$ 开始递推

$$N_{r+1} = \lfloor r^2 N_r \rfloor, \quad I_r = \{N_r + 1, \dots, N_{r+1}\}, \quad L_r = N_{r+1} - N_r.$$

这样 I_r 非空且两两不交; 从 S_{N_r} 到 $S_{N_{r+1}}$ 的增量恰好由 I_r 中变量组成。定义

$$G_r = \{X_j \neq -1, j \in I_r\}, \quad H_r = \{\#\{j \in I_r : X_j = -1\} \geq 0.4L_r\}.$$

G_r, H_r 都只依赖同一个区块 I_r 。

先估计无大跳概率。利用 $\log(1-x) = -x + O(x^2)$,

$$\mathbb{P}(G_r) = \prod_{j=N_r+1}^{N_{r+1}} \left(1 - \frac{1}{8j}\right),$$

$$\log \mathbb{P}(G_r) = -\frac{1}{8} \sum_{j=N_r+1}^{N_{r+1}} \frac{1}{j} + O(N_r^{-1}) = -\frac{1}{8} \log \frac{N_{r+1}}{N_r} + O(1/N_r).$$

由于 $N_{r+1}/N_r \sim r^2$, 存在常数 $c, C > 0$ 使充分大 r 时

$$cr^{-1/4} \leq \mathbb{P}(G_r) \leq Cr^{-1/4}.$$

在给定 G_r 后, 各坐标仍相互独立, 并且对 $j \in I_r$,

$$\mathbb{P}(X_j = -1 \mid X_j \neq 4j) = \frac{1/2}{1 - 1/(8j)} \geq \frac{1}{2}.$$

因此负值个数随机占优于 $\text{Bin}(L_r, 1/2)$ 。Hoeffding–Chernoff 下尾界给出

$$\mathbb{P}(H_r^c \mid G_r) \leq \exp[-2(0.5 - 0.4)^2 L_r] = e^{-L_r/50}.$$

于是

$$\mathbb{P}(G_r \cap H_r) = \mathbb{P}(G_r)\mathbb{P}(H_r \mid G_r) \geq cr^{-1/4}(1 - e^{-L_r/50})$$

对充分大 r 成立, 其级数发散。由于 $G_r \cap H_r$ 分别只依赖不交区块, 这些事件相互独立; 第二 Borel–Cantelli 说明它们几乎处处发生无穷多次。

还需控制每个区块起点。直接计算

$$\mathbb{E}X_j = -\frac{1}{2} + 4j \frac{1}{8j} = 0, \quad \text{Var}(X_j) = \mathbb{E}X_j^2 = \frac{1}{2} + 2j \leq 3j.$$

故 $\text{Var}(S_n) \leq \frac{3}{2}n(n+1)$ 。Chebyshev 给出

$$\sum_{r \geq 2} \mathbb{P}(S_{N_r} > rN_r) \leq \sum_{r \geq 2} \frac{CN_r^2}{r^2 N_r^2} < \infty.$$

由第一 Borel–Cantelli, 几乎处处最终 $S_{N_r} \leq rN_r$ 。

在同时满足 $G_r \cap H_r$ 且 $S_{N_r} \leq rN_r$ 的无穷多个区块上, 区块中没有正大跳、至少 $0.4L_r$ 个 -1 , 其余只能为 0 , 所以

$$S_{N_{r+1}} = S_{N_r} + \sum_{j \in I_r} X_j \leq rN_r - 0.4L_r.$$

因此

$$\frac{S_{N_{r+1}}}{N_{r+1}} \leq \frac{rN_r - 0.4(N_{r+1} - N_r)}{N_{r+1}} \longrightarrow -0.4,$$

因为 $N_{r+1}/N_r \sim r^2$ 。故 $\liminf S_n/n \leq -0.4 < 0$, 与正向结论合并即证。

易错点

本题最容易出现四种索引错误：

1. 从 $r = 1$ 开始会令 $N_2 = \lfloor N_1 \rfloor$ ，可能产生空区块；
2. G_r 与 H_r 若使用不同端点，就不能把区块增量精确写成其和；
3. 只证明 $\mathbb{P}(G_r) \asymp r^{-1/4}$ 而不控制 H_r ，不足以保证负值数量；
4. 使用第二引理前必须指出区块事件因 I_r 不交而相互独立。

答题方法

证明强大数律或振荡时，先辨认波动来源：

1. 方差均匀有限：最大不等式加几何子列；
2. 只有一阶矩：按随时间增长的阈值截断，分别处理大跳、中心波动、均值漂移；
3. 稀有大跳决定上极限：对单点事件使用第二 Borel–Cantelli；
4. 需要持续负漂移：构造不交区块，确保“无正跳”和“负项足够多”都只依赖同一区块；
5. 最后把区块端点估计除以正确的终点 N_{r+1} 。

本节结论

Kolmogorov 最大不等式把独立中心增量的方差转成整段路径控制。有限方差 SLLN 由几何尺度上的可求和尾界得到；仅一阶矩时，截断先排除有限多个大跳，再以方差可和与 Kronecker 引理控制中心部分。非常规三角阵列则需把稀有大跳和区块背景分开分析。

自测

1. 在定理 6.1 中说明独立性具体用在哪个等式。
2. 证明若 $\sum_n \text{Var}(X_n) < \infty$ 且 X_n 独立中心，则 $\sum_n X_n$ 的 L^2 极限与几乎处处极限相同。
3. 补出估计 $\sum_{n \geq [x]} n^{-2} \leq C/x$ ($x \geq 1$)。
4. 在 2025 秋 Q4 中，若把 N_{r+1} 改为 $2N_r$ ，说明为什么 $\sum_r \mathbb{P}(G_r)$ 虽发散，但终点负向不等式未必压过 rN_r 。

自测核对要点 第 1 题用于令未来增量与过去停止事件的交叉项为零。第 2 题利用 L^2 收敛蕴含依概率收敛，而同一序列的几乎处处极限与依概率极限必须一致。第 3 题与积分 $\int_{x-1}^{\infty} t^{-2} dt$ 比较。第 4 题中负区块长度仅与 N_r 同阶，无法压过允许的起点上界 rN_r ；选择 $r^2 N_r$ 正是为此。

6.6 Kolmogorov 最大不等式的停时分割

设独立、均值零且方差有限的 $X_1, \dots, X_n$, $S_k = \sum_{i=1}^k X_i$. 令

$$A_k = \{|S_1| < \lambda, \dots, |S_{k-1}| < \lambda, |S_k| \geq \lambda\}.$$

A_k 两两不交, 且由独立性, A_k 与 $S_n - S_k$ 独立。于是

$$\begin{aligned}\mathbb{E}[S_n^2 \mathbf{1}_{A_k}] &= \mathbb{E}[(S_k + S_n - S_k)^2 \mathbf{1}_{A_k}] \\ &= \mathbb{E}[S_k^2 \mathbf{1}_{A_k}] + \mathbb{E}[(S_n - S_k)^2] \mathbb{P}(A_k),\end{aligned}$$

交叉项因未来增量均值零而消失。因此

$$\mathbb{E}[S_n^2 \mathbf{1}_{A_k}] \geq \lambda^2 \mathbb{P}(A_k).$$

求和得

$$\mathbb{P}\left(\max_{k \leq n} |S_k| \geq \lambda\right) \leq \frac{\text{Var}(S_n)}{\lambda^2}.$$

事件按“首次越界时刻”分割, 是证明中不能省略的结构。

6.7 方差可和级数与 Kronecker 引理

若独立均值零 X_n 满足

$$\sum_n \text{Var}(X_n) < \infty,$$

最大不等式与 Borel–Cantelli 可证明 $\sum_n X_n$ a.s. 收敛。

Kronecker 引理: 若 $b_n \uparrow \infty$ 且 $\sum_n x_n/b_n$ 收敛, 则

$$\frac{1}{b_n} \sum_{k=1}^n x_k \rightarrow 0.$$

取 $b_n = n$ 。若独立均值零且

$$\sum_n \frac{\text{Var}(X_n)}{n^2} < \infty,$$

则 $\sum X_n/n$ a.s. 收敛, 进而

$$\frac{1}{n} \sum_{k=1}^n X_k \rightarrow 0 \quad a.s.$$

这给非同分布强大数律的一条标准路线。

6.8 仅一阶矩下截断法的完整主线

设 X_n i.i.d., $\mathbb{E}|X_1| < \infty$, 令

$$Y_n = X_n \mathbf{1}_{\{|X_n| \leq n\}}.$$

因为 X_n 与 X_1 同分布,

$$\sum_n \mathbb{P}(X_n \neq Y_n) = \sum_n \mathbb{P}(|X_1| > n) \leq \mathbb{E}|X_1| < \infty.$$

所以 $X_n = Y_n$ 最终 a.s.。

又

$$\sum_n \frac{\text{Var}(Y_n)}{n^2} \leq \sum_n \frac{\mathbb{E}[Y_n^2]}{n^2}.$$

按 $|X_1|$ 所处整数层交换求和, 可证右侧有限。于是

$$\frac{1}{n} \sum_{k=1}^n (Y_k - \mathbb{E}Y_k) \rightarrow 0 \quad a.s.$$

同时必须严格写成

$$\mathbb{E}Y_n = \mathbb{E}[X_n \mathbf{1}_{\{|X_n| \leq n\}}] = \mathbb{E}[X_1 \mathbf{1}_{\{|X_1| \leq n\}}] \rightarrow \mathbb{E}X_1.$$

这里用的是 X_n 与 X_1 同分布; 绝不能说 Y_n 与 X_n 同分布。Cesàro 平均再给

$$\frac{1}{n} \sum_{k=1}^n \mathbb{E}Y_k \rightarrow \mathbb{E}X_1.$$

合并并利用最终相等, 得到 SLLN。

6.9 重尾下普通均值为什么失败

若尾部如此重以至 $\mathbb{E}|X| = \infty$, 截断均值可能随截断水平漂移, 单个极端值也可能与前 n 项和同阶。此时普通 $n^{-1}S_n$ 未必收敛。诊断顺序是:

1. 检查 $\mathbb{E}|X|$;
2. 计算 $n\mathbb{P}(|X| > n)$ 或相应大跳频率;
3. 选择稳定分布尺度或截断/中心化;
4. 若需证明振荡, 用互不相交区块建立独立事件。

易错点

区块证明必须使事件只依赖各自区块。对 2025 秋 Q4, 取

$$I_r = \{N_r + 1, \dots, N_{r+1}\},$$

并令 G_r, H_r 都只依赖 I_r , 递推从 $r \geq 2$ 开始。这样从 S_{N_r} 到 $S_{N_{r+1}}$ 的增量与区块精确对应, 区块独立、乘积概率与 Chernoff 界才合法。

自测

1. 最大不等式证明中交叉项为何为零?
2. Kronecker 引理如何把随机级数收敛转成平均收敛?
3. 一阶矩 SLLN 截断中, Y_n 与哪个变量同分布、与哪个不一定同分布?
4. 重尾区块法为何必须让事件精确依赖不相交区块?

第七章 中心极限定理与三角阵列

学习目标

本章回答“总和为何近似正态，以及应按什么尺度归一化”。需要会证明独立同分布中心极限定理，理解 Lindeberg 条件如何排除单项主导，掌握 Lyapunov 的可检验充分条件，并能把稀有大跳与产生 Gaussian 波动的主体拆开。

7.1 从大数律到中心极限定理

大数律说明样本均值趋近常数，却不描述误差的形状。若 $S_n = \sum_{k=1}^n X_k, \mathbb{E}X_k = \mu, \text{Var}(X_k) = \sigma^2$ ，则 $S_n - n\mu$ 的方差是 $n\sigma^2$ ，自然尺度是 $\sqrt{n}$ ：

$$\frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

中心极限定理说明，在有限方差和适当独立性下，这个量趋于标准正态。

定理 7.1 (Lévy 连续性定理). 若随机变量 X_n 的特征函数 $\phi_n(t)$ 逐点收敛到函数 $\phi(t)$ ，且 ϕ 在 0 处连续并且是某个随机变量 X 的特征函数，则 $X_n \Rightarrow X$ 。反之，依分布收敛推出特征函数逐点收敛。

该定理把分布收敛转化为复函数的乘积极限，是独立和 CLT 的标准工具。

定理 7.2 (Lindeberg–Lévy CLT). 若 X_n 独立同分布， $\mathbb{E}X_1 = \mu$ ， $0 < \sigma^2 = \text{Var}(X_1) < \infty$ ，则

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sigma\sqrt{n}} \Rightarrow N(0, 1).$$

证明. 中心化并除以 σ ，只需考虑 $\mathbb{E}X_1 = 0, \mathbb{E}X_1^2 = 1$ 。由二阶矩有限，特征函数在 0 处有展开

$$\phi_X(u) = 1 - \frac{u^2}{2} + o(u^2).$$

确切地，

$$\frac{e^{iuX} - 1 - iuX}{u^2} \longrightarrow -\frac{X^2}{2}$$

逐点成立，且左边绝对值可由 CX^2 控制；控制收敛给出该展开。独立性于是得到

$$\phi_{S_n/\sqrt{n}}(t) = \left[\phi_X\left(\frac{t}{\sqrt{n}}\right) \right]^n = \left[1 - \frac{t^2}{2n} + o(n^{-1}) \right]^n \longrightarrow e^{-t^2/2}.$$

右端是标准正态特征函数且在 0 连续，由 [定理 7.1](#) 得结论。□

7.2 三角阵列与 Lindeberg 条件

不同年份或不同位置的随机变量分布可能改变, 此时写成三角阵列

$$X_{n1}, \dots, X_{nk_n}.$$

每一行内独立, 令 $\mathbb{E}X_{nk} = 0$,

$$s_n^2 = \sum_{k=1}^{k_n} \text{Var}(X_{nk}).$$

标准化后希望 $\sum_k X_{nk}/s_n \Rightarrow N(0, 1)$ 。仅有总方差归一化还不够; 如果一个单项占据固定比例方差, 极限会保留它的非正态形状。

定义 7.1 (Lindeberg 条件). 对每个 $\varepsilon > 0$,

$$L_n(\varepsilon) = \frac{1}{s_n^2} \sum_{k=1}^{k_n} \mathbb{E}[X_{nk}^2 \mathbf{1}_{\{|X_{nk}| > \varepsilon s_n\}}] \rightarrow 0.$$

定理 7.3 (Lindeberg–Feller CLT: 充分性). 设每行变量独立、中心、 $s_n^2 > 0$ 。若 [定义 7.1](#) 成立, 则

$$\frac{\sum_{k=1}^{k_n} X_{nk}}{s_n} \Rightarrow N(0, 1).$$

证明. 令 $Y_{nk} = X_{nk}/s_n$, 则 $\sum_k \mathbb{E}Y_{nk}^2 = 1$ 。Lindeberg 条件先推出无穷小性:

$$\max_k \mathbb{P}(|Y_{nk}| > \varepsilon) \leq \varepsilon^{-2} \sum_k \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| > \varepsilon\}}] \rightarrow 0,$$

并且

$$\max_k \mathbb{E}Y_{nk}^2 \leq \varepsilon^2 + \sum_j \mathbb{E}[Y_{nj}^2 \mathbf{1}_{\{|Y_{nj}| > \varepsilon\}}],$$

所以先令 $n \rightarrow \infty$ 再令 $\varepsilon \downarrow 0$, 得到 $\max_k \mathbb{E}Y_{nk}^2 \rightarrow 0$ 。

固定 t , 记 $\phi_{nk}(t) = \mathbb{E}e^{itY_{nk}}$ 。因为 $\mathbb{E}Y_{nk} = 0$,

$$\phi_{nk}(t) = 1 - \frac{t^2}{2} \mathbb{E}Y_{nk}^2 + r_{nk}.$$

对任意 $\delta > 0$, Taylor 余项满足某个常数 C_t 下

$$|r_{nk}| \leq \delta \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| \leq \eta\}}] + C_t \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| > \eta\}}],$$

其中先取足够小的 η 使局部余项系数不超过 δ 。对 k 求和并用 $\sum_k \mathbb{E}Y_{nk}^2 = 1$ 与 Lindeberg 条件, 得到

$$\limsup_{n \rightarrow \infty} \sum_k |r_{nk}| \leq \delta.$$

再令 $\delta \downarrow 0$, 故 $\sum_k |r_{nk}| \rightarrow 0$ 。同时 $\max_k |\phi_{nk}(t) - 1| \rightarrow 0$ 。令 $u_{nk} = \phi_{nk}(t) - 1$ 。上式和余项估计还给 $\sum_k |u_{nk}| = O_t(1)$, 故充分大 n 时 $\max_k |u_{nk}| < 1/2$, 可以在以 1 为中心且不穿过负实轴

的圆盘内统一取主值复对数。因每行独立，

$$\log \phi_{\sum_k Y_{nk}}(t) = \sum_k \log \phi_{nk}(t).$$

利用 $\log(1+z) = z + O(|z|^2)$ ，以及

$$\sum_k |\phi_{nk}(t) - 1|^2 \leq \max_k |\phi_{nk}(t) - 1| \sum_k |\phi_{nk}(t) - 1| \rightarrow 0,$$

且 $\sum_k r_{nk} \rightarrow 0$ ，可得

$$\log \phi_{\sum_k Y_{nk}}(t) \rightarrow -\frac{t^2}{2} \sum_k \mathbb{E} Y_{nk}^2 = -\frac{t^2}{2}.$$

故特征函数趋于 $e^{-t^2/2}$ ，由 Lévy 连续性定理完成证明。□

注 7.1 (Feller 条件与必要性). Feller 无穷小条件

$$\max_k \frac{\text{Var}(X_{nk})}{s_n^2} \rightarrow 0$$

比 Lindeberg 弱。若三角阵列 CLT 成立并加上 Feller 条件，则 Lindeberg 条件也是必要的。考试中若只要求证明正态极限，通常验证 Lindeberg 或更强的 Lyapunov 条件即可。

7.3 Lyapunov 条件

定理 7.4 (Lyapunov CLT). 若存在 $\delta > 0$ 使

$$\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{k_n} \mathbb{E} |X_{nk}|^{2+\delta} \rightarrow 0,$$

则 Lindeberg 条件成立，从而标准化行和趋于 $N(0, 1)$ 。

证明. 在 $|X_{nk}| > \varepsilon s_n$ 上，

$$X_{nk}^2 \leq \frac{|X_{nk}|^{2+\delta}}{(\varepsilon s_n)^\delta}.$$

代入 定义 7.1:

$$L_n(\varepsilon) \leq \varepsilon^{-\delta} s_n^{-(2+\delta)} \sum_k \mathbb{E} |X_{nk}|^{2+\delta} \rightarrow 0.$$

□

例题 7.1 (加权求和的 CLT). 设 Z_k 独立同分布、中心、方差 1，且 $\mathbb{E}|Z_1|^{2+\delta} < \infty$ 。对系数 a_{nk} ，令 $s_n^2 = \sum_k a_{nk}^2$ 。若

$$\frac{\max_k |a_{nk}|}{s_n} \rightarrow 0,$$

则

$$\frac{\sum_k a_{nk} Z_k}{s_n} \Rightarrow N(0, 1).$$

因为

$$\frac{\sum_k \mathbb{E}|a_{nk}Z_k|^{2+\delta}}{s_n^{2+\delta}} \leq \mathbb{E}|Z_1|^{2+\delta} \left(\frac{\max_k |a_{nk}|}{s_n} \right)^\delta.$$

7.4 方差尺度不是机械相加：先拆稀有大跳

有些数组的方差被极少发生、幅度极大的项主导，但这些项几乎处处只出现有限次，并不影响以较小尺度观察的长期分布。此时若盲目把总方差作为归一化尺度，会掩盖真正形成 Gaussian 波动的主体。正确做法是：

$$X_j = Y_j + R_j,$$

对主体 Y_j 验证 CLT，对稀有部分 R_j 用 Borel–Cantelli 证明可忽略。

正式真题：2026 年春季概率统计 Q2

设 $X_1, X_2, \dots$ 相互独立，且

$$\mathbb{P}(X_n = n^2) = \mathbb{P}(X_n = -n^2) = \frac{1}{6n^2},$$

$$\mathbb{P}(X_n = n) = \mathbb{P}(X_n = -n) = \frac{1}{6}, \quad \mathbb{P}(X_n = 0) = \frac{2}{3} - \frac{1}{3n^2}.$$

令 $S_n = \sum_{j=1}^n X_j$ 。求一列 $b_n > 0$ ，使 $S_n/b_n \Rightarrow N(0, 1)$ 。

完整解答（独立核验稿） 定义

$$Y_j = X_j \mathbf{1}_{\{|X_j|=j\}}, \quad R_j = X_j \mathbf{1}_{\{|X_j|=j^2\}},$$

则 $X_j = Y_j + R_j$ 。两部分都中心，但不需要在同一 j 内相互独立。主体满足

$$\text{Var}(Y_j) = \mathbb{E}Y_j^2 = j^2 \left(\frac{1}{6} + \frac{1}{6} \right) = \frac{j^2}{3}.$$

因此取

$$b_n^2 = \sum_{j=1}^n \text{Var}(Y_j) = \frac{1}{3} \sum_{j=1}^n j^2 = \frac{n(n+1)(2n+1)}{18} \sim \frac{n^3}{9}.$$

验证 Lindeberg 条件。由于 $|Y_j| \leq j \leq n$ 且 $b_n \asymp n^{3/2}$,

$$\max_{j \leq n} \frac{|Y_j|}{b_n} \leq \frac{n}{b_n} \rightarrow 0.$$

故对任意固定 $\varepsilon > 0$ ，充分大 n 时所有指标函数 $\mathbf{1}_{\{|Y_j| > \varepsilon b_n\}}$ 都为零。由 [定理 7.3](#),

$$\frac{\sum_{j=1}^n Y_j}{b_n} \Rightarrow N(0, 1).$$

稀有大跳满足

$$\sum_{j=1}^{\infty} \mathbb{P}(R_j \neq 0) = \sum_{j=1}^{\infty} \frac{1}{3j^2} < \infty.$$

第一 Borel–Cantelli 说明几乎处处只有有限多个 R_j 非零，所以 $\sum_{j \leq n} R_j$ 最终是一个有限随机常数。因 $b_n \rightarrow \infty$,

$$\frac{\sum_{j=1}^n R_j}{b_n} \rightarrow 0 \quad \text{几乎处处.}$$

最后由 Slutsky,

$$b_n = \sqrt{\frac{n(n+1)(2n+1)}{18}}, \quad \frac{S_n}{b_n} \Rightarrow N(0, 1).$$

易错点

上题若直接使用 $\text{Var}(X_j)$, j^4 级大跳会给出完全不同的总方差尺度。但这些大跳概率可求和，样本路径上最终不再出现；“方差贡献巨大”与“在分布极限中持续出现”不是同一件事。拆分后也不应声称 Y_j, R_j 独立，Slutsky 只需稀有部分依概率趋零。

7.5 CLT 之后的变换

若已经得到

$$\frac{T_n - a_n}{b_n} \Rightarrow N(0, 1),$$

再处理光滑函数、随机分母或可忽略余项，应使用第 四 章的 Delta 方法和 Slutsky，而不必重复证明 CLT。典型结构为

$$\sqrt{n}[g(\bar{X}_n) - g(\mu)] = g'(\mu)\sqrt{n}(\bar{X}_n - \mu) + o_{\mathbb{P}}(1).$$

若研究向量函数，则用 Jacobian:

$$\sqrt{n}[g(T_n) - g(\theta)] \Rightarrow N(0, Dg(\theta)\Sigma Dg(\theta)^{\top}).$$

答题方法

中心极限定理题的作答流程:

1. 中心化并计算候选方差尺度;
2. 检查是否有单项或稀有大跳主导，必要时先拆分;
3. 同分布有限方差用经典 CLT; 非同分布用 Lindeberg 或 Lyapunov;
4. 写出每个条件的数值验证，不能只报定理名称;
5. 可忽略项用几乎处处或依概率收敛，再由 Slutsky 合并;
6. 非线性统计量最后用 Delta 方法，并核验导数与定义域。

本节结论

CLT 的尺度来自累计的主体方差，正态性来自“每个增量都渐近可忽略”。Lindeberg 条件在二阶矩层面排除大项，Lyapunov 用更高阶矩提供简便验证。若高方差来自概率可求和的稀有跳跃，应先用 Borel–Cantelli 拆除，再对主体使用三角阵列 CLT。

自测

1. 在经典 CLT 的特征函数证明中，指出有限二阶矩用于何处。
2. 构造满足 Feller 条件但不满足 Lindeberg 条件的三角阵列，并解释极限为什么不是标准正态。
3. 若 $\sum_k a_{nk}^2 = 1$ 且 $\max_k |a_{nk}| \rightarrow 0$ ，证明 Rademacher 加权和趋于标准正态。
4. 在 2026 春 Q2 中说明为何 $\sum_{j \leq n} R_j/b_n \rightarrow 0$ 的结论比仅证明其依概率趋零更强。

自测核对要点 第 1 题用于获得 $\phi_X(u) = 1 - u^2/2 + o(u^2)$ 的控制收敛展开。第 2 题可让每行保留总方差中不可忽略的稀有固定幅度项。第 3 题用 Lindeberg：标准化后最大项绝对值趋零。第 4 题 Borel–Cantelli 给出样本路径上稀有和最终固定，因而得到几乎处处收敛，当然也蕴含依概率收敛。

7.6 经典 CLT 的特征函数证明

设 X_i i.i.d., $\mathbb{E}X_i = 0$, $\text{Var}(X_i) = \sigma^2$ 。记

$$S_n^* = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n X_i.$$

单个标准化变量的特征函数在零附近满足

$$\varphi(t) = 1 - \frac{t^2}{2} + o(t^2).$$

因此

$$\varphi_{S_n^*}(t) = \left[\varphi\left(\frac{t}{\sqrt{n}}\right) \right]^n = \left[1 - \frac{t^2}{2n} + o(n^{-1}) \right]^n \rightarrow e^{-t^2/2}.$$

右侧是标准正态特征函数，Lévy 连续性定理给

$$S_n^* \xrightarrow{d} N(0, 1).$$

复对数写法需确认因子最终位于 1 的邻域并选同一主支；或直接用 $(1 + a_n/n)^n \rightarrow e^a$ 避免支路问题。

7.7 Lindeberg 条件为何排除单个大项

三角阵列 Y_{nk} 独立、均值零，行方差归一化为

$$\sum_k \mathbb{E} Y_{nk}^2 = 1.$$

Lindeberg 条件要求对每个 $\varepsilon > 0$,

$$\sum_k \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| > \varepsilon\}}] \rightarrow 0.$$

它推出 Feller 条件

$$\max_k \text{Var}(Y_{nk}) \rightarrow 0,$$

因为对固定 ε

$$\mathbb{E} Y_{nk}^2 \leq \varepsilon^2 + \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| > \varepsilon\}}],$$

先取行最大与极限，再让 $\varepsilon \downarrow 0$ 。

特征函数 Taylor 余项必须写为两部分之和：

$$|r_{nk}| \leq \delta \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| \leq \eta\}}] + C_t \mathbb{E}[Y_{nk}^2 \mathbf{1}_{\{|Y_{nk}| > \eta\}}].$$

求和后第一项至多 δ ，第二项由 Lindeberg 消失；再令 $\delta \downarrow 0$ 。随后 Feller 条件保证每个特征函数因子接近一、二次余项总和可控，复对数展开才合法。

7.8 Lyapunov 条件与 Lindeberg

若存在 $\delta > 0$ ，总方差

$$s_n^2 = \sum_k \text{Var}(X_{nk})$$

且

$$\frac{1}{s_n^{2+\delta}} \sum_k \mathbb{E}|X_{nk} - \mathbb{E}X_{nk}|^{2+\delta} \rightarrow 0,$$

则对任意 $\varepsilon > 0$

$$X^2 \mathbf{1}_{\{|X| > \varepsilon s_n\}} \leq \frac{|X|^{2+\delta}}{(\varepsilon s_n)^\delta}.$$

求和即得 Lindeberg 条件。Lyapunov 更易核验但更强；没有高于二阶的矩时，Lindeberg 仍可能成立。

7.9 Berry–Esseen 与正态近似误差

若 i.i.d. 中心变量满足 $\mathbb{E}|X_1|^3 < \infty$ ，则存在普适常数 C ：

$$\sup_x \left| \mathbb{P}\left(\frac{S_n}{\sigma\sqrt{n}} \leq x\right) - \Phi(x) \right| \leq \frac{C \mathbb{E}|X_1|^3}{\sigma^3 \sqrt{n}}.$$

CLT 只给极限，不给有限 n 误差；Berry-Esseen 给 $n^{-1/2}$ 量级。若偏度很大，常数前因子可能使中等样本正态近似仍较差。

7.10 Cramér-Wold 与多维 CLT

随机向量 $Z_n \xrightarrow{d} Z$ 当且仅当对每个固定 $a \in \mathbb{R}^d$,

$$a^\top Z_n \xrightarrow{d} a^\top Z.$$

对 i.i.d. 向量 X_i 均值 μ 、协方差 Σ ，任意 a 上的一维 CLT 给

$$a^\top \sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, a^\top \Sigma a).$$

Cramér-Wold 推出

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N_d(0, \Sigma).$$

答题方法

CLT 题先中心化、归一化总方差，再检查最大项。i.i.d. 用经典 CLT；非同分布阵列核验 Lindeberg/Lyapunov；向量用 Cramér-Wold；非线性统计量再接 Delta 方法。

自测

1. 从单变量特征函数二阶展开证明经典 CLT。
2. Lindeberg 如何推出最大单项方差消失？
3. Lyapunov 为什么比 Lindeberg 强？
4. CLT 与 Berry-Esseen 的结论差别是什么？
5. 多维 CLT 中 Cramér-Wold 具体降成哪些一维变量？

第二部分

Stochastic Processes

第八章 鞅、停时与可选停止

学习目标

本章从滤过和适应性定义鞅，建立条件 Jensen、鞅差、指数鞅、停时与 Doob 最大不等式；重点是准确陈述可选停止的条件并识别失败反例。

8.1 滤过、适应性与停时

滤过 $(\mathcal{F}_n)$ 是递增的 σ -代数族；过程 X_n 适应是指 X_n 对 $\mathcal{F}_n$ 可测。随机时刻 τ 是停时若 $\{\tau \leq n\} \in \mathcal{F}_n$ 。首次进入可测集合通常是停时，而“最后一次到达”通常不是。

停时 σ -代数定义为

$$\mathcal{F}_\tau = \{A \in \mathcal{F} : A \cap \{\tau \leq n\} \in \mathcal{F}_n \forall n\}.$$

若 X_n 适应，则停下过程 $X_{n \wedge \tau}$ 仍适应。连续时间定义把 n 换成 t ；对连续适应过程，闭集首次达时在通常条件下是停时。

8.2 鞅、构造与不等式

定义 8.1 (鞅). 若 $X_n \in L^1$ 、适应且 $\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n$ a.s., 则称 $(X_n, \mathcal{F}_n)$ 为鞅。把等号换为 $\leq$ 或 $\geq$ 得上鞅或下鞅。

基本构造包括： $M_n = \mathbb{E}[X | \mathcal{F}_n]$ ；独立零均值和 $S_n = \sum_{i \leq n} X_i$ ；若 $\mathbb{E}e^{\theta X_1} < \infty$ ，则

$$M_n(\theta) = \exp\{\theta S_n - n \log \mathbb{E}e^{\theta X_1}\}$$

是鞅；平方补偿 $S_n^2 - \sum_{i \leq n} \mathbb{E}(X_i^2 | \mathcal{F}_{i-1})$ 是鞅。

定理 8.1 (Doob 最大不等式). 若 (X_k) 是非负下鞅，则对 $\lambda > 0$,

$$\lambda \mathbb{P}\left(\max_{k \leq n} X_k \geq \lambda\right) \leq \mathbb{E}[X_n \mathbf{1}_{\{\max_{k \leq n} X_k \geq \lambda\}}] \leq \mathbb{E}X_n.$$

若 $p > 1$ ，则存在共轭指数 $q = p/(p-1)$ 使 $\|\max_{k \leq n} |M_k|\|_p \leq q \|M_n\|_p$ 。

证明. 令

$$A_k = \{X_1 < \lambda, \dots, X_{k-1} < \lambda, X_k \geq \lambda\}.$$

A_k 不交且 $A_k \in \mathcal{F}_k$ 。下鞅性质给

$$\mathbb{E}[X_n \mathbf{1}_{A_k}] \geq \mathbb{E}[X_k \mathbf{1}_{A_k}] \geq \lambda \mathbb{P}(A_k).$$

求和得第一式。 L^p 版对尾积分使用第一式，再以 Hölder 控制

$$\mathbb{E} \left[|M_n| \max_{k \leq n} |M_k|^{p-1} \right].$$

□

定理 8.2 (鞅收敛). 若上鞅 (X_n) 满足 $\sup_n \mathbb{E} X_n^- < \infty$, 则 X_n a.s. 收敛到有限随机变量。特别地, 非负鞅 a.s. 收敛。若鞅一致可积, 则存在 $X_\infty \in L^1$ 使 $X_n = \mathbb{E}[X_\infty | \mathcal{F}_n]$ 且 $X_n \rightarrow X_\infty$ 于 L^1 。

非负鞅的 a.s. 极限可能损失期望, 因此不能自动推出 L^1 收敛; 指数鞅是常见反例来源。

8.3 可选停止与 Wald 等式

定理 8.3 (可选停止: 常用安全版本). 若 (M_n) 是鞅, $\sigma \leq \tau$ 是停时, 则在以下任一条件下 $\mathbb{E} M_\tau = \mathbb{E} M_\sigma$:

- (1) τ 有界;
- (2) $(M_{n \wedge \tau})_n$ 一致可积;
- (3) $\mathbb{E} \tau < \infty$ 且增量满足 $|M_n - M_{n-1}| \leq C$ a.s.;
- (4) M_n 一致有界且 $\tau < \infty$ a.s.

