📓 概率与统计 [补]

一、引言：为什么需要概率？（教材第一章 + 课程背景）

◇ 1.1 不确定性是现实世界的核心

在科学研究、工程决策、金融预测、医学诊断等领域，我们面对的现象往往**无法精确预测**，原因包括：

- **变量太多或不可控**（如天气、股市）
- **内在随机性**（如量子现象、放射性衰变）
- **测量误差**（仪器精度限制）
- **人为行为的不可预测性**（如消费者选择）

✅ **核心思想**：我们不能“确定”结果，但可以**量化不确定性**——这正是**概率论**的使命。

◇ 1.2 概率建模的三大要素（Three Ingredients of a Probabilistic Model）

来自教材第一章与 *25-09-03.pdf* 的整合：

要素	定义	数学表示	示例
1. 样本空间（Sample Space）	所有可能结果的集合	Ω	抛一枚硬币：$\Omega=\{H,T\}$
2. 事件（Events）	样本空间的子集，表示感兴趣的“结果集合”	$A\subseteq\Omega$	“出现正面”：$A=\{H\}$
3. 概率测度（Probability Measure）	给每个事件赋予一个 $[0,1]$ 区间内的数值，表示其发生的可能性	$P(A)$	$P(\{H\})=0.5$

⚠️ 注意：**样本空间 ≠ 所有事件**。事件是样本点的集合。

◇ 1.3 样本空间的类型

类型	特点	示例
有限样本空间	结果数量有限	抛 2 枚硬币：$\Omega=\{(H,H),(H,T),(T,H),(T,T)\}$
无限可数样本空间	结果无限但可列	掷骰子直到第一次出现 6：$\Omega=\{1,2,3,\dots\}$
无限不可数样本空间	结果连续，不可列	测量人的身高：$\Omega=[0,\infty)$ 或 $[0,300]$ cm

📌 来自 *25-09-03.pdf* 的例子：

- 抛 7 次硬币：$\Omega=\{(\omega_1,\dots,\omega_7)\mid\omega_i\in\{H,T\}\}$，共 $2^7=128$ 个结果

- 掷一个骰子：$\Omega = \{1,2,3,4,5,6\}$
- 轮盘赌（roulette）：$\Omega = \{0,1,2,\ldots,36\}$

二、事件与集合论（Set Theory for Events）

来自 *25-09-03.pdf* 和教材第 2 章

事件是**集合**，因此我们可以使用**集合运算**来描述复杂事件。

◇ 2.1 基本集合运算（Venn Diagrams）

运算	符号	含义	解释
并集 (Union)	$A \cup B$	A 或 B 发生	至少一个发生
交集 (Intersection)	$A \cap B$	A 且 B 发生	同时发生
补集 (Complement)	A^c 或 $\overline{A}$	A 不发生	所有不在 A 中的结果
差集 (Set Difference)	$A \backslash B$	A 发生但 B 不发生	$A \cap B^c$
空集	$\emptyset$	不可能事件	无任何结果满足
全集	Ω	必然事件	所有可能结果

📌 重要性质：

- $A \cup \emptyset = A$
- $A \cap \emptyset = \emptyset$
- $A \cup \Omega = \Omega$
- $A \cap \Omega = A$
- $A \cup A^c = \Omega$，$A \cap A^c = \emptyset$

◇ 2.2 互斥事件（Disjoint Events）

- 若 $A \cap B = \emptyset$，则称 A 与 B **互斥**（mutually exclusive）
- 表示：两个事件**不可能同时发生**

✅ 示例：掷一个骰子

- $A =$ “偶数” = $\{2,4,6\}$
- $B =$ “奇数” = $\{1,3,5\}$
- 则 $A \cap B = \emptyset$，互斥

◇ 2.3 德摩根定律（De Morgan's Laws）

来自 *25-09-03.pdf* 的重点公式

公式	含义
$(A \cup B)^c = A^c \cap B^c$	“A 或 B 不发生” = “A 不发生 且 B 不发生”
$(A \cap B)^c = A^c \cup B^c$	“A 且 B 不发生” = “A 不发生 或 B 不发生”

✅ 推广到多个事件：

$$\left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c, \quad \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c$$

💡 应用场景：证明“至少一个发生”的补事件是“全部都不发生”

◇ 2.4 事件的代数表达（来自 *25-09-03.pdf* 练习）

设抛 7 次硬币，定义事件：

- B_m = “第 m 次抛出正面”（$m = 1, \ldots, 7$）

尝试用集合运算表达以下事件：

事件描述	集合表达
第一次是正面，第二次是反面	$B_1 \cap B_2^c$
前两次都是正面	$B_1 \cap B_2$
至少有一次是正面	$\bigcup_{m=1}^{7} B_m$
全是反面	$\bigcap_{m=1}^{7} B_m^c = (\bigcup_{m=1}^{7} B_m)^c$
恰好两次是正面	所有恰好包含两个 H 的 7 元组集合（较复杂，需组合）

⚠️ 提示：这类问题需要先写出样本空间结构，再筛选满足条件的结果。

三、概率的公理化定义（Kolmogorov Axioms）

来自教材第 2 章 + *25-09-03.pdf* 的思想延伸

概率是一个函数 P:Events $\to [0,1]$，满足以下三条公理：

☑ 概率公理（Three Axioms）

1. **非负性**：对任意事件 A，有 $P(A) \geq 0$
2. **规范性**：$P(\Omega) = 1$（必然事件概率为 1）
3. **可加性**（有限或可列）：
   - 若 $A_1, A_2, \ldots$ 两两互斥，则

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

📌 推论（教材与课堂共同强调）：

- $P(\emptyset) = 0$
- $P(A^c) = 1 - P(A)$
- 若 $A \subseteq B$，则 $P(A) \leq P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$（容斥原理）

◇ 3.1 经典概率模型（等可能结果）

当样本空间有限且**每个结果等可能**时，使用：

$$P(A) = \frac{|A|}{|\Omega|}$$

✅ 示例：掷两个骰子，$\Omega = \{(i,j) \mid i,j = 1,\dots,6\}$，共 36 种结果

- $A =$ “点数和为 7” $= \{(1,6),(2,5),\dots,(6,1)\}$，共 6 个
- $P(A) = \frac{6}{36} = \frac{1}{6}$

◇ 3.2 概率的解释

解释	描述
频率解释	重复实验中事件发生的相对频率趋于稳定值
主观解释	个人对事件发生可能性的信念程度
古典解释	基于对称性和等可能性的推理（如骰子、硬币）

📌 教材强调：**概率模型是对现实的近似**，不是绝对真理。

四、基本计数原理（Elementary Combinatorics）

> 来自课程大纲（Week 1-2）和教材第 2 章

在计算概率时，常需计算样本空间大小或事件中结果的数量。

◇ 4.1 乘法原理（Multiplication Rule）

若一个过程分 k 步，第 i 步有 n_i 种选择，则总方案数为：

$$n_1 \times n_2 \times \cdots \times n_k$$

✅ 示例：抛 3 次硬币 → $2 \times 2 \times 2 = 8$ 种结果

◇ 4.2 排列与组合

概念	公式	说明
排列（有序）	$P(n,k) = \frac{n!}{(n-k)!}$	从 n 个中选 k 个并排序
组合（无序）	$\binom{n}{k} = \frac{n!}{k!\,(n-k)!}$	从 n 个中选 k 个不排序

✅ 示例：从 10 人中选 3 人组成委员会（无序）→ $\binom{10}{3} = 120$

◇ 4.3 应用：生日问题（Birthday Problem）

教材经典例子：房间中有 n 人，至少两人同一天生日的概率？

- 假设 365 天，忽略闰年
- 总方式：365^n
- 无人同一天：$P(365,n) = 365 \times 364 \times \cdots \times (365 - n + 1)$
- 无人同一天的概率：$\frac{P(365,n)}{365^n}$
- 至少两人同一天：$1 - \frac{P(365,n)}{365^n}$

📌 惊人结论：当 $n = 23$ 时，概率 > 50%！

五、现实案例与建模思想（来自教材第一章）

◇ 5.1 挑战者号航天飞机灾难（Challenger Disaster）

来自 *AMODER~1.PDF* 第 1 章与 *25-09-03.pdf* 提及

- 事故原因：低温下 O-ring 失效
- 数据显示：发射温度越低，O-ring 损伤越严重
- 但决策者忽略了**概率性趋势**，仅因“过去低温也成功发射”而冒险

✅ 教训：**不能仅凭“没出事”就认为安全**，必须分析**风险随条件变化的趋势**

◇ 5.2 老忠实间歇泉（Old Faithful Geyser）

来自教材多次提及（如 p.207）

- 记录喷发持续时间与下次喷发等待时间
- 数据呈现**双峰分布**（bimodal）
  - 短喷发 → 短等待
  - 长喷发 → 长等待
- 说明：不能只用“平均值”概括，需用**图形化分析**（如直方图、散点图）

✅ 教训：**数据可视化是探索性分析的第一步**

六、本阶段核心能力总结

能力	说明
✅ 定义样本空间	能为实验写出所有可能结果
✅ 描述事件为集合	用集合语言表达“至少”、“恰好”、“都不”等逻辑
✅ 使用集合运算	并、交、补、差、德摩根律
✅ 计算简单概率	特别是等可能模型下的计数
✅ 理解概率公理	明白概率的数学基础

能力	说明
✅ 应用组合数学	解决排列组合问题
✅ 建模现实问题	将语言描述转化为数学事件

七、学习建议与资源

◇ 1. 教材使用方法（来自 *AMODER~1.PDF* 前言）

- **Quick Exercises**：每章穿插的 2-3 分钟小题，务必动手做
- **Exercises**：每章约 13 题，一半有答案（Appendix C），一半有完整解答（Appendix D）
- **Remark 框**：可跳过，适合进阶学习者
- **数据集**：官网提供所有数据：*Springer 链接*

◇ 2. 课程支持资源（来自 *25-09-01.pdf*）

- **图书馆电子资源**：24/7 访问电子书、数据库
- **ARC 学术资源中心**：提供数学辅导
- **Office Hours**：Mo 4-5PM，W925，或预约 Zoom
- **考试政策**：开卷，可带手写笔记，**不可带打印材料或书**

八、下一步学习方向（预告）

根据课程大纲（Tentative Calendar）：

周次	主题
Week 3-4	条件概率、独立性、贝叶斯定理
Week 5-6	随机变量、期望、方差
Week 7-8	联合分布、协方差
Week 9-10	大数定律、中心极限定理

> ⚠️ 条件概率是**概率论的核心转折点**，务必打好基础！

☑ 总结：本阶段知识图谱

```
概率建模三要素
    ├── 样本空间 Ω（所有可能结果）
    ├── 事件 A（Ω 的子集，用集合表示）
    └── 概率 P(A)（满足三条公理的函数）

事件运算
    ├── 并 ∪、交 ∩、补 ^c
    ├── 互斥 A∩B=∅
    └── 德摩根律 (A∪B)^c = A^c ∩ B^c
```

