



Probability Theory and Statistics



22 Summary and Projects

张绍群

2024/12/17

关于成绩评定

□ 成绩构成

- 40% 平时成绩 + 20% 期中成绩 + 40% 期末成绩
- 平时成绩：每次作业百分制评定，13次作业取均值作为平时成绩
- 期中成绩：笔试
- 期末成绩：笔试

□ 关于期末考试

- 时间：1月8日上午 10:30 — 12:30
- 地点：南雍-西410 (22级+23级 ≤ 178),
南雍-西446 ($179 \leq 23级 \leq 332$),
南雍-西426 ($333 \leq 23级$),
- 题型：A/B 卷, 10 道选择题 + 5 道填空题 + 5 道计算题
- 考试范围：10 章 (去除第七章集中不等式) 所有的内容，包含例题、作业、思考题
- 难度比例：60%：30%：10% = 利用：组合：探索

章节重点

□ 第一章 组合计数：古典概型、几何概型、十二重计数

- **[夫妻匹配问题]**将 n 对夫妻任意分成 n 组,每组一男一女/每组2人,不限男女,问
 - ✓ 所有夫妻都没有分到同一组的概率?
 - ✓ 至少有一对夫妻被分到同一组的概率?
 - ✓ 每一对夫妻都恰好分到同一组的概率?
 - ✓ 用 x 表示夫妻两人被分到一组的对数, 求 x 的期望 (期中考试)
- **[生日问题]**有 K 个人($K < 365$), 每个人的生日等可能地出现于365天中的任意一天.
- **[抽签问题 (是否放回)]**袋中有 a 个不同的白球, b 个不同的红球,假设有 k 个人依次随机有放回地/无放回地从袋中取一个球,问第 i 个人($i \leq k$) 取出红球的概率?
- **[会面问题]**两银行经理约定中午12:00 - 13:00到某地会面,两人到达时间随机,先到者等另一人15分钟后离开.求两人见面的概率.
- **[十二重计数]** (直线、环、全) 排列、整数的 (有序、无序) 分拆

章节重点

□ 第二章 条件概率与独立性：条件概率、全概率公式、贝叶斯公式、独立性

- **条件概率公式:** 缩小了有效样本空间
 - ✓ 容斥原理
 - ✓ 乘法形式
- **独立性**
 - ✓ 独立性 vs 互斥, 如何判断独立性?
 - ✓ 小概率原理
- **全概率公式:** 若知道各种原因 $P(A_i)$ 、在该原因下事件B发生的概率 $P(B|A_i)$,此时利用全概率公式计算概率 $P(B)$.
- **贝叶斯公式:** 若知道各种原因 $P(A_i)$ 、在该原因下事件B发生的概率 $P(B|A_i)$,若结果事件B已经发生, 利用贝叶斯公式探讨是由某原因 A_i 导致该结果的概率 $P(A_i|B)$.
- **典型例题:** 相关计算、三囚徒问题、多项式相等、大矩阵乘法

章节重点

□ 第三章 离散型随机变量：分布列、数字特征、常用变量

- **分布列**: 随机变量的取值和概率, 可以完全刻画其概率属性

- **数字特征**

 - 期望

 - 方差

X	x_1	x_2	$\dots$	x_n	$\dots$
P	p_1	p_2	$\dots$	p_n	$\dots$

- **常用的变量**

 - 0-1分布 $X \sim \text{Ber}(p)$

 - 二项分布 $X \sim B(n, p)$

 - 泊松分布 $X \sim P(\lambda)$

 - 几何分布 $X \sim G(p)$: 无记忆性

- **典型案例**: 相关计算、估计德国坦克数量、集卡活动、Gambling

章节重点

□ 第四章 连续型随机变量：分布函数、密度函数、数字特征、常用变量

- 分布函数: $F(x) = P(X \leq x)$, 单调性、规范性、右连续性
- 密度函数: 指向连续型随机变量的概率, 概率密度 $f(x)$ 越大, 则 X 在 x 附近取值的概率越大.
- 数字特征: 积分中的变量代换、几个重要的不等式估算
 - 期望
 - 方差
- 常用的变量
 - 均匀分布 $X \sim U(a, b)$
 - 指数分布 $X \sim e(\lambda)$: 无记忆性, 是唯一具有无记忆性的连续型随机变量
 - 正态分布 $X \sim N(\mu, \sigma^2)$
- 典型案例: 相关计算、估计德国坦克数量、集卡活动、Gambling

$$P(|x - \mu| < \sigma) = 0.6826$$

$$P(|x - \mu| < 2\sigma) = 0.9544$$

$$P(|x - \mu| < 3\sigma) = 0.9974$$

章节重点

□ 第五/六章 多维随机向量：

- 联合分布函数和边缘分布函数: $F(x, y) = P(X \leq x, Y \leq y)$, $F_x(x) = P(X \leq x, y < +\infty)$
 - ✓ 性质: 单调性、规范性、右连续性、有界概率性 (矩形运算)
 - ✓ 边缘分布和联合分布的计算关系
- 分布列和密度函数:
 - ✓ 分布列: 逐行和逐列的计算、根据独立性有边缘概率乘积构成分布列
 - ✓ 密度函数: 积分中的变量代换、
- 独立性
 - ✓ 性质 $f(x, y) = f_X(x)f_Y(y)$ 和独立性判定 $f(x, y) = g(x)h(y)$
- 数字特征: 期望、协方差、相关系数
- 条件分布和条件期望: 密度函数的贝叶斯公式、全期望公式
- 多维随机变量的运算: 加减乘除 (卷积公式)、最大值和最小值、复合函数
- 典型案例: 相关计算、二维正态分布