证明. 有界停时下, 写 $M_\tau = M_0 + \sum_{k=1}^N (M_k - M_{k-1}) \mathbf{1}_{\{\tau \geq k\}}$, 其中指标对 $\mathcal{F}_{k-1}$ 可测, 故每项期望为零。一般版本先对 $\tau \wedge n$ 使用有界情形, 再分别用一致可积或支配收敛取极限。□

定理 8.4 (Wald 第一等式). 若 X_i i.i.d., $\mathbb{E}|X_1| < \infty$, τ 是关于自然滤过的停时、 $\mathbb{E} \tau < \infty$, 则

$$\mathbb{E} \sum_{i=1}^{\tau} X_i = \mathbb{E} \tau \mathbb{E} X_1.$$

证明. 写随机和为 $\sum_{i \geq 1} X_i \mathbf{1}_{\{\tau \geq i\}}$ 。绝对可积性由 $\sum_i \mathbb{E}|X_i| \mathbf{1}_{\{\tau \geq i\}} = \mathbb{E}|X_1| \mathbb{E} \tau < \infty$ 保证, 可交换求和与期望。因 $\{\tau \geq i\} \in \mathcal{F}_{i-1}$ 且与 X_i 独立, 每项期望为 $\mathbb{E} X_1 \mathbb{P}(\tau \geq i)$; 求和得到结论。□

例题 8.1 (赌徒破产). 对简单随机游动 S_n , 令 $\tau = \inf\{n : S_n = -a \text{ 或 } b\}$ 。公平情形 $S_{n \wedge \tau}$ 与 $S_{n \wedge \tau}^2 - (n \wedge \tau)$ 是鞅。因停下值有界, $\mathbb{P}(S_\tau = b) = a/(a+b)$, 且 $\mathbb{E} \tau = ab$ 。有漂移时使用指数鞅 r^{S_n} , 其中 $pr + q/r = 1$, 再解首达概率。

答题方法

停时题先证明事件 $\{\tau \leq n\} \in \mathcal{F}_n$, 再检查停后过程是否可积。使用可选停止前必须逐条核验有界停时、一致可积或有界增量加可积停时等条件。

自测

1. 证明停后过程 $M_{n \wedge \tau}$ 仍为鞅。
2. 由条件 Jensen 证明凸函数作用于鞅得到次鞅。

3. 证明 Doob 的 L^1 最大不等式, 并说明事件分层的关键一步。
4. 给出简单对称随机游走无界首达时使朴素可选停止失败的例子。

本节结论

鞅把“给定过去的公平性”转化为条件期望恒等式; 停时不会自动保留期望, 所有结论都依赖可积性与极限交换条件。

8.4 鞅的三种构造

条件期望鞅。 若 $X \in L^1$,

$$M_n = \mathbb{E}[X \mid \mathcal{F}_n]$$

满足

$$\mathbb{E}[M_{n+1} \mid \mathcal{F}_n] = \mathbb{E}[X \mid \mathcal{F}_n] = M_n.$$

中心独立和。 若 X_n 独立、可积、均值零, 令 $\mathcal{F}_n = \sigma(X_1, \dots, X_n)$, 则

$$S_n = \sum_{k=1}^n X_k$$

为鞅。

似然比鞅。 若 P, Q 在 $\mathcal{F}_n$ 上有密度, 令

$$L_n = \frac{dQ|_{\mathcal{F}_n}}{dP|_{\mathcal{F}_n}},$$

则在 P 下 L_n 是非负鞅。顺序检验与 change of measure 都使用这一结构。

8.5 Doob 最大不等式

若 X_k 是非负次鞅, 令

$$A_k = \{X_0 < \lambda, \dots, X_{k-1} < \lambda, X_k \geq \lambda\}.$$

$A_k \in \mathcal{F}_k$ 且两两不交。次鞅性质给对 $k \leq n$

$$\mathbb{E}[X_n \mathbf{1}_{A_k}] \geq \mathbb{E}[X_k \mathbf{1}_{A_k}] \geq \lambda \mathbb{P}(A_k).$$

求和得

$$\lambda \mathbb{P}\left(\max_{k \leq n} X_k \geq \lambda\right) \leq \mathbb{E}\left[X_n \mathbf{1}_{\{\max_{k \leq n} X_k \geq \lambda\}}\right] \leq \mathbb{E}X_n.$$

对 $p > 1$, 结合尾积分与 Hölder 可得

$$\left\| \max_{k \leq n} |M_k| \right\|_p \leq \frac{p}{p-1} \|M_n\|_p.$$

8.6 可选停止的安全版本

鞅 M_n 与停时 τ 并不自动满足 $\mathbb{E}M_\tau = \mathbb{E}M_0$ 。常用充分条件包括:

1. τ 有界;
2. $M_{n \wedge \tau}$ 一致可积;
3. 鞅增量有界且 $\mathbb{E}\tau < \infty$ 。

有界停时情形, 把

$$M_\tau = M_0 + \sum_{k=1}^N (M_k - M_{k-1}) \mathbf{1}_{\{\tau \geq k\}}$$

展开, 其中 $\{\tau \geq k\} \in \mathcal{F}_{k-1}$, 故每项期望为零。

反例 8.1. 简单对称随机游走从零出发, 令 $\tau = \inf\{n : S_n = 1\}$ 。一维随机游走 a.s. 命中 1, 故 $S_\tau = 1$, 而 $\mathbb{E}S_0 = 0$ 。这里 τ 无界且停止鞅族不 UI, 不能直接用可选停止。

8.7 赌徒破产与指数鞅

简单随机游走 S_n , 步长 +1 概率 p 、-1 概率 $q = 1 - p$ 。若 $p \neq q$,

$$M_n = \left(\frac{q}{p}\right)^{S_n}$$

是鞅, 因为条件乘子期望为

$$p(q/p) + q(p/q) = p + q = 1.$$

从 $i \in \{0, \dots, N\}$ 出发, 停在 $\tau = \inf\{n : S_n \in \{0, N\}\}$ 。有界状态确保停止值可积, 得到

$$\mathbb{P}_i(S_\tau = N) = \frac{1 - (q/p)^i}{1 - (q/p)^N}.$$

$p = q = 1/2$ 时对 S_n 本身停止, 得到

$$\mathbb{P}_i(S_\tau = N) = \frac{i}{N}.$$

8.8 Wald 等式与条件

若 X_i i.i.d. 可积, τ 是关于自然滤过的停时, $\mathbb{E}\tau < \infty$, 并有足够可积性, 则

$$\mathbb{E} \sum_{k=1}^{\tau} X_k = \mathbb{E}X_1 \mathbb{E}\tau.$$

证明写成

$$\sum_{k=1}^{\tau} X_k = \sum_{k \geq 1} X_k \mathbf{1}_{\{\tau \geq k\}},$$

而 $\{\tau \geq k\} \in \mathcal{F}_{k-1}$ 与 X_k 独立, 所以

$$\mathbb{E}[X_k \mathbf{1}_{\{\tau \geq k\}}] = \mathbb{E}X_1 \mathbb{P}(\tau \geq k).$$

求和并用 $\mathbb{E}\tau = \sum_{k \geq 1} \mathbb{P}(\tau \geq k)$ 。交换求和与期望需可积性, 不能省略。

答题方法

停时题先核验 $\{\tau \leq n\} \in \mathcal{F}_n$, 再写停止过程 $M_{n \wedge \tau}$, 最后逐条检查可选停止条件。不要从“它是鞅”直接跳到“停时前后期望相同”。

自测

1. 构造条件期望鞅与似然比鞅。
2. 用首次越界事件证明 Doob 弱型最大不等式。
3. 给出三个保证可选停止合法的充分条件。
4. 推导有偏赌徒破产的命中概率。
5. Wald 等式中停时事件为什么与当前增量独立?

第九章 鞅收敛与一致可积

学习目标

本章解释鞅为何以及以何种方式收敛：从上穿不等式得到路径极限，用非负性、 L^p 有界或一致可积控制极限交换，并用反例区分 a.s.、 L^1 与期望守恒。

9.1 上穿不等式与几乎处处收敛

对次鞅 X_n 和 $a < b$ ，记 $U_N[a, b]$ 为路径在时刻 N 前从 $(-\infty, a]$ 上穿到 $[b, \infty)$ 的完成次数。通过“低于 a 时买入、达到 b 时卖出”的可预测策略 $H_n \in \{0, 1\}$ ，累积收益 $(H \cdot X)_N$ 满足

$$(b - a)U_N[a, b] \leq (X_N - a)^- - (X_0 - a)^- + (H \cdot X)_N.$$

次鞅变换的期望非负，故

$$(b - a)\mathbb{E}U_N[a, b] \leq \mathbb{E}(X_N - a)^- + \mathbb{E}(X_0 - a)^+.$$

若 $\sup_n \mathbb{E}X_n^- < \infty$ ，右端一致有界，令 $N \rightarrow \infty$ 得每个有理 $a < b$ 的无限上穿概率为零。若路径的 $\liminf$ 小于 $\limsup$ ，必存在有理 $a < b$ 被无限上穿，故 X_n 几乎处处收敛。

定理 9.1 (非负鞅收敛). 若 $M_n \geq 0$ 是鞅，则存在有限随机变量 M_∞ 使 $M_n \rightarrow M_\infty$ 几乎处处，且 $\mathbb{E}M_\infty \leq \mathbb{E}M_0$ 。

证明. 非负性给 $\sup_n \mathbb{E}M_n^- = 0$ ，上穿论证给有限极限；Fatou 引理给 $\mathbb{E}M_\infty \leq \liminf_n \mathbb{E}M_n = \mathbb{E}M_0$ 。一般不能得到 L^1 收敛，因为质量可能逃向稀有值。□

9.2 L^p 、一致可积与闭鞅

定理 9.2 (L^p 鞅收敛). 若 $p > 1$ 且 $\sup_n \mathbb{E}|M_n|^p < \infty$ ，则存在 $M_\infty \in L^p$ ，使 $M_n \rightarrow M_\infty$ 几乎处处且于 L^p 。

证明. Doob L^p 不等式控制尾部最大函数：

$$\mathbb{E} \max_{n \leq k \leq N} |M_k - M_n|^p \leq q^p \mathbb{E}|M_N - M_n|^p, \quad q = \frac{p}{p-1}.$$

L^p 有界鞅由反射性得到弱紧性，并可表示为 $M_n = \mathbb{E}[M_\infty | \mathcal{F}_n]$ ；条件期望逼近与 Doob 不等式给 L^p Cauchy，从而得到所述收敛。考试若不允引用反射性，可先用上穿得到 a.s. 极限，再由 L^p 有界推出一致可积，结合 Vitali 得 L^1 ，最后对 $|M_n - M_\infty|^p$ 用条件期望近似完成。□

对 L^1 鞅, 以下三者等价: $\{M_n\}$ 一致可积; 存在 $M_\infty \in L^1$ 使 $M_n \rightarrow M_\infty$ 于 L^1 ; 存在 $Y \in L^1$ 使 $M_n = \mathbb{E}[Y | \mathcal{F}_n]$ 。这说明 UI 正是把 a.s. 收敛升级为守恒期望的条件。

反例 9.1 (非负鞅不必 L^1 收敛). 在二叉树上令 $M_n = 2^n \mathbf{1}_{\{\text{前}n\text{次均为正面}\}}$ 。则 $\mathbb{E}M_n = 1$, 且 M_n 是非负鞅; 但几乎处处最终出现反面, 所以 $M_n \rightarrow 0$ a.s., 而 $\mathbb{E}|M_n - 0| = 1$ 。该族不一致可积。

答题方法

判断鞅收敛先问负部或 L^p 是否一致有界, 再问是否需要 L^1 收敛。只需 a.s. 极限可用上穿; 要交换期望必须补 UI。

自测

1. 证明非负鞅的极限期望可以严格小于初始期望。
2. 证明 L^p 有界 ($p > 1$) 推出一致可积。
3. 说明闭鞅表示 $M_n = \mathbb{E}[Y | \mathcal{F}_n]$ 如何给出 L^1 收敛。
4. 在稀有倍增鞅中直接验证鞅性并证明不一致可积。

本节结论

上穿控制振荡, 矩或 UI 控制质量逃逸。二者合起来才给出可用于可选停止和极限交换的强收敛。

9.3 上穿不等式的买低卖高证明

对次鞅 $X_0, \dots, X_n$ 和 $a < b$, 令 $U_n[a, b]$ 为路径完成的上穿次数。构造可预测持仓 $H_k \in \{0, 1\}$: 当过程到达 $(-\infty, a]$ 后买入, 直到到达 $[b, \infty)$ 卖出。交易收益

$$(H \cdot X)_n = \sum_{k=1}^n H_k (X_k - X_{k-1})$$

至少为

$$(b - a)U_n[a, b] - (X_n - a)^-.$$

因为 H_k 是 $\mathcal{F}_{k-1}$ -可测且非负,

$$\mathbb{E}[H_k(X_k - X_{k-1})] \geq 0.$$

故

$$(b - a)\mathbb{E}U_n[a, b] \leq \mathbb{E}[(X_n - a)^-] + \mathbb{E}[(X_0 - a)^+]$$

(不同起始约定可得到等价边界)。若次鞅的负部期望一致有界, 则任意有理 $a < b$ 的上穿总数有限 a.s., 从而路径极限存在。

9.4 非负超鞅收敛

若 $X_n \geq 0$ 为超鞅, 则

$$\mathbb{E}X_n \leq \mathbb{E}X_0.$$

$-X_n$ 是次鞅且其负部为 X_n , 一致 L^1 有界。上穿论证给

$$X_n \rightarrow X_\infty \quad a.s.$$

Fatou 给

$$\mathbb{E}X_\infty \leq \liminf_n \mathbb{E}X_n \leq \mathbb{E}X_0.$$

但不一定 L^1 收敛, 期望可能在极限处下降; 要升级到 L^1 , 需要一致可积。

反例 9.2. 令 $M_n = 2^n \mathbf{1}_{A_n}$, 其中 $A_n \downarrow A_\infty$ 且 $\mathbb{P}(A_n) = 2^{-n}$, 并按适当二叉滤过构造使 M_n 为非负鞅。则 $\mathbb{E}M_n = 1$, 但 $M_n \rightarrow 0$ a.s., 所以不 L^1 收敛, 且族不 UI。

9.5 闭鞅与一致可积

若存在 $X \in L^1$ 使

$$M_n = \mathbb{E}[X | \mathcal{F}_n],$$

则称 M_n 被 X 闭合。条件期望族 $\{\mathbb{E}[X | \mathcal{G}]\}$ 一致可积, 因此 M_n UI。若 $\mathcal{F}_\infty = \sigma(\cup_n \mathcal{F}_n)$, 则

$$M_n \rightarrow \mathbb{E}[X | \mathcal{F}_\infty] \quad a.s. \text{ 且 } L^1.$$

反过来, 若鞅 M_n UI, 则 a.s. 与 L^1 收敛到某 $M_\infty \in L^1$, 且

$$M_n = \mathbb{E}[M_\infty | \mathcal{F}_n].$$

所以离散时间中 “UI 鞅” “ L^1 收敛鞅” “被可积终值闭合” 形成等价链。

9.6 L^p 有界鞅

若 $p > 1$ 且

$$\sup_n \mathbb{E}|M_n|^p < \infty,$$

则该族 UI, 并由 Doob L^p 不等式

$$\mathbb{E} \left[\sup_{k \leq n} |M_k|^p \right] \leq \left(\frac{p}{p-1} \right)^p \mathbb{E}|M_n|^p.$$

让 $n \rightarrow \infty$, 获得可积上界与 a.s. 收敛, 再由 UI 得

$$M_n \rightarrow M_\infty \quad \text{in } L^1.$$

事实上在适当论证下还可得 L^p 收敛。

9.7 Doob 分解

任何可积适应过程 X_n 可唯一写成

$$X_n = M_n + A_n,$$

其中 M_n 是鞅, $A_0 = 0$ 且 A_n 可预测:

$$A_n = \sum_{k=1}^n \mathbb{E}[X_k - X_{k-1} \mid \mathcal{F}_{k-1}],$$

$$M_n = X_n - A_n.$$

若 X_n 是次鞅, 则 A_n 递增。它把随机波动与可预测漂移分开, 是连续时间 Doob-Meyer 分解的离散原型。

答题方法

鞅收敛题依次判断: a.s. 收敛需要上穿/非负/负部有界; L^1 收敛需要 UI; L^p 有界 $p > 1$ 自动给 UI。不要把“期望有界”当成 UI。

自测

1. 上穿交易策略为何必须可预测?
2. 非负鞅 a.s. 收敛后, 为什么期望可能下降?
3. 证明 UI 鞅由极限变量闭合。
4. L^p 有界为何在 $p > 1$ 时推出 UI?
5. 写出任意适应过程的 Doob 分解。

第十章 Markov 链、Poisson 过程与首达递推

学习目标

本章用转移核统一有限与可数状态链，学习类结构、常返/暂留、平稳分布和遍历；同时保留 Poisson 过程的独立平稳增量，训练首达概率和期望的差分方程。

10.1 齐次 Markov 性、条件独立与半群

定义 10.1 (齐次离散时间 Markov 链). 设 S 为有限或可数状态空间, $\mathcal{F}_n = \sigma(X_0, \dots, X_n)$ 。若存在随机矩阵 $P = (p_{ij})_{i,j \in S}$, 使对每个 $n \geq 0$ 、 $j \in S$ 都有

$$\mathbb{P}(X_{n+1} = j \mid \mathcal{F}_n) = p_{X_n j} \quad \text{a.s.},$$

则称 (X_n) 为转移矩阵 P 的齐次 Markov 链。齐次性是右端不依赖 n ；Markov 性是给定现在 X_n 后，一步未来不再依赖更早历史。

定理 10.1 (给定现在后，过去与未来条件独立). 对任意 $r, s \geq 0$ 、过去事件 $A \in \sigma(X_0, \dots, X_{n-1})$ 与未来柱事件

$$B = \{X_{n+1} = j_1, \dots, X_{n+s} = j_s\},$$

有

$$\mathbb{P}(A \cap B \mid X_n) = \mathbb{P}(A \mid X_n) \mathbb{P}(B \mid X_n) \quad \text{a.s.}$$

因此 $\sigma(X_0, \dots, X_{n-1})$ 与 $\sigma(X_{n+1}, X_{n+2}, \dots)$ 在给定 X_n 后条件独立。

完整证明. 先固定 $X_n = i$ 。反复使用定义中的一步条件概率与塔式法则，得到

$$\mathbb{P}(B \mid \mathcal{F}_n) = p_{ij_1} p_{j_1 j_2} \cdots p_{j_{s-1} j_s},$$

右端只依赖 $X_n = i$ ，记为 $g(i)$ 。于是

$$\begin{aligned} \mathbb{E}[\mathbf{1}_A \mathbf{1}_B \mid X_n] &= \mathbb{E}[\mathbf{1}_A \mathbb{E}(\mathbf{1}_B \mid \mathcal{F}_n) \mid X_n] \\ &= \mathbb{E}[\mathbf{1}_A g(X_n) \mid X_n] = \mathbb{E}[\mathbf{1}_A \mid X_n] g(X_n). \end{aligned}$$

而 $g(X_n) = \mathbb{E}[\mathbf{1}_B \mid X_n]$ ，故柱事件结论成立。有限并、补与单调极限生成全部未来 σ -代数，单调类定理把结论扩张到一般未来事件。 $\square$

定理 10.2 (Chapman–Kolmogorov 公式). 记 $p_{ij}^{(n)} = \mathbb{P}_i(X_n = j)$. 对 $m, n \geq 0$,

$$p_{ij}^{(m+n)} = \sum_{k \in S} p_{ik}^{(m)} p_{kj}^{(n)}, \quad P^{m+n} = P^m P^n.$$

完整证明. 按中间时刻 m 的状态分割, 并用 定理 10.1:

$$\mathbb{P}_i(X_{m+n} = j) = \sum_k \mathbb{P}_i(X_m = k) \mathbb{P}_i(X_{m+n} = j \mid X_m = k) = \sum_k p_{ik}^{(m)} p_{kj}^{(n)}.$$

可数状态时各项非负, 故求和次序由 Tonelli 定理保证合法. $\square$

10.2 停时、强 Markov 性与首达边值问题

定理 10.3 (离散时间强 Markov 性). 设 T 是 $(\mathcal{F}_n)$ -停时. 对任意有界可测 $\varphi : S^m \rightarrow \mathbb{R}$, 在 $\{T < \infty\}$ 上

$$\mathbb{E}[\varphi(X_{T+1}, \dots, X_{T+m}) \mid \mathcal{F}_T] = G_\varphi(X_T),$$

其中 $G_\varphi(i) = \mathbb{E}_i[\varphi(X_1, \dots, X_m)]$.

完整证明. 先验证任意 $A \in \mathcal{F}_T$:

$$\begin{aligned} \mathbb{E}[\mathbf{1}_{A \cap \{T < \infty\}} \varphi(X_{T+1}, \dots, X_{T+m})] &= \sum_{n \geq 0} \mathbb{E}[\mathbf{1}_{A \cap \{T=n\}} \varphi(X_{n+1}, \dots, X_{n+m})] \\ &= \sum_{n \geq 0} \mathbb{E}[\mathbf{1}_{A \cap \{T=n\}} G_\varphi(X_n)] \\ &= \mathbb{E}[\mathbf{1}_{A \cap \{T < \infty\}} G_\varphi(X_T)]. \end{aligned}$$

第二行用的是确定时刻的 Markov 性, 且 $A \cap \{T = n\} \in \mathcal{F}_n$. 这正是条件期望的定义. 一般非负 φ 由单调收敛, 有界有符号情形由正负部分得到. $\square$

令 $T_A = \inf\{n \geq 0 : X_n \in A\}$. 对不交集合 A, B , 命中概率

$$h_i = \mathbb{P}_i(T_A < T_B)$$

满足

$$h_i = \sum_j p_{ij} h_j \quad (i \notin A \cup B), \quad h|_A = 1, \quad h|_B = 0. \quad (10.1)$$

若 $m_i = \mathbb{E}_i T_A < \infty$, 则

$$m_i = 1 + \sum_j p_{ij} m_j \quad (i \notin A), \quad m|_A = 0. \quad (10.2)$$

两式均由第一步分解得到. 反过来, 若状态空间无限, 方程可能有多个非负解; 真实命中概率和期望分别是相应边界问题的最小非负解. 证明方法是先截断 $T_A \wedge n$, 迭代方程, 再令 $n \uparrow \infty$ 并用单调收敛.

例题 10.1 (赌徒破产: 概率与期望停时的完整递推). 在 $\{0, 1, \dots, N\}$ 上, 内部状态以概率 p 向右、 $q = 1 - p$ 向左, $0, N$ 吸收。令 $\tau = \inf\{n : X_n \in \{0, N\}\}$ 。命中 N 的概率 $h_i = \mathbb{P}_i(X_\tau = N)$ 满足

$$h_i = ph_{i+1} + qh_{i-1}, \quad h_0 = 0, \quad h_N = 1.$$

当 $p \neq q$ 时, 特征根为 $1, q/p$, 故

$$h_i = \frac{1 - (q/p)^i}{1 - (q/p)^N};$$

当 $p = q = 1/2$ 时, 二阶差分为零, 故 $h_i = i/N$ 。

期望停时 $m_i = \mathbb{E}_i \tau$ 满足

$$m_i = 1 + pm_{i+1} + qm_{i-1}, \quad m_0 = m_N = 0.$$

当 $p \neq q$ 时, 常数特解为 $i/(q-p)$, 代入边界得

$$m_i = \frac{i}{q-p} - \frac{N}{q-p} \frac{1 - (q/p)^i}{1 - (q/p)^N}.$$

当 $p = q = 1/2$ 时, $\Delta^2 m_i = -2$, 故

$$m_i = i(N-i).$$

有限状态吸收链从任一内部点都有正概率在至多 N 步内吸收, 因此 $\mathbb{P}_i(\tau > kN) \leq (1-c)^k$ 对某个 $c > 0$ 成立, 保证上述期望有限, 不是只求到了形式解。

例题 10.2 (两状态链的平稳分布与周期). 令

$$P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}, \quad a, b \in (0, 1].$$

解 $\pi P = \pi$ 得

$$\pi = \left(\frac{b}{a+b}, \frac{a}{a+b} \right).$$

若 $a = b = 1$, 链在两个状态间交替, 周期为 2, 故 $P^n(0, \cdot)$ 不收敛, 尽管平稳分布唯一。只要 $a < 1$ 或 $b < 1$, 存在自环; 不可约性再推出共同周期为 1。

答题方法

首达题先写清停时与边界, 再做第一步分解。概率量满足调和方程, 期望时间多出常数 1。无限状态空间还要说明所取的是最小非负解; 有限吸收问题则用统一的分块吸收概率证明期望有限。

10.3 常返、暂留与沟通类

若存在 $n \geq 0$ 使 $p_{ij}^{(n)} > 0$, 记 $i \rightarrow j$; 若 $i \rightarrow j$ 且 $j \rightarrow i$, 记 $i \leftrightarrow j$ 。互达是等价关系, 其等价类称沟通类。令

$$T_i^+ = \inf\{n \geq 1 : X_n = i\}, \quad f_{ii}^{(n)} = \mathbb{P}_i(T_i^+ = n).$$

若 $\mathbb{P}_i(T_i^+ < \infty) = 1$, 称 i 常返; 否则称暂留。常返状态在 $\mathbb{E}_i T_i^+ < \infty$ 时称正常返, 在期望为无穷时称零常返。

定理 10.4 (Green 级数的常返判据). 对任意状态 i ,

$$i \text{ 常返} \iff \sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty.$$

若 i 暂留, 则从 i 出发访问 i 的总次数均值有限。

完整证明. 记 $u_n = p_{ii}^{(n)}$, 其中 $u_0 = 1$, 并令 $f_n = f_{ii}^{(n)}$ 。按第一次正回返发生在时刻 k 分解, 强 Markov 性给出更新方程

$$u_n = \sum_{k=1}^n f_k u_{n-k} \quad (n \geq 1).$$

对 $0 \leq s < 1$ 定义

$$U(s) = \sum_{n \geq 0} u_n s^n, \quad F(s) = \sum_{n \geq 1} f_n s^n.$$

非负项允许交换求和, 故 $U(s) = 1 + F(s)U(s)$, 即

$$U(s) = \frac{1}{1 - F(s)}.$$

令 $s \uparrow 1$ 。由单调收敛, $F(s) \uparrow \mathbb{P}_i(T_i^+ < \infty)$ 。因此回返概率为 1 当且仅当 $U(s) \uparrow \infty$, 等价于 $\sum_n u_n = \infty$ 。访问次数 $V_i = \sum_{n \geq 0} \mathbf{1}_{\{X_n = i\}}$ 满足 $\mathbb{E}_i V_i = \sum_n u_n$, 故暂留时其均值有限。 $\square$

定理 10.5 (常返性与周期是沟通类性质). 若 $i \leftrightarrow j$, 则 i, j 同为常返或同为暂留, 且周期

$$d(i) = \gcd\{n \geq 1 : p_{ii}^{(n)} > 0\}$$

满足 $d(i) = d(j)$ 。

完整证明. 取 r, s 使 $p_{ji}^{(r)} > 0, p_{ij}^{(s)} > 0$ 。由 Chapman-Kolmogorov,

$$p_{jj}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ij}^{(s)}.$$

若 i 常返, 则右侧对 n 求和发散, 故由 **定理 10.4**, j 常返; 交换 i, j 得反向结论。对周期, $r+s$ 是 j 的一个回返长度; 若 n 是 i 的回返长度, 则 $r+n+s$ 也是 j 的回返长度。因此 $d(j)$ 同时整除 $r+s$ 与 $r+n+s$, 从而整除每个这样的 n , 故 $d(j) \mid d(i)$ 。交换两状态得 $d(i) \mid d(j)$, 于是二者相等。 $\square$

定理 10.6 (有限不可约链正常返). 有限不可约 Markov 链的每个状态都正常返。

完整证明. 固定目标状态 j 。有限性与不可约性保证：对每个 i 可选一条至多 $|S| - 1$ 步到达 j 的正概率路径。令 L 为这些路径长度的最大值， $\varepsilon > 0$ 为其路径概率的最小值。由 Markov 性，从任意状态出发在接下来 L 步内命中 j 的条件概率至少为 ε 。故

$$\mathbb{P}_i(T_j > kL) \leq (1 - \varepsilon)^k.$$

尾和公式给

$$\mathbb{E}_i T_j = \sum_{n \geq 0} \mathbb{P}_i(T_j > n) \leq L \sum_{k \geq 0} (1 - \varepsilon)^k = \frac{L}{\varepsilon} < \infty.$$

从 j 出发先走一步再使用同一界，得到 $\mathbb{E}_j T_j^+ < \infty$ 。 $\square$

反例 10.1 (常返不等于正常返，也不自动给平稳概率)。整数格上的简单对称随机游走满足

$$p_{00}^{(2n)} = \binom{2n}{n} 2^{-2n} \sim (\pi n)^{-1/2}.$$

故 Green 级数发散，状态 0 常返；但其平均回返时间无穷，是零常返。更具体地，首次回返分布满足

$$\mathbb{P}_0(T_0^+ = 2n) = \frac{1}{2n-1} \binom{2n}{n} 2^{-2n} \asymp n^{-3/2},$$

故 $\mathbb{E}_0 T_0^+ = \sum_{n \geq 1} 2n \mathbb{P}_0(T_0^+ = 2n) = \infty$ 。计数测度满足不变性，却不能归一化为 $\mathbb{Z}$ 上的概率分布。这说明“常返”“正常返”和“存在平稳概率分布”不可混用。

10.4 平稳分布的构造、唯一性与详细平衡

定理 10.7 (回返周期构造平稳分布)。设链不可约且状态 i 正常返。定义一个回返周期内的访问量

$$\nu_j = \mathbb{E}_i \left[\sum_{n=0}^{T_i^+-1} \mathbf{1}_{\{X_n=j\}} \right].$$

则 $\nu P = \nu$ 、 $\sum_j \nu_j = \mathbb{E}_i T_i^+$ ，从而

$$\pi_j = \frac{\nu_j}{\mathbb{E}_i T_i^+}$$

是平稳概率分布；它与基准状态 i 的选择无关，并且是唯一平稳概率分布。特别地，

$$\pi_i = \frac{1}{\mathbb{E}_i T_i^+}.$$

完整证明. 非负项允许把期望与求和交换。对每个 k ,

$$\begin{aligned}\sum_j \nu_j p_{jk} &= \mathbb{E}_i \sum_{n=0}^{T_i^+-1} p_{X_n k} \\ &= \mathbb{E}_i \sum_{n=0}^{T_i^+-1} \mathbb{E}_i[\mathbf{1}_{\{X_{n+1}=k\}} \mid \mathcal{F}_n] \\ &= \mathbb{E}_i \sum_{n=1}^{T_i^+} \mathbf{1}_{\{X_n=k\}}.\end{aligned}$$