概率计算

├—— 公理：非负、归一、可加

├—— 等可能模型：P(A) = |A| / |Ω|

└—— 计数工具：乘法原理、排列、组合

现实建模

├—— 挑战者号：忽略概率趋势的代价

└—— 老忠实泉：数据可视化的重要性

20250908

一、概率模型的三大基本要素

任何概率实验的数学建模都由三个核心成分构成，通常记为三元组 $(\Omega, \mathcal{F}, P)$，即**概率空间**。

1. 样本空间（Sample Space）Ω

- **定义**：样本空间 Ω 是一个实验所有可能结果（outcomes）的集合，也称为“全集”或“宇宙”（universe）。
- **表示**：$\Omega = \{\text{所有可能的结果}\}$
- **示例**：
  - 抛一枚硬币：$\Omega = \{H, T\}$
  - 掷一枚骰子：$\Omega = \{1,2,3,4,5,6\}$
  - 连续型实验（如轮盘赌）：$\Omega = [0,1]$ 或 $[0,2\pi]$ 等区间

✅ 文档中出现的 *(H,1),(4, i), 25, by 1/ (T,T)* 应为笔误或识别错误，推测原意是描述抛两枚硬币的结果，如 $(H,H),(H,T),(T,H),(T,T)$。

---

2. 事件空间（Event Space / σ-代数）$\mathcal{F}$

- **定义**：事件是样本空间 Ω 的子集（subset），表示我们关心的某种结果集合。
- **事件空间** $\mathcal{F}$：是所有“可测事件”的集合，即我们能为其分配概率的子集族。
- **构造方式**：
  - **有限样本空间**：若 Ω 有限（如掷骰子），通常取 $\mathcal{F} = \mathcal{P}(\Omega)$，即 Ω 的幂集（所有子集）。
    - 示例：$\Omega = \{1,2,3,4,5,6\}$，则事件如 $\{1\},\{1,3,5\},\{2,4,6\},\emptyset,\Omega$ 都属于 $\mathcal{F}$。
  - **无限样本空间**（尤其是不可数无限）：不能取所有子集（会引发测度论悖论），需通过 σ-代数限制，通常由区间生成（如 Borel σ-代数）。
    - 示例：$\Omega = [0,1]$，事件可以是区间 $[a,b)$，并通过对补集、可数并、可数交等运算封闭生成 $\mathcal{F}$。

✅ 文档中 “EVENTS MANDCED THROUGH OPERATIONS” 应为 “**EVENTS MANAGED THROUGH OPERATIONS**”，指通过集合运算（补、并、交）来构造复杂事件。

- **符号说明**：文档中强调 “Always With &3” 可能是笔误，应为强调事件始终是 Ω 的子集。

---

3. 概率测度（Probability Measure）P

- **定义**：一个函数 $P: \mathcal{F} \to [0,1]$，将每个事件 A 映射到一个实数 $P(A)$，表示该事件发生的可能性。
- **直观解释**：$P(A)$ 可理解为在大量重复实验中，事件 A 出现的频率的极限（频率学派观点），或对事件发生信念的量化（贝叶斯观点）。

二、概率测度的公理化定义（Kolmogorov 公理）

概率函数 P 必须满足以下三条公理：

公理 P1：归一性（Normalization）

$$P(\Omega) = 1$$

- 整个样本空间的概率为 1，表示必然事件。

公理 P2：非负性（Non-negativity）

$$\forall A \in \mathcal{F}, \quad P(A) \geq 0$$

- 任何事件的概率不能为负。

公理 P3：可列可加性（Countable Additivity）

若 $\{A_1, A_2, \ldots\}$ 是一列**两两互斥**（mutually disjoint）的事件（即 $A_i \cap A_j = \emptyset$ 当 $i \neq j$），则：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- 这是概率论中最关键的公理，保证了概率在无限并集下的良好行为。

✅ 文档中"P1) P(u)= 1"中的"u"应为 Ω；"P2) A, Ay:-re ARE MUTRACCY) DISTOIM EVENT,S"明显为"**mutually disjoint events**"的识别错误。

三、基本性质与推论（Theorems）

由上述公理可推导出若干重要性质：

T1：不可能事件的概率为 0

$$P(\emptyset) = 0$$

- **证明**：Ω 与 $\emptyset$ 互斥，且 $\Omega \cup \emptyset = \Omega$，由可加性：
$$P(\Omega) = P(\Omega) + P(\emptyset) \Rightarrow P(\emptyset) = 0$$

T2：补事件的概率

$$P(A^c) = 1 - P(A)$$

- 其中 $A^c = \Omega \backslash A$ 是 A 的补集。
- **证明**：A 与 A^c 互斥且 $A \cup A^c = \Omega$，故：
$$1 = P(\Omega) = P(A) + P(A^c) \Rightarrow P(A^c) = 1 - P(A)$$
- 文档中"VENd:-T Proof?2 1= P(y)= P(b)+ P(B) eP(BY= 1- P1B)"明显为此性质的尝试证明，其中"y"应为 Ω，"b"应为 B。

T3：单调性

若 $A \subseteq B$，则 $P(A) \leq P(B)$

- 因为 $B = A \cup (B \setminus A)$，且 $A \cap (B \setminus A) = \emptyset$，所以：

$$P(B) = P(A) + P(B \setminus A) \geq P(A)$$

T4：加法公式（容斥原理）

对任意两个事件 A, B：

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- 当 $A \cap B = \emptyset$（互斥）时，退化为 $P(A \cup B) = P(A) + P(B)$

  ✅ 文档中 “DISJ0Im: EVENIS = 1) DIE A 1B=q= P(AUB)= PA)+ P(P)A -” 应为 “**Disjoint events:** $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$”

四、集合运算与事件代数（Venn Diagrams & De Morgan's Laws）

1. 集合运算符号

- 并（Union）：$A \cup B$ → “A 或 B 发生”
- 交（Intersection）：$A \cap B$ → “A 且 B 发生”
- 补（Complement）：A^c → “A 不发生”
- 差：$A \setminus B = A \cap B^c$

2. 德摩根定律（De Morgan's Laws）

文档中 “DE MORGAR'S LAWS LAUB)"= An iS 11” 应为：

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

- 这些定律在处理“至少一个不发生”、“都不发生”等逻辑时非常有用。

3. 文氏图（Venn Diagrams）

- 用于直观表示事件之间的关系（包含、交、并、补等）。
- 文档强调：“**NOT a Proof!**” —— 文氏图有助于猜测恒等式，但不能作为严格证明。

五、经典概率模型（等可能概型）

当样本空间有限且所有基本结果等可能发生时，称为**古典概型**。

定义

设 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$，且 $P(\{\omega_i\}) = \frac{1}{n}$ 对所有 i 成立。

则对任意事件 $A \subseteq \Omega$：

$$P(A) = \frac{|A|}{|\Omega|} = \frac{\text{有利结果数}}{\text{总结果数}}$$

示例：公平骰子（Fair Die）

- $\Omega = \{1,2,3,4,5,6\}$，$P(\{i\}) = \frac{1}{6}$
- 事件 $A = \{2,4,6\}$（掷出偶数）：

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

- 文档中 "↑ COUTCOME Every= P132, 4,(3)= P(12sUSY(2351/ (p, P(3D)+ P((y))+1(3x)= E+tz= / TOTAL I OUTCOMES # favorable outcomes" 明显为此类计算的混乱表达。

✅ 此类问题涉及**计数问题（Counting Issues）**，需借助**组合数学（Combinatorics）** 工具（排列、组合、乘法原理等）。

六、非均匀概率模型（非公平实验）

当基本结果不等可能时，不能使用古典概型。

示例：非公平骰子（Unfair Die）

- 设 $P(\{1\}) = p_1, P(\{2\}) = p_2, \ldots, P(\{6\}) = p_6$
- 必须满足归一性条件：

$$\sum_{i=1}^{6} p_i = 1$$

- 文档中给出具体数值：
  - $P(\{1\}) = \frac{1}{4}$, $P(\{2\}) = \frac{1}{8}$, $P(\{3\}) = \frac{1}{16}$, $P(\{4\}) = ?$, $P(\{5\}) = \frac{1}{16}$, $P(\{6\}) = \frac{1}{6}$
  - 注意：这些值之和未必为 1，可能是教学示例或待补全。

七、连续型概率模型

当样本空间为不可数无限集（如区间）时，需引入**概率密度函数（PDF）**。

示例：均匀分布（Fair [0,1]）

- $\Omega = [0,1]$
- 对区间 $[a,b) \subseteq [0,1]$，定义：

$$P([a,b)) = b - a$$

- 这对应于均匀密度函数 $f(x) = 1$ on $[0,1]$

✅ 文档中 "P([a, b))=e(b-a) e! B+(1= P(M= P(10,2)= el-of e? 1= e) =0= t" 明显为严重识别错误，应为：

$$P([a,b)) = b - a, \quad P([0,1]) = 1, \quad P(\{a\}) = 0$$

关键性质

- 单点事件的概率为 0：$P(\{a\}) = 0$
- 区间概率由积分给出：$P([a,b]) = \int_a^b f(x)dx$
- 密度函数满足：$\int_{-\infty}^{\infty} f(x)dx = 1$

✅ 文档中 “"DENSITey" TECHNIQUES I TOICUS] FOR COMPUTATION OF PROBAB12/TRES” 应为 “**Density techniques for computation of probabilities**”

八、全概率公式（Law of Total Probability）初步

文档末尾尝试引入全概率思想：

> “In A TOWN PROB SICK AND VACCINATED: to PROD Sick And Not rac PROB OF BEING siCk?”

设：

- A：生病（Sick）
- B：已接种疫苗（Vaccinated）

则：

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

因为 $A = (A \cap B) \cup (A \cap B^c)$，且两者互斥。

这是**全概率公式**的最简单形式（划分 $\Omega = B \cup B^c$）。

九、总结：概率模型三元组 $(\Omega, \mathcal{F}, P)$

组成部分	数学对象	含义	示例
样本空间 Ω	集合	所有可能结果	$\{1,2,3,4,5,6\}$
事件空间 $\mathcal{F}$	σ-代数	可测事件集合	所有子集（有限情形）
概率测度 P	函数 $P: \mathcal{F} \to [0,1]$	分配概率	$P(A) = \frac{

结语

该文档虽因转录质量问题导致大量文本混乱，但其核心内容完整覆盖了**概率论入门的关键概念**：从样本空间、事件、概率测度的定义，到公理系统、基本性质、古典与连续模型，再到集合运算与初步的全概率思想。掌握这些内容是深入学习概率论与统计学的基础。

20250910

一、概率模型的基本框架：概率空间 (Probabilistic Model / Probability Space)

概率论的数学基础建立在**概率空间**三元组上，记为 $(M, \mathcal{F}, P)$。

1. 样本空间 (Sample Space) M

- **定义**：所有可能基本结果（样本点）的集合。
- **示例**：抛一枚骰子，$M = \{1,2,3,4,5,6\}$。

2. 事件 (Events)

- **定义**：样本空间 M 的某些子集称为**事件**。
- 在有限样本空间中，通常取所有子集构成的集合族（即幂集）作为事件域 $\mathcal{F}$。
- **文档说明**："EVENTS: ALL SUBSETS OF M" —— 这适用于**有限样本空间**的情况。

3. 概率测度 (Probability Measure) P

- **定义**：一个从事件集合 $\mathcal{F}$ 映射到区间 $[0,1]$ 的函数：
$$P: \mathcal{F} \to [0,1]$$
- 满足以下两条**柯尔莫哥洛夫公理 (Kolmogorov Axioms)**：

公理 P1：规范性 (Normalization)

$$P(M) = 1$$

- 整个样本空间的概率为 1，表示必然事件。

公理 P2：可列可加性 (Countable Additivity)