表 5.1 二维随机向量的概率分布表

X Y	y_1	y_2	$\cdots$	y_j	$\cdots$	$p_{i\cdot} = \sum_j p_{ij}$
x_1	p_{11}	p_{12}	$\cdots$	p_{1j}	$\cdots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	$\cdots$	p_{2j}	$\cdots$	$p_{2\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	$\cdots$	p_{ij}	$\cdots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\ddots$	$\vdots$
$p_{\cdot j} = \sum_i p_{ij}$	$p_{\cdot 1}$	$p_{\cdot 2}$	$\cdots$	$p_{\cdot j}$	$\cdots$	1

章节重点

□ 第八章 大数定律及中心极限定理：

• 四种收敛方式

收敛方式小结 (思考题)

考虑 $X_n = f_n : \mathcal{X} \rightarrow \mathcal{Y}$, 有如下四种收敛方式

- 一致收敛: $f_n \rightarrow f$
- 点态收敛: $f_n \xrightarrow{\cdot} f$
- 依概率收敛: $X_n \xrightarrow{P} X$
- 依分布收敛: $X_n \xrightarrow{d} X$

• 大数定律: $X_n \xrightarrow{P} a$

✓ 大数定律基本式 $\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i]$

✓ 大数定律小结

• Markov 大数定律:

若随机变量序列 $\{X_i\}$ 满足 $\frac{\mathbb{V}\mathbb{A}\mathbb{R}(\sum_{i=1}^n X_i)}{n^2} \rightarrow 0$, 则满足大数定律.

• Chebyshev 大数定律:

若独立随机变量序列 $\{X_i\}$ 满足 $\mathbb{V}\mathbb{A}\mathbb{R}(X_i) \leq c$, 则满足大数定律.

• Khintchine 大数定律:

若独立同分布随机变量序列 $\{X_i\}$ 期望存在, 则满足大数定律.

• Bernoulli 大数定律:

对二项分布 $X_n \sim \text{Ber}(n, p)$, 有 $\frac{X_n}{n} \xrightarrow{P} p$.

• 中心极限定理: $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} \xrightarrow{d} N(0,1)$

• 棣莫弗-拉普拉斯中心极限定理:

随机变量独立且同伯努利/二项分布. 若 $X_n \sim B(n, p)$, 则

$$X_n \xrightarrow{d} \mathcal{N}(np, np(1-p))$$

• 林德贝格-勒维中心极限定理:

随机变量独立同分布. 若 $\mathbb{E}[X_k] = \mu$ 和 $\mathbb{V}\mathbb{A}\mathbb{R}(X_k) = \mathbb{V}\mathbb{A}\mathbb{R}^2$, 则

$$\sum_{k=1}^n X_i \xrightarrow{d} \mathcal{N}(n\mu, n\mathbb{V}\mathbb{A}\mathbb{R}^2)$$

• 李雅普诺夫定理:

随机变量独立不同分布.

章节重点

□ 第九/十/十一章 数理统计：

- **常用统计量**: 样本均值、样本方差 (无偏)、样本矩 (原点矩、中心矩)、次序统计量
- **三大抽样分布**: 来源 (统计量构造)、数字特征、上侧分位数 (正态分布、三大抽样分布)
- **参数估计 — 点估计**
 - ✓ 矩估计 (总体分布形式未知的情况) 和极大似然估计 (总体分布形式已知、需要假设)
 - ✓ 如何评价估计? 无偏性 (原点矩、中心矩)、有效性 (有效估计量)、一致性 (判别定理)
- **参数估计 — 区间估计**
 - ✓ 置信区间、置信度 (双侧和单侧的计算区别)
 - ✓ 总体为正态分布下的区间估计 (μ 和 σ 分别已知和未知的情况)
 - ✓ 总体为二维正态分布下的均值之差、方差之比的区间估计 (μ 和 σ 分别已知和未知的情况)
 - ✓ 非正态的区间估计: 集中不等式和中心极限定理
- **假设检验**:
 - ✓ 建立假设 — 给出拒绝域 — 由样本统计量做判断
 - ✓ 检验的两类错误
 - ✓ 非参假设检验 — 分布的 χ^2 拟合优度检验

关于期末考试

□ 考试内容及分值比例

- 第 1-4 章: 20 分左右
- 第 5-6 章: 30 ~ 40 分
- 第 8 章: 10 分左右
- 第 9-11 章: 30 ~ 40 分

□ 倒计时还有？天

- 12.24 习题课
- 12.26 全助教答疑
- 12.26 – 01.08



一些建议

□ 关于学业.

- 专业体系方面没有什么可说的
- 掌握自学能力、实践能力 (比如自学《概率论与数理统计》等, 自学编程等)
- 大量地阅读, 做好通识教育, ...
- 锻炼好身体 (一把辛酸泪)

□ 关于思维.

- 保持好奇、保持行动力. 感兴趣的事情要尽快去了解、尝试
- 不要被课堂束缚！不要被校区束缚！不要被南大束缚！要被专业束缚！
- 要敢想敢做 ...

□ 关于心态. 任何时候不要自我放弃！

写在最后