因为 $X_0 = X_{T_i^+} = i$, 最后一个随机和与 $\sum_{n=0}^{T_i^+-1} \mathbf{1}_{\{X_n=k\}}$ 逐路径相等, 故 $\nu P = \nu$ 。再对 j 求和得 $\sum_j \nu_j = \mathbb{E}_i T_i^+$, 所以归一化后得到平稳分布。一个周期内只有时刻 0 访问基准状态 i , 故 $\nu_i = 1$, 推出 $\pi_i = 1/\mathbb{E}_i T_i^+$ 。

为说明唯一性, 从任意状态 j 的相继回返时刻切分轨道。强 Markov 性使完整回返周期独立同分布; 更新报酬定理于是给出每个状态 k 的长期访问频率

$$\frac{\mathbb{E}_j \sum_{n=0}^{T_j^+-1} \mathbf{1}_{\{X_n=k\}}}{\mathbb{E}_j T_j^+}.$$

若初始分布本身平稳, 则该频率的期望在每个时刻都等于该平稳分布的第 k 个分量。上式只由 P 决定, 故任意两个平稳概率分布逐坐标相等。这也说明从不同基准状态构造得到同一个 π 。 $\square$

命题 10.8 (详细平衡推出平稳性). 若概率分布 π 满足

$$\pi_i p_{ij} = \pi_j p_{ji} \quad (i, j \in S),$$

则 $\pi P = \pi$ 。

证明. 固定 j 并对 i 求和:

$$(\pi P)_j = \sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j.$$

这里最后一个求和是转移矩阵第 j 行之和。详细平衡是充分条件而非必要条件。 $\square$

例题 10.3 (有限 birth-death 链的手算). 在 $\{0, 1, 2\}$ 上令

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/4 & 1/2 & 1/4 \\ 0 & 1/2 & 1/2 \end{pmatrix}.$$

详细平衡递推给 $\pi_1/\pi_0 = p_{01}/p_{10} = 2$ 与 $\pi_2/\pi_1 = p_{12}/p_{21} = 1/2$, 归一化得

$$\pi = (1/4, 1/2, 1/4).$$

链不可约且有自环, 故非周期; 因此任意初值下的时刻分布都收敛到 π 。

10.5 周期、时刻分布与时间平均

反例 10.2 (周期链的时刻分布不收敛). 在三个状态上令 $0 \rightarrow 1 \rightarrow 2 \rightarrow 0$ 为确定转移。链有限不可约, 唯一平稳分布为 $(1/3, 1/3, 1/3)$, 共同周期为 3。从 0 出发时, X_n 的分布在三个点质量之间循环, 因而没有时刻分布极限; 但每个状态在前 n 步中的比例都趋于 $1/3$ 。

定理 10.9 (分布收敛与遍历结论). 设可数状态链不可约且正常返, 平稳分布为 π 。

1. 若链非周期, 则对任意 i, j ,

$$p_{ij}^{(n)} \rightarrow \pi_j.$$

2. 对任意满足 $\sum_j \pi_j |f(j)| < \infty$ 的 f , 无论周期如何,

$$\frac{1}{n} \sum_{k=0}^{n-1} f(X_k) \xrightarrow{\text{a.s.}} \sum_j \pi_j f(j).$$

证明思路

第一条的完整证明通常使用更新定理或耦合。以状态 i 的回返时刻为更新点, 非周期性保证回返时间分布的跨度为 1, 关键更新定理给 $p_{ij}^{(n)} \rightarrow \nu_j / \mathbb{E}_i T_i^+ = \pi_j$ 。若周期为 $d > 1$, 更新质量只落在固定剩余类, 故一般不能逐步收敛。第二条把轨道分成回返周期, 对每周期报酬 $\sum f(X_k)$ 与周期长度分别用强大数律, 再取比值; 截断处理 $\pi(|f|) < \infty$ 的一般情形。这是考试中应写出的证明路线, 深层的关键更新定理本身不属于本考纲的独立证明要求。

10.6 Poisson 过程: 指数等待时间、条件到达与分裂

令 $W_1, W_2, \dots$ 独立同分布为 $\text{Exp}(\lambda)$, $S_n = \sum_{k=1}^n W_k$, 并定义

$$N(t) = \max\{n : S_n \leq t\}.$$

由于 S_n 具有 Erlang 密度,

$$\mathbb{P}(N(t) = n) = \mathbb{P}(S_n \leq t < S_{n+1}) = e^{-\lambda t} \frac{(\lambda t)^n}{n!}.$$

更一般地, 对互不相交区间 $I_1, \dots, I_r$, 直接对指数联合密度在相应到达单纯形上积分, 得到各区间计数相互独立, 且

$$N(I_\ell) \sim \text{Poisson}(\lambda |I_\ell|).$$

因此指数到达间隔构造确实产生独立平稳增量的 Poisson 过程; 反过来, 小时间条件与独立平稳增量给出指数等待时间的无记忆性。

定理 10.10 (给定总到达数的次序统计量). 给定 $N(t) = n$, 向量 $(S_1, \dots, S_n)$ 的条件密度为

$$\frac{n!}{t^n} \mathbf{1}_{\{0 < s_1 < \dots < s_n < t\}},$$

即与 n 个独立 $\text{Unif}(0, t)$ 的次序统计量同分布。

完整证明. 在单纯形 $0 < s_1 < \cdots < s_n < t$ 上, 恰有这 n 次到达且 t 前无下一次到达的联合密度是

$$\lambda^n e^{-\lambda s_n} \mathbb{P}(W_{n+1} > t - s_n) = \lambda^n e^{-\lambda t}.$$

除以 $\mathbb{P}(N(t) = n) = e^{-\lambda t} (\lambda t)^n / n!$, 得到 $n! / t^n$ 。这正是均匀样本次序统计量的联合密度。 $\square$

命题 10.11 (叠加、稀疏化与条件分裂). 独立 Poisson 过程之和仍为 Poisson, 速率相加。若每次到达独立以概率 p 标为第一类, 则两类计数过程独立, 速率分别为 $p\lambda$ 与 $(1-p)\lambda$ 。给定总数 $N(t) = n$, 各类计数服从多项分布。

证明. 对固定区间长度 u , 设两类计数为 K, L 。条件于 $N(u)$ 后取概率母函数:

$$\begin{aligned} \mathbb{E}[z^K w^L] &= \mathbb{E}[(pz + (1-p)w)^{N(u)}] \\ &= \exp\{\lambda u[p(z-1) + (1-p)(w-1)]\} \\ &= \exp\{p\lambda u(z-1)\} \exp\{(1-p)\lambda u(w-1)\}. \end{aligned}$$

最后一行分解为两个 Poisson 母函数, 证明独立性与各自速率。对互不相交时间区间重复同一计算, 得到过程层面的独立增量。叠加结论由独立 Poisson 母函数相乘得到; 给定总数时, 标记独立性直接给出多项分布。 $\square$

命题 10.12 (补偿 Poisson 鞅). 对自然滤过 $\mathcal{F}_t = \sigma(N(s) : s \leq t)$,

$$M_t = N(t) - \lambda t$$

是鞅。

证明. 若 $s < t$, 独立增量给

$$\mathbb{E}[N(t) - N(s) \mid \mathcal{F}_s] = \lambda(t-s).$$

因此

$$\mathbb{E}[M_t \mid \mathcal{F}_s] = N(s) + \lambda(t-s) - \lambda t = M_s.$$

且 $\mathbb{E}|M_t| < \infty$, 故鞅定义的可积性也满足。 $\square$

10.7 连续时间 Markov 链与生成元入口

有限状态齐次连续时间 Markov 链的转移半群 $P(t) = (p_{ij}(t))$ 满足 $P(t+s) = P(t)P(s)$ 。若小时间展开为

$$P(h) = I + hQ + o(h), \quad h \downarrow 0,$$

则 $Q = (q_{ij})$ 为生成元。概率非负与行和为一分别给出

$$q_{ij} \geq 0 \ (i \neq j), \quad q_{ii} = -\sum_{j \neq i} q_{ij}.$$

从状态 i 出发的停留时间参数为 $-q_{ii}$, 离开时跳到 $j \neq i$ 的条件概率为 $q_{ij}/(-q_{ii})$ 。

命题 10.13 (Kolmogorov 前后向方程). 有限状态且生成元有限时,

$$P'(t) = P(t)Q = QP(t), \quad P(0) = I,$$

因而 $P(t) = e^{tQ}$. 平稳分布满足 $\pi^\top Q = 0$.

证明. 由半群性质,

$$\frac{P(t+h) - P(t)}{h} = P(t) \frac{P(h) - I}{h} \rightarrow P(t)Q.$$

也可写 $P(t+h) = P(h)P(t)$, 得到左乘形式 $QP(t)$. 矩阵常微分方程的唯一解为 e^{tQ} . 若 $\pi^\top P(t) = \pi^\top$ 对所有 t 成立, 对 $t=0$ 求导得 $\pi^\top Q = 0$; 反之 $\pi^\top e^{tQ} = \pi^\top$. $\square$

易错点

四个概念必须分开: 常返只问是否最终返回; 正常返再要求平均回返时间有限; 平稳分布是一步演化不变; 从任意初值的时刻分布收敛还需要非周期等条件. 详细平衡只是一种便捷的充分条件, 也不是所有平稳链都可逆。

答题方法

Markov 题依次检查状态空间、沟通类、常返类型与周期; 平稳分布先解 $\pi P = \pi$, 再判断能否用详细平衡. 首达概率写调和方程, 首达期望写带常数项的 Poisson 方程. Poisson 题在“独立增量、指数等待、条件次序统计量、标记分裂”四个入口间切换. 连续时间链先从生成元读出离开率和跳转概率。

自测

1. 用更新方程证明 $\sum_n p_{ii}^{(n)} = \infty$ 等价于常返。
2. 证明互达状态周期相同, 并给出一个周期为 3 的有限链。
3. 从回返周期访问量构造平稳分布, 解释归一化常数为何是平均周期长度。
4. 分别求赌徒破产的命中概率与期望吸收时间。
5. 推导给定 $N(t) = n$ 后到达时刻的联合密度。
6. 用联合概率母函数证明 Poisson 稀疏化后的两类过程独立。
7. 从半群性质推出 Kolmogorov 前后向方程。

本节结论

离散链的主线是“Markov 性 $\rightarrow$ 强 Markov 性 $\rightarrow$ 首达递推与回返周期 $\rightarrow$ 平稳和遍历”; Poisson 过程把指数等待、独立增量、条件次序统计量与补偿鞅连成同一结构; 连续时间链则由生成元编码瞬时跳转率。

第十一章 Brownian motion、反射原理与指数鞅

学习目标

本章从 Gaussian 独立增量和连续路径定义 Brownian motion, 推导缩放、反射原理、最大值与首达分布, 建立指数鞅、Brownian bridge 和协方差核计算。

11.1 Brownian motion、反射原理与二次变差

定义 11.1 (Brownian motion). $B_0 = 0$; 路径连续; 独立平稳增量; $B_t - B_s \sim N(0, t - s)$ 。因此它是均值零、协方差 $\mathbb{E}B_s B_t = \min(s, t)$ 的高斯过程。

尺度性: $(c^{-1/2}B_{ct})_{t \geq 0} \stackrel{d}{=} (B_t)$; 强 Markov 性: 对 a.s. 有限停时 T , $(B_{T+t} - B_T)$ 是独立于 $\mathcal{F}_T$ 的 Brownian motion。指数过程 $\exp\{\theta B_t - \theta^2 t/2\}$ 是鞅。

定理 11.1 (反射原理). 令 $M_t = \max_{s \leq t} B_s$ 。对 $a > 0$, $x < a$,

$$\mathbb{P}(M_t \geq a, B_t \leq x) = \mathbb{P}(B_t \geq 2a - x).$$

特别地 $\mathbb{P}(M_t \geq a) = 2\mathbb{P}(B_t \geq a)$, 故 $M_t \stackrel{d}{=} |B_t|$ 。

证明. 在首达 a 的时刻 T_a 后反射路径: $\tilde{B}_s = B_s$ 对 $s \leq T_a$, 而 $\tilde{B}_s = 2a - B_s$ 对 $s > T_a$ 。强 Markov 性与对称性说明该变换保持 Wiener 测度, 并把事件 $\{M_t \geq a, B_t \leq x\}$ 双射到 $\{B_t \geq 2a - x\}$ 。□

首达时 Laplace 变换由指数鞅得到

$$\mathbb{E}e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}, \quad a, \lambda > 0,$$

进而令 $\lambda \downarrow 0$ 得 $\mathbb{P}(T_a < \infty) = 1$ 。沿划分网格 Π_n 网长趋零, $\sum (B_{t_{k+1}} - B_{t_k})^2 \xrightarrow{\mathbb{P}} t$ (适当嵌套划分可 a.s.), 即 Brownian 二次变差为 t , 而总变差 a.s. 无穷。

若只需要确定性被积函数 $f \in L^2[0, T]$, Itô 积分由阶梯函数在 L^2 中完备化定义, 并满足

$$\mathbb{E} \int_0^T f(t) dB_t = 0, \quad \mathbb{E} \left(\int_0^T f(t) dB_t \right)^2 = \int_0^T f(t)^2 dt.$$

第二式即 Itô 等距。一般适应被积过程与 Itô 公式只作随机积分的延伸接口, 考试主要掌握定义、等距和 Brownian 鞅的标准计算。

答题方法

Brownian 题先用缩放和独立增量标准化；最大值/首达用反射，条件端点用 bridge，指数泛函先找鞅或做 Cameron–Martin 型倾斜。

自测

1. 由反射原理推导 $\mathbb{P}(\sup_{s \leq t} B_s \geq a) = 2\mathbb{P}(B_t \geq a)$ 。
2. 求首达时 τ_a 的分布函数和 Laplace 变换。
3. 验证 $\exp(\theta B_t - \theta^2 t/2)$ 是鞅。
4. 计算 Brownian bridge 的均值与协方差，并说明它不是独立增量过程。

本节结论

Brownian motion 的核心工具是 Gaussian 线性代数、强 Markov 性、反射与指数鞅；路径直觉必须由这些可计算结构支撑。

11.2 Brownian motion 的有限维分布与协方差

标准 Brownian motion 满足 $B_0 = 0$ 、连续路径、独立平稳高斯增量：

$$B_t - B_s \sim N(0, t - s), \quad 0 \leq s < t.$$

因此它是中心高斯过程，协方差

$$\text{Cov}(B_s, B_t) = \min(s, t).$$

若 $s \leq t$ ，写

$$B_t = B_s + (B_t - B_s),$$

后项与 B_s 独立，故协方差为 $\text{Var}(B_s) = s$ 。

缩放性：

$$\{c^{-1/2}B_{ct} : t \geq 0\} \stackrel{d}{=} \{B_t : t \geq 0\}.$$

时间反演在有限区间也产生 Brownian 结构。这些是过程层面的有限维分布等式，若需要路径事件还要结合连续版本。

11.3 反射原理与最大值分布

令

$$M_t = \max_{0 \leq s \leq t} B_s, \quad \tau_a = \inf\{s : B_s = a\}.$$

对 $b < a$ ，在首次到达 a 后把路径增量反射，得到

$$\mathbb{P}(M_t \geq a, B_t \leq b) = \mathbb{P}(B_t \geq 2a - b).$$

取 $b = a$ 并分解终点，得到

$$\mathbb{P}(M_t \geq a) = 2\mathbb{P}(B_t \geq a), \quad a > 0.$$

所以

$$M_t \stackrel{d}{=} |B_t|,$$

但二者不是同一个随机变量，联合结构也不同。

命中时间分布：

$$\mathbb{P}(\tau_a \leq t) = 2 \left[1 - \Phi \left(\frac{a}{\sqrt{t}} \right) \right].$$

密度为

$$f_{\tau_a}(t) = \frac{a}{\sqrt{2\pi t^3}} \exp \left(-\frac{a^2}{2t} \right).$$

其期望发散，提醒命中 a.s. 有限不等于平均命中时间有限。

11.4 指数鞅与漂移命中

对任意 $\theta \in \mathbb{R}$,

$$M_t = \exp \left(\theta B_t - \frac{1}{2} \theta^2 t \right)$$

是鞅。因为对 $s < t$,

$$\mathbb{E}[e^{\theta(B_t - B_s) - \theta^2(t-s)/2}] = 1$$

且增量独立。

它可用于带漂移 Brownian $X_t = B_t + \mu t$ 的 hitting probability 或 change of measure。停止前必须先截断 $\tau \wedge n$ ，再检查 UI/有界性让 $n \rightarrow \infty$ ；不能对任意无界命中时刻直接使用指数鞅等式。

11.5 二次变差与非经典可微性

在分割 $\Pi_n : 0 = t_0 < \dots < t_m = t$ 网格宽趋零时，

$$\sum_i (B_{t_{i+1}} - B_{t_i})^2 \xrightarrow{\mathbb{P}} t.$$

因为期望为 $\sum_i \Delta t_i = t$ ，方差在独立高斯增量下为

$$2 \sum_i (\Delta t_i)^2 \rightarrow 0.$$

这称二次变差 $[B]_t = t$ 。光滑函数路径的二次变差为零，说明 Brownian 路径不可能像普通可微函数那样变化。

Itô 公式中的二阶修正正来自非零二次变差：

$$df(t, B_t) = f_t dt + f_x dB_t + \frac{1}{2} f_{xx} dt.$$

即使考纲只要求基础过程，也应知道为什么随机链式法则多出二阶项。

11.6 Brownian bridge 与条件高斯

给定 $B_T = 0$ ，过程的条件均值为零，条件协方差

$$\text{Cov}(B_s, B_t | B_T = 0) = \min(s, t) - \frac{st}{T}.$$

可构造

$$\beta_t = B_t - \frac{t}{T} B_T,$$

则 β_t 与 B_T 联合高斯且协方差为零，因此独立； β 具有 Brownian bridge 分布。它是 Gaussian 条件分布在过程层面的直接应用。

答题方法

Brownian 题先用独立高斯增量求协方差；最大值/命中用反射；指数函数用指数鞅；路径平方增量用二次变差。每次停止都先截断再放极限。

自测

1. 推导 Brownian 协方差 $\min(s, t)$ 。
2. 用反射原理求 M_t 的分布。
3. 命中时间为何 a.s. 有限但期望无穷？
4. 证明网格二次变差在 L^2 中趋于 t 。
5. 推导 Brownian bridge 的条件协方差。

第十二章 Galton–Watson 分枝过程与树上临界

学习目标

本章用概率生成函数迭代描述 Galton–Watson 过程，证明灭绝概率是固定点，并把树渗流的临界值化为平均后代数问题。

12.1 Galton–Watson 分枝过程

设 $Z_0 = 1$ 、 $Z_{n+1} = \sum_{j=1}^{Z_n} X_{n,j}$ ，后代 pgf 为 $f(s) = \mathbb{E}s^X$ 、均值 $m = f'(1)$ 。则

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = mZ_n, \quad \mathbb{E}Z_n = m^n, \quad \mathbb{E}[s^{Z_{n+1}} | \mathcal{F}_n] = f(s)^{Z_n}, \quad \mathbb{E}s^{Z_n} = f^{\circ n}(s).$$

所以 Z_n/m^n ($m > 0$) 是非负鞅。

定理 12.1 (灭绝概率). 灭绝概率 q 是 $f(s) = s$ 在 $[0, 1]$ 上的最小根。在 $p_0 > 0$ 且分布非退化条件下，若 $m \leq 1$ 则 $q = 1$ ；若 $m > 1$ 则存在唯一 $q < 1$ 。

证明. $q_n = \mathbb{P}(Z_n = 0) = f^{\circ n}(0)$ 单调上升，极限 q 满足 $f(q) = q$ 。若 r 是任一固定点，则由单调性 $q_n \leq r$ ，故 $q \leq r$ 。根的数目由 f 的凸性、 $f(1) = 1$ 与 $f'(1) = m$ 决定。□

分枝树每条边以概率 p 保留后，新后代均值为 pm ，故超临界原树存活条件下渗流临界值为 $p_c = 1/m$ 。严格的 quenched 结论需遗传事件 0–1 性；完整证明在 2026 春 Q6 解答中给出。

12.2 综合题与易错点

例题 12.1 (Brownian 首次时). 对 $T_a = \inf\{t : B_t = a\}$ ，停指数鞅于 $T_a \wedge n$ ，取参数 $u = \sqrt{2\lambda}$ ： $1 = \mathbb{E}e^{uB_{T_a \wedge n} - \lambda(T_a \wedge n)}$ 。通过有界区域先停于 $[-b, a]$ 、再令 $b, n \rightarrow \infty$ 可严格控制一致可积，得到 $\mathbb{E}e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}$ 。直接无界停时上套 OST 是不完整的。

易错点

OST 绝非“对任意停时都成立”；必须逐条写明有界、一致可积、增量有界加有限均值等条件。非负鞅 a.s. 收敛不保证期望不丢失。不可约正常返保证平稳分布，非周期才保证逐时分布收敛。Brownian 的每个固定时刻正态不等于路径可微；事实上路径 a.s. 处处不可微。

复习检查

能够构造指数鞅和平方补偿鞅；证明 Doob 最大不等式；列出四个安全 OST 版本；证明 Wald 等式；解赌徒破产；分类有限/可数 Markov 链并求平稳分布；使用详细平衡；计算 Poisson 条件到达；用反射原理和指数鞅求首次；写出分枝过程 pgf 递推和灭绝固定点。

答题方法

分枝题先写后代 pgf f ，再用 $f^{\circ n}$ 表示第 n 代。灭绝概率取最小固定点；临界判断看 $f^{(2)}(1)$ 与均值 $f^{(1)}(1)$ ，不要把某个非最小固定点误作灭绝率。

自测

1. 证明灭绝概率是 $f(s) = s$ 在 $[0, 1]$ 的最小解。
2. 讨论均值小于、等于和大于一时的灭绝结论。
3. 把 m -叉树边渗流化为 $\text{Binomial}(m, p)$ 分枝过程并求 p_c 。
4. 说明临界情形平均规模不变却仍几乎必然灭绝。

本节结论

pgf 把随机树的递归独立性变成函数迭代；均值决定临界方向，固定点决定精确灭绝概率。

12.3 生成函数迭代与灭绝概率

Galton-Watson 过程中，每个个体后代数 ξ i.i.d.，生成函数

$$f(s) = \mathbb{E}s^\xi.$$

从单个祖先出发，

$$\mathbb{E}[s^{Z_{n+1}} | Z_n] = f(s)^{Z_n},$$

故

$$\mathbb{E}s^{Z_n} = f^{\circ n}(s).$$

灭绝事件为

$$\{\exists n : Z_n = 0\},$$

概率

$$q = \lim_{n \rightarrow \infty} \mathbb{P}(Z_n = 0) = \lim_{n \rightarrow \infty} f^{\circ n}(0).$$

连续性给 $f(q) = q$ ，且单调迭代表明 q 是 $[0, 1]$ 中最小不动点。

令 $m = f'(1) = \mathbb{E}\xi$ 。在非退化情形，

$$m \leq 1 \Rightarrow q = 1, \quad m > 1 \Rightarrow q < 1.$$

临界 $m = 1$ 仍 a.s. 灭绝, 但平均规模 $\mathbb{E}Z_n = 1$ 不下降, 显示期望不决定典型路径。

12.4 归一化人口鞅

若 $0 < m < \infty$,

$$W_n = \frac{Z_n}{m^n}$$

是非负鞅, 因为

$$\mathbb{E}[Z_{n+1} | \mathcal{F}_n] = mZ_n.$$

故 $W_n \rightarrow W$ a.s.。但要有 $\mathbb{E}W = 1$ 和 L^1 收敛, 需要 UI; 经典 $L \log L$ 条件

$$\mathbb{E}[\xi \log^+ \xi] < \infty$$

在超临界情形起关键作用。只由非负鞅收敛不能断言期望保持。

12.5 总后代数与随机游走编码

总人口

$$T = \sum_{n \geq 0} Z_n$$

可用队列探索编码: 每处理一个个体, 待处理数变化 $\xi_i - 1$ 。从一个祖先出发,

$$T = \inf \left\{ n \geq 1 : 1 + \sum_{i=1}^n (\xi_i - 1) = 0 \right\}.$$

这把分枝总规模转化为随机游走首次达时间。亚临界 $m < 1$ 时

$$\mathbb{E}T = \sum_{n \geq 0} \mathbb{E}Z_n = \sum_{n \geq 0} m^n = \frac{1}{1-m}.$$

12.6 树渗流的 boundary 身份

在 d -叉树上, 每条边以概率 p 独立开放。从根出发的开放子树是后代分布 $\text{Bin}(d, p)$ 的 Galton-Watson 过程, 平均后代 dp 。因此存在无限开放簇的临界值

$$p_c = \frac{1}{d}.$$

这一联系可用于历史真题, 但树渗流/GFF 等不自动扩成当前统计主线; 在覆盖矩阵中应标为历史题或 boundary, 并保留唯一完整解答。

答题方法

分枝题先写后代生成函数, 再用迭代求第 n 代分布, 用最小不动点求灭绝, 用 Z_n/m^n 研究规模。期望、灭绝概率与归一化极限是三个不同对象。

自测

1. 为什么灭绝概率是 $f(s) = s$ 的最小不动点?
2. 临界过程中 $\mathbb{E}Z_n = 1$ 如何与 a.s. 灭绝相容?
3. $W_n = Z_n/m^n$ 的 UI 为什么重要?
4. 用随机游走编码解释总后代数。
5. 推导 d -叉树独立边渗流的临界值。

第三部分

Mathematical Statistics

第十三章 次序统计量、充分性、完备性与 Basu

学习目标

本章建立统计推断的压缩原则：从次序统计量与联合密度进入 factorization、minimal sufficiency、complete statistics、Rao–Blackwell、Lehmann–Scheffé 与 Basu。

13.1 统计模型、统计量与充分性

统计模型是分布族 $\{P_\theta : \theta \in \Theta\}$ 。统计量是样本的可测函数，不含未知参数。充分性表达“给定统计量后，样本剩余随机性不再含参数信息”。

定理 13.1 (Fisher–Neyman 因子分解). 在由同一 σ -有限测度支配的模型中，统计量 $T(X)$ 对 θ 充分，当且仅当存在非负函数 g_θ, h 使

$$f_\theta(x) = g_\theta(T(x))h(x)$$

几乎处处成立，其中 h 不依赖 θ 。

证明. 若分解成立，则给定 $T = t$ 后，条件分布中 $g_\theta(t)$ 约去，故不依赖参数。反向若 T 充分，取由 T 诱导的边缘密度 $g_\theta(t)$ 与一个固定参数下的条件核 $K(dx | t)$ ；充分性说明该核可选为与参数无关，Radon–Nikodym 分解即给 $f_\theta(x) = g_\theta(T(x))h(x)$ 。离散情形可直接按每个纤维 $\{x : T(x) = t\}$ 归一化。□

命题 13.2 (极小充分的似然比判据). 在支配模型中，若对任意 x, y ，比值 $f_\theta(x)/f_\theta(y)$ 与 θ 无关当且仅当 $T(x) = T(y)$ ，则 T 极小充分。支集随参数变化时，比值为 $0/0$ 的点要按共同支集谨慎处理。

例题 13.1 (指数族). k 参数自然指数族

$$f_\eta(x) = h(x) \exp\{\eta^T T(x) - A(\eta)\}$$

中 T 充分。若自然参数空间含有 $\mathbb{R}^k$ 中的非空开集，且相应矩母函数在邻域内有限，则 T 的族通常完备；“满秩”条件不可省略。

13.2 完备性、辅助统计量与 Basu

定义 13.1 (完备与辅助). 统计量 T 完备是指对所有 θ ， $\mathbb{E}_\theta g(T) = 0$ 蕴含 $g(T) = 0$ a.s. (对所有参数)。统计量 A 辅助是指其分布不依赖参数。

定理 13.3 (Basu). 若 T 是完备充分统计量, A 是辅助统计量, 则 T 与 A 独立。

证明. 对任意有界 $h(A)$, 令 $g(T) = \mathbb{E}[h(A) | T] - \mathbb{E}h(A)$ 。充分性使条件期望可选为不依赖参数的同一函数; 辅助性使 $\mathbb{E}_\theta h(A)$ 也不依赖参数。故对所有 θ 有 $\mathbb{E}_\theta g(T) = 0$, 完备性推出 $g(T) = 0$ a.s., 即 $\mathbb{E}[h(A) | T] = \mathbb{E}h(A)$ 。对指标函数使用单调类定理得独立。 $\square$

正态样本中 $(\bar{X}, S^2)$ 独立可由 Basu 得到: 当方差已知时 $\bar{X}$ 对均值完备充分, 而残差方向/样本方差适当标准化为辅助; 双参数情形更常用正交变换与 Cochran 定理。

答题方法

充分性用因子分解; 极小充分性比较似然比是否与参数无关; 完备性从 $\mathbb{E}_\theta g(T) = 0$ 对全部参数推出 $g = 0$ 。UMVU 先 Rao-Blackwell, 再用完备性保证唯一。

自测

1. 求指数族的自然充分统计量并说明自然参数空间含开集时的完备性。
2. 对 $\text{Uniform}(0, \theta)$ 求极小充分统计量并讨论完备性。
3. 证明 Rao-Blackwell 不增 MSE, 并写出等号条件。
4. 用 Basu 定理判断正态样本均值与样本方差的独立性。

本节结论

充分性负责不丢参数信息, 完备性负责消除无偏估计歧义, 二者共同把 UMVU 问题转化为条件期望计算。

13.3 因子分解定理的双向证明

统计量 $T(X)$ 对参数 θ 充分, 直观上指给定 T 后样本的条件分布不再含 θ 。在有共同支配测度时, Neyman-Fisher 因子分解定理为

$$f_\theta(x) = g_\theta(T(x))h(x)$$

当且仅当 T 充分。

若有分解, 对离散模型,

$$\mathbb{P}_\theta(X = x | T = t) = \frac{g_\theta(t)h(x)}{\sum_{y: T(y)=t} g_\theta(t)h(y)} = \frac{h(x)}{\sum_{y: T(y)=t} h(y)},$$

与 θ 无关。连续情形用正则条件分布/支配测度作相同消去。

反向若条件分布 $q(x | T(x))$ 不含参数, 则

$$f_\theta(x) = q(x | T(x)) f_{\theta, T}(T(x)),$$

取 $h = q$ 、 $g_\theta = f_{\theta, T}$ 即得分解。

13.4 极小充分统计量的似然比判据

在密度共同支配且支撑不随参数改变的常见情形，若统计量 T 满足

$$\frac{f_{\theta}(x)}{f_{\theta}(y)} \text{ 与 } \theta \text{ 无关} \iff T(x) = T(y),$$

则 T 极小充分：任何充分统计量 S 的等值类都细于 T ，所以 T 是 S 的函数。

例题 13.2 (均匀分布的支撑含参). 若 $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$,

$$f_{\theta}(x) = \theta^{-n} \mathbf{1}\{0 < x_{(1)}, x_{(n)} < \theta\}.$$

$T = X_{(n)}$ 充分且极小充分。支撑依赖 θ ，不能把常规 score 的期望为零与 Cramér–Rao 正则公式无条件套用。

13.5 一参数指数族与完备性

满秩自然指数族写为

$$f_{\eta}(x) = h(x) \exp\{\eta T(x) - A(\eta)\}, \quad \eta \in \mathcal{N},$$

其中自然参数空间 $\mathcal{N}$ 含开区间。 T 充分。

若可测 g 满足

$$\mathbb{E}_{\eta} g(T) = 0 \quad \forall \eta \in \mathcal{N},$$

则

$$\int g(t) e^{\eta t} \nu(dt) = 0$$

在一个开区间成立。Laplace 变换唯一性推出 $g(T) = 0$ a.s., 故 T 完备。开自然参数空间很关键；若参数只取有限点，有限个矩方程一般不足以强迫 $g = 0$ 。

例题 13.3 (Poisson). $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\lambda)$,

$$T = \sum_i X_i \sim \text{Poisson}(n\lambda)$$

充分且完备。若 $\mathbb{E}_{\lambda} g(T) = 0$, 则

$$e^{-n\lambda} \sum_{t=0}^{\infty} g(t) \frac{(n\lambda)^t}{t!} = 0 \quad \forall \lambda > 0.$$

幂级数唯一性逐项给 $g(t) = 0$ 。

13.6 Rao–Blackwell、Lehmann–Scheffé 与 Basu

若 U 估计 $g(\theta)$, T 充分, 则

$$\tilde{U} = \mathbb{E}(U | T)$$

仍有相同期望；平方损失下

$$\text{Var}(U) = \mathbb{E}[\text{Var}(U | T)] + \text{Var}(\tilde{U}) \geq \text{Var}(\tilde{U}).$$

若 T 还完备, 任何两个基于 T 的无偏估计几乎处处相等。因此 $\tilde{U}$ 是唯一 UMVU, 这就是 Lehmann–Scheffé 定理。

Basu 定理: 完备充分统计量与任何辅助统计量独立。辅助统计量的分布不依赖 θ 。正态样本中

$$\bar{X} \text{ 与 } \frac{\sum_i (X_i - \bar{X})^2}{\sum_i (X_i - \bar{X})^2 + \text{适当位置量}}$$

等结构需具体辨认; 最经典结论是 $\bar{X}$ 与样本方差独立, 可由正交 Gaussian 分解直接证明, 也与指数族结构相容。

13.7 次序统计量的联合密度

连续 i.i.d. 样本分布函数 F 、密度 f 。第 k 个次序统计量密度为

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

理由是: 一个样本落在 $[x, x+dx]$, $k-1$ 个落左侧, $n-k$ 个落右侧, 再乘排列数。

对 $i < j$, 联合密度为

$$\frac{n!}{(i-1)!(j-i-1)!(n-j)!} F(x)^{i-1} [F(y)-F(x)]^{j-i-1} [1-F(y)]^{n-j} f(x)f(y)$$

在 $x < y$ 上成立。

答题方法

充分性题先因子分解; 极小充分用似然比等值类; 完备性用变换唯一性; UMVU 用 Rao–Blackwell 加完备; 辅助统计量独立用 Basu。五步的条件不能互相替代。

自测

1. 证明因子分解定理的两个方向。
2. 为什么 $\text{Uniform}(0, \theta)$ 的最大次序统计量充分?
3. 指数族完备性为何需要自然参数空间有内点?
4. Lehmann–Scheffé 如何从 Rao–Blackwell 与完备性得到唯一性?
5. 推导第 k 个次序统计量密度。

第十四章 参数估计、Fisher 信息与风险

学习目标

本章比较矩估计、MLE、无偏估计和决策论风险，推导 score、Fisher information 与 Cramér–Rao 下界，并区分正则与参数依赖支集的非正则模型。

14.1 矩估计、MLE 与存在性

矩估计以理论矩 $m_j(\theta) = \mathbb{E}_\theta X^j$ 匹配样本矩。MLE 为 $\hat{\theta} \in \arg \max_{\theta \in \Theta} L(\theta; x)$ 。MLE 的不变性指若 $\hat{\theta}$ 最大化似然，则 $g(\hat{\theta})$ 是 $g(\theta)$ 的 MLE（多对一时在原像上取剖面最大值）。

求 MLE 的标准链：写支集与参数空间；写对数似然；找内部驻点；检查 Hessian/单调性；检查边界；最后给出完整 $\arg \max$ 集。解 score 方程并不保证最大值存在或唯一，尤其在参数依赖支集、开参数空间、混合模型和尺度趋零问题中。

例题 14.1 ($\text{Uniform}(0, \theta)$)。似然 $L(\theta) = \theta^{-n} \mathbf{1}_{\{\theta \geq X_{(n)}\}}$ ，故 MLE 为 $X_{(n)}$ 。矩估计为 $2\bar{X}$ 。前者非正则：支集依赖参数，不能直接套通常 Fisher 信息与 MLE 渐近正态定理；事实上 $n(\theta - X_{(n)})/\theta \xrightarrow{d} \text{Exp}(1)$ 。

14.2 偏差、MSE 与 Rao–Blackwell

估计量 δ 的偏差和均方误差为

$$\text{Bias}_\theta(\delta) = \mathbb{E}_\theta \delta - g(\theta), \quad \text{MSE}_\theta(\delta) = \text{Var}_\theta(\delta) + \text{Bias}_\theta(\delta)^2.$$

定义 14.1 (一致最小方差无偏估计). 设目标为 $g(\theta)$ 。若 δ 对每个 θ 都满足 $\mathbb{E}_\theta \delta = g(\theta)$ ，且对任意另一个同一目标的无偏估计 U 都有

$$\text{Var}_\theta(\delta) \leq \text{Var}_\theta(U) \quad \text{对所有 } \theta,$$

则称 δ 为 $g(\theta)$ 的一致最小方差无偏估计 (UMVU/UMVUE；部分讲义记 BUE)。比较类必须限定为同一目标函数的无偏估计，不能拿有偏估计的更小方差作反例。

定理 14.1 (Rao–Blackwell). 若 T 对 θ 充分， U 是 $g(\theta)$ 的无偏估计且平方可积，则 $U^* = \mathbb{E}[U | T]$ 仍无偏，并且 $\text{Var}_\theta(U^*) \leq \text{Var}_\theta(U)$ ；等号当且仅当 $U = U^*$ a.s.