若 $\{A_i\}_{i=1}^{\infty}$ 是一组**互不相交**（两两互斥）的事件序列，则：

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

- 文档中写作："DISTONT Everis [filAjeOiFe] THEN NNAi)=EPA"，即"Disjoint events ⇒ Probability of union = sum of probabilities"。

> ⚠️ **注意**：文档中"Error64 10WIRCase"可能是笔误或上下文缺失，但后续内容表明讨论的是**有限样本空间**情形。

二、有限样本空间下的概率模型

当样本空间 M 有限时，模型可进一步简化。

1. 概率权重 (Probability Weights)

- 设 $M = \{a_1, a_2, \ldots, a_n\}$，每个样本点 a_i 被赋予一个非负权重 $p(a_i) \geq 0$。
- 总权重归一化：

$$\sum_{i=1}^{n} p\,(a_i) = 1$$

2. 任意事件的概率计算

- 对任意事件 $A \subseteq M$，其概率为：

$$P(A) = \sum_{a_i \in A} p\,(a_i)$$

- 即：事件概率等于其所含样本点的概率权重之和。

3. 特殊情形：古典概型（等可能样本点）

- 若所有样本点等可能发生，即：

$$p(a_i) = \frac{1}{n}, \quad \forall i$$

- 则对任意事件 A：

$$P(A) = \frac{|A|}{|M|} = \frac{\text{有利结果数}}{\text{总结果数}}$$

- 文档中写作："p(ai)= 1/m And p # aA=ATTA"，其中"ATTA"应为 $|A|$（集合 A 的元素个数），$m = |M|$。

三、概率的基本性质 (Properties of Probability)

由概率公理可推导出以下重要性质。

性质 R1：补事件的概率

$$P(A^c) = 1 - P(A)$$

- 其中 $A^c = M \setminus A$ 是 A 的补集。
- **证明思路**：A 与 A^c 互斥且 $A \cup A^c = M$，故：

$$P(M) = P(A) + P(A^c) = 1 \Rightarrow P(A^c) = 1 - P(A)$$

性质 R2：条件分解公式（全概率公式雏形）

若 B 是一个事件，则对任意事件 A，有：

$$P(A) = P(A \cap B) + P(A \cap B^c)$$

- **证明**：将 A 分解为两个互斥部分：

$$A = (A \cap B) \cup (A \cap B^c), \quad \text{且 } (A \cap B) \cap (A \cap B^c) = \emptyset$$

- 由可加性得证。

文档中图示："A=(AnR)VLA1BY [atiy wanby: AniBuisi -00" 应为 "A = (A ∩ B) ∪ (A ∩ B^c)"，且交集为空。

四、划分与全概率公式

定义：划分 (Partition)

- 事件族 $\{B_1, B_2, \dots, B_n\}$ 是样本空间 M 的一个**划分**，若满足：
    a. $B_i \cap B_j = \emptyset$ 对所有 $i \neq j$（互斥）；
    b. $\bigcup_{i=1}^{n} B_i = M$（完备）。

全概率公式 (Law of Total Probability)

若 $\{B_1, \dots, B_n\}$ 是 M 的一个划分，则对任意事件 A：

$$P(A) = \sum_{i=1}^{n} P\left(A \cap B_i\right)$$

- **证明**：由划分性质，
$$A = A \cap M = A \cap \left(\bigcup_{i=1}^{n} B_i\right) = \bigcup_{i=1}^{n} (A \cap B_i)$$
- 由于 B_i 互斥，$A \cap B_i$ 也互斥，故可加性适用。

  文档中写作：“THEN lace Plat Food: A=(nB,) U ... VIAB B DISTONY UNI HENCE, By R2y · P(A)=& PLAN Bil B”，即使用 R2 推广至多个划分块。

五、单调性与包含关系

性质 R3：单调性 (Monotonicity)

若 $A \subseteq B$，则：

$$P(A) \leq P(B)$$

- **证明**：将 B 分解为互斥部分：
$$B = A \cup (B \setminus A), \quad A \cap (B \setminus A) = \emptyset$$
- 故：
$$P(B) = P(A) + P(B \setminus A) \geq P(A) \quad (\text{因 } P(B \setminus A) \geq 0)$$

  文档中“P(A1B) 1 DIA) ⑭ P(A1m)[P(B) A” 应为 “If $A \subseteq B$, then $P(A) \leq P(B)$”。

推论：

- $P(A \cap B) \leq \min(P(A), P(B))$
- 因为 $A \cap B \subseteq A$ 且 $A \cap B \subseteq B$，故由单调性得证。

六、并事件的概率：容斥原理

两事件的容斥原理 (Inclusion-Exclusion Principle)

对于任意两个事件 A 和 B：

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

证明:

- 分解:

$$A = (A \cap B) \cup (A \backslash B), \quad B = (A \cap B) \cup (B \backslash A)$$

- 且:

$$A \cup B = (A \backslash B) \cup (B \backslash A) \cup (A \cap B)$$

三者互斥。

- 故:

$$P(A \cup B) = P(A \backslash B) + P(B \backslash A) + P(A \cap B)$$

- 又:

$$P(A) = P(A \backslash B) + P(A \cap B) \Rightarrow P(A \backslash B) = P(A) - P(A \cap B)$$
$$P(B) = P(B \backslash A) + P(A \cap B) \Rightarrow P(B \backslash A) = P(B) - P(A \cap B)$$

- 代入即得:

$$P(A \cup B) = [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] + P(A \cap B) = P(A) + P(B) - P(A \cap B)$$

文档中用"↑ROP"标记，并给出代数推导过程。

应用示例:

- 某城镇 52% 有儿子，55% 有女儿。
- 若直接相加 $P(A) + P(B) = 0.52 + 0.55 = 1.07 > 1$，显然错误。
- 正确做法需减去交集:

$$P(\text{有子女}) = P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- 若无更多信息，无法确定 $P(A \cap B)$。

推广至三事件:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C)$$

- 文档中练习题给出此式。

七、独立性 (Independence)

1. 实验的独立性（笛卡尔积模型）

定义:

考虑两个独立实验:

- 实验 1: 样本空间 $M_1 = \{a_1, \dots, a_m\}$，概率 $P_1(a_i) = p(a_i)$
- 实验 2: 样本空间 $M_2 = \{b_1, \dots, b_n\}$，概率 $P_2(b_j) = q(b_j)$

若两个实验**相互独立**，则复合实验的样本空间为**笛卡尔积**:

$$M = M_1 \times M_2 = \{(a_i, b_j) \mid i = 1, \dots, m; j = 1, \dots, n\}$$

且每个结果的概率为:

$$P((a_i, b_j)) = p(a_i) \cdot q(b_j)$$

示例：

- 同时掷一个公平骰子和抛一枚公平硬币。
- $M = \{(1,H),(1,T),\dots,(6,H),(6,T)\}$，共 12 个样本点。
- $P((3,H)) = P(骰子 = 3)\cdot P(硬币 = H) = \frac{1}{6}\cdot\frac{1}{2} = \frac{1}{12}$

  文档中“PROB(DIE>,5 AND JOIN HEAD” 应为 “P(Die > 5 AND Coin = Head)”，即：

$$P(\{6\}\times\{H\}) = P(\text{Die} = 6)\cdot P(\text{Head}) = \frac{1}{6}\cdot\frac{1}{2}$$

---

2. 事件的独立性

定义：

在概率空间 $(M,\mathcal{F},P)$ 中，两个事件 A 和 B 称为**独立的**，如果：

mutual pairwise

$$P(A\cap B) = P(A)\cdot P(B)$$

推广：多个事件的独立性 更严谨的说法是整体独立（整体独立 ≠ 两两独立）

事件 $A_1,A_2,\dots,A_k$ **相互独立** **(mutually independent)**，如果对任意子集 $\{i_1,\dots,i_r\}\subseteq\{1,\dots,k\}$，有：

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

两者充分不必要条件 pairwise

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

eg: A.B.C三人两两关系好

但不代表三人在一起就和谐

mutual

文档中写作：“PAt,)---PLAN)” 应为 $P(A_1\cap\cdots\cap A_n) = P(A_1)\cdots P(A_n)$。

示例：重复抛硬币

- 每次抛硬币是独立实验。
- 设 B_i 表示“第 i 次抛出正面”。
- 若硬币不公平，$P(H) = p$，$P(T) = 1-p$。
- 则：

$$P(3 次反面后接 2 次正面) = P(T,T,T,H,H) = (1-p)^3\cdot p^2$$

- 形式化为：

$$P(B_1^c\cap B_2^c\cap B_3^c\cap B_4\cap B_5) = P(B_1^c)P(B_2^c)P(B_3^c)P(B_4)P(B_5) = (1-p)^3p^2$$

八、无限样本空间的初步思想

文档末尾提及“FLIP A COIN FOREVER”，引出无限序列的样本空间。

示例：首次出现正面的时间

- 定义事件 A_i：第 i 次抛硬币是**首次**出现正面。
- 则 A_i 对应序列：前 $i-1$ 次为反面，第 i 次为正面。
- $P(A_i) = (1-p)^{i-1}p$

事件“最终会出现正面”

- $E = \bigcup_{i=1}^{\infty} A_i$，所有 A_i 互斥。
- 故：
$$P(E) = \sum_{i=1}^{\infty} P(A_i) = \sum_{i=1}^{\infty} (1-p)^{i-1} p = p \cdot \frac{1}{1-(1-p)} = 1 \quad (\text{当 } p > 0)$$
- 即：只要 $p > 0$，正面**几乎必然**（with probability 1）会最终出现。

文档中“PRODIN WILL EVENTRALLY APPEAR)” 即 “Probability that Head will eventually appear”。

总结：核心概念图谱

概念	关键公式/定义
概率空间	$(M, \mathcal{F}, P)$，$P(M) = 1$，可列可加性
古典概型	$P(A) = \frac{
补事件	$P(A^c) = 1 - P(A)$
全概率公式	$P(A) = \sum P(A \cap B_i)$，$\{B_i\}$ 为划分
包含关系	$A \subseteq B \Rightarrow P(A) \leq P(B)$
容斥原理	$P(A \cup B) = P(A) + P(B) - P(A \cap B)$
独立事件	$P(A \cap B) = P(A)P(B)$
独立实验	$P((a_i, b_j)) = p(a_i)q(b_j)$

📓 概率论练习题解析（六道题完整版）

☑ P1：证明

$$\max\{0, P(A)+P(B)-1\} \leq P(A\cap B) \leq \min\{P(A), P(B)\}$$

🔍 题目解析：

求两个事件交集概率的上下界。

🧩 解题过程：

上界：$P(A\cap B) \leq \min\{P(A), P(B)\}$

- 因为 $A\cap B \subseteq A \;\Rightarrow\; P(A\cap B) \leq P(A)$
- 同理 $A\cap B \subseteq B \;\Rightarrow\; P(A\cap B) \leq P(B)$
- 所以 $P(A\cap B) \leq \min\{P(A), P(B)\}$

下界：$P(A\cap B) \geq \max\{0, P(A)+P(B)-1\}$

由容斥原理：

$$P(A\cup B) = P(A) + P(B) - P(A\cap B)$$

而 $P(A\cup B) \leq 1$，所以：

$$P(A) + P(B) - P(A\cap B) \leq 1 \Rightarrow P(A\cap B) \geq P(A) + P(B) - 1$$

又因概率 $\geqslant 0$，故取最大值：

$$P(A\cap B) \geq \max\{0, P(A)+P(B)-1\}$$

✅ 得证。

💡 思考方式：

利用集合包含关系和容斥公式，从“整体”推导“局部”，注意边界条件（如概率不能为负）。

☑ P2：已知 $A \triangle B = (A\backslash B)\cup(B\backslash A)$，证明

$$P(A\triangle B) = P(A) + P(B) - 2P(A\cap B)$$

🔍 题目解析：

对称差是“只属于其中一个”的部分，要求其概率表达式。

🧩 解题过程：

分解：

$$A\triangle B = (A\backslash B)\cup(B\backslash A)$$

且两部分互不相交 $\Rightarrow$ 可直接加概率。

$P(A \triangle B) = P(A \backslash B) + P(B \backslash A) = [P(A) - P(A \cap B)] + [P(B) - P(A \cap B)] = P(A) + P(B) - 2P(A \cap B)$

✅ 得证。

💡 思考方式:

将复杂集合拆解为简单部分，利用补集思想（$A \backslash B = A \cap B^c$），再代入公式。

☑ P3：若 A 和 B 独立，证明 A^c 和 B 独立

🔍 题目解析:

独立性是否传递到补事件?