证明. 塔式性质给无偏性。全方差公式给 $\text{Var}(U) = \text{Var}(\mathbb{E}[U | T]) + \mathbb{E} \text{Var}(U | T)$ 。充分性保证条件期望可作为只依赖 T 而不依赖未知参数的统计量。□

定理 14.2 (Lehmann–Scheffé). 若 T 完备充分, $h(T)$ 是 $g(\theta)$ 的无偏估计, 则它是唯一 (a.s.) UMVU 估计。

证明. 任一无偏 U 经 Rao–Blackwell 得到 $U^* = \mathbb{E}[U | T]$, 方差不增。 $U^* - h(T)$ 对所有参数期望为零, 完备性使其 a.s. 为零。因此 $h(T)$ 达到所有无偏估计中的最小方差, 且唯一。 $\square$

14.3 Fisher 信息与 Cramér–Rao 下界

在支集不依赖参数、可对参数微分并交换积分、score 平方可积且信息非退化的正则模型中,

$$U_\theta(X) = \partial_\theta \log f_\theta(X), \quad \mathbb{E}U_\theta = 0, \quad I(\theta) = \mathbb{E}U_\theta^2 = -\mathbb{E}\partial_\theta^2 \log f_\theta(X).$$

i.i.d. 样本信息为 $nI_1(\theta)$ 。

定理 14.3 (一参数 Cramér–Rao). 若 T 无偏估计 $g(\theta)$, 且上述正则条件成立, 则

$$\text{Var}_\theta(T) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}.$$

等号当且仅当 $T - g(\theta) = a(\theta)U_\theta$ a.s.

证明. 由换序,

$$g'(\theta) = \partial_\theta \mathbb{E}_\theta T = \mathbb{E}_\theta[TU_\theta] = \text{Cov}(T, U_\theta).$$

Cauchy–Schwarz 给

$$[g'(\theta)]^2 \leq \text{Var}(T) \text{Var}(U_\theta) = \text{Var}(T)I_n(\theta).$$

等号条件来自 Cauchy–Schwarz 等号。 $\square$

多参数情形若信息矩阵 $I(\theta)$ 正定, 则对标量 g , $\text{Var}(T) \geq \nabla g(\theta)^T I(\theta)^{-1} \nabla g(\theta)$ 。边界参数、参数依赖支集和不可识别模型不应机械套用。

答题方法

估计题先写参数空间和支集, 再写似然。使用 score 恒等式、CRLB 或 MLE 渐近理论前, 必须检查微分与积分交换、信息有限和真参数是否为内点。

自测

1. 推导一参数正则模型的 Cramér–Rao 下界并写出等号条件。
2. 给出 $\text{Uniform}(0, \theta)$ 违反正则 score 恒等式的计算。
3. 比较无偏、相合、有效、可容许和 minimax 的逻辑关系。
4. 对一个一参数指数族计算矩估计、MLE、信息和渐近方差。

本节结论

点估计没有单一“最好”标准：有限样本偏差/MSE、无偏类 UMVU、渐近效率和决策风险回答的是不同问题。

14.4 矩估计、MLE 与参数空间边界

矩估计把理论矩 $m_j(\theta) = \mathbb{E}_\theta X^j$ 与样本矩匹配。它计算简单，但不自动有效或唯一。MLE 定义为

$$\hat{\theta} \in \arg \max_{\theta \in \Theta} L(\theta; x).$$

必须检查：

1. 参数空间开/闭与边界；
2. 似然是否连续、上确界是否取到；
3. 驻点是否为全局最大；
4. 支撑是否依赖参数。

例题 14.2. $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Unif}(0, \theta)$ 时

$$L(\theta) = \theta^{-n} \mathbf{1}\{\theta \geq X_{(n)}\},$$

在 $\theta \geq X_{(n)}$ 上递减，故 $\hat{\theta} = X_{(n)}$ 。用 score 方程会漏掉支撑指示和边界。

14.5 偏差、方差、MSE 与风险

平方损失下

$$\text{MSE}_\theta(\hat{\theta}) = \mathbb{E}_\theta(\hat{\theta} - \theta)^2 = \text{Var}_\theta(\hat{\theta}) + \text{Bias}_\theta(\hat{\theta})^2.$$

无偏并不等于 MSE 最小。若 $\hat{\theta}_c = c\bar{X}$ 估计正态均值，可能通过轻微缩减降低某些参数区间的 MSE，却引入偏差。比较估计量必须说明是逐点风险、一致支配、Bayes 风险还是 minimax 风险。

14.6 Fisher 信息与 score 恒等式

score

$$U_\theta(X) = \partial_\theta \log f_\theta(X).$$

在可交换微分积分、支撑不随参数变化等正则条件下，

$$\mathbb{E}_\theta U_\theta = 0.$$

Fisher 信息有两种等价形式

$$I(\theta) = \mathbb{E}_\theta U_\theta^2 = -\mathbb{E}_\theta \partial_\theta^2 \log f_\theta(X).$$

i.i.d. 样本信息相加:

$$I_n(\theta) = nI_1(\theta).$$

若支撑含参, 这些恒等式可能失败, 必须回到定义。

14.7 Cramér–Rao 下界的完整证明

设 T 无偏估计 $g(\theta)$ 。正则条件下

$$g'(\theta) = \partial_\theta \mathbb{E}_\theta T = \mathbb{E}_\theta [TU_\theta] = \text{Cov}_\theta(T, U_\theta),$$

因为 $\mathbb{E}U_\theta = 0$ 。Cauchy–Schwarz 给

$$[g'(\theta)]^2 \leq \text{Var}_\theta(T) I_n(\theta),$$

故

$$\text{Var}_\theta(T) \geq \frac{[g'(\theta)]^2}{I_n(\theta)}.$$

等号当且仅当

$$T - g(\theta) = c(\theta)U_\theta \quad a.s.$$

这通常要求 score 对样本的依赖可由同一个统计量表达。

定理 14.4 (多参数 Cramér–Rao 矩阵界). 设 $\theta \in \mathbb{R}^p$, score 向量 $U_\theta = \nabla_\theta \log f_\theta(X)$ 满足交换微分与积分等正则条件, 信息矩阵 $I(\theta) = \mathbb{E}_\theta U_\theta U_\theta^\top$ 正定。若 $T \in \mathbb{R}^q$ 无偏估计 $g(\theta)$, 且 $G(\theta) = \nabla_\theta g(\theta)^\top$, 则

$$\text{Cov}_\theta(T) \succeq G(\theta)I(\theta)^{-1}G(\theta)^\top.$$

证明. 正则性给 $\mathbb{E}U_\theta = 0$, 对 $\mathbb{E}_\theta T = g(\theta)$ 求导得 $\text{Cov}(T, U_\theta) = G(\theta)$ 。令 $A = GI^{-1}$, 随机向量 $(T - g) - AU_\theta$ 的协方差半正定。展开后交叉项用 $\text{Cov}(T, U) = G$, 得到 $\text{Cov}(T) - GI^{-1}G^\top \succeq 0$ 。若信息矩阵奇异, 必须限制到可辨识切空间或使用广义逆, 不能直接写普通逆。□

例题 14.3 (Bernoulli). $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$,

$$I_n(p) = \frac{n}{p(1-p)}.$$

$\bar{X}$ 无偏, 方差

$$\frac{p(1-p)}{n} = \frac{1}{I_n(p)},$$

达到估计 $g(p) = p$ 的 Cramér–Rao 下界。

14.8 置信区间与枢轴量

枢轴量的分布不含未知参数。正态样本:

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

在 σ 已知时给

$$\bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

σ 未知时

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

依赖正态样本均值与样本方差独立, 以及 $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ 。

置信区间的频率解释是重复抽样覆盖率:

$$\mathbb{P}_\theta\{\theta \in C(X)\} = 1 - \alpha.$$

对已观察到的固定区间, 不在频率学派中说“参数以 0.95 概率位于其中”。

14.9 正则 MLE 的渐近效率

score 方程在真值附近展开:

$$0 = U_n(\hat{\theta}) = U_n(\theta_0) + (\hat{\theta} - \theta_0)U'_n(\tilde{\theta}).$$

若

$$\frac{1}{\sqrt{n}}U_n(\theta_0) \xrightarrow{d} N(0, I(\theta_0)), \quad -\frac{1}{n}U'_n(\tilde{\theta}) \xrightarrow{\mathbb{P}} I(\theta_0),$$

且 $\hat{\theta} \xrightarrow{\mathbb{P}} \theta_0$, 则

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

一致性、信息非奇异与余项控制缺一不可; “MLE 所以正态” 不是证明。

答题方法

估计题先检查存在性, 再算偏差/风险; 无偏最优用充分完备; 信息下界先核验正则性; 渐近 MLE 用 score 展开、CLT 与 LLN 三件套。

自测

1. Uniform 模型为什么不能只解 score 方程?
2. 推导 Fisher 信息的两种形式及所需正则条件。
3. 完整证明 Cramér–Rao 下界并写等号条件。
4. 频率置信区间的 $1 - \alpha$ 概率针对什么随机对象?
5. MLE 渐近正态证明的三个主要极限步骤是什么?

第十五章 假设检验、似然比与置信区间

学习目标

本章从两类错误与功效出发，证明 Neyman–Pearson，引入 MLR/Karlin–Rubin、UMP/UMPU、LRT、Wald/score，并通过检验反演构造置信区间。

15.1 置信区间、枢轴量与正态模型

随机区间 $C(X)$ 的覆盖率是 $\mathbb{P}_\theta(\theta \in C(X))$ ，它是重复抽样概率；观测后频率区间不是“参数有 $1 - \alpha$ 概率落入”。枢轴量的分布不含未知参数，通过分位数反演构造区间。

若 $X_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$ ：

$$\begin{aligned} \sigma \text{ 已知：} & \quad \mu \in \bar{X} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}; \\ \sigma \text{ 未知：} & \quad \mu \in \bar{X} \pm t_{n-1, 1-\alpha/2} \frac{S}{\sqrt{n}}; \\ \mu \text{ 未知：} & \quad \frac{(n-1)S^2}{\chi_{n-1, 1-\alpha/2}^2} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{n-1, \alpha/2}^2}. \end{aligned}$$

这里 $S^2 = (n-1)^{-1} \sum (X_i - \bar{X})^2$ 。双样本正态问题必须先判断独立、等方差与配对结构，再选择 pooled t 、Welch 或配对 t 。

例题 15.1 (Poisson 的 UMVU). 若 $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Poisson}(\eta)$, $T = \sum X_i \sim \text{Poisson}(n\eta)$ 完备充分。要估计 η^2 ，因为 $\mathbb{E}[T(T-1)] = (n\eta)^2$ ，故 $T(T-1)/n^2$ 是 UMVU。若参数写作 $\eta = \sqrt{\lambda}$ ，这正给出 λ 的 UMVU，对应 2026 春 Q10。

例题 15.2 (指数分布尾概率). $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$, $T = \sum X_i$ 完备充分。要估计 $\phi = e^{-\lambda x} = \mathbb{P}(X_1 > x)$ ，先取无偏估计 $\mathbf{1}_{\{X_1 > x\}}$ ，再对 T 条件化。给定 $T = t$, $X_1/t \sim \text{Beta}(1, n-1)$ ，故

$$\hat{\phi}_{\text{UMVU}} = \left(1 - \frac{x}{T}\right)^{n-1} \mathbf{1}_{\{T > x\}}.$$

对应 2025 秋 Q8。

易错点

充分不等于完备；极小充分也不自动完备。MLE 不自动无偏、存在或唯一。CRLB 的换序、共同支集和信息正定条件不可省略。Rao–Blackwell 的条件期望必须能写成不含未知参数的统计量。置信度是程序覆盖率，不是参数后验概率。

复习检查

应能使用因子分解找充分统计量；用似然比判据找极小充分；证明一参数指数族完备；使用 Basu；完整求 MLE 并检查边界；用 Rao–Blackwell/Lehmann–Scheffé 求 UMVU；计算 Fisher 信息和 CRLB；写出正态单/双样本枢轴区间。

15.2 检验、错误概率、功效与 p 值

检验函数 $\phi(X) \in [0, 1]$ 表示拒绝概率。对 $H_0 : \theta \in \Theta_0$, size 是 $\sup_{\theta \in \Theta_0} \mathbb{E}_\theta \phi$ ；level α 表示 size 不超过 α 。功效函数是 $\beta_\phi(\theta) = \mathbb{E}_\theta \phi$ 。第二类错误概率为 $1 - \beta_\phi(\theta)$ ($\theta \in \Theta_1$)。

p 值是：在 H_0 下，得到“至少同样不利于原假设”的统计量值的最小显著性水平。它不是 $\mathbb{P}(H_0 | X)$ ，也不是效应大小。复合原假设下通常取最不利参数的尾概率上确界。

15.3 Neyman–Pearson 引理

定理 15.1 (Neyman–Pearson). 对简单假设 $H_0 : f_0$ 对 $H_1 : f_1$ ，若存在 $k \geq 0$ 与边界随机化 γ 使

$$\phi^*(x) = \begin{cases} 1, & f_1(x) > k f_0(x), \\ \gamma, & f_1(x) = k f_0(x), \\ 0, & f_1(x) < k f_0(x), \end{cases} \quad \mathbb{E}_0 \phi^* = \alpha,$$

则 ϕ^* 是所有 size 不超过 α 检验中功效最大的。反之，任一最强检验在 $f_0 + f_1 > 0$ 处必须具有这一形式（除零测集与边界随机化外）。

证明. 对任一 ϕ 满足 $\mathbb{E}_0 \phi \leq \mathbb{E}_0 \phi^*$ ，由构造可知 $(\phi^* - \phi)(f_1 - k f_0) \geq 0$ 逐点成立。积分得

$$\mathbb{E}_1 \phi^* - \mathbb{E}_1 \phi \geq k(\mathbb{E}_0 \phi^* - \mathbb{E}_0 \phi) \geq 0.$$

这证充分性。必要性若最强检验在 $f_1 > k f_0$ 的正测集上未充分拒绝、同时在较小似然比区域拒绝，则可把相同的 f_0 -概率质量从后者移向前者而严格增加 f_1 -功效，矛盾；边界上只能由 size 校准决定。□

简单对简单检验没有“UMP”必要，直接称最强检验；UMP 是对每个备择参数同时最强。

15.4 单调似然比、UMP 与检验反演

若一参数族对统计量 T 有单调似然比： $\theta_2 > \theta_1$ 时 $f_{\theta_2}(x)/f_{\theta_1}(x)$ 是 $T(x)$ 的非降函数，则 Karlin–Rubin 定理给单侧检验的 UMP 结构： $H_0 : \theta \leq \theta_0$ 对 $H_1 : \theta > \theta_0$ 在 T 大时拒绝；反向单侧在 T 小时拒绝。临界值按边界参数 θ_0 校准，离散情形可能需要随机化。

双侧复合备择通常不存在 UMP 检验；可采用无偏 UMPU、似然比或等尾检验。置信集与检验反演： $C(x) = \{\theta : H_{0,\theta} \text{ 未被 level } \alpha \text{ 检验拒绝}\}$ ，则覆盖率至少 $1 - \alpha$ 。

15.5 似然比、Wald 与 score

广义似然比记作

$$\Lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(\theta; x)}{\sup_{\theta \in \Theta} L(\theta; x)} \leq 1,$$

小的 Λ 拒绝。另一常用约定用倒数 $R = 1/\Lambda \geq 1$ 并在大值拒绝；两者完全等价。本书统一采用 $\Lambda \leq 1$ ，遇到倒数约定时会显式说明。

定理 15.2 (Wilks). 在真参数位于 Θ_0 内点、模型可识别、对数似然足够光滑、信息矩阵正定、约束正则且维数差为 r 等条件下，

$$-2 \log \Lambda \xrightarrow{d} \chi_r^2.$$

边界参数、不可识别、混合模型或参数依赖支集时结论可能失效。

对一维 $H_0: \theta = \theta_0$ ，三统计量为

$$W = nI(\hat{\theta})(\hat{\theta} - \theta_0)^2, \quad S = \frac{U_n(\theta_0)^2}{nI(\theta_0)}, \quad LR = -2 \log \Lambda.$$

正则条件下三者均趋于 χ_1^2 且差为 $o_p(1)$ 。Wald 只需非约束 MLE 但参数化敏感；score 只需原假设拟合；LR 参数化不变但需受限与非受限优化。

15.6 卡方检验与拟合优度

多项计数 O_j 、原假设期望 $E_j = np_j(\hat{\eta})$ 下 Pearson 统计量 $X^2 = \sum_j (O_j - E_j)^2 / E_j$ 。若各期望格数足够、模型正则，则极限自由度为“格数 -1 - 从数据估计的独立参数个数”。合并小期望格需在看数据前按合理规则进行。独立性列联表自由度为 $(r-1)(c-1)$ 。

答题方法

检验题先写原假设/备择、显著性水平和拒绝域。简单对简单用似然比排序；单参数单侧找 MLR；双侧问题不能默认存在 UMP，应考虑无偏约束、LRT 或检验反演。

自测

1. 完整证明 Neyman-Pearson 引理并处理边界随机化。
2. 从 MLR 推导单侧 UMP 检验。
3. 给出双侧正态均值不存在 UMP 的理由，并构造常用双侧检验。
4. 通过检验反演推导正态均值的 t 区间。
5. 比较 LRT、Wald 与 score 在有限样本和渐近层面的差别。

本节结论

检验与区间是同一族接受域的一种读法。最关键的不是临界值本身，而是似然排序、大小控制和参数空间方向。

15.7 Neyman–Pearson 引理的交换证明

检验简单假设

$$H_0 : f_0, \quad H_1 : f_1.$$

取

$$\phi^*(x) = \begin{cases} 1, & f_1(x) > k f_0(x), \\ \gamma, & f_1(x) = k f_0(x), \\ 0, & f_1(x) < k f_0(x), \end{cases}$$

使 $\mathbb{E}_0 \phi^* = \alpha$ 。对任意尺寸不超过 α 的检验 ϕ ，点态有

$$(\phi^* - \phi)(f_1 - k f_0) \geq 0.$$

积分得

$$\mathbb{E}_1 \phi^* - \mathbb{E}_1 \phi \geq k(\mathbb{E}_0 \phi^* - \mathbb{E}_0 \phi) \geq 0.$$

因此似然比检验在所有 size- α 检验中功效最大。随机化只在离散分布无法精确达到 α 时需要。

15.8 单调似然比与 Karlin–Rubin

定理 15.3 (Karlin–Rubin 单侧 UMP 定理). 若一参数族对统计量 T 有单调似然比: $\theta_1 > \theta_0$ 时

$$\frac{f_{\theta_1}(x)}{f_{\theta_0}(x)}$$

是 $T(x)$ 的非减函数, 则对单侧

$$H_0 : \theta \leq \theta_0, \quad H_1 : \theta > \theta_0$$

存在基于 T 大值拒绝的 level- α UMP 检验; 离散分布在临界点可能需要随机化。

证明. 先在边界参数 θ_0 下选 c, γ , 使

$$\phi^*(x) = \mathbf{1}_{\{T(x) > c\}} + \gamma \mathbf{1}_{\{T(x) = c\}}, \quad \mathbb{E}_{\theta_0} \phi^* = \alpha.$$

对任意固定 $\theta_1 > \theta_0$, 单调似然比说明 $f_{\theta_1}/f_{\theta_0}$ 随 T 非减, 因此 Neyman–Pearson 引理的似然比拒绝域正可取为 T 的上尾。故同一个 ϕ^* 对每个 $\theta_1 > \theta_0$ 都是最强检验。

还需验证复合零假设的水平。单调似然比蕴含 T 随 θ 单调随机增大: 对任意非减函数 $g, \mathbb{E}_{\theta} g(T)$ 随 θ 非减。取非减的 $g = \phi^*$, 对所有 $\theta \leq \theta_0$ 有

$$\mathbb{E}_{\theta} \phi^* \leq \mathbb{E}_{\theta_0} \phi^* = \alpha.$$

所以 ϕ^* 是 level- α , 并同时每个备择参数最强, 即 UMP。 □

双侧备择通常没有 UMP: 提高两侧功效的资源相互竞争。此时可能考虑 UMPU、似然比或等尾检验。

15.9 检验反演与置信集

对每个 θ_0 构造 size- α 检验 $\phi_{\theta_0}(X)$, 定义

$$C(X) = \{\theta_0 : \phi_{\theta_0}(X) = 0\}.$$

则

$$\mathbb{P}_{\theta}(\theta \in C(X)) = 1 - \mathbb{P}_{\theta}(\phi_{\theta}(X) = 1) \geq 1 - \alpha.$$

因此置信区间可由一族假设检验反演。双侧 t 区间正是反演

$$|T_{\mu_0}| \leq t_{n-1, 1-\alpha/2}.$$

15.10 似然比、Wald 与 score 的局部关系

广义似然比统计量

$$\Lambda = \frac{\sup_{\theta \in \Theta_0} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}.$$

正则条件下, 若约束维数差为 r , Wilks 定理给

$$-2 \log \Lambda \xrightarrow{d} \chi_r^2.$$

Wald 检验用估计量距零假设的标准化距离, score 检验只在零假设受限估计处计算梯度。三者局部正则备择下一阶等价, 但有限样本表现不同。

若真值位于参数空间边界、信息奇异、模型不可识别或支撑含参, Wilks 的卡方极限可能失效。例如方差分量检验的极限常为卡方混合。

15.11 p 值、尺寸与功效

p 值是零假设下“至少同样极端”的尾概率, 它是数据的函数。若

$$p(X) \leq \alpha$$

则拒绝 size- α 检验。p 值不是 $\mathbb{P}(H_0 | X)$, 也不衡量效应大小。

功效函数

$$\beta(\theta) = \mathbb{P}_{\theta}(\text{拒绝 } H_0).$$

在零假设参数上上确界不超过 α 是尺寸控制; 在备择上越大越好。增加样本量通常提高固定效应下功效, 但也会使非常小的实际效应达到统计显著。

15.12 卡方拟合优度与参数估计扣自由度

多项计数 O_j , 零假设期望 $E_j = np_j$, Pearson 统计量

$$X^2 = \sum_{j=1}^K \frac{(O_j - E_j)^2}{E_j}.$$

若 p_j 完全给定, 渐近自由度 $K - 1$, 因为计数总和固定。若从数据估计 r 个正则参数, 通常再扣 r , 自由度

$$K - 1 - r.$$

期望频数太小时卡方近似不可靠, 应合并类别或使用精确/模拟校准。

答题方法

检验题固定顺序: 写 H_0, H_1 与参数空间; 选择统计量; 给出零假设分布; 用尺寸校准临界值; 写功效; 说明单/双侧和是否随机化。p 值放在最后解释。

自测

1. 用点态符号证明 Neyman-Pearson 引理。
2. MLR 为什么容易给单侧 UMP, 却一般不给双侧 UMP?
3. 从检验反演证明置信集覆盖率。
4. Wilks 定理在哪些非正则情形可能失效?
5. 拟合优度检验为什么要扣除已估参数数目?

15.13 精确离散置信区间与最短性边界

若 $X \sim \text{Bin}(n, p)$, Clopper-Pearson 区间通过反演两个精确尾检验定义: 观测 x 时端点满足

$$\mathbb{P}_{p_L}(X \geq x) = \alpha/2, \quad \mathbb{P}_{p_U}(X \leq x) = \alpha/2,$$

并在 $x = 0, n$ 采用相应边界端点。它保证覆盖率至少 $1 - \alpha$, 但因离散性通常保守。若 $X \sim \text{Poisson}(\lambda)$, Garwood 区间为

$$\left[\frac{1}{2} \chi_{2x, \alpha/2}^2, \frac{1}{2} \chi_{2x+2, 1-\alpha/2}^2 \right],$$

其中 $x = 0$ 时下端取零。

例题 15.3 (边界样本的精确区间). 若二项样本观测到 $x = 0$, Clopper-Pearson 等尾区间为

$$\left[0, 1 - (\alpha/2)^{1/n} \right];$$

若 $x = n$, 则为

$$\left[(\alpha/2)^{1/n}, 1 \right].$$

例如 $n = 10, \alpha = 0.05, x = 0$ 时上端约为 0.3085。对 Poisson 计数 $x = 0$, Garwood 区间为

$$\left[0, \frac{1}{2} \chi_{2, 1-\alpha/2}^2 \right] = [0, -\log(\alpha/2)],$$

在 $\alpha = 0.05$ 时上端约为 3.6889。这些计算展示了边界端点, 但并没有求解“所有精确区间中最短者”的离散优化问题。

“精确”指有限样本覆盖保证, 不等于在所有参数下长度最短。离散模型中接受域跳跃使等尾、最短期望长度和无偏检验反演可能给不同区间; 若题目要求最短区间, 必须明确优化准则和允许的随机化, 不能把等尾端点直接称为最短。

第十六章 Bayesian 推断、共轭与决策

学习目标

本章从先验、似然和 posterior 的严格归一化开始，比较 posterior mean/median/MAP，处理参数变换、共轭族、improper prior 与 generalized Bayes，并用后验风险连接 Bayes 与 minimax。

16.1 Bayesian 后验、决策与共轭

Bayes 公式为

$$\pi(\theta | x) = \frac{f(x | \theta)\pi(\theta)}{\int f(x | u)\pi(u) du}.$$

平方损失下 Bayes action 是后验均值；绝对损失下是后验中位数；0-1 损失下是后验众数（连续参数需谨慎解释密度众数）。后验预测为 $p(x_{\text{new}} | x) = \int p(x_{\text{new}} | \theta)\pi(\theta | x)d\theta$ 。

似然	共轭先验	后验参数更新
Binomial(n, p)	Beta(a, b)	Beta($a + x, b + n - x$)
Poisson(λ)	Gamma(a, b) (率)	Gamma($a + \sum x_i, b + n$)
Exponential(λ)	Gamma(a, b) (率)	Gamma($a + n, b + \sum x_i$)
$N(\mu, \sigma^2)$, σ^2 已知	$N(m_0, s_0^2)$	精度相加，均值为精度加权

定义 16.1 (Bayes 与 minimax). 风险 $R(\theta, \delta) = \mathbb{E}_\theta L(\theta, \delta(X))$ 。Bayes rule 最小化先验平均风险；minimax rule 最小化 $\sup_\theta R(\theta, \delta)$ 。若某 Bayes rule 的风险恒为常数且其 Bayes 风险等于该常数，则它 minimax；更一般可用最不利先验极限。

定理 16.1 (等风险 Bayes 规则的 minimax 证书). 设先验 π 为概率分布， δ_π 是 Bayes 规则，且

$$R(\theta, \delta_\pi) = c \quad \text{对所有 } \theta \text{ 成立.}$$

则 δ_π 是 minimax 规则。若参数空间有限、 π 对每个参数点赋正质量，且 δ_π 的 Bayes 风险有限，则它还是可容许的。

完整证明. 对任意规则 δ ，先验平均风险不超过其最大风险，故

$$\sup_\theta R(\theta, \delta) \geq \int R(\theta, \delta) \pi(d\theta) \geq \int R(\theta, \delta_\pi) \pi(d\theta) = c.$$

而 $\sup_\theta R(\theta, \delta_\pi) = c$ ，所以 minimax 值恰为 c ， δ_π 达到它。

再证有限参数空间下的可容许性。若存在规则 δ' 在所有参数点风险不大于 δ_π ，并至少在一个参数点严格更小，则因每个点的先验质量都为正，

$$\sum_{\theta} \pi(\theta) R(\theta, \delta') < \sum_{\theta} \pi(\theta) R(\theta, \delta_\pi),$$

与 δ_π 的 Bayes 最优性矛盾。正质量条件不能删除：若被改进的参数点先验质量为零，平均风险看不见该改进。□

例题 16.1 (两点先验与绝对损失). $Y | \mu \sim N(\mu, 1)$ ，先验在 $\{-1, 1\}$ 上各半。后验赔率 $\mathbb{P}(\mu = 1 | y) / \mathbb{P}(\mu = -1 | y) = e^{2y}$ 。绝对损失且行动也为 $\{-1, 1\}$ 时，Bayes rule 为 $\delta(y) = 1$ 当 $y > 0$ ，否则 -1 ($y = 0$ 任意随机化)。两端风险相等，且该先验使其成为等风险 Bayes rule，故 minimax。对应 2026 春 Q9。