🧩 解题过程:

已知：$P(A \cap B) = P(A)P(B)$

计算:

$$P(A^c \cap B) = P(B) - P(A \cap B) = P(B) - P(A)P(B) = P(B)(1 - P(A)) = P(B)P(A^c)$$

所以 $P(A^c \cap B) = P(A^c)P(B) \Rightarrow A^c$ 与 B 独立。

✅ 得证。

💡 思考方式:

从定义出发，用补集公式转化，验证乘法法则成立。这是“独立性保持在补事件中”的关键性质。

☑ P4：构造一对不相交事件 A, B，使得它们不独立

注：提示 “Independent ≠ disjoint”

🔍 题目解析:

说明“不相交”≠“独立”。

🧩 解题过程:

设样本空间为抛硬币：Ω = {H, T}

令:

- $A = \{H\}, P(A) = \frac{1}{2}$
- $B = \{T\}, P(B) = \frac{1}{2}$

显然 $A \cap B = \emptyset \Rightarrow P(A \cap B) = 0$

但 $P(A)P(B) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \neq 0$

⇒ 不满足独立性条件。

所以 A 和 B **不独立**。

✅ 构造完成。

💡 思考方式:

找一个非零概率的事件对，使其交集为空。此时 $P(A \cap B) = 0$，但 $P(A)P(B) > 0$，矛盾 ⇒ 不独立。

☑ P5：从 52 张牌中抽两次，A: 第一张是 A，B: 第二张是 A，判断是否独立？

🔍 题目解析:

比较有放回 vs 无放回下的独立性。

🧩 解题过程:

Case 1: With replacement（放回）

- $P(A) = \frac{4}{52}$
- $P(B) = \frac{4}{52}$
- $P(A \cap B) = \frac{4}{52} \cdot \frac{4}{52} = P(A)P(B)$

⇒ **独立**

Case 2: Without replacement（不放回）

- $P(A) = \frac{4}{52}$
- $P(B) = \frac{4}{52}$
- $P(A \cap B) = \frac{4}{52} \cdot \frac{3}{51} = \frac{12}{2652}$
- $P(A)P(B) = \frac{4}{52} \cdot \frac{4}{52} = \frac{16}{2704}$

比较:

$$\frac{12}{2652} \approx 0.00452, \quad \frac{16}{2704} \approx 0.00592 \Rightarrow P(A \cap B) \neq P(A)P(B)$$

⇒ **不独立**

✅ 结论：放回时独立，不放回时不独立。

💡 思考方式:

看信息是否影响后续概率。不放回改变了样本空间 ⇒ 依赖；放回则保持不变 ⇒ 独立。

⚠ **Challenge：证明**

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P\left(A_i\right)$$

提示：数学归纳法

🔍 题目解析：

Boole's Inequality，用于估计多个事件至少一个发生的概率上限。

🧩 解题过程：

Base case (n=1)：

$$P(A_1) \leq P(A_1) \quad \text{成立}$$

Inductive step：

假设对 n 成立，即：

$$P\left(\bigcup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} P\left(A_i\right)$$

考虑 $n+1$：

$$\bigcup_{i=1}^{n+1} A_i = \left(\bigcup_{i=1}^{n} A_i\right) \cup A_{n+1}$$

由概率加法：

$$P\left(\bigcup_{i=1}^{n+1} A_i\right) = P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) - P\left(\left(\bigcup_{i=1}^{n} A_i\right) \cap A_{n+1}\right) \leq P\left(\bigcup_{i=1}^{n} A_i\right) + P(A_{n+1}) \leq \sum_{i=1}^{n} P\left(A_i\right) + P(A_{n+1})$$
$$= \sum_{i=1}^{n+1} P\left(A_i\right)$$

✅ 归纳完成。

💡 思考方式：

使用数学归纳法处理有限并集，利用"并集可拆分为前 n 个和第 n+1 个"的结构，结合减去交集项来控制误差。

📌 知识点总结

类别	内容
核心概念	事件独立性、对称差、不相交、补事件、联合概率
重要公式	容斥原理、对称差概率、Boole 不等式
关键技巧	分解集合、利用补集、归纳法、反例构造

🔍 思考方式提炼（简洁版）

4. **先理解定义**：独立 ⇔ $P(A \cap B) = P(A)P(B)$
5. **拆解复杂事件**：用集合运算（并、交、补）化简
6. **利用对称性/不变性**：如放回 vs 不放回
7. **构造反例**：找非零概率且不相交的事件
8. **归纳法处理递推问题**：从 n 到 n+1 推广
9. **边界检查**：概率范围 [0,1]，避免越界

20250915

概率模型与组合数学基础:

一、概率空间的基本构成 (The Probabilistic Model)

本部分定义了概率论的公理化框架，即**概率空间** $(\Omega, \mathcal{F}, P)$ 的三大要素。

1. 样本空间 (Sample Space)

- **定义**：样本空间 Ω 是一个实验所有可能结果（outcomes）的集合。
- **示例**：
    - 抛一枚硬币：$\Omega = \{H, T\}$
    - 掷一个六面骰子：$\Omega = \{1,2,3,4,5,6\}$
    - 连续抛 50 次公平硬币：$\Omega = \{H, T\}^{50}$，其元素个数为 $|\Omega| = 2^{50}$。

2. 事件 (Events)

- **定义**：事件是样本空间 Ω 的某些子集（subsets），表示我们关心的一类结果的集合。
- **注意**：并非所有子集都自动成为事件；在初等概率中，通常考虑所有子集构成的幂集作为事件域 $\mathcal{F}$。
- **事件域** $\mathcal{F}$ 应满足一定的代数结构（如σ-代数），但在有限样本空间下可简化处理。

3. 概率测度 / 分布 (Probability Measure / Distribution)

- **定义**：概率测度 P 是从事件集合到区间 $[0,1]$ 的函数，记作 $P: \mathcal{F} \to [0,1]$，满足以下两条**柯尔莫哥洛夫公理**:

公理 (Axioms of Probability):

10. **归一性**：$P(\Omega) = 1$
    （整个样本空间发生的概率为 1）
11. **可列可加性（σ-可加性）**：若 $A_1, A_2, \dots$ 是两两互斥（disjoint）的事件（即 $A_i \cap A_j = \emptyset$ 对所有 $i \neq j$ 成立），则：

$$P\left(\bigcup_{j=1}^{\infty} A_j\right) = \sum_{j=1}^{\infty} P\left(A_j\right)$$

注：这两条公理是现代概率论的基石。任何满足它们的函数 P 都称为一个合法的概率测度。

二、概率的基本性质 (Properties of Probability)

由上述公理可以推导出若干重要性质，这些性质在计算和推理中极为常用。

性质 1：补事件的概率

$$P(B^c) = 1 - P(B), \quad \text{特别地, } P(\emptyset) = 0$$

- 推导：因为 $B \cup B^c = \Omega$ 且 $B \cap B^c = \emptyset$，所以 $P(B) + P(B^c) = P(\Omega) = 1$。

性质 2：全概率公式（基于划分 Partition）

若 $B_1, B_2, \dots, B_n$ 构成样本空间的一个**划分**（partition），即：

- $B_i \cap B_j = \emptyset$ for $i \neq j$,
- $\bigcup_{i=1}^{n} B_i = \Omega$,

则对任意事件 A，有：

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

> 注意：此处原文写作 $P(A) = \sum P(A \cap B_i)$，这是正确的，但更完整的表达应引入条件概率形式。此式也称“**全概率法则**”。

性质 3：并事件的概率（容斥原理初步）

对于任意两个事件 A 和 B：

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

- 当 $A \cap B = \emptyset$ 时退化为加法公式。
- 此公式可推广至多个事件（高级容斥原理）。

性质 4：独立性 (Independence)

两个事件 A 和 B 被称为**相互独立**，当且仅当：

$$P(A \cap B) = P(A) \cdot P(B)$$

- **关键点**：独立性是一个定义出来的数学关系，不一定对应直观上的“无关”。
- 在多次独立试验中（如重复抛硬币、掷骰子），各次试验的结果通常被建模为独立事件。

> 示例：连续抛两次公平硬币，“第一次正面”与“第二次反面”是独立事件。

三、组合数学基础：计数原理（Combinatorics: Counting Principles)

概率计算往往依赖于“有利情况数 / 总情况数”，因此精确计数至关重要。本节介绍几种基本计数方法。

1. 基本概念与符号

- 设 A 是一个集合，其元素个数称为**基数**（cardinality），记作：
$$|A|,\ \#A,\ 或\ n(A)$$
- 若 $|A| = n < \infty$，称 A 为**有限集**。
- 若 $|A| = |\mathbb{N}|$，称 A 为**可数无限集**。
- 若 $|A| > |\mathbb{N}|$（如 $[0,1]$），称 A 为**不可数无限集**。

---

2. 乘法原理（Multiplication Principle / Product Rule)

定理（乘积空间计数）：

若一个过程由 k 个连续步骤组成，第 i 步有 n_i 种选择方式，则总方案数为：

$$|\Omega| = n_1 \times n_2 \times \cdots \times n_k = \prod_{i=1}^{k} n_i$$

应用实例：

- **掷 5 个六面骰子**：每个骰子有 6 种结果，总样本空间大小为：
$$|\Omega| = 6^5$$
- **抛 50 次公平硬币**：每次有 2 种结果，总样本空间大小为：
$$|\Omega| = 2^{50}$$
若硬币公平，则每种特定序列的概率为 $P(\omega) = \frac{1}{2^{50}}$。

> 文档中提到 Galileo 的经典问题：“三个骰子点数之和为 9 vs 10 哪个更容易？”
> 实际上，虽然两种和的整数解数量相同，但由于不同排列数不同（例如 (1,2,6) 有 6 种排列，而 (3,3,3) 只有 1 种），导致总实现方式数不同，从而影响概率。这体现了计数的重要性。

---

3. 排列（Permutations)

定义：

从 n 个不同元素中取出 k 个进行有序排列的方式数。

(P1) 全排列（Full Permutations)

- n 个不同元素的全排列数为：

$$n! = n \cdot (n-1) \cdot (n-2) \cdots 2 \cdot 1$$

- 特别地，$0! = 1$（约定）。

应用示例：

- 英文字母表共有 26 个字母，其全排列数为：

$$26!$$

- 一副标准扑克牌（52 张）的所有洗牌方式数为：

$$52!$$

若洗牌是公平的（uniform random shuffle），则任一特定排列出现的概率为 $\frac{1}{52!}$。

(P1') 部分排列（k-permutations of n）

- 从 n 个不同元素中取 k 个进行有序排列的数量为：

$$P(n,k) = n \cdot (n-1) \cdot \cdots \cdot (n-k+1) = \frac{n!}{(n-k)!}$$

应用示例：

- 用英文字母构造长度为 3 的“单词”（允许非实际词汇），且字母不重复：

$$P(26,3) = 26 \times 25 \times 24$$

(P2) 含有不可区分元素的排列（Permutations with Indistinguishable Objects）

当集合中有重复元素时，全排列数需修正。

一般命题：

若有 n 个对象，其中：

- n_1 个类型 1 的对象（彼此相同），
- n_2 个类型 2 的对象（彼此相同），
- ...
- n_k 个类型 k 的对象（彼此相同）， 且 $n_1 + n_2 + \cdots + n_k = n$，

则不同的排列总数为**多重组合数**（multinomial coefficient）：

$$\frac{n!}{n_1! \cdot n_2! \cdot \cdots \cdot n_k!}$$

应用示例 1：单词 "BOOK"

- 字母：B, O, O, K → 共 4 个字母，其中 O 出现 2 次。
- 不同排列数为：

$$\frac{4!}{2!} = \frac{24}{2} = 12$$

（若两个 O 视为不同，则为 4! = 24；但因它们不可区分，需除以 2!）

应用示例 2：单词 "HAVANA"

- 字母统计：H(1), A(3), V(1), N(1) → 总共 6 个字母

- 不同排列数为:

$$\frac{6!}{3!\cdot 1!\cdot 1!\cdot 1!}=\frac{720}{6}=120$$

---

4. 复合计数问题示例：球放入线性孔位

问题描述:

有 7 个球要放入一排 7 个孔中，其中:

- 3 个黑球（black）→ 彼此不可区分
- 2 个红球（red）→ 彼此不可区分
- 1 个蓝球（blue）
- 1 个绿球（green）

问：没有任何两个黑球相邻的概率是多少?

解法思路:

第一步：计算总的排列方式数（Total Arrangements）

使用含不可区分元素的排列公式:

$$\text{Total}=\frac{7!}{3!\cdot 2!\cdot 1!\cdot 1!}=\frac{5040}{6\cdot 2}=420$$

第二步：计算“无两个黑球相邻”的有利情况数（Favorable Cases）

采用**插空法**（gap method）：

12. 先安排其他 4 个非黑球（2 红、1 蓝、1 绿）。由于红球不可区分，其余可区分:
    - 非黑球排列数：$\frac{4!}{2!}=12$
13. 这 4 个非黑球形成 5 个“间隙”（包括首尾）可用于放置黑球:

    $$\_\ \ O\ \ \_\ \ O\ \ \_\ \ O\ \ \_\ \ O\ \ \_$$

    共 5 个空位（_），从中选 3 个放黑球，且每空最多一个（保证不相邻）：
    - 选法数：$\binom{5}{3}=10$
14. 因此，有利情况数为:

$$\text{Favorable}=12\times 10=120$$

第三步：计算概率

$$P(\text{no two black balls adjacent})=\frac{\text{Favorable}}{\text{Total}}=\frac{120}{420}=\frac{2}{7}$$

原文中的计算似乎不完整或存在笔误（如“4.3”等），但思路指向插空法。

---

5. 补充练习：圆形排列（Circular Arrangement）

文档末尾提及“circular array of holes”，提示考虑环形排列。

环形排列特点：

- **n 个可区分对象围成一圈的不同排列数为 $(n-1)!$。**
    - 原因：固定一人位置消除旋转对称性，剩下 $n-1$ 人相对排列。
- 若对象中有不可区分者，则进一步除以其阶乘。

示例：4 个学生坐圆桌，其中 2 人穿蓝衣（不可区分），2 人穿红衣（不可区分）：

- 固定一人位置后，其余 3 人排列，但要考虑颜色对称。
- **更稳妥的方法是枚举或使用 Burnside 引理（群作用下的轨道计数），属于进阶内容。**

四、综合应用示例：扑克牌顶四张为 A 的概率

文档中提出一个问题（虽未完全写出）：

> “Shuffling is fair, probability 4 aces on top?”

问题重述：

在一副 52 张的标准扑克牌中，随机洗牌（均匀分布）。求最上面四张牌恰好是四张 A（顺序不限）的概率。

解答：

方法一：基于排列

- 总排列数：52!
- 有利情况：前 4 位是四张 A 的某种排列，后 48 位是其余牌的任意排列。
    - 四张 A 在前四位的排列数：4!
    - 后 48 张牌排列数：48!
- 所以有利情况总数：$4! \times 48!$
- 概率为：

$$P = \frac{4! \cdot 48!}{52!} = \frac{24}{52 \cdot 51 \cdot 50 \cdot 49} = \frac{1}{270725}$$

方法二：逐步抽取（条件概率）

- 第一张是 A 的概率：$\frac{4}{52}$
- 第二张是 A 的概率（已知第一张是 A）：$\frac{3}{51}$
- 第三张是 A 的概率：$\frac{2}{50}$
- 第四张是 A 的概率：$\frac{1}{49}$
- 联合概率：

$$P = \frac{4}{52} \cdot \frac{3}{51} \cdot \frac{2}{50} \cdot \frac{1}{49} = \frac{24}{52 \cdot 51 \cdot 50 \cdot 49} = \frac{1}{270725}$$

结果一致。说明即使顺序不同，只要关注集合即可。

五、总结与展望

核心知识点回顾:

模块	关键内容
概率模型	样本空间、事件、概率测度（公理）、补集、全概率、并事件、独立性
组合计数	乘法原理、阶乘、排列（全/部分）、含重复元素的排列、插空法
应用场景	骰子、硬币、洗牌、字母排列、球盒问题

下一步建议（Next Wed 提示）:

文档结尾标注“DEFINITION: NEXT WED”，预示将引入新的定义，可能是:

- 组合（Combinations, $\binom{n}{k}$）
- 二项式系数与二项分布
- 条件概率的深入讨论
- 随机变量的初步定义

建议后续学习围绕“组合数 $\binom{n}{k} = \frac{n!}{k!(n-k)!}$”展开，并将其与概率分布（如超几何分布、二项分布）联系起来。

20250918

一、组合数学（Combinatorics）

1. 基本计数原理

- **乘法原理**：若完成一件事需分两步，第一步有 m 种方式，第二步有 n 种方式，则共有 $m \times n$ 种方式。
- **加法原理**：若完成一件事有两种互斥方法，分别有 m 和 n 种方式，则总共有 $m + n$ 种方式。

2. 排列（Permutations）

从 n 个不同元素中取出 k 个进行有序排列：

$$P(n,k) = \frac{n!}{(n-k)!}$$

3. 组合（Combinations）

从 n 个不同元素中取出 k 个组成集合（无序）：

$$\binom{n}{k} = C(n,k) = \frac{n!}{k!\,(n-k)!}$$

二、二项式定理与二项系数

1. 二项式展开

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

2. 二项系数性质

- 对称性：$\binom{n}{k} = \binom{n}{n-k}$
- 递推关系（帕斯卡恒等式）：$\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$
- 总和公式：

$$\sum_{k=0}^{n} \binom{n}{k} = 2^n, \quad \sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}$$

$(C_n^0)^2 + (C_n^1)^2 + \cdots + (C_n^n)^2 = C_{2n}^n$

三、概率基础（Probability Fundamentals）

1. 概率公理

二项分布 $C_n^k p^k (1-p)^{n-k}$

设样本空间为 S，事件 $A \subseteq S$，则：

- $0 \le P(A) \le 1$
- $P(S) = 1$
- 若 $A \cap B = \emptyset$，则 $P(A \cup B) = P(A) + P(B)$

2. 古典概型

若所有结果等可能，则：

$$P(A) = \frac{\text{有利结果数}}{\text{总结果数}}$$

示例：公平硬币抛掷

- 单次抛掷：$P(\text{正面}) = P(\text{反面}) = 0.5$
- 抛 n 次，出现 k 次正面的概率服从**二项分布**

四、条件概率与贝叶斯定理

贝叶斯：

公式二：贝叶斯定理（Bayes' Theorem）

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{\sum_{i=1}^{n} P(B|A_i) \cdot P(A_i)} \quad \text{if } P(B) > 0$$

$$P(A|B) = \frac{p(B|A) \times p(A)}{\sum_{i=1}^{n} p(B|A_i) p(A_i)}$$

1. 条件概率定义

在事件 B 发生的前提下，事件 A 发生的概率：

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

前提：AB两个不独立

2. 乘法法则

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

3. 全概率公式（Law of Total Probability）

若 $B_1, B_2, \dots, B_n$ 构成样本空间的一个划分，则对任意事件 A：

$$P(A)=\sum_{i=1}^{n} P\,(A|B_i)P(B_i)$$

应用实例（疾病检测问题）

- 设某病发病率 $P(D)=0.05$
- 检测灵敏度（真阳性率）：$P(+|D)=0.95$
- 特异性（真阴性率）：$P(-|\neg D)=0.95$，故假阳性率 $P(+|\neg D)=0.05$
- 求：某人检测阳性时实际患病的概率 $P(D|+)$

使用贝叶斯定理:

$$P(D|+)=\frac{P(+|D)P(D)}{P(+|D)P(D)+P(+|\neg D)P(\neg D)}=\frac{0.95\times 0.05}{0.95\times 0.05+0.05\times 0.95}=\frac{0.0475}{0.095}=0.5$$

五、独立性与联合概率

1. 事件独立

事件 A 与 B 独立当且仅当:

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

或等价地:

$$P(A|B)=P(A),\quad P(B|A)=P(B)$$ $=\frac{P(A)\times P(B)}{P(A)}=P(B)$

2. 贝叶斯网络思想初现

原文: “KVEMS TI COwDITJouk aovA pilit! OF B rS PlA)9)=((403CokEiS”

推测原句: “Given the conditional probability of A given B is... used in Bayes' rule”

表明正在应用**条件概率推理链**。

六、二项分布（Binomial Distribution）

定义

进行 n 次独立伯努利试验，每次成功概率为 p，令随机变量 X 表示成功次数，则:

$$X\sim \text{Binomial}(n,p)$$

其概率质量函数为:

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k},\quad k=0,1,2,\dots,n$$

期望与方差

$$E[X]=np,\quad \mathrm{Var}(X)=np(1-p)$$

七、实际应用案例解析

案例 1：医学诊断（Medical Testing）

符号	含义	值
D	患病	$P(D)=0.05$
$+$	检测阳性	—
$ P(+	D) $	灵敏度
$ P(+	\neg D) $	假阳性率

计算：

$$P(D|+)=\frac{P(+|D)P(D)}{P(+)}=\frac{0.95\times 0.05}{0.95\times 0.05+0.05\times 0.95}=50\%$$