可信区间满足后验概率条件；频率置信区间满足重复抽样覆盖条件，二者不能互换。非信息先验可能不正规；必须检查后验是否可归一化。

答题方法

Bayes 题先验证后验是否正规，再按损失选择动作：平方损失取均值，绝对损失取中位数，0-1/小邻域损失对应众数。MAP 在参数变换下不具不变性，必须连同基准测度说明。

自测

1. 推导 Beta-Binomial、Gamma-Poisson 和 Normal-Normal 共轭后验。
2. 证明平方损失下 Bayes 动作为后验均值。
3. 给出 improper prior 产生 proper posterior 的例子，并区分 generalized Bayes。
4. 说明 MAP 在非线性重参数化下如何因 Jacobian 改变。
5. 用 least favorable prior 构造一个 minimax 规则。

本节结论

Bayesian 推断的三个层次不能混淆：后验可归一化、所选动作存在、后验风险有限。只有三者都核验，决策才完整。

16.2 后验分布与后验预测

Bayes 公式以密度写为

$$\pi(\theta | x) = \frac{p(x | \theta) \pi(\theta)}{\int p(x | u) \pi(u) du}.$$

分母是边缘似然

$$m(x) = \int p(x | \theta) \pi(\theta) d\theta,$$

既负责归一化，也用于模型比较。后验预测为

$$p(x_{\text{new}} | x) = \int p(x_{\text{new}} | \theta) \pi(\theta | x) d\theta,$$

它把参数不确定性积分掉，不等于简单代入后验均值。

16.3 共轭族的三个完整例子

Beta–Binomial.

$$X | p \sim \text{Bin}(n, p), \quad p \sim \text{Beta}(a, b).$$

后验

$$p | X = x \sim \text{Beta}(a + x, b + n - x).$$

后验均值

$$\mathbb{E}[p | x] = \frac{a + x}{a + b + n}$$

是先验均值与样本比例的加权平均。

Gamma–Poisson. 采用 rate 参数化

$$X_i | \lambda \sim \text{Poisson}(\lambda), \quad \lambda \sim \text{Gamma}(a, b),$$

则

$$\lambda | x \sim \text{Gamma}\left(a + \sum_i x_i, b + n\right).$$

必须注明 b 是 rate 还是 scale，否则后验公式会差倒数。

Normal–Normal. σ^2 已知,

$$X_i | \mu \sim N(\mu, \sigma^2), \quad \mu \sim N(m_0, s_0^2).$$

后验精度相加:

$$\frac{1}{s_n^2} = \frac{1}{s_0^2} + \frac{n}{\sigma^2}, \quad \frac{m_n}{s_n^2} = \frac{m_0}{s_0^2} + \frac{n\bar{X}}{\sigma^2}.$$

16.4 Bayes 决策与常见损失

观察 x 后, Bayes 动作最小化后验风险

$$\rho(a | x) = \mathbb{E}[L(\theta, a) | x].$$

平方损失下最优动作是后验均值；绝对损失下是后验中位数；0–1 邻域损失的极限对应后验众数。MAP 依赖参数化：密度众数在非线性变换下不保持，而后验均值也会随损失与参数目标改变。因此必须先声明损失。

定理 16.2 (三种常见损失的逐点 Bayes 动作). 给定观测 x , 设后验分布具有有限一阶矩。则平方损失下任一 Bayes 动作为 $\mathbb{E}(\theta | x)$, 绝对损失下任一后验中位数为 Bayes 动作。若参数空间离散, 则 0-1 损失下任一后验概率最大点为 Bayes 动作。

完整证明. 记 $m = \mathbb{E}(\theta | x)$ 。平方损失的后验风险满足

$$\mathbb{E}[(\theta - a)^2 | x] = \mathbb{E}[(\theta - m)^2 | x] + (a - m)^2,$$

交叉项因 $\mathbb{E}(\theta - m | x) = 0$ 消失, 故且仅当 $a = m$ 时最小。

对绝对损失令 $g(a) = \mathbb{E}(|\theta - a| | x)$ 。其左右导数分别为

$$g'_-(a) = \mathbb{P}(\theta < a | x) - \mathbb{P}(\theta \geq a | x), \quad g'_+(a) = \mathbb{P}(\theta \leq a | x) - \mathbb{P}(\theta > a | x).$$

凸函数在 a 处最小时需且只需 $g'_-(a) \leq 0 \leq g'_+(a)$, 这正等价于 $\mathbb{P}(\theta \leq a | x) \geq 1/2$ 且 $\mathbb{P}(\theta \geq a | x) \geq 1/2$, 即 a 为后验中位数。

离散参数下, 0-1 损失的后验风险为

$$\mathbb{E}[\mathbf{1}_{\{\theta \neq a\}} | x] = 1 - \mathbb{P}(\theta = a | x),$$

所以选择后验概率最大点。连续参数对单点的 0-1 损失几乎处处恒为 1, 必须改用邻域损失并声明基准测度, 不能把密度众数无条件当作同一定理。□

16.5 可信区间与置信区间

Bayesian $1 - \alpha$ 可信集满足

$$\mathbb{P}(\theta \in C(x) | x) = 1 - \alpha$$

, 概率来自后验。频率置信区间则在重复样本下覆盖真参数

$$\mathbb{P}_\theta(\theta \in C(X)) = 1 - \alpha.$$

数值上两者在某些共轭/平坦先验模型中可重合, 但解释仍不同。

最高后验密度区间 (HPD) 在给定后验概率下选择密度不低于区间外点的区域; 多峰后验的 HPD 集可能不连通。

16.6 非信息先验与 proper posterior

不正规先验如位置参数的 $\pi(\mu) \propto 1$ 不是概率分布。若后验归一化常数有限, 仍可得到 proper posterior; 但边缘似然中的任意比例常数使 Bayes factor 不可定义。每次使用不正规先验都要检查后验 proper, 而不能把“常数抵消”当作普遍规则。

16.7 Bayes factor 与先验敏感性

模型 M_0, M_1 的 posterior odds 为

$$\frac{\mathbb{P}(M_1 | x)}{\mathbb{P}(M_0 | x)} = \underbrace{\frac{m_1(x)}{m_0(x)}}_{\text{Bayes factor}} \frac{\mathbb{P}(M_1)}{\mathbb{P}(M_0)}.$$

边缘似然对先验在整个参数空间的质量分配敏感，自动产生复杂度惩罚；但若先验尺度极宽，可能显著压低复杂模型证据。报告 Bayes factor 时应做合理的先验敏感性分析。

答题方法

Bayes 题依次写似然、先验、核函数、归一化常数、后验目标和损失。共轭只简化代数，不替代参数化说明；MAP、后验均值和 posterior predictive 是三个不同输出。

自测

1. 后验预测为什么不是一般地把参数换成后验均值？
2. 推导三个共轭模型的后验参数。
3. 平方、绝对与 0-1 型损失分别对应什么 Bayes 动作？
4. 不正规先验何时仍可使用，何时 Bayes factor 失去定义？
5. 可信区间与置信区间的概率针对什么？

第十七章 渐近统计：一致性、正态性与非正则极限

学习目标

本章把第 4、7 章的收敛与 CLT 用到估计量：统一大数律给一致性，score/Taylor 展开给正则渐近正态，经验分位数和端点模型展示 Delta 方法与非正则速率。

17.1 一致性与 MLE 渐近理论

估计量 $\hat{\theta}_n$ 相合是 $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$ 。M-estimator 的标准证明：归一化目标 $M_n(\theta)$ 一致收敛到 $M(\theta)$ ； M 在真参数唯一最大；参数空间或远离真值区域有紧性/分离；于是 argmax 连续性给相合。

定理 17.1 (MLE 渐近正态). 设真参数 θ_0 是参数空间内点，模型可识别，MLE 相合；对数密度在邻域二次可微，可交换微分积分；score 有二阶矩，Fisher 信息 $I(\theta_0)$ 正定；观测 Hessian 满足一致 LLN，Taylor 余项可忽略。则

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

证明. score 方程在 θ_0 展开：

$$0 = U_n(\hat{\theta}) = U_n(\theta_0) + H_n(\tilde{\theta})(\hat{\theta} - \theta_0).$$

于是

$$\sqrt{n}(\hat{\theta} - \theta_0) = [-n^{-1}H_n(\tilde{\theta})]^{-1}n^{-1/2}U_n(\theta_0).$$

由相合性与 Hessian LLN，第一因子依概率趋于 $I(\theta_0)^{-1}$ ；score CLT 给第二因子趋于 $N(0, I(\theta_0))$ ；Slutsky 完成证明。□

渐近有效性指极限协方差达到信息下界，不能与有限样本 CRLB 等号混为一谈。U 统计量的 Hoeffding 分解和 bootstrap 不属于当前考试主干，只要求掌握定义、基本用途与标准结论的适用条件。

17.2 M-estimation 的一致性模板

设 $\hat{\theta}_n$ 近似最小化经验准则 $M_n(\theta) = n^{-1} \sum_{i=1}^n m_\theta(X_i)$, 总体准则 $M(\theta) = \mathbb{E}m_\theta(X)$ 在 θ_0 唯一最小。若

$$\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \xrightarrow{\mathbb{P}} 0$$

且总体准则在 θ_0 外有正分离, 则 argmin 连续性给 $\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0$ 。证明时不能只写逐点大数律; 关键是统一收敛和唯一识别。

17.3 score 展开与渐近正态

对内点正则真参数, MLE 满足

$$0 = \ell'_n(\hat{\theta}_n) = \ell'_n(\theta_0) + \ell''_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta_0).$$

若 $n^{-1/2} \ell'_n(\theta_0) \xrightarrow{d} N(0, I(\theta_0))$, 且 $-n^{-1} \ell''_n(\tilde{\theta}_n) \xrightarrow{\mathbb{P}} I(\theta_0)$, 则

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{d} N(0, I(\theta_0)^{-1}).$$

这里一致性保证 $\tilde{\theta}_n \rightarrow \theta_0$, 信息非奇异保证可逆; 参数依赖支集或边界真参数会破坏该链。

17.4 样本分位数与非正则速率

若 F 在 $q_p = F^{-1}(p)$ 处有连续正密度 $f(q_p) > 0$, 经验分位数满足

$$\sqrt{n}(\hat{q}_p - q_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{f(q_p)^2}\right).$$

可由经验分布 CLT 与逆映射 Delta 方法得到。相反, Uniform(0, \theta) 端点 MLE $X_{(n)}$ 具有 n 而非 $\sqrt{n}$ 速率:

$$n(\theta - X_{(n)}) \xrightarrow{d} \text{Exp}(1/\theta).$$

因此“所有 MLE 都渐近正态”是错误命题。

17.5 典型题型与易错点

例题 17.1 (正态均值单侧 UMP). $X_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, σ 已知, 检验 $H_0: \mu \geq \mu_0$ 对 $H_1: \mu < \mu_0$ 。族对 $\bar{X}$ 有 MLR, 故在小值拒绝: $\bar{X} < \mu_0 + z_\alpha \sigma / \sqrt{n}$ 。size 的上确界在边界 μ_0 取得。

易错点

size 是原假设上的功效上确界, level 是允许的上界。NP 引理只直接解决简单对简单。GLR 的两种记号方向相反, 本书始终用小 Λ 拒绝。Wilks、MLE 渐近正态和 Fisher 信息公式均需正则条件。p 值不是原假设为真的概率。后验均值只在平方损失下是 Bayes action。

复习检查

能够完整证明 NP；由 MLR 写单侧 UMP 拒绝域并校准边界；区分随机化、size 和 level；写 GLRT 并确定自由度；从 score Taylor 展开证明 MLE 渐近正态；比较 LR/Wald/score；计算常见共轭后验、预测和不同损失下 Bayes action；使用等风险 Bayes rule 证明 minimax。

答题方法

渐近题固定写出标准化速率、中心、极限分布和正则条件。先证明一致性再在随机中间点展开；若支集依赖参数、真值在边界或信息奇异，应主动寻找极值/Poisson/混合极限。

自测

1. 用统一收敛和识别证明一个 M-estimator 的一致性。
2. 从 score 方程推导一维 MLE 渐近正态并标出每个 Slutsky 步骤。
3. 推导样本中位数的渐近方差。
4. 证明 Uniform 端点 MLE 具有 n 速率并求极限。
5. 给出 Delta 方法一阶导数为零时需要二阶展开的例子。

本节结论

渐近统计不是套用 $\sqrt{n}$ 正态模板；速率和极限由局部识别、支集、导数阶数与随机波动共同决定。

17.6 估计量一致性的三条路线

直接概率界。 若 $\mathbb{E}T_n = \theta + o(1)$ 且 $\text{Var}(T_n) \rightarrow 0$ ，Chebyshev 给 $T_n \xrightarrow{\mathbb{P}} \theta$ 。

连续映射。 若 $T_n = g(\bar{X}_n)$ ，由大数律 $\bar{X}_n \xrightarrow{\mathbb{P}} \mu$ ，且 g 在 μ 连续，则

$$T_n \xrightarrow{\mathbb{P}} g(\mu).$$

argmax 定理。 若随机准则 $M_n(\theta)$ 在紧参数集上一致收敛到 $M(\theta)$ ，而 M 在 θ_0 有唯一最大值，且 $\hat{\theta}_n$ 近似最大化 M_n ，则

$$\hat{\theta}_n \xrightarrow{\mathbb{P}} \theta_0.$$

“逐点收敛”通常不够，因为随机极大点随 n 移动；需要一致控制或可替代的紧性/局部化论证。

17.7 M-estimator 的渐近线性化

设估计量解

$$\frac{1}{n} \sum_{i=1}^n \psi(X_i, \hat{\theta}) = 0.$$

在真值 θ_0 附近展开：

$$0 = \frac{1}{n} \sum_i \psi(X_i, \theta_0) + A_n(\tilde{\theta})(\hat{\theta} - \theta_0),$$

其中

$$A_n(\theta) = \frac{1}{n} \sum_i \partial_{\theta} \psi(X_i, \theta).$$

若

$$A_n(\tilde{\theta}) \xrightarrow{\mathbb{P}} A, \quad \frac{1}{\sqrt{n}} \sum_i \psi(X_i, \theta_0) \xrightarrow{d} N(0, B),$$

且 A 非奇异，则

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, A^{-1}BA^{-\top}).$$

这是 sandwich covariance；MLE 正则模型中 information identity 使 $A = -I, B = I$ ，化为 I^{-1} 。

17.8 渐近等价、Slutsky 与 studentization

若

$$\sqrt{n}(T_n - \theta) \xrightarrow{d} N(0, \sigma^2)$$

且 $\hat{\sigma} \xrightarrow{\mathbb{P}} \sigma > 0$ ，Slutsky 给

$$\frac{\sqrt{n}(T_n - \theta)}{\hat{\sigma}} \xrightarrow{d} N(0, 1).$$

这称 studentization。估计标准误必须一致；仅把未知 σ 换成任意样本量依赖量不够。

若

$$T_n - S_n = o_{\mathbb{P}}(n^{-1/2}),$$

则两者具有同一 $\sqrt{n}$ 极限。渐近线性表示

$$T_n - \theta = \frac{1}{n} \sum_{i=1}^n \varphi(X_i) + o_{\mathbb{P}}(n^{-1/2})$$

直接给 CLT， φ 称影响函数。

17.9 样本分位数的 Bahadur 线性化

设 $q_p = F^{-1}(p)$ ，且 $f(q_p) > 0$ 。经验分布 F_n ，样本分位数 $\hat{q}_p$ 满足

$$F_n(\hat{q}_p) \approx p.$$

在 q_p 附近线性化：

$$0 \approx F_n(q_p) - p + f(q_p)(\hat{q}_p - q_p).$$

因此

$$\sqrt{n}(\hat{q}_p - q_p) \xrightarrow{d} N\left(0, \frac{p(1-p)}{f(q_p)^2}\right).$$

若 $f(q_p) = 0$ 或分布在分位数处有跳跃， $\sqrt{n}$ 正态率可能失效。

17.10 Bootstrap 的目标与失效

非参数 bootstrap 从经验分布 F_n 重抽样，计算 T_n^* ，用条件分布

$$\mathcal{L}(\sqrt{n}(T_n^* - T_n) \mid X_{1:n})$$

近似 $\mathcal{L}(\sqrt{n}(T_n - \theta))$ 。对平滑渐近线性统计量通常有效。

对边界参数、最大值、模型选择后统计量或非标准速率问题，普通 bootstrap 可能失效。例如 $X_i \sim \text{Unif}(0, \theta)$ 的最大值以 n 而非 $\sqrt{n}$ 缩放，经验分布的端点固定在样本最大值，无法正确复制真端点外的尾部。

17.11 局部备择与检验功效

固定备择下许多一致检验功效趋一，无法比较精细差异。取局部备择

$$\theta_n = \theta_0 + \frac{h}{\sqrt{n}},$$

正则模型中 log-likelihood ratio 常展开为

$$\log \frac{L_n(\theta_n)}{L_n(\theta_0)} = h^\top \Delta_n - \frac{1}{2} h^\top I(\theta_0) h + o_{\mathbb{P}}(1),$$

其中 $\Delta_n \xrightarrow{d} N(0, I)$ 。这解释 Wald、score、LRT 的局部一阶等价与非中心卡方功效。

答题方法

渐近推断固定六步：一致性；主方程；一阶展开；随机项 CLT；导数项 LLN；Slutsky/连续映射。非正则模型先识别边界、支撑含参、不可识别或导数为零，再决定尺度。

自测

1. argmax 一致性为什么需要一致而非逐点收敛？
2. 推导 M-estimator 的 sandwich covariance。
3. studentization 中标准误需要满足什么条件？
4. 推导样本分位数的渐近方差。
5. 给出普通 bootstrap 失效的一个非正则例子。

第十八章 线性模型：投影、可估性与 t/F 推断

18.1 考纲版本与模型设定

本章是 2027 春起新版考纲新增模块；旧版概率、随机过程与统计主干不因它而缩减。值得注意的是，2024 秋 Q8 已直接考查 Gauss–Markov 与信息矩阵，说明“新版新增”不等于“过去从未命题”。本章采用固定设计矩阵的标准线性模型框架，区分满列秩、秩亏、可估函数与正态误差下的精确推断。

线性模型写作

$$Y = X\beta + \varepsilon, \quad \mathbb{E}\varepsilon = 0, \quad \text{Cov}(\varepsilon) = \sigma^2 I_n,$$

其中 $Y \in \mathbb{R}^n$ 、 $X \in \mathbb{R}^{n \times p}$ 为固定设计、 $\beta \in \mathbb{R}^p$ 。Gauss–Markov 假设不要求正态；正态性只在精确 $t/F/\chi^2$ 推断时加入。

满列秩 $\text{rank}(X) = p$ 时参数可唯一识别。秩亏时 β 不唯一，但拟合值 $X\hat{\beta}$ 仍是 Y 到列空间 $\mathcal{C}(X)$ 的唯一正交投影；线性函数 $c^T\beta$ 可估当且仅当 $c \in \mathcal{C}(X^T)$ 。

18.2 最小二乘、正规方程与投影

最小化 $Q(\beta) = \|Y - X\beta\|_2^2$ ，一阶条件给正规方程

$$X^T X \hat{\beta} = X^T Y.$$

满列秩时

$$\hat{\beta} = (X^T X)^{-1} X^T Y, \quad H = X(X^T X)^{-1} X^T, \quad \hat{Y} = HY, \quad e = (I - H)Y.$$

H 对称幂等，是到 $\mathcal{C}(X)$ 的正交投影； $M = I - H$ 也是对称幂等，且 $X^T e = 0$ 。杠杆值 h_{ii} 满足 $0 \leq h_{ii} \leq 1$ 、 $\sum h_{ii} = p$ 。

命题 18.1 (OLS 均值与协方差). 满列秩 Gauss–Markov 模型下， $\mathbb{E}\hat{\beta} = \beta$ 、 $\text{Cov}(\hat{\beta}) = \sigma^2(X^T X)^{-1}$ ， $\mathbb{E}e = 0$ 、 $\text{Cov}(e) = \sigma^2 M$ 。

证明. 把 $Y = X\beta + \varepsilon$ 代入 $\hat{\beta} = \beta + (X^T X)^{-1} X^T \varepsilon$ ，取期望与协方差。残差为 $e = M\varepsilon$ ，用 $M^2 = M = M^T$ 。□

18.3 Gauss–Markov 定理与 BLUE

定理 18.2 (Gauss–Markov). 在固定设计、 $\text{rank}(X) = p$ 、 $\mathbb{E}\varepsilon = 0$ 、 $\text{Cov}(\varepsilon) = \sigma^2 I$ 下, 对任意 $c \in \mathbb{R}^p$, $c^T \hat{\beta}$ 是 $c^T \beta$ 的最佳线性无偏估计 (BLUE): 在所有形如 $a^T Y$ 且对每个 β 无偏的估计中方差最小。

证明. 无偏要求 $a^T X = c^T$ 。OLS 对应 $a_0 = X(X^T X)^{-1}c$, 满足 $a_0^T X = c^T$ 。任意其他 a 写为 $a = a_0 + d$, 则 $d^T X = 0$ 。因 $a_0 \in \mathcal{C}(X)$, 有 $a_0^T d = 0$, 故

$$\text{Var}(a^T Y) = \sigma^2 \|a\|^2 = \sigma^2 \|a_0\|^2 + \sigma^2 \|d\|^2 \geq \text{Var}(a_0^T Y).$$

这只比较线性无偏估计, 不声称优于所有可能的有偏或非线性估计; 也不需要正态性。□

若 $\text{Cov}(\varepsilon) = \sigma^2 V$ 且 V 已知正定, 广义最小二乘为 $(X^T V^{-1} X)^{-1} X^T V^{-1} Y$ 。这不属于最基础的同方差模型, 但用于理解协方差结构条件为何关键。

18.4 正态误差下的分布理论

加入 $\varepsilon \sim N(0, \sigma^2 I)$ 后,

$$\hat{\beta} \sim N(\beta, \sigma^2 (X^T X)^{-1}), \quad \frac{SSE}{\sigma^2} = \frac{Y^T M Y}{\sigma^2} \sim \chi_{n-p}^2, \quad \hat{\beta} \perp SSE.$$

独立性来自 $H\varepsilon$ 与 $M\varepsilon$ 联合高斯且协方差为零。无偏方差估计为 $s^2 = SSE/(n-p)$ 。

对单个可估对比 $c^T \beta$,

$$T = \frac{c^T \hat{\beta} - c^T \beta}{s \sqrt{c^T (X^T X)^{-1} c}} \sim t_{n-p}.$$

置信区间由此直接得到。

18.5 一般线性假设、F 检验与置信椭圆

检验 $H_0: C\beta = d$, 其中 C 为 $q \times p$ 满行秩矩阵, 统计量

$$F = \frac{(C\hat{\beta} - d)^T [C(X^T X)^{-1} C^T]^{-1} (C\hat{\beta} - d)/q}{s^2} \sim F_{q, n-p} \quad (H_0).$$

等价地, 对于嵌套模型, $F = ((SSE_R - SSE_F)/q)/(SSE_F/(n-p))$ 。单个约束时 $F = T^2$ 。

参数置信椭圆可写为

$$\left\{ \beta : (\hat{\beta} - \beta)^T X^T X (\hat{\beta} - \beta) \leq p s^2 F_{p, n-p, 1-\alpha} \right\}.$$

该公式要求满列秩、正态同方差误差且 $n > p$ 。左端二次型除以 σ^2 服从 χ_p^2 , 并与 s^2 独立, 因而与 $p s^2$ 之比服从 $F_{p, n-p}$ 。

18.6 ANOVA、回归与预测

含截距模型下 $SST = \sum(Y_i - \bar{Y})^2 = SSR + SSE$, 自由度为 $n - 1 = (p - 1) + (n - p)$ 。整体回归检验比较 $F = [SSR/(p - 1)]/[SSE/(n - p)]$ 。没有截距时这一中心化分解及通常 R^2 解释可能失效。

简单回归 $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ 中

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}, \quad \hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{x}, \quad \text{Var}(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}.$$

在 x_0 处均值响应与新观测预测的标准误分别为

$$s\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \quad s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}.$$

后者多出的“1”来自新观测自身的噪声方差。

方差分析是线性模型的特殊情形。单因素 ANOVA 的原假设为各处理均值相等，要求独立、正态、同方差（大样本近似另论）。组间与组内平方和的正交分解产生 F 比。

18.7 秩亏、可估函数与广义逆

若 $\text{rank}(X) = r < p$, 正规方程有多解。任何广义逆解给相同拟合值 $\hat{Y} = P_{\mathcal{C}(X)}Y$, 但 $\hat{\beta}$ 本身依约束/广义逆而变。 $c^T\beta$ 可估的充要条件是 $c = X^T a$ 对某个 a 成立；此时 $c^T\hat{\beta}$ 与解的选择无关。

定理 18.3 (可估函数与秩亏拟合值). 在线性模型 $\mathbb{E}Y = X\beta$ 中, 线性函数 $c^T\beta$ 存在线性无偏估计当且仅当 $c \in \mathcal{C}(X^T)$ 。此外, 所有最小二乘解给出同一个拟合值 $P_{\mathcal{C}(X)}Y$; 若 $c^T\beta$ 可估, 则 $c^T\hat{\beta}$ 也与所选最小二乘解无关。

完整证明. 若 a^TY 对每个 β 都无偏, 则

$$\mathbb{E}(a^TY) = a^TX\beta = c^T\beta \quad \forall \beta,$$

所以 $X^T a = c$, 即 $c \in \mathcal{C}(X^T)$; 反向取满足 $X^T a = c$ 的 a 即得线性无偏估计。

最小二乘是在子空间 $\mathcal{C}(X)$ 中寻找离 Y 最近的向量。正交投影唯一, 故虽然参数解可写成 $\hat{\beta}_0 + v, v \in \mathcal{N}(X)$, 拟合值始终为 $X\hat{\beta} = P_{\mathcal{C}(X)}Y$ 。若 $c = X^T a$, 则对任意 $v \in \mathcal{N}(X), c^T v = a^T Xv = 0$, 所以 $c^T\hat{\beta}$ 不随解改变。反之 $c \notin \mathcal{C}(X^T) = \mathcal{N}(X)^\perp$ 时存在 $v \in \mathcal{N}(X)$ 使 $c^T v \neq 0$, 参数 β 与 $\beta + v$ 产生同一分布均值却有不同 $c^T\beta$, 故该函数不可识别、更不可能无偏估计。□

章末补充用薄 SVD 给出 Moore–Penrose 解、拟合值不变性与可估函数的完整线性代数证明。考试若给具体因子模型, 应优先使用识别约束（如效应和为零）或直接转写为可估对比。一般秩亏 F 检验和同时区间只需掌握标准思路。

多重共线性意味着 $X^T X$ 接近奇异, 导致估计方差大和符号不稳定, 但 OLS 在精确满秩下仍无偏。方差膨胀因子、AIC/BIC/交叉验证只作概念比较; 复习重点是识别病态、解释方差放大和区分预测问题与参数解释问题。

例题 18.1 (小矩阵 OLS). 令 $X = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}$, $Y = (1, 2, 2)^T$. 则 $X^T X = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}$, $(X^T X)^{-1} = \frac{1}{6} \begin{pmatrix} 5 & -3 \\ -3 & 3 \end{pmatrix}$, $X^T Y = (5, 6)^T$, 故 $\hat{\beta} = (7/6, 1/2)^T$. 拟合值为 $(7/6, 5/3, 13/6)^T$, 残差和为零且与设计两列正交。考试必须检查维数和 $X^T e = 0$.

易错点

Gauss–Markov 不需要正态；BLUE 的“最佳”只在同一目标的线性无偏类中比较。 s^2 分母是残差自由度 $n - p$ 。均值响应区间与新观测预测区间相差一个噪声方差。秩亏时参数估计不唯一，但拟合值唯一；只有可估函数具有不依赖约束的推断意义。

复习检查

能够从最小二乘推导正规方程；证明 H 对称幂等；计算 OLS 均值协方差；完整证明 Gauss–Markov；推导正态模型下 t/F ；填 ANOVA 表；区分均值区间与预测区间；判断秩亏模型中一个线性函数是否可估。

18.8 投影矩阵与最小二乘几何

线性模型

$$y = X\beta + \varepsilon, \quad \mathbb{E}\varepsilon = 0, \quad \text{Cov}(\varepsilon) = \sigma^2 I.$$

若 X 满列秩，最小二乘

$$\hat{\beta} = (X^T X)^{-1} X^T y.$$

拟合值与残差为

$$\hat{y} = Hy, \quad e = (I - H)y,$$

$$H = X(X^T X)^{-1} X^T.$$

$H = H^T = H^2$ 是投影到 $\mathcal{C}(X)$ 的正交投影，故

$$X^T e = 0, \quad \|y - X\hat{\beta}\|^2 = \|e\|^2 + \|X\hat{\beta} - X\beta\|^2.$$

18.9 Gauss–Markov 定理的矩阵证明

任一线性无偏估计 Ay 估计 β 需

$$AX = I.$$

写

$$A = (X^T X)^{-1} X^T + D, \quad DX = 0.$$

则

$$\begin{aligned}\text{Cov}(Ay) &= \sigma^2 AA^\top \\ &= \sigma^2(X^\top X)^{-1} + \sigma^2 DD^\top,\end{aligned}$$

交叉项因 $DX = 0$ 消失。因此

$$\text{Cov}(Ay) - \text{Cov}(\hat{\beta}) = \sigma^2 DD^\top \succeq 0.$$

$\hat{\beta}$ 是所有线性无偏估计中的 BLUE。该定理不要求误差正态；正态性用于精确 t/F 分布。

18.10 正态模型下的独立分解

若 $\varepsilon \sim N(0, \sigma^2 I)$, 则

$$\hat{\beta} \sim N\left(\beta, \sigma^2(X^\top X)^{-1}\right).$$

残差平方和

$$\text{SSE} = \|e\|^2 = y^\top (I - H)y.$$

$(I - H)$ 是秩 $n - p$ 的幂等投影, 所以

$$\frac{\text{SSE}}{\sigma^2} \sim \chi_{n-p}^2.$$

而 Hy 与 $(I - H)y$ 协方差为

$$\sigma^2 H(I - H) = 0,$$

联合正态故独立。因此

$$s^2 = \frac{\text{SSE}}{n - p}$$

独立于 $\hat{\beta}$ 。

对对比 $c^\top \beta$,

$$\frac{c^\top \hat{\beta} - c^\top \beta}{s \sqrt{c^\top (X^\top X)^{-1} c}} \sim t_{n-p}.$$

18.11 一般线性假设与 F 检验

检验

$$H_0 : R\beta = r, \quad \text{rank}(R) = q.$$

Wald 型平方距离

$$W = (R\hat{\beta} - r)^\top \left[R(X^\top X)^{-1} R^\top \right]^{-1} (R\hat{\beta} - r).$$

在 H_0 下

$$\frac{W/q}{s^2} \sim F_{q, n-p}.$$

等价地，以受限模型和完全模型残差平方和比较：

$$F = \frac{(\text{SSE}_R - \text{SSE}_F)/q}{\text{SSE}_F/(n-p)}.$$

分子测量加入被检验方向后减少的残差，分母估计噪声方差。

18.12 ANOVA 与平方和分解

含截距模型中

$$\text{SST} = \sum_i (y_i - \bar{y})^2 = \text{SSR} + \text{SSE},$$

其中

$$\text{SSR} = \sum_i (\hat{y}_i - \bar{y})^2.$$

分解来自中心化响应在模型子空间与残差子空间上的正交投影。若模型不含截距，常规定义下交叉项不一定消失，不能机械使用 $R^2 = 1 - \text{SSE} / \text{SST}$ 的全部性质。

18.13 秩亏、可估函数与广义逆

若 X 不满列秩， β 不唯一可识别。线性函数 $c^\top \beta$ 可估当且仅当

$$c \in \mathcal{C}(X^\top).$$

这保证所有满足同一均值 $X\beta$ 的参数给相同 $c^\top \beta$ 。任取广义逆得到一个最小二乘解，但只有可估函数的估计与广义逆选择无关。

18.14 异方差与广义最小二乘

若

$$\text{Cov}(\varepsilon) = \Sigma$$

已知且正定，白化

$$\Sigma^{-1/2}y = \Sigma^{-1/2}X\beta + \Sigma^{-1/2}\varepsilon$$

给 GLS

$$\hat{\beta}_{\text{GLS}} = (X^\top \Sigma^{-1} X)^{-1} X^\top \Sigma^{-1} y.$$

若错误地仍用 OLS，在线性无偏条件下 OLS 仍可能无偏，但不再 BLUE，常规同方差标准误也错误。实践可用 heteroskedasticity-consistent sandwich covariance，但小样本精确 t/F 结论不再自动成立。

18.15 预测区间与均值响应区间

在新协变量 x_0 处, 均值响应 $x_0^\top \beta$ 的估计方差为

$$\sigma^2 x_0^\top (X^\top X)^{-1} x_0.$$

新个体

$$Y_0 = x_0^\top \beta + \varepsilon_0$$

还含独立噪声, 预测误差方差为

$$\sigma^2 \left[1 + x_0^\top (X^\top X)^{-1} x_0 \right].$$

所以预测区间严格宽于同一点的均值响应置信区间。

18.16 同时推断、拟合不足与诊断边界

若同时报告 m 个预先指定的线性组合, Bonferroni 法把每个双侧区间的尾概率改为 $\alpha/(2m)$, 由 union bound 保证联合覆盖率至少 $1 - \alpha$, 不要求各区间独立。Scheffé 法同时覆盖全部 $c^\top \beta$:

$$\frac{|c^\top (\hat{\beta} - \beta)|}{s \sqrt{c^\top (X^\top X)^{-1} c}} \leq \sqrt{p F_{p, n-p; 1-\alpha}} \quad \text{对所有 } c \neq 0.$$

它适合事后选择的任意对比, 但对少量预先指定对比通常更宽。

若同一设计点 x_i 有 m_i 次独立重复观测, 令组均值为 $\bar{Y}_i$ 。残差平方和有正交分解

$$\text{SSE} = \underbrace{\sum_i \sum_{j=1}^{m_i} (Y_{ij} - \bar{Y}_i)^2}_{\text{SS}_{\text{PE}}} + \underbrace{\sum_i m_i (\bar{Y}_i - \hat{Y}_i)^2}_{\text{SS}_{\text{LOF}}}.$$

若有 g 个不同设计点、回归模型秩为 p , 自由度分别为 $n - g$ 与 $g - p$ 。正态同方差且模型正确时,

$$F_{\text{LOF}} = \frac{\text{SS}_{\text{LOF}}/(g - p)}{\text{SS}_{\text{PE}}/(n - g)} \sim F_{g-p, n-g}.$$

没有重复观测就不能把普通 SSE 再精确拆成 pure error 与 lack of fit。

残差图、杠杆值和影响度是模型诊断工具, 不是上述精确 F 检验的替代。多重共线性放大估计方差但不自动造成有偏; 调整 R^2 、AIC、BIC 与交叉验证分别含不同的复杂度惩罚或预测目标, 不能用同一数值阈值互换。

答题方法

线性模型题先写设计矩阵的秩; 用投影分解求估计与自由度; 正态时再导出卡方、 t 、 F ; 秩亏时改问可估函数; 异方差时更换协方差与标准误。

自测

1. 证明帽子矩阵是正交投影。
2. 完成 Gauss–Markov 的协方差差分证明。
3. SSE 为什么有 $n - p$ 个自由度，且与 $\hat{\beta}$ 独立？
4. 一般线性假设的 F 统计量如何由受限/完全模型平方和得到？
5. 为什么个体预测区间比均值响应区间多一个 1？

第四部分

概率论与数理统计历年真题与完整解答

第十九章 2023 年秋季博士资格考试

易错点

除特别注明为官方答案的内容外,本部分解答均为复习用整理稿,不代表官方评分标准。原题若存在标准化、参数化或记号歧义,解答会先说明采用的解释。

考试范围提醒

本卷共 8 道大题、100 分,要求全部作答;以下按原题顺序完整收入。

19.1 第 1 题: 二维高斯、边缘与极坐标 (原卷 p.1, 10 分)

2023 秋 Q1

English original. Let (X, Y) have joint density

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

(a) Prove that f is a density; find the marginals and the condition for independence. (b) Put $R = \sqrt{X^2 + Y^2}$ and $\Phi \in [0, 2\pi)$ with $X = R \cos \Phi, Y = R \sin \Phi$. Determine the laws of (R, Φ) and Φ .