📌 **结论**：即使检测准确率高达 95%，由于疾病稀有性，阳性者中只有一半真正患病。

案例 2：并联系统可靠性（Parallel System Reliability）

两个组件并联工作，只要至少一个正常即可运行。

- 设每个组件正常工作的概率为 p
- 则系统失效概率为 $(1-p)^2$
- 故系统正常工作概率为：

$$P(\text{system works})=1-(1-p)^2$$

结合后续："P(A|B)=0.7, P(B)=0.4, P(A∩B)=0.28"，验证了交集计算：

$$P(A\cap B)=P(A|B)P(B)=0.7\times 0.4=0.28$$

进一步给出：

$$P(A)=P(A\cap B)+P(A\cap\neg B)=0.28+0.12=0.4$$

说明事件不独立。

八、指数衰减模型（Radioactive Decay Model）还原核心信息：

这是一个**放射性衰变存活模型**，假设衰变服从指数分布。

⇒ 无记忆性

- 衰变率 $\lambda = 0.03$ /单位时间
- 存活概率函数：

$$P(\text{survive } t \text{ units}) = e^{-\lambda t}$$

- 若 $t = 1$，则 $P = e^{-0.03} \approx 0.9704$
- 若给定某粒子未在前一段时间衰变，求后续存活概率 → 典型**无记忆性**问题

文中尝试计算某个条件概率：

$$P(A) = 0.012 \times (\text{some factor}) \quad \Rightarrow \quad \text{可能涉及似然比或后验估计}$$

最终表达式近似于：

$$P(A) \approx \frac{0.012}{0.043} \approx 0.279$$

👉 推测为某种**基于观测数据的参数估计或假设检验雏形**。

📓 20250922 概率论核心笔记：条件概率、贝叶斯定理与独立性

一、条件概率（Conditional Probability）

1. 定义

给定事件 B 发生的前提下，事件 A 发生的**条件概率**定义为：

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}, \quad \text{其中 } P(B) > 0$$

> ⚠️ 注意：此定义仅在 $P(B) > 0$ 时成立。

---

2. 条件概率的乘法法则（Multiplication Rule）

由上述定义可得：

$$P(A \cap B) = P(A \mid B) \cdot P(B)$$

若 A 和 B 不独立，则必须使用条件概率进行计算。

---

3. 条件概率作为概率测度的性质

尽管是“条件”下的概率，$P(\cdot \mid B)$ 本身仍满足概率公理，即：

（1）非负性：

$$P(A \mid B) \geq 0$$

（2）规范性：

$$P(\Omega \mid B) = 1$$

（3）可列可加性（σ-additivity）：

若 $A_1, A_2, \ldots$ 是互不相交的事件集，则：

$$P\left(\bigcup_{i=1}^{\infty} A_i \,\middle|\, B\right) = \sum_{i=1}^{\infty} P(A_i \mid B)$$

☑️ **结论**：对于固定的 B，映射 $A \mapsto P(A \mid B)$ 构成一个合法的概率测度。

---

4. 推论：条件概率的基本性质

性质	公式
补集规则	$P(A^c \mid B) = 1 - P(A \mid B)$
分解规则	$P(A \mid B) = P(A \cap C \mid B) + P(A \cap C^c \mid B)$
并集规则	$P(A \cup C \mid B) = P(A \mid B) + P(C \mid B) - P(A \cap C \mid B)$

这些性质表明：我们可以像操作普通概率一样操作条件概率。

二、全概率公式（Law of Total Probability）

1. 背景思想："分而治之"

将复杂事件 A 按照某个划分 $B_1, B_2, \dots, B_n$ 分解，分别计算每个子情形下 A 的发生概率，再加权求和。

2. 数学形式

设 $B_1, B_2, \dots, B_n$ 是样本空间 Ω 的一个**完备划分**（即互斥且并集为全集），且 $P(B_i) > 0$，则对任意事件 A：

$$P(A) = \sum_{i=1}^{n} P\left(A \mid B_i\right) \cdot P(B_i)$$

✅ 这被称为"Divide and Conquer"策略在概率中的体现。

三、贝叶斯定理（Bayes' Theorem）

1. 核心目标

从已知的 $P(B \mid A)$ 反推 $P(A \mid B)$ —— 即"后验概率"。

2. 基本形式

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

结合全概率公式，常用于如下场景：

$$P(B_j \mid A) = \frac{P(A \mid B_j) \cdot P(B_j)}{\sum_i P\left(A \mid B_i\right) \cdot P(B_i)}$$

其中 $\{B_i\}$ 是对样本空间的一个划分。

四、经典案例分析

◇ 案例一：周六夜晚交通事故（Saturday Night Accident）

已知信息：

- D：司机酒驾（Drunk）
- A：发生事故（Accident）
- $P(D) = 0.05$（5%的人酒驾）
- $P(A) = 0.1$（每 10 次出行中有 1 次事故）
- 酒驾者导致的事故占总事故的 20%

目标：求 $P(A \mid D)$，即酒驾条件下发生事故的概率。

解法步骤：

1. 计算酒驾相关的联合概率：

$$P(A \cap D) = \text{酒驾引起的事故比例} \times P(A) = 0.2 \times 0.1 = 0.02$$

2. 利用条件概率公式：

$$P(A \mid D) = \frac{P(A \cap D)}{P(D)} = \frac{0.02}{0.05} = 0.4 = 40\%$$

✅ 结论：酒驾者发生事故的概率为 **40%**。

◇ 案例二：医学检测（Medical Test for Mad Cow Disease）

设定变量：

- B：患病（Sick）
- T：检测呈阳性（Test Positive）
- $P(B) = 0.02$（人群中有 2%患病）
- $P(T \mid B^c) = 0.1$（假阳性率：健康人被误判为阳性的概率）
- $P(T^c \mid B) = 0.3 \Rightarrow P(T \mid B) = 0.7$（真阳性率）

目标：评估检测的有效性 → 求 $P(B \mid T)$

第一步：使用全概率公式计算 $P(T)$

$$\begin{aligned} P(T) &= P(T \mid B)P(B) + P(T \mid B^c)P(B^c) \\ &= (0.7)(0.02) + (0.1)(0.98) \\ &= 0.014 + 0.098 = 0.112 \end{aligned}$$

第二步：应用贝叶斯公式

$$
\begin{aligned}
P(B \mid T) &= \frac{P(T \mid B)P(B)}{P(T)} \\
&= \frac{0.7 \times 0.02}{0.112} = \frac{0.014}{0.112} \approx 0.125 = 12.5\%
\end{aligned}
$$

📌 **惊人发现**：即使检测结果为阳性，实际患病的概率仅为 **12.5%**！

改进建议：

3. 降低假阳性率（False Positive Rate）
4. 多次重复检测（假设条件独立）

◇ 案例三：重复检测的贝叶斯推广（Repeated Testing）

假设两次检测相互独立（在给定健康状态条件下），即：

$$
P(T_1 \cap T_2 \mid B) = P(T_1 \mid B) \cdot P(T_2 \mid B)
$$

继续上例，计算两次阳性后的后验概率：

$$
P(B \mid T_1 \cap T_2) = \frac{P(T_1 \cap T_2 \mid B)P(B)}{P(T_1 \cap T_2)}
$$

分子：

$$
P(T_1 \cap T_2 \mid B)P(B) = (0.7)^2 \cdot 0.02 = 0.49 \cdot 0.02 = 0.0098
$$

分母（全概率展开）：

$$
\begin{aligned}
P(T_1 \cap T_2) &= P(T_1 \cap T_2 \mid B)P(B) + P(T_1 \cap T_2 \mid B^c)P(B^c) \\
&= (0.7)^2 \cdot 0.02 + (0.1)^2 \cdot 0.98 \\
&= 0.0098 + 0.00098 = 0.01078
\end{aligned}
$$

最终：

$$
P(B \mid T_1 \cap T_2) = \frac{0.0098}{0.01078} \approx 0.909 = 90.9\%
$$

✅ 结论：**两次阳性后，患病概率跃升至约 91%**，显著提升诊断可信度。

◇ 案例四：司法谬误（Prosecutor's Fallacy）——稀有疾病与犯罪现场血迹

场景描述：

- 城市人口：10,000,000

- 罕见病发病率：$1/1{,}000{,}000 \Rightarrow$ 预计有 10 名患者
- 犯罪现场遗留具有该疾病的生物标记
- 警方逮捕一名携带该标记的嫌疑人，并据此主张其极可能为罪犯

关键问题：这种推理是否合理？

定义事件：

- D：个体患有该疾病
- E：个体无辜（Innocent）

警方声称：$P(E \mid D)$ 很小？
但真实问题是：**在已知某人有疾病的情况下，他是无辜的概率有多大？**

假设：

- 所有 10 名患者都可能出现在现场附近
- 没有其他证据指向特定个体

则，在这 10 名患者中随机选一人被捕，他真正犯罪的概率最多为 $1/10 = 10\%$，即 **无辜概率高达 90%**！

❌ 错误逻辑（检察官谬误）：

> “这种病极其罕见，所以这个人几乎不可能是无辜的。”
>
> ❌ 实际上混淆了 $P(D \mid E)$ 和 $P(E \mid D)$

📌 正确做法应使用贝叶斯推理，考虑先验罪责和证据的似然比。

五、事件的独立性（Independence）

1. 两事件独立的定义

事件 A 和 B 称为**独立**，当且仅当：

$$P(A \cap B) = P(A) \cdot P(B)$$

等价地（若 $P(B) > 0$）：

$$P(A \mid B) = P(A)$$

> 💡 直观含义：知道 B 发生与否，不影响 A 发生的可能性。

---

2. 独立 ≠ 互斥

特性	独立（Independent）	互斥（Mutually Exclusive）
是否能同时发生	可以	否
$P(A \cap B)$	$P(A)P(B)$	0
若 $P(A), P(B) > 0$，能否独立又互斥？	❌ 不可能！因为 $P(A)P(B) > 0$，但 $P(A \cap B) = 0$	

---

3. 多个事件的独立性

不能仅凭“两两独立”断言整体独立！

定义：三个事件 A, B, C 相互独立，需满足以下全部条件：

1. $P(A \cap B) = P(A)P(B)$
2. $P(A \cap C) = P(A)P(C)$
3. $P(B \cap C) = P(B)P(C)$
4. $P(A \cap B \cap C) = P(A)P(B)P(C)$

✅ 缺一不可！仅满足前 3 条称为“pairwise independent”，但未必整体独立。

六、总结框架图

```
            ┌──────────────┐
            │ P(A ∩ B)    │
            └──────┬───────┘
                   ↓ 使用条件定义
          ┌───────────────────┐
          │ P(A | B)          │ ← 条件概率
          └─────────┬─────────┘
                    ↓ 结合先验
    ┌──────────────────────────────────────┐
    │ 全概率公式: P(A) = Σ P(A|Bᵢ)P(Bᵢ)    │
    └────────────────┬─────────────────────┘
                     ↓ 联合使用
         ┌────────────────────────────────┐
         │ 贝叶斯定理                     │
         │ P(B|A) = [P(A|B)P(B)] / P(A)   │
         └────────────────────────────────┘
                     ↓ 应用于
      医学诊断 · 法律推理 · 机器学习 · 数据科学
```

七、关键要点回顾

概念	公式/说明
条件概率	$P(A \mid B) = \dfrac{P(A \cap B)}{P(B)}$
乘法法则	$P(A \cap B) = P(A \mid B)P(B)$
全概率公式	$P(A) = \sum_i P(A \mid B_i)P(B_i)$
贝叶斯定理	$P(B_j \mid A) = \dfrac{P(A \mid B_j)P(B_j)}{\sum_i P(A \mid B_i)P(B_i)}$
独立性	$P(A \cap B) = P(A)P(B)$
检察官谬误	混淆 P(证据 | 无辜) 与 P(无辜 | 证据)

八、建议与思考

1. **避免直觉陷阱**：人类对概率直觉常常出错（如低估基数率），务必依赖数学工具。
2. **重视先验概率（Prior）**：贝叶斯分析中，忽略先验会导致严重偏差。
3. **多次验证的重要性**：单一测试不可靠，尤其在低患病率人群中。
4. **法律与统计交叉领域需谨慎**：不得将稀有性等同于罪责。

专题笔记：Sally Clark 案与“检察官谬误” – Prosecutor's Fallacy

——当数学被滥用时，正义如何失守？

资料来源：Leila Schneps & Coralie Colmez, *Math on Trial* (Chapter 1)
主题：统计误用、独立性假设错误、检察官谬误、Meadow's Law、司法公正中的数学责任

一、案件背景概述

◇ 基本信息

项目	内容
当事人	Sally Clark（英国律师）及其丈夫 Steve Clark
子女	Christopher（1996 年出生），Harry（1997 年出生）
事件时间线	• 1996 年 12 月：长子 Christopher 猝死，诊断为肺部感染 • 1998 年 1 月：次子 Harry 猝死，尸检发现视网膜出血、肋骨骨折等疑似虐待迹象 • 1999 年：Sally 被控谋杀两名婴儿，以“双重摇晃婴儿综合征”定罪 • 判决结果：1999 年 11 月，陪审团 10:2 裁定有罪 → 终身监禁 • 平反时间：2003 年 1 月 29 日，刑事复审委员会推翻原判，无罪释放 |

二、关键人物：Sir Roy Meadow

身份与影响力

- 英国著名儿科医生
- “代理型孟乔森综合症”（Munchausen Syndrome by Proxy, MSbP）提出者
- 曾任皇家儿科学会主席，1997 年获封爵士
- 在多起儿童死亡案中作为专家证人出庭

核心观点（被称为 "Meadow's Law"）

> **“一个摇篮死是悲剧，两个是可疑，三个就是谋杀。”**
> (*One cot death is a tragedy, two is suspicious, three is murder.*)

⚠️ 这并非科学定律，而是未经验证的经验主义断言，在司法实践中被当作“权威结论”。

三、庭审中的致命统计论证

Prosecution's Argument（控方逻辑链）

5. **前提**：无法确定婴儿确切死因 → 属于“摇篮死”（SIDS, Sudden Infant Death Syndrome）
6. **引用数据**：

- CESDI 报告显示：像 Clark 家庭这样无风险因素的家庭（高收入、非吸烟、母亲年龄>26 岁），SIDS 发生率为 $\frac{1}{8{,}543}$

7. **计算方法**：
   - 控方声称：两次 SIDS 是独立事件
   - 因此，连续发生两起的概率为：

$$P(\text{双 SIDS}) = \left(\frac{1}{8{,}543}\right)^2 \approx \frac{1}{73{,}000{,}000}$$

8. **结论陈述**：

   "这种自然发生的可能性只有七千三百万分之一，几乎不可能。所以她一定是凶手。"

四、两大根本性数学错误（Math Errors）

❌ 错误一：将非独立事件视为独立（Multiplying Non-Independent Probabilities）

☑ 正确认知：

- SIDS **不是完全随机、彼此独立的事件**
- 已知存在遗传、环境、免疫系统等潜在共同诱因
- 若一个家庭已出现一例 SIDS，第二例的风险**显著升高**（即存在正相关性）

📌 类比说明：

就像一个人得心脏病的概率是 1%，但如果他的父亲也得过，那他自己患病的概率远高于 1%。
你不能简单地把两个 1% 相乘得出"家族连发心脏病"的概率！

⚖ 统计学批评（来自皇家统计学会，2001 年正式致函）：

"该计算仅在 SIDS 家庭内完全独立的前提下成立，而这一假设缺乏实证支持；相反，有很强的先验理由认为它是错误的。"

❌ 错误二：检察官谬误（Prosecutor's Fallacy）

🔁 混淆了两个不同概率：

概率表达式	含义	实际值
$P(\text{双 SIDS} \mid \text{无辜})$	一个清白的母亲遭遇两次 SIDS 的概率	$\approx \frac{1}{73\text{ million}}$（假设独立）
$P(\text{无辜} \mid \text{双 SIDS})$	遇到双 SIDS 的母亲其实是无辜的概率	❗ 这才是我们需要的答案！

🧠 错误推理链条：

"这件事极不可能自然发生 ⇒ 所以它一定不是自然发生的 ⇒ 所以她是凶手。"

这相当于说：

"中彩票的概率极低 ⇒ 所以每个中奖的人都作弊了。"

📌 **正确逻辑应使用贝叶斯定理**：

$$P(\text{无辜} \mid \text{双 SIDS}) = \frac{P(\text{双 SIDS} \mid \text{无辜}) \cdot P(\text{无辜})}{P(\text{双 SIDS})}$$

但控方跳过了这个推理，直接用分子代替整个分数，犯了典型的条件概率倒置错误。

五、其他关键问题与缺失环节

1. 忽视替代医学解释

- 次子 Harry 尸体内发现 **8 种金黄色葡萄球菌（Staphylococcus aureus）** 和炎症细胞（polymorphs）
- 多位独立医学专家指出：**婴儿很可能死于严重细菌感染，甚至可能是脑膜炎**
- 死亡根本不应归类为“原因不明的 SIDS”，更谈不上“谋杀”

✅ 关键事实：这些医疗记录**从未提交给辩护方**，构成严重程序不公。

---

2. CONI 计划的实际数据反驳

- Care of Next Infants (CONI) 项目追踪了约 5,000 名曾有 SIDS 家庭的新生儿
- 其中已有 **8 名婴儿再次死于类似原因**
- 表明“双 SIDS”在现实中每几年就会发生一次，而非“百年一遇”

👉 这说明：即使在低风险家庭中，也不能排除重复发生的可能。

---

3. 社会偏见与刻板印象的影响

控方描绘 Sally 为：

- “职业野心勃勃”
- “过分整洁、控制欲强”
- “不像正常母亲那样悲伤”

这些主观判断被包装成“行为证据”，强化了“她不像无辜者”的偏见，进一步影响陪审团判断。

六、案件反转与平反之路

🕵 Steve Clark 的坚持

- 卖掉房子，搬到监狱附近
- 自费聘请律师、申请信息公开
- 最终获取被隐瞒的医疗报告

🔬 新证据揭示真相

- Harry 死于严重感染，非窒息

- 视网膜出血和肋骨骨折可能是急救过程中造成（CPR 常见副作用）

📢 皇家统计学会公开谴责（2001 年）

- 明确指出 Meadow 的统计方法“在统计上无效”
- 强调必须考虑家族聚集性和未知遗传因素

⚖️ 结果

- 2003 年 1 月 29 日：刑事复审委员会（CCRC）宣布撤销定罪
- Sally 获释，但身心严重受损
- 2007 年 3 月 16 日：因长期抑郁与酒精依赖去世，年仅 42 岁

七、后续影响与其他冤案

🔁 类似案件曝光

案件	结果
Angela Cannings	三名子女相继夭折 → 被判终身监禁 → 2003 年平反
Trupti Patel	三孩夭折 → Meadow 出庭作证 → 2003 年陪审团宣判无罪
Donna Anthony	两孩死亡 → 服刑多年后平反

所有案件均涉及 Meadow 的“1 in 73 million”式统计误导。

🏛 对 Sir Roy Meadow 的追责

- 2005 年：英国医学总会（GMC）裁定其“严重职业失当”
- 名字从医生注册簿中移除（后经上诉恢复，但已退休）
- GMC 批评：“他涉足了超出其专业领域的统计领域，且结论毫无依据。”

八、理论总结：三大核心教训

教训	内容
1. 独立性不能随意假设	在没有充分证据的情况下，不得将罕见事件视为相互独立。尤其对于生物现象，往往存在隐藏变量或遗传倾向。
2. 条件概率不可颠倒	$P(E \mid H) \neq P(H \mid E)$。这是“检察官谬误”的根源。必须使用贝叶斯框架进行推理。
3. 专家不应越界发言	医生可以判断病理，但无权做统计学家的工作。跨学科主张需谨慎，否则极易误导法庭。

九、补充知识：什么是 Munchausen Syndrome by Proxy (MSbP)?

项目	说明
定义	一种心理障碍，照顾者（通常是母亲）故意制造或伪造孩子的疾病症状，以获取医疗关注

项目	说明
提出者	Roy Meadow, 1977 年发表于《柳叶刀》
争议点	• 缺乏明确诊断标准 • 易被滥用为“无法解释死亡”的万能借口 • 导致数千儿童被强行带离家庭，酿成“合法绑架”悲剧 |

📌 如今医学界普遍认为：MSbP 应谨慎诊断，不能成为忽视真实病因的挡箭牌。

十、结语：数字的力量与危险

“When statistics are misused, justice is not served — it is perverted.” —— Leila Schneps

Sally Clark 案是一个沉痛的警示：

- 数学本身是中立的工具
- 但在法庭上，一旦被误解或滥用，就会变成最有力的定罪武器
- 我们需要的不仅是“数字”，更是对数字背后逻辑的深刻理解

🔚 一句话总结

“七千三百万分之一”的奇迹不是罪行，而是科学无知与司法傲慢合谋下的悲剧。

📓 概率论核心笔记：条件概率、独立性与随机变量

整理日期：2025 年 4 月
原始材料来源：*25-09-24.pdf*
主题：条件概率（Conditional Probability）、贝叶斯公式、事件独立性、随机变量（Random Variables）及其分布

一、条件概率（Conditional Probability）

1. 定义

设 A 和 B 是样本空间 Ω 中的两个事件，且 $P(B) > 0$，则 **在事件 B 发生的条件下，事件 A 发生的条件概率**定义为：

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

> ⚠️ 注意：此定义仅当 $P(B) > 0$ 时成立。

---

2. 补集与并集的条件概率性质

（1）补集规则：

$$P(A^c \mid B) = 1 - P(A \mid B)$$

（2）并集规则（容斥原理）：

$$P(A \cup C \mid B) = P(A \mid B) + P(C \mid B) - P(A \cap C \mid B)$$

---

3. 全概率公式（Divide and Conquer / Law of Total Probability）

若 $B_1, B_2, \ldots$ 构成样本空间 Ω 的一个**划分**（即互斥且完备），即：

- $B_i \cap B_j = \emptyset$ 对所有 $i \neq j$
- $\bigcup_i B_i = \Omega$

则对任意事件 A，有：

$$P(A) = \sum_i P\left(A \mid B_i\right) P(B_i)$$

这是处理复杂事件概率的基本工具，常用于"分情况讨论"。

---

4. 贝叶斯公式（Bayes' Theorem）

基本形式：

$$P(B \mid A) = \frac{P(A \mid B)P(B)}{P(A)}, \quad \text{其中 } P(A) > 0$$

推广形式（使用全概率公式展开分母）：

若 $B_1, B_2, \ldots$ 是 Ω 的一个划分，则：

$$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_i P(A \mid B_i)P(B_i)}$$

✅ 应用场景：从观测结果反推原因的概率（如医学诊断、机器学习分类等）

二、事件的独立性（Independence）

1. 两事件独立的等价定义

以下两种定义在 $P(B) > 0$ 下等价：

定义 1（条件概率视角）：

$$P(A \mid B) = P(A)$$

定义 2（乘积视角）：

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

✅ 若满足上述任一条件，则称事件 A 与 B **相互独立**，记作 $A \perp B$ 或 $A \perp\!\!\!\perp B$。

---

2. 多个事件的独立性：不能仅靠两两独立！考虑三个事件 A_1, A_2, A_3。

❌ 错误认知：

若 $P(A_i \cap A_j) = P(A_i)P(A_j)$ 对所有 $i \neq j$ 成立（即**两两独立**），是否意味着三者整体独立？

答案：否！

☑️ 正确定义：**相互独立（Mutual Independence）**

事件 $A_1, A_2, \ldots, A_n$ 称为**相互独立**，当且仅当对任意子集索引 $i_1 < i_2 < \cdots < i_k \leq n$，都有：