中文译述. 验证给定相关二维标准高斯密度, 求两边缘及独立条件; 再作极坐标变换, 求半径-角度联合分布和角度边缘分布。

完整解答

配方 $x^2 - 2\rho xy + y^2 = (x - \rho y)^2 + (1 - \rho^2)y^2$, 先对 x 积分得 $f_Y(y) = (2\pi)^{-1/2}e^{-y^2/2}$, 再积分为一; 对称地 $X \sim N(0, 1)$ 。联合高斯中独立等价于协方差为零, 或直接比较密度, 故当且仅当 $\rho = 0$ 。

极坐标 Jacobian 为 r , 且 $x^2 - 2\rho xy + y^2 = r^2(1 - \rho \sin 2\phi)$, 所以

$$f_{R,\Phi}(r, \phi) = \frac{r}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{r^2(1-\rho \sin 2\phi)}{2(1-\rho^2)}\right],$$

$r \geq 0, 0 \leq \phi < 2\pi$ 。用 $\int_0^\infty r e^{-ar^2/2} dr = 1/a$ 得

$$f_\Phi(\phi) = \frac{\sqrt{1-\rho^2}}{2\pi(1-\rho \sin 2\phi)}.$$

当 $\rho = 0$ 时 R 与 Φ 独立、 R 为 Rayleigh、角度均匀；一般情形不独立。易错点是漏 Jacobian 或误以为边缘标准正态便自动独立。

19.2 第 2 题：Laplace 增量的指数鞅（原卷 p.1, 10 分）

2023 秋 Q2

English original. Let X_n be i.i.d. with density $\frac{1}{2}e^{-|x|}$ and natural filtration $\mathcal{F}_n$. For $M_n(\theta) = \exp\{\theta \sum_{k=1}^n X_k - A_n(\theta)\}$, determine $A_n(\theta)$ so that $(M_n(\theta))$ is a martingale for every $|\theta| < 1$.

中文译述. 对 i.i.d. Laplace 增量，选择规范化项使指数过程成为鞅。

完整解答

对 $|\theta| < 1$,

$$\mathbb{E}e^{\theta X_1} = \frac{1}{2} \int_{-\infty}^0 e^{(1+\theta)x} dx + \frac{1}{2} \int_0^{\infty} e^{-(1-\theta)x} dx = \frac{1}{1-\theta^2}.$$

条件期望为

$$\mathbb{E}[M_n | \mathcal{F}_{n-1}] = M_{n-1} \exp\{A_{n-1} - A_n\}(1-\theta^2)^{-1}.$$

故需 $A_n - A_{n-1} = -\log(1-\theta^2)$, 即 $A_n(\theta) = A_0(\theta) - n \log(1-\theta^2)$ 。标准取 $A_0 = 0$, 使 $M_0 = 1$ 。必须说明 $|\theta| < 1$ 保证可积。

19.3 第 3 题：随机重抛硬币的 Markov 链（原卷 p.1, 10 分）

2023 秋 Q3

English original. There are N coins. At each round choose one uniformly and flip it independently. Let X_n be the number of heads. (a) Prove it is a Markov chain and give transitions. (b) Prove a unique stationary distribution exists and identify it.

中文译述. 每步均匀选择一枚硬币独立重抛，以正面数为状态，证明 Markov 性、写转移并求唯一平稳分布。

完整解答

给定当前正面数 k , 选中反面并抛成正面的概率为 $(N-k)/(2N)$; 选中正面并抛成反面的概率为 $k/(2N)$; 其余概率为 $1/2$ 。故

$$P_{k,k+1} = \frac{N-k}{2N}, \quad P_{k,k-1} = \frac{k}{2N}, \quad P_{k,k} = \frac{1}{2}.$$

下一状态只依赖 k , 故 Markov。有限链在 $\{0, \dots, N\}$ 上不可约且有自环, 所以有唯一平稳分布且非周期。详细平衡给 $\pi_{k+1}/\pi_k = (N-k)/(k+1)$, 归一化得 $\pi_k = 2^{-N} \binom{N}{k}$, 即 $\text{Bin}(N, 1/2)$ 。另一解释是平稳时每枚硬币独立公平。

19.4 第 4 题：三骰等待时间最优策略（原卷 p.2, 10 分）

2023 秋 Q4

English original. Alice has three fair dice. Each round she may resample any subset and keep the other values. Let T be the first round displaying $(6, 6, 6)$. (a) Find the strategy minimizing T . (b) Compute $\mathbb{E}T$ under it.

中文译述. 每轮可保留任意骰子并重掷其余骰子，求首次出现三个 6 的最优策略及期望轮数。

完整解答

动态规划/单调耦合表明已经出现的 6 应保留，非 6 应重掷；任何重掷一个 6 都使下一轮“已收集 6 的数目”随机意义下降。令 T_i 为第 i 枚骰首次出现 6 的轮次，则最优策略下 $T = \max_i T_i$, T_i i.i.d. 几何分布且 $\mathbb{P}(T_i \geq n) = (5/6)^{n-1}$ 。尾和公式给

$$\mathbb{E}T = \sum_{n \geq 1} \{1 - [1 - (5/6)^{n-1}]^3\} = 18 - \frac{108}{11} + \frac{216}{91} = \frac{10566}{1001}.$$

可替代方法是按当前已有 0、1、2 个 6 建三状态递推。

19.5 第 5 题：正则树渗流（原卷 p.2, 15 分）

2023 秋 Q5

English original. On an infinite d -regular tree ($d \geq 3$), independently open each edge with probability p . Let $C(o)$ be the root cluster and $\theta(p) = \mathbb{P}(|C(o)| = \infty)$. (a) Find p_c separating zero and positive survival probability. (b) For $p < p_c$, prove $\mathbb{P}(|C(o)| \geq k) \leq Ce^{-ck}$.

原卷歧义. 原卷 (a) 两处印成 $\theta(p_c)$ ；按上下文应为 $\theta(p)$ 。本书不改写这一事实，只按标准问题求解。

中文译述. 在无限 d -正则树上作 Bernoulli 边渗流，求根簇无限的临界概率，并证明亚临界簇大小指数尾。

完整解答

离开根后，每个顶点有 $d-1$ 条向外边，故后代分布为 $\text{Bin}(d-1, p)$ ，均值 $m = (d-1)p$ ；根的 d 个分支只改变首代，不改变临界点。因此 $p_c = 1/(d-1)$ 。若 $p \leq p_c$ ，相应 Galton-Watson 过程灭绝 a.s.；若 $p > p_c$ ，单个分支存活概率正，故根簇无限概率正。

亚临界时总后代 W 的 pgf 满足 $G(s) = s(1-p+pG(s))^{d-1}$ 。因右端在 $(s, G) = (1, 1)$ 的 G 导数为 $m < 1$ ，隐函数定理给某个 $s_0 > 1$ 使 $G(s_0) < \infty$ 。根有有限个独立分支，所以 $\mathbb{E}s_0^{|C(o)|} < \infty$ 。Markov 不等式给 $\mathbb{P}(|C(o)| \geq k) \leq \mathbb{E}s_0^{|C(o)|} s_0^{-k} = Ce^{-k \log s_0}$ 。

19.6 第 6 题：高斯随机矩阵 CLT (原卷 pp.2–3, 15 分)

2023 秋 Q6

English original. Let X be $n \times n$ with i.i.d. $N(0,1)$ entries. (a) For deterministic orthogonal U, V , prove $UXV \stackrel{d}{=} X$. (b) For a unit u , find mean/variance of $Y_n = u^T X^T X u$ and prove its CLT. (c) For unit u, v , do the same for $Z_n = u^T X^T X u + \sqrt{n} v^T X v$.

中文译述. 证明标准高斯矩阵双侧正交不变；计算两个二次/线性组合的均值方差和中心极限定理。

完整解答

向量化后 $\text{vec}(UXV) = (V^T \otimes U) \text{vec}(X)$, Kronecker 矩阵正交, 故标准高斯向量分布不变。

取正交矩阵把 u 旋到 e_1 。由不变性 $Y_n \stackrel{d}{=} \sum_{i=1}^n X_{i1}^2 \sim \chi_n^2$, 故 $\mathbb{E}Y_n = n$, $\text{Var}(Y_n) = 2n$, 并且 $(Y_n - n)/\sqrt{2n} \xrightarrow{d} N(0,1)$ 。

线性项 $L = v^T X v$ 是方差 $\sum_{ij} v_i^2 v_j^2 = 1$ 的标准正态。高斯奇数矩为零, 故 $\text{Cov}(Y_n, L) = 0$, 从而 $\mathbb{E}Z_n = n$, $\text{Var}(Z_n) = 2n + n = 3n$ 。为证明 CLT, 可把 u 旋到 e_1 , 对每行的 $X_{i1}^2 - 1$ 与线性高斯项组成三角阵; Lindeberg 条件由高斯矩统一有界和最大系数 $O(n^{-1/2})$ 得到, 或用高斯积分分部。于是 $(Z_n - n)/\sqrt{3n} \xrightarrow{d} N(0,1)$ 。注意零协方差本身不说明有限 n 独立, 但足以计算方差。

19.7 第 7 题：正态模型的最强/UMP 检验 (原卷 p.3, 15 分)

2023 秋 Q7

English original. For i.i.d. $N(\mu, \sigma^2)$: (a) find a size- α most powerful test of (μ_0, σ_0^2) versus (μ_1, σ_1^2) when $\mu_0 < \mu_1$, $\sigma_0^2 = \sigma_1^2$; (b) when $\mu_0 = \mu_1$, $\sigma_0^2 < \sigma_1^2$; (c) with $\sigma^2 = 1$, find the UMP level- α test of $H_0: \mu \geq 1$ versus $H_1: \mu < 1$.

中文译述. 对正态样本分别处理同方差均值变化、同均值方差变化的简单假设, 以及均值单侧复合检验。

完整解答

(a) 似然比随 $\sum X_i$ 增大, 故拒绝域为 $\bar{X} > \mu_0 + z_{1-\alpha} \sigma_0 / \sqrt{n}$ 。(b) 似然比随 $Q = \sum (X_i - \mu_0)^2$ 增大, H_0 下 $Q/\sigma_0^2 \sim \chi_n^2$, 故拒绝当 $Q > \sigma_0^2 \chi_{n,1-\alpha}^2$ 。(c) 正态族对 $\bar{X}$ 有 MLR, 备择在左, 故在小值拒绝: $\bar{X} < 1 + z_\alpha / \sqrt{n}$ 。原假设上的拒绝概率在边界 $\mu = 1$ 最大, 故 size 为 α 且为 UMP。

19.8 第 8 题：Cauchy 极小充分与双样本 MLE (原卷 p.3, 15 分)

2023 秋 Q8

English original. For the Cauchy location density $f(x | \theta) = \{\pi[1 + (x - \theta)^2]\}^{-1}$: (a) find a minimal sufficient statistic. (b)(c) For two observations, put $\Delta = (x_1 - x_2)/2$; prove the MLE is unique if $|\Delta| \leq 1$ and nonunique if $|\Delta| > 1$, and find all maximizers.

中文译述. 求 Cauchy 位置族的极小充分统计量, 并完整分析两个观测时 MLE 的唯一性分

岔。

完整解答

(a) 似然比 $\prod_i [1 + (y_i - \theta)^2] / \prod_i [1 + (x_i - \theta)^2]$ 与 θ 无关, 当且仅当两组样本作为多重集合相同 (比较两边关于 θ 的多项式根/系数)。故次序统计量向量 $(X_{(1)}, \dots, X_{(n)})$ 极小充分; 一般没有低维压缩。

(b)(c) 令 $m = (x_1 + x_2)/2$ 、 $d = \Delta$ 、 $t = \theta - m$ 。最大化似然等价于最小化

$$D(t) = [1 + (t - d)^2][1 + (t + d)^2], \quad D'(t) = 4t(1 + t^2 - d^2).$$

若 $|d| \leq 1$, 唯一极小点为 $t = 0$, 故 $\hat{\theta} = m$ 。若 $|d| > 1$, $t = 0$ 变为局部极大点, 而两个全局极小点为 $t = \pm\sqrt{d^2 - 1}$, 故 MLE 集为

$$\left\{ m - \sqrt{\Delta^2 - 1}, m + \sqrt{\Delta^2 - 1} \right\}.$$

边界 $|d| = 1$ 虽二阶退化, $t = 0$ 仍唯一最小。

第二十章 2024 年春季博士资格考试

考试范围提醒

Q2 的打印题面与唯一完整解答见 第 5 章的指数极值真题 (第 34 页)。

考试范围提醒

本卷共 11 题, 考试时选做 8 题; 4 题为 15 分, 其余为 10 分。

20.1 第 1 题: 无限方差与非标准 CLT (原卷 p.1, 10 分)

2024 春 Q1

English original. Let X_i be i.i.d. with density $f(x) = |x|^{-3}$ for $|x| > 1$ and 0 otherwise; let $\varphi(t) = \mathbb{E}e^{itX_1}$ and $S_n = \sum_{i=1}^n X_i$. (a) Prove $\mathbb{E}X_1^2 = \infty$. (b) Given $\lim_{t \rightarrow 0} (1 - \varphi(t))/(t^2 \log |t|) = -1$, find a_n and prove $S_n/a_n \xrightarrow{d} N(0, 1)$.

中文译述. 对尾密度 $|x|^{-3}$, 证明二阶矩发散; 利用已给特征函数渐近寻找带对数修正的正态归一化。

完整解答

(a) $\mathbb{E}X^2 = 2 \int_1^\infty x^2 x^{-3} dx = \infty$. (b) 取 $a_n = \sqrt{n \log n}$. 固定 t ,

$$n\{1 - \varphi(t/a_n)\} \sim -n \frac{t^2}{a_n^2} \log \frac{|t|}{a_n} = t^2 \frac{\log a_n - \log |t|}{\log n} \rightarrow \frac{t^2}{2}.$$

又 $1 - \varphi(t/a_n) \rightarrow 0$, 故 $\varphi(t/a_n)^n \rightarrow e^{-t^2/2}$. Lévy 连续性定理给标准正态极限。常见错误是取 $\sqrt{n}$ 或漏掉 $\log a_n \sim \frac{1}{2} \log n$ 。

20.2 第 3 题: 乘积收敛与反例 (原卷 p.1, 10 分)

2024 春 Q3

English original. (a) If $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$, prove $X_n Y_n \xrightarrow{\mathbb{P}} XY$. (b) Is the analogous statement for L^1 convergence true? Prove or give a counterexample.

中文译述. 证明依概率收敛对乘法封闭, 并判断 L^1 版本。

完整解答

(a) $(X_n, Y_n) \xrightarrow{\mathbb{P}} (X, Y)$, 而乘法连续, 故由连续映射定理得结论。也可用 $|X_n Y_n - XY| \leq |X_n| |Y_n - Y| + |Y| |X_n - X|$, 并先证明依概率收敛序列有界于概率。

(b) 错。令同一事件 A_n 满足 $\mathbb{P}(A_n) = n^{-2}$, 取 $X_n = Y_n = n \mathbf{1}_{A_n}$, $X = Y = 0$ 。则 $\mathbb{E}|X_n| = 1/n \rightarrow 0$, 两者都 L^1 收敛; 但 $\mathbb{E}|X_n Y_n| = n^2 \mathbb{P}(A_n) = 1$, 乘积不 L^1 收敛。缺少的是更高矩或一致可积控制。

20.3 第 4 题: 偏置随机游动首达 (原卷 pp.1–2, 15 分)

2024 春 Q4

English original. Let $\xi_i \in \{-1, 1\}$ with $\mathbb{P}(\xi_i = 1) = p$, $S_n = \sum_{i=1}^n \xi_i$, and $T_k = \inf\{n \geq 1 : S_n = k\}$. Put $M(t) = \mathbb{E}e^{t\xi_1}$. (a) For $p \geq 1/2$, prove $e^{tk}\mathbb{E}[M(t)^{-T_k}] = 1$ for $t > 0$. (b) For $p < 1/2$, compute $\mathbb{P}(T_k < \infty)$. (c) For $p < 1/2$, find the law of $Y = 1 + \sup_{n \geq 0} S_n$.
中文译述. 使用指数鞅求首达时生成函数、逆漂移命中概率和最大高度分布。

完整解答

$L_n = e^{tS_n}/M(t)^n$ 是非负鞅。停于 $T_k \wedge n$, 并用 $S_{T_k} = k$ 、 $p \geq 1/2$ 时 $T_k < \infty$ a.s. 及非负鞅的一致控制取极限, 得到 $1 = e^{tk}\mathbb{E}M(t)^{-T_k}$ 。

若 $p < q = 1 - p$, 取 $r = q/p > 1$, 则 r^{S_n} 是鞅。停于首次到达 k 或 $-m$, 解得命中 k 概率; 令 $m \rightarrow \infty$, $\mathbb{P}(T_k < \infty) = (p/q)^k$ 。令 $a = p/q < 1$, 则 $\mathbb{P}(Y \geq j) = \mathbb{P}(\sup S_n \geq j - 1) = a^{j-1}$, 故

$$\mathbb{P}(Y = j) = (1 - a)a^{j-1}, \quad j = 1, 2, \dots,$$

即参数 $1 - a$ 的几何分布。停时取极限必须先用双边有界停止保证合法。

20.4 第 5 题: Markov 链的递推表示 (原卷 p.2, 10 分)

2024 春 Q5

English original. On countable S , let Z_n be i.i.d. in $(E, \mathcal{E})$, X_0 independent, and $X_n = f(X_{n-1}, Z_n)$. (a) Prove (X_n) is Markov and give its transition. (b) Prove every discrete time-homogeneous Markov chain on S has such a representation.
中文译述. 证明“状态加独立噪声”的递推产生 Markov 链, 并用逆变换表示任意可数状态齐次链。

完整解答

(a) 由 Z_n 与过去独立, $\mathbb{P}(X_n = y \mid X_0, \dots, X_{n-1}) = \mathbb{P}(f(X_{n-1}, Z_n) = y) = P(X_{n-1}, y)$, 其中 $P(x, A) = \mathbb{P}(f(x, Z_1) \in A)$ 。

(b) 枚举 $S = \{s_1, s_2, \dots\}$, 令 $Z_n \stackrel{\text{i.i.d.}}{\sim} U(0, 1)$ 。对每个 x 定义累积和 $F_x(j) = \sum_{\ell \leq j} P(x, s_\ell)$, 并令 $f(x, u) = s_j$ 当 $F_x(j-1) \leq u < F_x(j)$ 。则 $\mathbb{P}(f(x, Z) = s_j) = P(x, s_j)$, 按递推即可构造给定转移核的链。

20.5 第 6 题: Brownian 首次时 (原卷 p.2, 10 分)

2024 春 Q6

English original. For standard Brownian motion B_t , $T_a = \inf\{t > 0 : B_t \geq a\}$. (a) Prove T_a is a stopping time. (b) Compute $\mathbb{E}e^{-\lambda T_a}$ and prove $T_a < \infty$ a.s.; hint $e^{uB_t - u^2t/2}$.

中文译述. 证明 Brownian 首次正水平是停时, 求其 Laplace 变换并证明有限。

完整解答

连续路径给 $\{T_a \leq t\} = \{\sup_{s \leq t} B_s \geq a\}$; 上确界可只取有理时刻, 故属于 $\mathcal{F}_t$ 。对指数鞅先停在 $T_a \wedge T_{-b}$, 再令 $b \rightarrow \infty$, 或停在有界区域以保证 OST, 得到

$$\mathbb{E}e^{-\lambda T_a} = e^{-a\sqrt{2\lambda}}.$$

最后 $\lambda \downarrow 0$, 单调收敛给 $\mathbb{P}(T_a < \infty) = \lim_{\lambda \downarrow 0} \mathbb{E}e^{-\lambda T_a} = 1$ 。

20.6 第 7 题: 高斯均值估计的 minimax 性 (原卷 p.2, 15 分)

2024 春 Q7

English original. If $X_i \sim N(\mu_i, 1)$ independently and loss is $\sum_i (a_i - \mu_i)^2$, prove $\delta(X) = X$ is minimax.

中文译述. 在 n 维标准高斯位置模型和平方损失下证明恒等估计是 minimax。

完整解答

$\delta_0 = X$ 的风险恒为 n , 故 minimax 风险至多 n 。取先验 $\mu \sim N(0, \tau^2 I_n)$, Bayes 估计为 $\delta_\tau(X) = \frac{\tau^2}{1+\tau^2}X$, 其 Bayes 风险为 $n\tau^2/(1+\tau^2)$ 。任意规则的最大风险不小于任一先验 Bayes 风险, 令 $\tau \rightarrow \infty$ 得下界 n 。上下界相同, 故 X minimax。此题不声称在 $n \geq 3$ 时 X admissible; James-Stein 现象与 minimax 可并存。

20.7 第 8 题: 平移均匀分布的相合与渐近 (原卷 p.2, 10 分)

2024 春 Q8

English original. If $X_i \stackrel{\text{i.i.d.}}{\sim} U(\theta, \theta + 10)$, $\theta > 0$: (a) prove $\hat{\theta} = \min_i X_i$ is consistent; (b) find the asymptotic law of $\sqrt{12n}(\bar{X} - 5 - \theta)/\hat{\theta}$.

中文译述. 证明样本最小值估计平移参数相合, 并用 CLT/Slutsky 求给定标准化的极限。

完整解答

对 $0 < \varepsilon < 10$, $\mathbb{P}(\hat{\theta} > \theta + \varepsilon) = (1 - \varepsilon/10)^n \rightarrow 0$, 且 $\hat{\theta} \geq \theta$, 故相合。又 $\mathbb{E}X = \theta + 5$, $\text{Var } X = 100/12$, 所以 $\sqrt{12n}(\bar{X} - \theta - 5) \xrightarrow{d} N(0, 100)$ 。由 $\hat{\theta} \xrightarrow{\mathbb{P}} \theta$, Slutsky 给极限 $N(0, 100/\theta^2)$ 。

20.8 第 9 题：均值方差同参数的 MLE (原卷 p.2, 10 分)

2024 春 Q9

English original. If $Y_i = \log X_i \stackrel{\text{i.i.d.}}{\sim} N(\theta, \theta)$ with $\theta > 0$, derive the MLE and its asymptotic distribution.

中文译述. 正态均值和方差均为同一正参数，求 MLE 与渐近分布。

完整解答

忽略常数， $\ell(\theta) = -\frac{n}{2} \log \theta - \frac{1}{2\theta} \sum Y_i^2 + \sum Y_i - \frac{n}{2} \theta$ 。score 方程等价于 $\theta^2 + \theta - \overline{Y^2} = 0$ ，正根为

$$\hat{\theta} = \frac{-1 + \sqrt{1 + 4\overline{Y^2}}}{2}.$$

其为唯一最大点。单观测 Fisher 信息 $I(\theta) = (1 + 2\theta)/(2\theta^2)$ ，故正则 MLE 理论给

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N\left(0, \frac{2\theta^2}{1 + 2\theta}\right).$$

可替代地对 $\overline{Y^2}$ 用 CLT 和 Delta。

20.9 第 10 题：Pareto 均值的 UMP 检验 (原卷 p.2, 10 分)

2024 春 Q10

English original. For Pareto density $f(x | \theta) = 100^\theta \theta x^{-(1+\theta)}$, $x > 100$, $\theta > 1$, let $\mu = \mathbb{E}X$. Give the rejection region of a UMP level- α test of $H_0 : \mu = \mu_0$ versus $H_1 : \mu > \mu_0$.

中文译述. 将 Pareto 均值假设转换为形状参数单侧假设并写 UMP 拒绝域。

完整解答

$\mu = 100\theta/(\theta - 1)$ 随 θ 递减，令 $\theta_0 = \mu_0/(\mu_0 - 100)$ (需 $\mu_0 > 100$)，则备择 $\mu > \mu_0$ 等价于 $\theta < \theta_0$ 。令 $T = \sum \log(X_i/100)$ ，则 $T \sim \text{Gamma}(n, \text{rate } \theta)$ ，小 θ 使 T 偏大。由 MLR，拒绝当

$$2\theta_0 T > \chi_{2n, 1-\alpha}^2.$$

边界参数校准 size。变量变换 $\log(X/100) \sim \text{Exp}(\theta)$ 是关键。

20.10 第 11 题：约束正态均值的 UMVU (原卷 p.2, 15 分)

2024 春 Q11

English original. Nonnegative $\theta_1, \theta_2, \theta_3$ satisfy $\theta_1 + \theta_2 + \theta_3 = 1$. Observe $X_{ik} = \theta_k + \varepsilon_{ik}$ with independent $N(0, 1)$ errors. Derive the UMVUE of θ_1 .

中文译述. 在三个均值和为一的正态模型中，利用约束改进 $\bar{X}_1$ 并求 UMVU。

完整解答

无偏控制变量为 $C = \bar{X}_1 + \bar{X}_2 + \bar{X}_3 - 1$ 。考虑 $\bar{X}_1 + aC$ ，其噪声系数为 $(1 + a, a, a)$ ，方差 $[(1 + a)^2 + 2a^2]/n$ ，在 $a = -1/3$ 最小。因此

$$\delta = \bar{X}_1 - \frac{1}{3}(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 - 1) = \frac{1 + 2\bar{X}_1 - \bar{X}_2 - \bar{X}_3}{3}.$$

参数空间相对内部是二维开放集；投影到约束平面的两维正态自然统计量构成满秩指数族，完备充分。 δ 是该统计量的无偏函数，故由 Lehmann–Scheffé 为 UMVU。只在“线性无偏类”最小化尚不足，必须补完备充分论证。

第二十一章 2024 年秋季博士资格考试

考试范围提醒

Q2、Q5 的打印题面与唯一完整解答分别见 第 4 章的 Delta 方法真题（第 26 页）、第 4 章的几乎处处与 L^1 极限真题（第 27 页）。

考试范围提醒

本卷共 11 题，考试时选做 8 题。Q9(c) 按原式不存在有限非退化极限，解答同时说明原式结论和自然标准化后的极限。

21.1 第 1 题：均匀和的四阶矩与 Irwin–Hall 分布（原卷 p.1, 10 分）

2024 秋 Q1

English original. Let $U_i \stackrel{\text{i.i.d.}}{\sim} U[0, 1]$, $S_n = \sum_{k=1}^n U_k$. (a) Compute $\mathbb{E}S_n^4$. (b) Determine the distribution of S_3 .

中文译述. 计算均匀变量和的四阶矩，并求三个均匀变量和的完整分布。

完整解答

用 $\mathbb{E}U^r = 1/(r+1)$ 并按指标重合类型展开：

$$\mathbb{E}S_n^4 = \frac{n}{5} + \frac{n(n-1)}{2} + \frac{n(n-1)}{3} + \frac{n(n-1)(n-2)}{2} + \frac{n(n-1)(n-2)(n-3)}{16}.$$

S_3 的密度由三次卷积或容斥得到

$$f(s) = \begin{cases} s^2/2, & 0 < s < 1, \\ (-2s^2 + 6s - 3)/2, & 1 \leq s < 2, \\ (3-s)^2/2, & 2 \leq s < 3, \\ 0, & \text{其他.} \end{cases}$$

积分可得 CDF；检查三段连续且总积分为一。常见错误是四阶展开漏掉 $6n(n-1)(n-2)$ 型项。

21.2 第 3 题：次指数尾、矩与 mgf 等价（原卷 pp.1–2, 10 分）

2024 秋 Q3

English original. Prove equivalence up to absolute constants among: (a) $\mathbb{P}(|X| \geq t) \leq 2e^{-t/K_1}$; (b) $\mathbb{E}|X|^p \leq (K_2 p)^p$ for all integers p ; (c) $\mathbb{E}e^{|X|/K_3} \leq 2$ for some K_3 .

中文译述. 证明次指数随机变量的指数尾、线性矩增长和某点指数矩有界三种刻画等价，常数只差绝对因子。

完整解答

(a) $\Rightarrow$ (b): 尾积分公式给

$$\mathbb{E}|X|^p = p \int_0^\infty t^{p-1} \mathbb{P}(|X| > t) dt \leq 2K_1^p p! \leq (CK_1 p)^p.$$

(b) $\Rightarrow$ (c): 展开指数, 取 $K_3 = cK_2$. 由 Stirling 下界 $p! \geq (p/e)^p$,

$$\mathbb{E}e^{|X|/K_3} \leq 1 + \sum_{p \geq 1} (e/c)^p,$$

选 $c \geq 3e$ 使右端不超过 2. (c) $\Rightarrow$ (a): Markov 给 $\mathbb{P}(|X| \geq t) \leq e^{-t/K_3} \mathbb{E}e^{|X|/K_3} \leq 2e^{-t/K_3}$. 三步同时给出常数因子关系。

21.3 第 4 题：正则树与 b 叉树随机游动（原卷 p.2, 10 分）

2024 秋 Q4

English original. For simple random walk on an infinite d -regular tree T_d , determine recurrence/transience for $d \geq 2$. Do the same for the rooted infinite b -ary tree T_b .

中文译述. 判断正则树和有根 b 叉树上简单随机游动的常返/暂留并证明。

完整解答

看离根距离。 T_d 上除根外, 距离以概率 $(d-1)/d$ 增一、 $1/d$ 减一。 $d=2$ 时等价于一维简单随机游动的绝对值, 常返; $d \geq 3$ 时有向外正漂移, 命中根的概率从距离 r 出发为 $(1/(d-1))^r < 1$, 故暂留。

T_b 除根外每点有 b 个孩子和一个父亲, 等价于 $d=b+1$ 的径向链。 $b=1$ 常返, $b \geq 2$ 暂留。根处有限修改不改变分类。

21.4 第 6 题: Brownian 存活条件与 Brownian bridge/GFF (原卷 pp.2-3, 15 分)

2024 秋 Q6

English original. (a) For $X_t = x + B_t$, $x > 0$, compute $\mathbb{P}(X_t \in [a, b] \mid \min_{s \leq t} X_s > 0)$.
 (b) A Brownian bridge $W_t = (B_t \mid B_1 = 0)$ and the zero-boundary 1D GFF have Green covariance G . Find G and prove the two processes have identical finite-dimensional laws.
中文译述. 用反射原理求 Brownian 在未触零条件下的终点分布; 证明 Brownian bridge 与一维零边界 GFF 同分布。

完整解答

(a) 吸收于零的转移密度由反射原理为

$$p_t^+(x, y) = \frac{1}{\sqrt{2\pi t}} \left(e^{-(y-x)^2/(2t)} - e^{-(y+x)^2/(2t)} \right), \quad y > 0.$$

存活概率为 $\int_0^\infty p_t^+(x, y) dy = 2\Phi(x/\sqrt{t}) - 1$, 故所求为

$$\frac{\int_a^b p_t^+(x, y) dy}{2\Phi(x/\sqrt{t}) - 1}.$$

(b) 解 $-\partial_t^2 G(t, s) = \delta_s(t)$ 、Dirichlet 边界得 $G(t, s) = \min(t, s) - ts$ 。而 $W_t = B_t - tB_1$, 故 $\mathbb{E}W_t W_s = \min(t, s) - ts$ 。两者都是零均值高斯过程, 协方差相同即所有有限维高斯分布相同。

21.5 第 7 题: 只记录符号后的推断 (原卷 p.3, 10 分)

2024 秋 Q7

English original. $X_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, 1)$ but only $Y_i = \mathbf{1}_{\{X_i < 0\}}$ are recorded. (a) Derive the MLE of μ . (b) Construct a size- α UMP test of $H_0 : \mu \leq \mu_0$ versus $H_1 : \mu > \mu_0$. (c) Describe a $(1 - \alpha)$ confidence interval.
中文译述. 正态样本只保留正负号后, 完成 MLE、单侧 UMP 与置信区间构造。

完整解答

$Y_i \sim \text{Bernoulli}(q), q = \Phi(-\mu)$ 。令 $S = \sum Y_i$ 。内部情形 $0 < S < n, \hat{q} = S/n$, 故 $\hat{\mu} = -\Phi^{-1}(S/n)$; $S = 0, n$ 时扩展 MLE 分别为 $+\infty, -\infty$ 。

μ 增大使 q 减小, 二项族对 S 有 MLR, 故在 S 小时拒绝; 选整数临界值和必要边界随机化使 $q_0 = \Phi(-\mu_0)$ 下拒绝概率为 α 。置信区间可先对二项参数 q 作 Clopper-Pearson 或反演 UMP 检验得到 $[q_L, q_U]$, 再单调反变换为 $[-\Phi^{-1}(q_U), -\Phi^{-1}(q_L)]$ 。

21.6 第 8 题：线性模型 BLUE 与信息矩阵（原卷 p.3, 10 分）

2024 秋 Q8

English original. In $Y = Z\beta + \varepsilon$ under Gauss–Markov assumptions, $\hat{\beta} = (Z^T Z)^{-1} Z^T Y$. For $\theta = b^T \beta$: (a) give mean/variance of $\hat{\theta} = b^T \hat{\beta}$ and prove it has minimum variance among all linear unbiased estimators. (b) If errors are i.i.d. $N(0, \sigma^2)$ with known σ^2 , derive $I(\beta)$.
中文译述. 证明一般线性函数的 OLS 估计是 BLUE，并求正态模型信息矩阵。

完整解答

$\mathbb{E}\hat{\theta} = b^T \beta$, $\text{Var}(\hat{\theta}) = \sigma^2 b^T (Z^T Z)^{-1} b$ 。任意线性无偏 $c^T Y$ 满足 $Z^T c = b$ 。写 $c = c_0 + d$, $c_0 = Z(Z^T Z)^{-1} b$, $Z^T d = 0$, 则 $c_0^T d = 0$, 故 $\text{Var}(c^T Y) = \sigma^2 (\|c_0\|^2 + \|d\|^2) \geq \text{Var}(c_0^T Y)$ 。正态对数似然 Hessian 为 $-Z^T Z/\sigma^2$, 所以 $I(\beta) = Z^T Z/\sigma^2$ 。对应理论第十八章。

21.7 第 9 题：两个估计量的渐近比较（原卷 pp.3–4, 10 分）

2024 秋 Q9

English original. $X_i \stackrel{\text{i.i.d.}}{\sim} U[0, \theta]$, fixed $t \in (0, \theta)$. Compare $F_n(t) = n^{-1} \sum \mathbf{1}_{\{X_i \leq t\}}$ and $T_n(t) = t/(2\bar{X})$: (a) find asymptotic laws; (b) when does F_n have smaller asymptotic variance? (c) with $\theta = 1$, find the asymptotic distribution of $nF_n(n^{-1/2}) - \sqrt{n}$.
中文译述. 比较经验分布估计与代入估计，并处理阈值随样本量变化的二项计数。

完整解答

令 $r = t/\theta$ 。则

$$\sqrt{n}(F_n(t) - r) \xrightarrow{d} N(0, r(1-r)).$$

又 $\bar{X}$ 均值 $\theta/2$ 、方差 $\theta^2/(12n)$, 对 $g(x) = t/(2x)$ 用 Delta:

$$\sqrt{n}(T_n(t) - r) \xrightarrow{d} N(0, r^2/3).$$

F_n 方差较小当 $r(1-r) < r^2/3$, 即 $t > 3\theta/4$ 。

(c) 令 $K_n = nF_n(n^{-1/2}) \sim \text{Bin}(n, n^{-1/2})$, 均值 $\sqrt{n}$ 、方差 $\sqrt{n} - 1$ 。按原题字面, $K_n - \sqrt{n}$ 方差发散且不紧, 故在 $\mathbb{R}$ 上没有有限非退化弱极限。自然的标准化是

$$\frac{K_n - \sqrt{n}}{n^{1/4}} \xrightarrow{d} N(0, 1).$$

此处记为原题的 minor ambiguity, 不擅自把缺失的 $n^{-1/4}$ 写入题面。

21.8 第 10 题：受限正态模型的 UMVU 与 admissibility (原卷 p.4, 15 分)

2024 秋 Q10

English original. $X_i \stackrel{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, $\mu \geq 0$, $\sigma > 0$, $n \geq 2$. (a) Show $\bar{X}$ and S^2 are independent. (b) Find a UMVUE of μ/σ if it exists. (c) Is $\bar{X}$ admissible for estimating μ under squared error?

中文译述. 在均值非负的正态双参数模型中证明均值方差独立, 讨论 μ/σ 的 UMVU 及样本均值的可容许性。

完整解答

(a) 对高斯向量作正交分解: 沿 $\mathbf{1}$ 的投影给 $\bar{X}$, 正交补给残差; 二者联合高斯且协方差零, 故独立, 且 $(n-1)S^2/\sigma^2 \sim \chi_{n-1}^2$ 。

(b) 对 $n > 2$,

$$\mathbb{E} \frac{\bar{X}}{S} = \frac{\mu}{\sigma} \sqrt{\frac{n-1}{2}} \frac{\Gamma((n-2)/2)}{\Gamma((n-1)/2)}.$$

故无偏估计为

$$\delta = \sqrt{\frac{2}{n-1}} \frac{\Gamma((n-1)/2)}{\Gamma((n-2)/2)} \frac{\bar{X}}{S}.$$

正态族的 $(\bar{X}, S^2)$ 在相对开放参数域上完备充分, 故该估计是唯一无偏函数。

若按有限方差定义 UMVU, 需 $n > 3$, 因为 $\mathbb{E}S^{-2} < \infty$ 当且仅当 $n-1 > 2$ 。 $n=2$ 时上述无偏函数不可积; 题目未说明低自由度约定, 应明确此边界。

(c) 不可容许。投影估计 $\delta_+(X) = \max(0, \bar{X})$ 对每个 $\mu \geq 0$ 逐点满足 $(\delta_+ - \mu)^2 \leq (\bar{X} - \mu)^2$, 并在 $\bar{X} < 0$ 上严格改善; 该事件有正概率。

21.9 第 11 题：参数依赖支集的矩法、MLE 与相合 (原卷 p.4, 15 分)

2024 秋 Q11

English original. $X_i \stackrel{\text{i.i.d.}}{\sim} U[\theta, \theta + |\theta|]$, $\theta \neq 0$. (a) Derive the method-of-moments estimator. (b) Derive the MLE. (c) Is it consistent?

中文译述. 分正负参数讨论均匀区间模型的矩估计、MLE 和相合性。

完整解答

若 $\theta > 0$, 支集 $[\theta, 2\theta]$ 、均值 $3\theta/2$; 若 $\theta < 0$, 支集 $[\theta, 0]$ 、均值 $\theta/2$ 。样本符号 a.s. 识别分支, 故矩估计为 $2\bar{X}/3$ (样本为正) 或 $2\bar{X}$ (样本为负)。

正分支似然为 $\theta^{-n} \mathbf{1}_{\{X_{(n)}/2 \leq \theta \leq X_{(1)}\}}$, 故 $\hat{\theta} = X_{(n)}/2$; 负分支似然为 $(-\theta)^{-n} \mathbf{1}_{\{\theta \leq X_{(1)} \leq X_{(n)} \leq 0\}}$, 故 $\hat{\theta} = X_{(1)}$ 。极值强律给正分支 $X_{(n)} \rightarrow 2\theta$, 负分支 $X_{(1)} \rightarrow \theta$, 所以 MLE 相合。

第二十二章 2025 年春季概率论与数理统计博士资格考试

本卷共 11 题, 考生选做 8 题; 多做时只批阅最先作答的 8 题。四题各 15 分, 其余各 10 分。以下按原卷顺序完整收录。

22.1 第 1 题: 样本极值的联合分布与条件期望 (原卷 p.1, 10 分)

2025 春 Q1

English original. Let $\{U_n\}$ be i.i.d. with uniform distribution on $[0, 1]$. For $n \geq 2$, define

$$X_n := \min\{U_1, \dots, U_n\}, \quad Y_n := \max\{U_1, \dots, U_n\}.$$

(1) Derive the joint distribution of (X_n, Y_n) . (2) Calculate $\mathbb{E}[X_n | \sigma(Y_n)]$.

中文翻译. 对独立同分布的 $U[0, 1]$ 样本, 求样本最小值与最大值的联合分布, 并计算给定最大值的最大值条件期望。

完整解答

对 $0 < x < y < 1$, 恰有一个观测落在 dx 、一个落在 dy , 其余 $n - 2$ 个落在 (x, y) , 故

$$f_{X_n, Y_n}(x, y) = n(n-1)(y-x)^{n-2} \mathbf{1}_{\{0 < x < y < 1\}}.$$

等价地, 可先写 $\Pr(X_n > x, Y_n \leq y) = (y-x)^n$ 再作混合偏导。边缘密度 $f_{Y_n}(y) = ny^{n-1} \mathbf{1}_{(0,1)}(y)$, 所以

$$f_{X_n|Y_n=y}(x) = \frac{(n-1)(y-x)^{n-2}}{y^{n-1}}, \quad 0 < x < y.$$

积分并令 $x = yu$ 得

$$\mathbb{E}[X_n | Y_n = y] = y(n-1) \int_0^1 u(1-u)^{n-2} du = \frac{y}{n}.$$

因此 $\mathbb{E}[X_n | \sigma(Y_n)] = Y_n/n$ a.s. 易错点是把条件分布误当成 $U(0, Y_n)$; 只有 $n = 2$ 时才如此。参见第一章。

22.2 第 2 题：高斯向量的线性表示（原卷 p.1, 10 分）

2025 春 Q2

English original. Suppose (X, Y, Z) is a three-dimensional mean-zero Gaussian vector with covariance matrix

$$\begin{pmatrix} 1 & \rho & \tau \\ \rho & 1 & 0 \\ \tau & 0 & 1 \end{pmatrix}.$$

(1) Show that $\rho^2 + \tau^2 \leq 1$. (2) Show that there exist real numbers a, b, c and a standard Gaussian random variable W independent of (Y, Z) such that $X = aY + bZ + cW$, and derive a, b, c .

中文翻译. 由协方差矩阵的半正定性导出参数约束，并把 X 分解为 (Y, Z) 的线性投影与独立高斯残差。

完整解答

协方差矩阵必须半正定，其行列式为 $1 - \rho^2 - \tau^2$ ，故第一问成立。又因 (Y, Z) 联合高斯且协方差为零，二者独立。令 $R = X - \rho Y - \tau Z$ ，则

$$\text{Cov}(R, Y) = \text{Cov}(R, Z) = 0, \quad \text{Var}(R) = 1 - \rho^2 - \tau^2.$$

联合高斯向量中零协方差推出独立。若 $1 - \rho^2 - \tau^2 > 0$ ，取

$$W = \frac{R}{\sqrt{1 - \rho^2 - \tau^2}}, \quad a = \rho, \quad b = \tau, \quad c = \sqrt{1 - \rho^2 - \tau^2}.$$

边界情形令 $c = 0$ ，另取任意独立标准正态 W ，同一表示仍成立。关键条件是“联合高斯”，一般随机向量中不相关不推出独立。

22.3 第 3 题：随机游走的矩母函数与退出时生成函数（原卷 p.1, 15 分）

2025 春 Q3

English original. Suppose $\{\xi_i\}$ are i.i.d. Bernoulli random variables with $\Pr(\xi_1 = 1) = \Pr(\xi_1 = -1) = 1/2$. Set $S_n = \sum_{j=1}^n \xi_j$ and define $\tau = \min\{n : S_n = -a \text{ or } b\}$ for $a, b \in \mathbb{Z}_{>0}$.

(1) Calculate $\mathbb{E}[e^{\lambda S_n}]$ for $\lambda \in \mathbb{R}$. (2) Calculate $\mathbb{E}[s^\tau]$ for $s \in (0, 1)$.

中文翻译. 对简单对称随机游走，计算位置的矩母函数，以及离开区间 $(-a, b)$ 的时间生成函数。

完整解答

独立性给出

$$\mathbb{E}e^{\lambda S_n} = \prod_{j=1}^n \mathbb{E}e^{\lambda \xi_j} = (\cosh \lambda)^n.$$

令 $u_k = \mathbb{E}_k[s^\tau]$ 。首步分析给出

$$u_{-a} = u_b = 1, \quad u_k = \frac{s}{2}(u_{k-1} + u_{k+1}), \quad -a < k < b.$$

取 $\gamma > 0$ 使 $\cosh \gamma = 1/s$ 。差分方程特征根为 $e^{\pm\gamma}$ ；满足两个边界条件的解为

$$u_k = \frac{\sinh((b-k)\gamma) + \sinh((k+a)\gamma)}{\sinh((a+b)\gamma)}.$$

从零出发因此

$$\mathbb{E}[s^\tau] = \frac{\sinh(b\gamma) + \sinh(a\gamma)}{\sinh((a+b)\gamma)}, \quad \cosh \gamma = s^{-1}.$$

把边界值误设为 s 是常见错误：到达边界后剩余退出时间为零，故边界生成函数为 1。参见第十章。

22.4 第 4 题：非一致有界增量鞅与停时（原卷 p.1, 10 分）

2025 春 Q4

English original. Suppose $\{\xi_i\}$ are i.i.d. with $\Pr(\xi_1 = 1) = \Pr(\xi_1 = -1) = 1/2$. Set $X_n = \sum_{j=1}^n 2^{j-1} \xi_j$. (1) Define $T = \min\{j \geq 1 : \xi_j = +1\}$. Derive the law of T and calculate $\mathbb{E}[T]$. (2) Show that $\{X_n\}$ is a martingale and calculate $\mathbb{E}[X_T]$.

中文翻译. 求第一次出现正号的几何停时分布，验证加权随机游走在鞅，并计算停时处的值。

完整解答

$\Pr(T = k) = 2^{-k}$, $k \geq 1$, 所以 T 为成功概率 $1/2$ 、取值从 1 开始的几何变量, $\mathbb{E}T = 2$ 。对自然过滤 $\mathcal{F}_n = \sigma(\xi_1, \dots, \xi_n)$,

$$\mathbb{E}[X_{n+1} | \mathcal{F}_n] = X_n + 2^n \mathbb{E}\xi_{n+1} = X_n,$$

且每个固定 n 下 X_n 可积，故为鞅。在事件 $T = k$ 上，前 $k-1$ 项均为 -1 ，第 k 项为 $+1$ ，于是

$$X_T = -\sum_{j=1}^{T-1} 2^{j-1} + 2^{T-1} = 1 \quad \text{a.s.},$$

故 $\mathbb{E}X_T = 1$ 。不能直接套用可选停止定理，因为被停过程缺乏一致可积等充分条件；本题应逐路径计算。

22.5 第 5 题: Brownian 对称退出与符号独立性 (原卷 pp.1–2, 10 分)

2025 春 Q5

English original. Let $\{B_t\}_{t \geq 0}$ be a one-dimensional Brownian motion starting from the origin. For $a \in (0, \infty)$, define $T_a := \inf\{t \geq 0 : |B_t| = a\}$. (1) Prove that $T_a < \infty$ a.s. (2) Prove that T_a and $\mathbf{1}_{\{B_{T_a}=a\}}$ are independent.

中文翻译. 证明 Brownian motion 几乎必然在有限时间退出 $(-a, a)$, 并证明退出时间与退出端点符号独立。

完整解答

对任意 $t > 0$, 若 $T_a > t$ 则 $|B_t| < a$, 因此 $\Pr(T_a > t) \leq \Pr(|B_t| < a) = \Pr(|N(0, 1)| < a/\sqrt{t}) \rightarrow 0$, 故 $\Pr(T_a = \infty) = 0$.

令路径反射 $\omega \mapsto -\omega$. Wiener 测度在此映射下不变, T_a 不变, 而退出端点 $+a$ 与 $-a$ 互换. 故对任意 Borel 集 $A \subset [0, \infty)$,

$$\Pr(T_a \in A, B_{T_a} = a) = \Pr(T_a \in A, B_{T_a} = -a) = \frac{1}{2} \Pr(T_a \in A).$$

又 $\Pr(B_{T_a} = a) = 1/2$, 所以联合概率分解, 得到独立性. 此证明同时使用路径连续性保证退出点只能是 $\pm a$. 参见第十章。

22.6 第 6 题: 条件二阶矩刻画正态分布 (原卷 p.2, 15 分)

2025 春 Q6

English original. Suppose X and Y are i.i.d. with $\mathbb{E}X = 0$ and $\text{Var}(X) = 1$. Suppose further that

$$\mathbb{E}[(X - Y)^2 \mid \sigma(X + Y)] = 2.$$

Show that X and Y are independent standard Gaussian random variables.

中文翻译. 用和给定时差的条件二阶矩恒定这一条件, 刻画标准正态分布。

完整解答

独立性已由 i.i.d. 假设包含, 需证明公共分布为 $N(0, 1)$. 令 $\phi(t) = \mathbb{E}e^{itX}$. 将条件等式乘 $e^{it(X+Y)}$ 后取期望; 因二阶矩有限, ϕ 二次可微, 且

$$\mathbb{E}[(X - Y)^2 e^{it(X+Y)}] = -2\phi''(t)\phi(t) + 2\phi'(t)^2,$$

右端为 $2\phi(t)^2$. 因此

$$-\phi''\phi + (\phi')^2 = \phi^2. \quad (*)$$

在 $\phi \neq 0$ 的连通区间上令 $g = \log \phi$, 则由 (*) 得 $g'' = -1$. 又 $g(0) = 0$, $g'(0) = i\mathbb{E}X = 0$, 故零点所在连通分支上 $\phi(t) = e^{-t^2/2}$. 该解不在有限端点趋于零, 故此分支不能有有限边界; 于是等式对全部 $t \in \mathbb{R}$ 成立. 特征函数唯一性给出 $X, Y \sim N(0, 1)$.

关键步骤不是“条件方差恒定”本身，而是把条件恒等式转成对所有 t 成立的特征函数微分方程。

22.7 第 7 题：均匀分布端点的广义 Bayes 估计（原卷 p.2, 10 分）

2025 春 Q7

English original. Let $X_1, \dots, X_n$ ($n \geq 2$) be i.i.d. $U(0, \theta)$, $0 < \theta < \infty$. (a) Under squared error loss, derive the generalized Bayes rule for the improper prior $\pi(\theta) = 1$, $0 < \theta < \infty$. (b) Is the rule consistent for θ ?

中文翻译. 在平方损失和常数非正常先验下求均匀端点参数的广义 Bayes 规则，并判断相合性。

完整解答

记 $M = X_{(n)}$ 。似然为 $\theta^{-n} \mathbf{1}_{\{\theta \geq M\}}$ ，故后验密度正比于该式， $n > 1$ 时可正规化。形式上的平方损失 Bayes 规则为后验均值：对 $n > 2$,

$$\delta_n(X) = \frac{\int_M^\infty \theta^{1-n} d\theta}{\int_M^\infty \theta^{-n} d\theta} = \frac{n-1}{n-2} M.$$

$n = 2$ 时后验均值发散，有限 Bayes 动作不存在；若要求后验平方损失本身有限，还需 $n > 3$ ， $n = 3$ 的上述形式估计具有无限后验二阶风险。作为随样本量变化的估计序列， $M \rightarrow \theta$ a.s. 且 $(n-1)/(n-2) \rightarrow 1$ ，故 $\delta_n \rightarrow \theta$ a.s.，即相合。不要把固定 n 的倍数因子误当成相合性的极限。

22.8 第 8 题：Gamma 尺度族的检验与 MLE 渐近分布（原卷 p.2, 10 分）

2025 春 Q8

English original. Let $X_1, \dots, X_n$ be i.i.d. $\text{Gamma}(2025, \theta)$ with density

$$f(x | \alpha, \theta) = \frac{x^{\alpha-1} e^{-x/\theta}}{\Gamma(\alpha) \theta^\alpha}, \quad x > 0.$$

(a) For $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$, derive a UMP level 0.05 test if it exists; otherwise explain. (b) Derive the MLE of $1/\theta$ and its asymptotic distribution.

中文翻译. 讨论 Gamma 尺度参数双侧检验是否存在 UMP，并求逆尺度参数的 MLE 及渐近分布。

完整解答

$T = \sum X_i \sim \text{Gamma}(2025n, \theta)$ ，该族对 T 具有单调似然比。对 $\theta > 1$ 的最强检验拒绝大 T ，对 $\theta < 1$ 的最强检验拒绝小 T ；同一非平凡水平检验不能同时在两个相反方向都最强，故双侧备择下一般不存在 UMP 检验（可进一步构造 UMPU 或似然比检验，但不是题问）。

对数似然关于 θ 最大化给 $\hat{\theta} = \bar{X}/2025$, 所以不变性给 $\widehat{1/\theta} = 2025/\bar{X}$ 。单个观测均值 2025θ 、方差 $2025\theta^2$, CLT 与 Delta 方法给

$$\sqrt{n} \left(\frac{2025}{\bar{X}} - \frac{1}{\theta} \right) \xrightarrow{d} N \left(0, \frac{1}{2025\theta^2} \right).$$

易错点是把 Gamma 的 scale 与 rate 参数化互换。参见第 十三章。

22.9 第 9 题：样本相关系数与最小充分性（原卷 p.2, 15 分）

2025 春 Q9

English original. Let (X_i, Y_i) be i.i.d. bivariate vectors and let ρ be their correlation. With the usual sample means and unbiased sample variances, define

$$\hat{\rho} = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y}.$$

- (a) If $\mathbb{E}|X_i|^4, \mathbb{E}|Y_i|^4 < \infty$, prove $\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, c)$ and identify c in terms of moments.
 (b) If $(X_i, Y_i) \sim N_2(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix})$, find a minimal sufficient statistic for ρ and determine completeness.

中文翻译. 推导一般双变量总体样本相关系数的渐近分布, 并研究标准边缘二元正态相关参数的最小充分统计量和完备性。

完整解答

设总体均值为 μ_X, μ_Y , 标准差为 $\sigma_X, \sigma_Y > 0$ 。样本一阶、二阶和交叉矩的向量满足多元 CLT, 相关系数是其可微函数, Delta 方法给渐近正态。其影响函数可写为

$$\psi(x, y) = \frac{(x - \mu_X)(y - \mu_Y) - \sigma_{XY}}{\sigma_X \sigma_Y} - \frac{\rho}{2} \left[\frac{(x - \mu_X)^2 - \sigma_X^2}{\sigma_X^2} + \frac{(y - \mu_Y)^2 - \sigma_Y^2}{\sigma_Y^2} \right].$$

四阶矩条件保证 $\mathbb{E}\psi^2 < \infty$, 且

$$\sqrt{n}(\hat{\rho} - \rho) \xrightarrow{d} N(0, c), \quad c = \mathbb{E}[\psi(X, Y)^2].$$

这已按题意用总体矩识别 c , 无需再展开。

正态子题的似然只通过

$$A = \sum_{i=1}^n (X_i^2 + Y_i^2), \quad B = \sum_{i=1}^n X_i Y_i$$

依赖样本。似然比判据表明 (A, B) 最小充分: 两样本的似然比与 ρ 无关, 当且仅当二者的 A, B 相同。它不完备, 因为 $A - 2n$ 对所有 $\rho \in (-1, 1)$ 期望均为零, 但不是几乎处处零。易错点是“一参数族”自动等同于“完备指数族”。

22.10 第 10 题: 绝对值矩估计与可容许性 (原卷 pp.2–3, 10 分)

2025 春 Q10

English original. Let $X_1, \dots, X_n$ be i.i.d. $N(0, \sigma^2)$, $\sigma > 0$, and $W = n^{-1} \sum_{i=1}^n |X_i|$. (a) Find an unbiased estimator of σ based on W and identify its limiting distribution. (b) Is it admissible under squared error loss? Justify.

中文翻译. 用样本绝对值均值构造尺度参数的无偏估计, 求渐近分布并判断平方损失下的可容许性。

完整解答

因 $\mathbb{E}|X_i| = \sigma\sqrt{2/\pi}$, $\text{Var}(|X_i|) = \sigma^2(1 - 2/\pi)$,

$$\delta = \sqrt{\frac{\pi}{2}}W, \quad \sqrt{n}(\delta - \sigma) \xrightarrow{d} N\left(0, \sigma^2\left(\frac{\pi}{2} - 1\right)\right).$$

关于可容许性, 令 $V = \sum X_i^2$. V 是完备充分统计量, Rao-Blackwell 化 $\delta^* = \mathbb{E}(\delta | V)$ 仍无偏且平方风险不增。给定 V 时方向在球面上随机, 而 δ 依赖各坐标绝对值, $n \geq 2$ 时并非 V 的函数, 故条件方差以正概率为正:

$$R(\sigma, \delta) = R(\sigma, \delta^*) + \mathbb{E}_\sigma \text{Var}(\delta | V) > R(\sigma, \delta^*).$$

所以该基于 W 的无偏估计不可容许。若 $n = 1$, 此 Rao-Blackwell 严格改进论证不成立, 应单独讨论; 原卷通常默认一般样本量, 完整答案必须声明此边界。

22.11 第 11 题: 带位置参数指数分布的非正则 MLE (原卷 p.3, 15 分)

2025 春 Q11

English original. Let $X_1, \dots, X_n$ be i.i.d. with density $f(x; \lambda, \alpha) = \lambda e^{-\lambda(x-\alpha)} \mathbf{1}_{\{x \geq \alpha\}}$, where $\lambda > 0$ and $\alpha \in \mathbb{R}$ are unknown. (a) Find the MLE $(\hat{\lambda}, \hat{\alpha})$. (b) To what distribution does

$$\frac{n(\hat{\lambda} - \lambda)}{n(\hat{\alpha} - \alpha)}$$

converge? Partial credit is available for identifying the numerator and/or denominator distributions.

中文翻译. 对未知率和未知左端点的移位指数族求 MLE, 并研究两个非同阶估计误差之比的极限分布。

完整解答

似然为 $L = \lambda^n \exp[-\lambda \sum_i (X_i - \alpha)] \mathbf{1}_{\{\alpha \leq X_{(1)}\}}$. 固定 λ 时它随 α 增大, 故 $\hat{\alpha} = X_{(1)}$; 再对 λ 最大化得

$$\hat{\lambda} = \frac{n}{\sum_i (X_i - X_{(1)})}.$$

写 $X_i = \alpha + E_i$, $E_i \stackrel{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$ 。指数次序统计量的无记忆性给

$$n(\hat{\alpha} - \alpha) = nE_{(1)} \sim \text{Exp}(\lambda),$$

且 $E_{(1)}$ 与 $\sum_i (E_i - E_{(1)}) \sim \text{Gamma}(n-1, \text{rate } \lambda)$ 独立。后者推出 $\sqrt{n}(\hat{\lambda} - \lambda) \xrightarrow{d} N(0, \lambda^2)$, 所以 $n(\hat{\lambda} - \lambda) = \sqrt{n}O_p(1)$ 并不紧。更精确地, 令 $Z \sim N(0, 1)$ 、 $E \sim \text{Exp}(\lambda)$ 独立, 则

$$\frac{1}{\sqrt{n}} \frac{n(\hat{\lambda} - \lambda)}{n(\hat{\alpha} - \alpha)} \xrightarrow{d} \frac{\lambda Z}{E}.$$

因此原题所写未再除以 $\sqrt{n}$ 的比值在实线上没有有限非退化弱极限; 它以 $\sqrt{n}$ 量级发散。此处保留原卷字面并标为 minor ambiguity, 不擅自补入标准化因子。

第二十三章 2025 年秋季概率论与数理统计博士资格考试

考试范围提醒

Q4 的打印题面与唯一完整解答见 第 6 章的稀有大跳与样本均值振荡真题 (第 44 页)。

本卷共 11 题, 选做 8 题; 四题各 15 分, 其余各 10 分。以下按原卷顺序完整收录。

23.1 第 1 题: 随机幂的矩 (原卷 p.1, 10 分)

2025 秋 Q1

English original. Let X and N be independent. Suppose X is standard exponential and $N \sim \text{Poisson}(\lambda)$, $\lambda > 0$. (a) Find all λ such that $\mathbb{E}[X^N] < \infty$. (b) Calculate $\mathbb{E}[X^N]$ under that condition. (c) Is there any $\lambda > 0$ such that $\text{Var}(X^N) < \infty$? Justify.

中文翻译. 对指数变量的 Poisson 随机次幂, 判断一阶、二阶矩的存在范围并计算期望。

完整解答

条件于 $N = n$, $\mathbb{E}X^n = \Gamma(n+1) = n!$, 故

$$\mathbb{E}X^N = e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} n! = \frac{e^{-\lambda}}{1-\lambda}$$

当且仅当 $0 < \lambda < 1$ 有限。二阶矩为 $e^{-\lambda} \sum_{n \geq 0} \lambda^n (2n)!/n!$, 相邻项比 $\lambda(2n+2)(2n+1)/(n+1) = 2\lambda(2n+1) \rightarrow \infty$, 故对每个 $\lambda > 0$ 均发散。因此不存在有限方差的正参数。Tonelli 定理保证非负项下交换求和与期望合法。

23.2 第 2 题: 三角区域上的条件分布 (原卷 p.1, 10 分)

2025 秋 Q2

English original. Let X, Y be independent $U[0, 1]$ and $Z = \max\{X, Y\}$. (a) Determine the distribution of Z conditioned on $\{X + Y \in [0, 1]\}$. (b) Determine it conditioned on $\{X + Y \in [1, 2]\}$.

中文翻译. 在单位正方形的下、上三角区域条件下, 求两坐标最大值的分布。

完整解答

两个条件事件面积均为 $1/2$ 。在下三角 $A = \{x + y \leq 1\}$ 上,

$$F_{Z|A}(z) = \begin{cases} 0, & z < 0, \\ 2z^2, & 0 \leq z \leq 1/2, \\ -2z^2 + 4z - 1, & 1/2 < z \leq 1, \\ 1, & z > 1. \end{cases}$$

前一段是边长 z 的正方形面积除以 $1/2$; 后一段从该正方形减去直角边 $2z - 1$ 的上三角。对上三角 $B = \{x + y \geq 1\}$, $z < 1/2$ 不可能, 而

$$F_{Z|B}(z) = \begin{cases} 0, & z < 1/2, \\ (2z - 1)^2, & 1/2 \leq z \leq 1, \\ 1, & z > 1. \end{cases}$$

相应密度分别为 $4z, 4(1 - z)$ 与 $4(2z - 1)$ (在分段内部)。

23.3 第 3 题: 鞅的严格凸等号情形 (原卷 pp.1–2, 10 分)

2025 秋 Q3

English original. Let $(X_n, \mathcal{F}_n)$ be a martingale. (a) Prove that if $(X_n^2, \mathcal{F}_n)$ is also a martingale, then $X_m = X_n$ a.s. for all $m, n \in \mathbb{N}$. (b) Prove that if $(|X_n|^p, \mathcal{F}_n)$ is a martingale for some $p > 1$, then $(|X_n|^q, \mathcal{F}_n)$ is also a martingale for every $1 \leq q \leq p$.

中文翻译. 证明鞅及其严格凸函数同时为鞅时过程必须退化, 并推出较低次幂也为鞅。

完整解答

若 $m < n$, 则 $\mathbb{E}(X_n X_m) = \mathbb{E}[X_m \mathbb{E}(X_n | \mathcal{F}_m)] = \mathbb{E}X_m^2$, 而平方鞅给 $\mathbb{E}X_n^2 = \mathbb{E}X_m^2$ 。所以 $\mathbb{E}(X_n - X_m)^2 = 0$, 即 $X_n = X_m$ a.s.

对 (b), 条件 Jensen 给

$$|X_m|^p = |\mathbb{E}(X_n | \mathcal{F}_m)|^p \leq \mathbb{E}(|X_n|^p | \mathcal{F}_m) = |X_m|^p.$$

两边相等; 因 $x \mapsto |x|^p$ 在 $p > 1$ 时严格凸, Jensen 等号条件推出 $X_n = X_m$ a.s.。故 $|X_n|^q$ 对所有 $q \in [1, p]$ 都逐时相同, 当然为鞅。这里不能只说“由 Jensen”, 必须使用严格凸函数的等号条件。

23.4 第 5 题: 指数 Brownian 鞅的积分 (原卷 p.2, 10 分)

2025 秋 Q5

English original. Let B_t be standard Brownian motion and $X_t = \exp(B_t - t/2)$. (a) Prove X_t converges a.s. as $t \rightarrow \infty$. (b) For $t > 0$, calculate $\mathbb{E} \int_0^t X_s ds$ and $\text{Var} \int_0^t X_s ds$. (c) Show $\int_0^\infty X_t dt < \infty$ a.s.

中文翻译. 研究指数 Brownian 鞅的几乎处处极限、有限时积分矩与无限时积分可积性。

完整解答

$B_t/t \rightarrow 0$ a.s., 故 $\log X_t/t = B_t/t - 1/2 \rightarrow -1/2$, 从而 $X_t \rightarrow 0$ a.s. (也可用非负鞅收敛定理)。因 $\mathbb{E}X_s = 1$, Tonelli 给 $\mathbb{E} \int_0^t X_s ds = t$ 。若 $s \leq u$, 独立增量给 $\mathbb{E}(X_s X_u) = \mathbb{E}X_s^2 = e^s$, 故

$$\text{Var} \left(\int_0^t X_s ds \right) = 2 \int_0^t (t-s)(e^s - 1) ds = 2e^t - 2t - 2 - t^2.$$

最后, 由 $B_t/t \rightarrow 0$, 几乎每条路径最终满足 $B_t \leq t/4$, 于是 $X_t \leq e^{-t/4}$; 有限初段连续、尾部由可积指数函数控制, 故无限积分有限。

23.5 第 6 题: 二阶 GOE 的特征值与特征向量 (原卷 p.2, 15 分)

2025 秋 Q6

English original. Let $M = \begin{pmatrix} X_1 & Y \\ Y & X_2 \end{pmatrix}$, where $X_1, X_2 \sim N(0, 2)$ and $Y \sim N(0, 1)$ are independent. Let $\lambda_1 \geq \lambda_2$ be the eigenvalues and v_1, v_2 corresponding unit eigenvectors. (a) Find the distribution of $\lambda_1 + \lambda_2$. (b) Find that of $\lambda_1 - \lambda_2$. (c) Are λ_1, λ_2 independent? (d) Prove the joint distribution of (v_1, v_2) is invariant under orthogonal transformations.

中文翻译. 求二阶高斯正交系综的迹、谱间距, 判断两个有序特征值的独立性, 并证明特征方向的正交不变性。

完整解答

令 $A = (X_1 + X_2)/2$, $D = (X_1 - X_2)/2$ 。则 A, D, Y 独立且均为 $N(0, 1)$, 并且

$$\lambda_{1,2} = A \pm R, \quad R = \sqrt{D^2 + Y^2}.$$

故 $\lambda_1 + \lambda_2 = 2A \sim N(0, 4)$ 。 R 为 Rayleigh 变量, 密度 $re^{-r^2/2}$, 所以 $G = \lambda_1 - \lambda_2 = 2R$ 的密度为 $f_G(g) = \frac{g}{4}e^{-g^2/8}\mathbf{1}_{\{g \geq 0\}}$ 。二者不独立, 例如

$$\text{Cov}(\lambda_1, \lambda_2) = \text{Var}(A) - \text{Var}(R) = 1 - \frac{4 - \pi}{2} = \frac{\pi - 2}{2} \neq 0.$$

矩阵密度正比于 $\exp\{-\text{tr}(M^2)/4\}$, 故 $M \stackrel{d}{=} OMO^T$ 对任意正交 O 成立; 于是特征方向随之变为 Ov_j , 得到不变性。严格地说, 单位特征向量有符号歧义; 结论对一维特征子空间成立, 或在给每个向量独立随机符号后对有序正交框成立。忽略此符号约定会使题述在形式上不完备。

23.6 第 7 题：充分统计量上的相合估计（原卷 p.2, 10 分）

2025 秋 Q7

English original. Suppose $\hat{\theta}_n$ is a real-valued consistent estimator of θ based on n observations, $\theta \in [-1, 1]$, and T_n is sufficient for θ . Show there exists a consistent estimator based on T_n .

中文翻译. 证明有界参数若存在相合估计，则也存在仅依赖任一充分统计量的相合估计。

完整解答

先截断 $\tilde{\theta}_n = \max(-1, \min(1, \hat{\theta}_n))$ 。投影不增加到 $\theta \in [-1, 1]$ 的距离，故仍相合且一致有界。定义

$$\delta_n(T_n) = \mathbb{E}_\theta(\tilde{\theta}_n | T_n).$$

由充分性，给定 T_n 的样本条件分布不依赖 θ ，所以右端确为同一个仅依赖 T_n 的统计量。又有界相合推出 L^2 相合，条件 Jensen 给

$$\mathbb{E}_\theta(\delta_n - \theta)^2 \leq \mathbb{E}_\theta(\tilde{\theta}_n - \theta)^2 \rightarrow 0.$$

因此 $\delta_n \rightarrow \theta$ 依概率。截断步骤不可省略，因为题设没有给原估计的可积性。

23.7 第 8 题：指数分布尾概率的 UMVU（原卷 p.3, 15 分）

2025 秋 Q8

English original. Let $X_1, \dots, X_n$ be i.i.d. $\text{Exp}(\lambda)$ and $\phi = \Pr_\lambda(X_1 > x) = e^{-\lambda x}$. (a) Find a complete sufficient statistic for λ . (b) Find the UMVUE of ϕ .

中文翻译. 找到指数率参数的完全充分统计量，并对尾概率作 Rao-Blackwell 化得到 UMVU。

完整解答

$T = \sum_i X_i \sim \text{Gamma}(n, \text{rate } \lambda)$ 。因子分解给充分性，Gamma 全秩一参数指数族的 Laplace 变换唯一性给完备性。 $\mathbf{1}_{\{X_1 > x\}}$ 是 ϕ 的无偏估计。给定 $T = t$ ，比例向量服从 $\text{Dirichlet}(1, \dots, 1)$ ，所以

$$\mathbb{E}(\mathbf{1}_{\{X_1 > x\}} | T = t) = \Pr(X_1/T > x/t | T = t) = \left(1 - \frac{x}{t}\right)^{n-1} \mathbf{1}_{\{t > x\}}.$$

由 Lehmann-Scheffé 定理，这就是唯一 UMVU。参见第 十四章。

23.8 第 9 题: Bernoulli 泛函的 UMVU 与可容许性 (原卷 p.3, 15 分)

2025 秋 Q9

English original. Let $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$, $0 < p < 1$. (a) For $g(p) = p^k + (1-p)^{n-k}$, $0 \leq k \leq n$, find the BLUE (UMVUE). (b) Prove $\bar{X}$ is admissible for p under loss $l(p, a) = (p - a)^2/[p(1-p)]$.

中文翻译. 求 Bernoulli 参数多项式泛函的最佳无偏估计, 并在加权平方损失下证明样本均值可容许。

完整解答

$T = \sum X_i \sim \text{Bin}(n, p)$ 是完全充分统计量。利用阶乘矩,

$$\delta(T) = \frac{\binom{T}{k}}{\binom{n}{k}} + \frac{\binom{n-T}{n-k}}{\binom{n}{n-k}}$$

分别无偏估计 p^k 与 $(1-p)^{n-k}$, 故其和为 UMVU (约定越界组合数为零)。

对 (b), $R(p, \bar{X}) = \text{Var}(\bar{X})/[p(1-p)] = 1/n$ 。在均匀先验的 Bayes 风险中, 损失权重把似然 $p^T(1-p)^{n-T}$ 变为 $p^{T-1}(1-p)^{n-T-1}$, 其广义 Bayes 动作为 T/n (边界 $T=0, n$ 由极限得到)。用 $\text{Beta}(\varepsilon, \varepsilon)$ 正常先验逼近, 所得 Bayes 规则趋于 T/n , Bayes 风险差趋于零; Blyth 可容许性准则遂排除任何处处不劣且某处严格改善的规则。因此 $\bar{X}$ 可容许。只写“常风险所以 minimax”不足以推出可容许, 必须补充广义 Bayes/Blyth 论证。

23.9 第 10 题: 对数正态先验下的端点 MAP (原卷 p.3, 10 分)

2025 秋 Q10

English original. Let $X_i \stackrel{\text{i.i.d.}}{\sim} U(0, \theta)$, $\theta > 0$, and suppose $\log \theta \sim N(\mu_0, \sigma_0^2)$. (a) Find the posterior density of $\log \theta$. (b) Define the Bayes estimate of θ as the maximizer of the posterior density of θ . Find it and determine consistency.

中文翻译. 求均匀端点模型在对数正态先验下的对数参数后验, 并区分对 $\eta = \log \theta$ 与对 θ 本身取 MAP 的差别。

完整解答

记 $M = X_{(n)}$ 、 $\eta = \log \theta$ 。后验关于 η 的密度为

$$\pi(\eta | x) \propto \exp \left[-\frac{(\eta - \mu_0)^2}{2\sigma_0^2} - n\eta \right] \mathbf{1}_{\{\eta \geq \log M\}},$$

即 $N(\mu_0 - n\sigma_0^2, \sigma_0^2)$ 在 $[\log M, \infty)$ 上的截断分布。

题目要求最大化 θ 的后验密度, 需计入对数正态先验的 Jacobian $1/\theta$ 。无约束众数为 $\exp\{\mu_0 -$

$(n+1)\sigma_0^2\}$, 受支持集限制后

$$\hat{\theta}_{\text{MAP}} = \max \left\{ M, e^{\mu_0 - (n+1)\sigma_0^2} \right\}.$$

第二项趋于零而 $M \rightarrow \theta$ a.s., 所以估计强相合。把 $n+1$ 写成 n 是混淆两个参数化下 MAP 的常见错误。

23.10 第 11 题: 离散均匀族的 UMP 检验 (原卷 p.3, 10 分)

2025 秋 Q11

English original. Let X_i be i.i.d. discrete uniform on $\{1, 2, \dots, \theta\}$, where θ is an unknown positive integer, and let $M = X_{(n)}$. For known positive integer θ_0 and level $0 < \alpha < 1$: (a) for $H_0 : \theta \leq \theta_0$ versus $H_1 : \theta > \theta_0$, show

$$\phi(x) = \begin{cases} 1, & M > \theta_0, \\ \alpha, & M \leq \theta_0 \end{cases}$$

is randomized UMP of size α . (b) for $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, show

$$\phi(x) = \begin{cases} 1, & M > \theta_0 \text{ or } M \leq \theta_0 \alpha^{1/n}, \\ 0, & \text{otherwise} \end{cases}$$

is UMP of size α when $\theta_0 \alpha^{1/n}$ is an integer.

中文翻译. 对支持集随整数参数变化的离散均匀族, 证明给定单侧及双侧复合备择的 UMP 检验。

完整解答

充分统计量为 M , 似然为 $\theta^{-n} \mathbf{1}_{\{M \leq \theta\}}$ 。(a) 在任意 $\theta \leq \theta_0$ 下 $M > \theta_0$ 不可能, 故尺寸为 α 。对 $\theta > \theta_0$, 功效为

$$1 - (1 - \alpha)(\theta_0/\theta)^n.$$

零假设不可能出现的区域必须全拒绝; 在共同支持 $M \leq \theta_0$ 上, 任一固定备择与边界零假设的似然比为常数, 故剩余水平任意分配具有同样功效, 所给检验因而对每个备择均最强。

(b) 令 $c = \theta_0 \alpha^{1/n} \in \mathbb{Z}$ 。在 θ_0 下拒绝概率 $\Pr(M \leq c) = (c/\theta_0)^n = \alpha$ 。对 $\theta < \theta_0$, 最强区域是尽量小的 M ; 对 $\theta > \theta_0$, $M > \theta_0$ 是零概率区域, 同时 $M \leq c$ 在共同支持上给出剩余最强区域。因嵌套支持结构使这一区域对每个两侧备择同时最强, 故为 UMP。其功效分别为 1 ($\theta \leq c$)、 $(c/\theta)^n$ ($c < \theta < \theta_0$) 及 $1 - (\theta_0/\theta)^n + (c/\theta)^n$ ($\theta > \theta_0$)。

第二十四章 2026 年春季概率论与数理统计博士资格考试

考试范围提醒

Q2 的打印题面与唯一完整解答见 第 7 章的三角阵列 CLT 真题（第 53 页）。

本卷共 11 题，选做 8 题；四题各 15 分，其余各 10 分。以下按原卷顺序完整收录。

24.1 第 1 题：高斯测度的指数倾斜（原卷 p.1, 10 分）

2026 春 Q1

English original. Let $X = (X_1, \dots, X_n)^T$ be an $\mathbb{R}^n$ -valued Gaussian vector under P , with mean μ and covariance V . For $\alpha \in \mathbb{R}^n$, define Q by

$$\frac{dQ}{dP} = \frac{\exp(\alpha^T X)}{\mathbb{E}_P[\exp(\alpha^T X)]}.$$

Show that X is still Gaussian under Q and identify its mean and covariance.

中文翻译. 对高斯向量作线性指数倾斜，证明新测度下仍为高斯并求其参数。

完整解答

高斯矩母函数为 $M_P(t) = \exp(t^T \mu + \frac{1}{2} t^T V t)$ 。对任意 $t \in \mathbb{R}^n$,

$$M_Q(t) = \mathbb{E}_Q e^{t^T X} = \frac{M_P(t + \alpha)}{M_P(\alpha)} = \exp \left[t^T (\mu + V \alpha) + \frac{1}{2} t^T V t \right].$$

这是均值 $\mu + V\alpha$ 、协方差 V 的高斯矩母函数，故结论成立。证明对奇异 V 也有效；不需要假设密度存在。

24.2 第 3 题：随机连接面条的环数（原卷 p.2, 10 分）

2026 春 Q3

English original. There are 100 noodles, hence 200 free ends. Repeatedly choose two available free ends uniformly and connect them until none remain. The noodles form disjoint loops. Compute the expected number of loops.

中文翻译. 随机配对 100 根面条的 200 个端点, 求最终闭环数量的期望。

完整解答

当尚有 m 条开链时共有 $2m$ 个自由端。选定第一个端点后, 余下 $2m - 1$ 个端点中恰有一个属于同一条开链; 选中它就闭合一个环, 概率为 $1/(2m - 1)$ 。无论闭合还是连接两条不同开链, 操作后开链数都从 m 减为 $m - 1$ 。由指示变量线性性,

$$\mathbb{E}L = \sum_{m=1}^{100} \frac{1}{2m-1}.$$

各阶段是否闭环并不独立, 但期望计算不需要独立性。

24.3 第 4 题: Brownian 指数积分的 Laplace 原理 (原卷 p.2, 10 分)

2026 春 Q4

English original. Let B_t be standard Brownian motion. Find and prove the distributional limit of

$$\left[\int_0^{T^2} \exp(B_t) dt \right]^{1/T} \quad \text{as } T \rightarrow \infty,$$

and write the answer as a function of a standard normal distribution.

中文翻译. 求长时间 Brownian 指数积分取 T 次根后的极限分布。

完整解答

Brownian 缩放给 $\{B_{T^2s}/T : 0 \leq s \leq 1\} \stackrel{d}{=} \{B_s : 0 \leq s \leq 1\}$, 所以对数除以 T 与

$$\frac{2 \log T}{T} + \frac{1}{T} \log \int_0^1 e^{TB_s} ds$$

同分布。对任意连续函数 w , Laplace 原理 $T^{-1} \log \int_0^1 e^{Tw(s)} ds \rightarrow \max_{s \leq 1} w(s)$: 上界由积分不超过 $e^{T \max w}$, 下界由最大点附近的正长度区间得到。故连续映射定理给原变量收敛到 $\exp(\sup_{s \leq 1} B_s)$ 。反射原理给 $\sup_{s \leq 1} B_s \stackrel{d}{=} |Z|$, $Z \sim N(0, 1)$, 最终

$$\left[\int_0^{T^2} e^{B_t} dt \right]^{1/T} \xrightarrow{d} e^{|Z|}.$$

24.4 第 5 题: Galton–Watson 过程基础 (原卷 p.2, 10 分)

2026 春 Q5

English original. For a Galton–Watson process $Z_0 = 1$, $Z_{n+1} = \sum_{j=1}^{Z_n} X(j, n)$, let offspring mean be $m < \infty$, probabilities p_k , with $p_0 > 0$, $p_0 + p_1 < 1$, and pgf $f(s) = \mathbb{E}s^X$. Let $\mathcal{F}_n = \sigma(Z_0, \dots, Z_n)$. (a) Compute $\mathbb{E}(Z_{n+1} | \mathcal{F}_n)$, $\mathbb{E}Z_n$, $\mathbb{E}(s^{Z_{n+1}} | \mathcal{F}_n)$ and $\mathbb{E}s^{Z_n}$. (b) Determine the number of solutions of $f(s) = s$ on $[0, 1]$. (c) For $\tau = \inf\{n \geq 1 : Z_n = 0\}$, compute $\Pr(\tau < \infty)$ as a function of the smallest fixed point q .

中文翻译. 计算分枝过程的条件矩与概率母函数迭代, 分析不动点并求灭绝概率。

完整解答

条件于 Z_n , 下一代是 Z_n 个独立子代数之和, 故

$$\mathbb{E}(Z_{n+1} | \mathcal{F}_n) = mZ_n, \quad \mathbb{E}Z_n = m^n,$$

$$\mathbb{E}(s^{Z_{n+1}} | \mathcal{F}_n) = f(s)^{Z_n}, \quad \mathbb{E}s^{Z_n} = f^{\circ n}(s).$$

令 $g = f - s$ 。假设保证 f 严格凸, $g(0) = p_0 > 0$, $g(1) = 0$, $g'(1) = m - 1$ 。若 $m \leq 1$, 严格凸性给 $[0, 1]$ 上 $g > 0$, 只有固定点 1; 若 $m > 1$, 恰有两个固定点 $q \in (0, 1)$ 与 1。最后 $\Pr(Z_n = 0) = f^{\circ n}(0) \uparrow q$, 而一旦为零即永远为零, 故

$$\Pr(\tau < \infty) = q.$$

24.5 第 6 题: 分枝树继承性质与渗流临界值 (原卷 pp.2–3, 15 分)

2026 春 Q6

English original. A rooted-tree property is inherited if it holds for every finite tree and, whenever it holds for a tree, it holds for every child-rooted subtree. (a) For a Galton–Watson tree T , prove that conditioned on non-extinction, an inherited property A has probability 0 or 1. (Hint: use $s \leq f(s)$ implies $s \in [0, q] \cup \{1\}$.) (b) If the offspring mean $m > 1$, perform independent bond percolation with density p and define

$$p_c(T) = \sup_p \{p : \Pr(|C_0| = \infty | T) = 0\}.$$

Compute $p_c(T)$ conditioned on non-extinction.

中文翻译. 证明分枝树继承事件在存活条件下的零一律, 并求超临界 Galton–Watson 树的根簇渗流临界概率。

完整解答

记 $s = \Pr(T \in A)$ 。所有有限树属于 A , 故灭绝事件包含于 A , 于是 $s \geq q$ 。若整树有 A , 则根的每棵子树都有 A ; 条件于子代数并用子树独立同分布, 得 $s \leq \mathbb{E}s^X = f(s)$ 。提示中的凸性结

论给 $s \in [0, q] \cup \{1\}$, 结合 $s \geq q$ 可知 $s = q$ 或 1。故

$$\Pr(A \mid \text{non-extinction}) = \frac{s-q}{1-q} \in \{0, 1\}.$$

对固定 p , 开放根簇在退火分布下是子代经 $\text{Bernoulli}(p)$ 稀疏后的分枝过程, 均值为 pm , 故其存活概率在 $pm \leq 1$ 时为零, 在 $pm > 1$ 时为正。事件“给定树后开放簇存活概率为零”是继承性质; 利用上面的零一律并先对有理 p 同时取概率一事件, 再用关于 p 的单调性, 得到在原树存活条件下

$$p_c(T) = 1/m \quad \text{a.s.}$$

24.6 第 7 题: 变换指数族的 MLR 与 UMP (原卷 p.3, 10 分)

2026 春 Q7

English original. Let X_i be i.i.d. with density $f(x; \theta) = (2\theta\sqrt{x})^{-1}e^{-\sqrt{x}/\theta}\mathbf{1}_{\{x>0\}}$, $\theta > 0$.

(a) Show the family has a monotone likelihood ratio in a statistic T . (b) Find a size- α UMP test for $H_0: \theta \geq 1$ versus $H_1: \theta < 1/2$, with an explicit rejection region.

中文翻译. 通过平方根变换识别指数尺度族, 证明单调似然比并构造复合单侧 UMP 检验。

完整解答

令 $Y_i = \sqrt{X_i}$, Jacobian 给 Y_i 密度 $\theta^{-1}e^{-y/\theta}$, 故 $T = \sum_i \sqrt{X_i}$ 且似然正比于 $\theta^{-n}e^{-T/\theta}$ 。对 $\theta_2 > \theta_1$, 似然比 $L(\theta_2)/L(\theta_1)$ 随 T 严格递增, 因此有 MLR。

小 θ 使 T 偏小, 故拒绝域为 $T \leq c_\alpha$ 。在零假设 $\theta \geq 1$ 中拒绝概率在边界 $\theta = 1$ 最大, 而 $2T/\theta \sim \chi_{2n}^2$, 所以

$$\text{拒绝 } H_0 \iff 2 \sum_{i=1}^n \sqrt{X_i} \leq \chi_{2n, \alpha}^2.$$

Karlin-Rubin 定理给 UMP; 备择写成 $\theta < 1/2$ 不改变最强方向。

24.7 第 8 题: 离散分布的样本分位数相合性 (原卷 p.3, 15 分)

2026 春 Q8

English original. Let $X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(\theta)$ and $\hat{\xi}_p = \inf\{x : F_n(x) \geq p\}$. (a) Does $\hat{\xi}_p$ converge in probability to the population quantile ξ_p for every $0 < p < 1$? Prove. (b) Can a non-degenerate discrete distribution have $\hat{\xi}_p \rightarrow \xi_p$ in probability for all $0 < p < 1$? Prove.

中文翻译. 检查样本分位数在分布函数跳点处的相合性, 并推广到一般非退化离散分布。

完整解答

Bernoulli 分布满足 $\xi_p = 0$ 当 $p \leq 1 - \theta$, 否则为 1; 样本分位数为零当且仅当 $p \leq 1 - \bar{X}_n$ 。

若 $p \neq 1 - \theta$, 大数律给相合; 但在 $p = 1 - \theta$,

$$\Pr(\hat{\xi}_p = 0) = \Pr(\bar{X}_n \leq \theta) \rightarrow \frac{1}{2}$$

(由 CLT; 格点处的原子质量趋零), 而总体分位数为 0, 故不相合。

一般非退化离散分布必存在 x 使 $0 < F(x) < 1$ 且 $\Pr(X = x) > 0$ 。取 $p = F(x)$, 则 $\xi_p \leq x$, 而经验比例 $F_n(x)$ 以均值 p 波动, $\Pr(F_n(x) < p) \rightarrow 1/2$; 在该事件上样本分位数严格大于 x 。故不可能对全部 p 都相合。

24.8 第 9 题: 两点先验的 Bayes 与 minimax 规则 (原卷 p.3, 10 分)

2026 春 Q9

English original. Let $Y | \mu \sim N(\mu, 1)$ and $\pi(\mu) = 0.5\mathbf{1}_{\{\mu=-1\}} + 0.5\mathbf{1}_{\{\mu=1\}}$. The action space is $\{-1, 1\}$. (a) Find the posterior and Bayes rule under absolute loss. (b) Is the rule minimax? Justify.

中文翻译. 在正态观测与对称两点先验下求绝对损失 Bayes 分类规则, 并验证 minimax 性。

完整解答

后验优势比为

$$\frac{\Pr(\mu = 1 | y)}{\Pr(\mu = -1 | y)} = \exp \left\{ -\frac{(y-1)^2}{2} + \frac{(y+1)^2}{2} \right\} = e^{2y}.$$

故 $\Pr(\mu = 1 | y) = e^{2y}/(1 + e^{2y})$ 。绝对损失下选后验中位数, 因而 $\delta(y) = 1$ 当 $y \geq 0$, 否则 -1 ($y = 0$ 任取)。在 $\mu = 1$ 下风险为 $2\Pr_1(Y < 0) = 2\Phi(-1)$; 在 $\mu = -1$ 下由对称性相同。该规则风险在参数空间上常数, 且其均匀先验 Bayes 风险等于最大风险, 所以 Bayes 下界与最大风险相等, 故为 minimax。

24.9 第 10 题: Poisson 平方均值参数的 UMVU 与 MLE (原卷 p.4, 10 分)

2026 春 Q10

English original. Let X_i be i.i.d. $\text{Poisson}(\sqrt{\lambda})$, $\lambda > 0$. (a) Find the BUE (UMVUE) $\tilde{\lambda}$ for λ . (b) If $\hat{\lambda}$ is the MLE, show $\sqrt{n}(\tilde{\lambda} - \hat{\lambda}) \rightarrow 0$ in probability.

中文翻译. 对 Poisson 均值的平方参数求 UMVU, 并证明它与不变 MLE 在根 n 尺度渐近等价。

完整解答

令 $T = \sum_i X_i \sim \text{Poisson}(n\sqrt{\lambda})$, 它是完全充分统计量。因 $\mathbb{E}[T(T-1)] = (n\sqrt{\lambda})^2$,

$$\tilde{\lambda} = \frac{T(T-1)}{n^2}$$

无偏且为 UMVU。MLE 不变性给 $\hat{\lambda} = (T/n)^2$ ，于是

$$\sqrt{n}(\tilde{\lambda} - \hat{\lambda}) = -\frac{T}{n^{3/2}} \xrightarrow{\mathbb{P}} 0,$$

因为 $T/n \xrightarrow{\mathbb{P}} \sqrt{\lambda}$ 。这里的差异是精确的 $-T/n^2$ ，无需作 Taylor 展开。

24.10 第 11 题：偏斜正态模型（原卷 p.4, 15 分）

2026 春 Q11

English original. Let $Y = \sigma(\rho|U| + \sqrt{1-\rho^2}V)$, where U, V are independent $N(0, 1)$, $\rho \in (-1, 1)$ and $\sigma > 0$. (a) Show

$$f(y | \rho, \sigma) = \frac{a}{\sqrt{2\pi\sigma^2}} e^{-y^2/(2\sigma^2)} \Phi\left(\frac{\rho y}{b\sigma\sqrt{1-\rho^2}}\right)$$

and find positive a, b . (b) For i.i.d. observations, does an MLE in $(-1, 1) \times (0, \infty)$ always exist? (c) For $H_0 : \rho = 0$ versus $\rho \neq 0$, find c such that rejection by $\sqrt{n}|\bar{Y}|/s_Y > c$ has size 0.05.

中文翻译. 推导偏斜正态密度，分析开放参数空间上的 MLE 存在性，并在对称零假设下给出精确 t 检验临界值。

完整解答

标准偏斜正态的条件积分给

$$f(y | \rho, \sigma) = \frac{2}{\sigma} \phi(y/\sigma) \Phi\left(\frac{\rho y}{\sigma\sqrt{1-\rho^2}}\right),$$

故 $a = 2, b = 1$ 。可由对 $(|U|, V)$ 作线性变换并对半正态变量积分直接验证。

MLE 不总存在。例如所有 $y_i > 0$ 时，对固定 σ ，每个 Φ 因子随 ρ 严格增加，似然上确界在 $\rho \uparrow 1$ 的边界取得，而参数空间不含 1；所有观测为正的事件具有正概率。因此不能声称总有内部最大值（全负样本类似趋于 -1 ）。

在 H_0 下 $Y = \sigma V \sim N(0, \sigma^2)$ 。若 $s_Y^2 = (n-1)^{-1} \sum (Y_i - \bar{Y})^2$ ，则 $\sqrt{n}\bar{Y}/s_Y \sim t_{n-1}$ ，所以

$$c = t_{n-1, 0.975}.$$

若采用分母为 n 的样本标准差，临界值需相应乘以 $\sqrt{n/(n-1)}$ ；原题未明示约定，答案必须说明。

索引

- Basu 定理, 86
- Bayes rule, 109
- Bayesian 推断, 86
- Borel–Cantelli 引理, 2
- Brownian motion, 59, 82
- Cramer–Rao 下界, 96
- Cramér–Rao 下界, 86
- Fisher 信息, 86
- Gauss–Markov 定理, 115
- Lehmann–Scheffe 定理, 96
- Markov 链, 59, 82
- minimax, 109
- MLE 渐近正态, 109
- Neyman–Pearson 引理, 86, 109
- Poisson 过程, 59
- Rao–Blackwell 定理, 96
- UMP 检验, 109
- Wald 等式, 82
- 一致可积, 2
- 一般线性假设, 115
- 中心极限定理, 2
- 似然比检验, 109
- 充分统计量, 86, 96
- 分枝过程, 59, 82
- 可估函数, 115
- 可选停止, 59
- 可选停止定理, 82
- 完备性, 86, 96
- 帽子矩阵, 115
- 强大数律, 2
- 方差分析, 115
- 条件期望, 2
- 概率空间, 2
- 渐近统计, 86
- 线性模型, 86, 115
- 置信区间, 96
- 鞅, 59, 82
- 鞅收敛, 59