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

特别地，必须包括三重交集：

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2)P(A_3)$$

3. 反例说明：两两独立 ≠ 相互独立 (Pairwise independence ≠ mutual independence)

实验设置：

掷两个公平骰子（或理解为两次独立抛硬币，但原文似指骰子）。定义如下事件：

- A_1：第一个骰子点数为偶数 → $P(A_1)=\frac{1}{2}$
- A_2：第二个骰子点数为偶数 → $P(A_2)=\frac{1}{2}$
- A_3：两骰子点数之和为偶数 → $P(A_3)=\frac{1}{2}$

验证两两独立：

- $P(A_1\cap A_2)=P(\text{都为偶})=\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{4}=P(A_1)P(A_2)$
- $P(A_1\cap A_3)=P(\text{第一个为偶，和为偶})=P(\text{第二个也为偶})=\frac{1}{4}=P(A_1)P(A_3)$
- 同理 $P(A_2\cap A_3)=\frac{1}{4}=P(A_2)P(A_3)$

✅ 所以是**两两独立**

检查三重交集：

$$P(A_1\cap A_2\cap A_3)=P(\text{两骰均为偶})=\frac{1}{4}$$

而

$$P(A_1)P(A_2)P(A_3)=\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{8}$$

❌ 不相等！因此 **不是相互独立**

---

4. 独立性的闭包性质

若 $A\perp B$，则 $A\perp B^c$

证明简述：

已知 $P(A\cap B)=P(A)P(B)$，我们验证 $P(A\cap B^c)=P(A)P(B^c)$

$$\begin{aligned}P(A\cap B^c)&=P(A)-P(A\cap B)\\&=P(A)-P(A)P(B)\\&=P(A)(1-P(B))=P(A)P(B^c)\end{aligned}$$

✅ 得证。

更一般地，若 $A\perp B$，则 $A\perp B^c$, $A^c\perp B$, $A^c\perp B^c$ 均成立。

三、随机变量（Random Variables, RVs）

1. 动机示例：抛两枚硬币，统计正面数

实验：连续抛两枚公平硬币。

样本空间：

$$\Omega = \{(H,H),(H,T),(T,H),(T,T)\}$$

定义随机变量 $X:\Omega \to \mathbb{R}$，表示出现正面的数量：

结果	$X(\omega)$
(H,H)	2
(H,T), (T,H)	1
(T,T)	0

对应的概率质量函数（PMF）为：

$$\begin{cases} P(X=0)=\dfrac{1}{4} \\ P(X=1)=\dfrac{1}{2} \\ P(X=2)=\dfrac{1}{4} \end{cases}$$

✅ 这种映射方式将抽象结果转化为数值，便于分析和建模。

---

2. 连续型例子：转盘实验（Continuous Roulette）

设转盘可均匀旋转，指针落在区间 [0,1] 上。

实验：转动两次，得到结果 $(\omega_1,\omega_2) \in [0,1]^2$

定义总和变量：

$$X(\omega_1,\omega_2)=\omega_1+\omega_2$$

注意：

- 单点概率 $P(X=a)=0$（因为连续）
- 正确提问方式应为："落在某个区间的概率"

例如：

$$P(a<X<b)=P((\omega_1,\omega_2)\in[0,1]^2: a<\omega_1+\omega_2<b)$$

这需要通过几何概率或积分计算。

---

3. 随机变量的形式定义

给定概率空间 $(\Omega, \mathcal{F}, P)$，一个 **随机变量（Random Variable, RV）** 是一个函数：

$$X: \Omega \to \mathbb{R}$$

使得对于任意实数区间 $(a, b]$，集合

$$\{\omega \in \Omega: X(\omega) \in (a, b]\}$$

是一个**事件**（属于 $\mathcal{F}$）。

> 🔁 等价地，要求所有形如 $\{X \leq x\}$ 的集合都是事件。

---

4. 离散随机变量（Discrete Random Variables）

定义：

若随机变量 X 的取值集合是有限集或可数无限集（如 $\{x_1, x_2, \ldots\}$），则称其为**离散型随机变量**。

示例：

掷一颗六面骰子，$X(\omega) = \omega \in \{1,2,3,4,5,6\}$

---

5. 概率质量函数（Probability Mass Function, PMF）

对于离散随机变量 X，其 PMF 定义为：

$$p_X(x) = P(X = x)$$

满足：

- $p_X(x) \geq 0$
- $\sum_{x \in \mathrm{Im}(X)} p_X(x) = 1$

示例：掷两颗骰子，求点数和的 PMF

令 $X = D_1 + D_2$，可能取值 $2,3,\ldots,12$

x	$p_X(x)$	计算方式
2	$\frac{1}{36}$	(1,1)
3	$\frac{2}{36}$	(1,2),(2,1)
4	$\frac{3}{36}$	(1,3),(2,2),(3,1)
…	…	…
7	$\frac{6}{36}$	最大值

x	$p_X(x)$	计算方式
...	...	...
12	$\frac{1}{36}$	(6,6)

---

6. 累积分布函数（Cumulative Distribution Function, CDF）

无论离散还是连续，任何随机变量 X 都有一个 CDF：

$$F_X(x) = P(X \le x), \quad x \in \mathbb{R}$$

性质：

- $F_X(x) \in [0,1]$
- 单调非减
- 右连续
- $\lim_{x\to-\infty} F_X(x) = 0$，$\lim_{x\to\infty} F_X(x) = 1$

---

7. CDF 与 PMF 的关系（针对离散 RV）

若 X 是离散随机变量，取值于 $\{x_1 < x_2 < \cdots\}$，则：

（1）由 PMF 求 CDF：

$$F_X(x) = \sum_{x_k \le x} p_X(x_k)$$

（2）由 CDF 求 PMF：

$$p_X(x_i) = F_X(x_i) - \lim_{x\to x_i^-} F_X(x) = F_X(x_i) - F_X(x_i^-)$$

在跳跃点处，PMF 就是 CDF 的跳跃高度。

四、总结与关键要点回顾

概念	核心公式/定义	备注
条件概率	$P(A \mid B) = \frac{P(A\cap B)}{P(B)}$	$P(B) > 0$
全概率公式	$P(A) = \sum_i P(A \mid B_i)P(B_i)$	$\{B_i\}$ 为划分
贝叶斯公式	$P(B_j \mid A) = \frac{P(A \mid B_j)P(B_j)}{\sum_i P(A \mid B_i)P(B_i)}$	逆向推理
两事件独立	$P(A\cap B) = P(A)P(B)$	或 $P(A \mid B) = P(A)$
多事件独立	所有子集交集概率等于各自概率乘积	不能只看两两
随机变量	$X:\Omega\to\mathbb{R}$，且 $\{X \le x\}$ 是事件	映射到实数

概念	核心公式/定义	备注
PMF（离散）	$p_X(x) = P(X = x)$	仅对离散有效
CDF（通用）	$F_X(x) = P(X \le x)$	所有 RV 都有
PMF ↔ CDF	$p_X(x_i) = F_X(x_i) - F_X(x_i^-)$	离散情形下

五、附加说明与建议

- **避免混淆**："两两独立" ≠ "相互独立"，务必检查高阶交集。
- **建模思想**：引入随机变量是为了将实验结果量化，便于数学分析。
- **连续变量注意**：单点概率为零，关注区间概率 $P(a < X < b)$。
- **CDF 是更普适的对象**：它适用于离散、连续、混合型随机变量，是统一框架下的核心工具。

✅ **下一步学习建议**：

9. 学习常见离散分布（伯努利、二项、泊松）
10. 引入期望与方差的概念
11. 探索联合分布与协方差
12. 深入理解连续型随机变量与概率密度函数（PDF）

📌 附录：符号说明

符号	含义
$P(A \mid B)$	条件概率
A^c	事件 A 的补集
$A \cap B$	事件交集
$A \cup B$	事件并集
X	随机变量
$p_X(x)$	概率质量函数
$F_X(x)$	累积分布函数
$\perp\!\!\!\perp$	独立符号（可省略为 $\perp$）

二、条件概率中的全概率公式（带额外条件）

现在我们考虑：在已知另一个事件 C 发生的前提下，求事件 A 的条件概率 $P(A \mid C)$。此时，如果存在一个关于 B_i 的划分，那么我们可以将全概率公式推广到条件概率的情形。

✅ **条件版本的全概率公式：**

设 $B_1, B_2, \ldots, B_n$ 是样本空间的一个划分，且 $P(B_i \cap C) > 0$ 或至少 $P(C) > 0$，则：

$$P(A \mid C) = \sum_{i=1}^{n} P(A \mid B_i \cap C) \cdot P(B_i \mid C)$$

离散型随机变量（Discrete Random Variables）

一、基本定义

1. 随机变量（Random Variable, RV）

- **定义**：设 $(\Omega, \mathcal{F}, P)$ 是一个概率空间，**一个随机变量是一个函数**

$$X: \Omega \to \mathbb{R}$$

  满足：对任意 $a, b \in \mathbb{R}$，集合 $\{\omega \in \Omega: a \le X(\omega) \le b\}$ 是事件（即属于 $\mathcal{F}$）。

- **核心思想**：我们只关心随机试验的**数值结果**，而非样本点本身。

---

2. 离散型随机变量

- **定义**：若随机变量 X 的**像集**（取值集合）$\mathrm{Im}(X)$ 是**有限集**或**可数无限集**，则称 X 为**离散型随机变量**。
  即：

$$\mathrm{Im}(X) = \{x_1, x_2, \dots\} \subseteq \mathbb{R}, \quad \text{其中 } |\mathrm{Im}(X)| \le \aleph_0$$

二、概率质量函数（PMF）

1. 定义

对离散型随机变量 X，其**概率质量函数**（Probability Mass Function, PMF）定义为：

$$p_X(x_i) = P(X = x_i) = P(\{\omega \in \Omega: X(\omega) = x_i\})$$

- 定义域：$\mathrm{Im}(X)$
- 值域：$[0,1]$
- 满足归一性：

$$\sum_i p_X(x_i) = 1$$

2. 区间概率计算

对任意实数 $a < b$：

$$P(a \le X \le b) = \sum_{x_i \in [a,b]} p_X(x_i)$$

三、累积分布函数（CDF）

1. 定义

随机变量 X 的**累积分布函数**（Cumulative Distribution Function, CDF）为：

$$F_X(x) = P(X \le x), \quad x \in \mathbb{R}$$

2. 离散型 RV 的 CDF 性质

- **等价性**：对于离散型 RV，知道 PMF p_X 与知道 CDF F_X 是等价的。
- **跳跃关系**：

$$p_X(x_i) = F_X(x_i) - \lim_{x \to x_i^-} F_X(x) = \text{在 } x_i \text{ 处的跳跃高度}$$

四、CDF 的通用性质（适用于所有 RV）

设 $F_X(x) = P(X \leq x)$，则：

性质	描述
F1	$\lim_{x\to-\infty} F_X(x) = 0$
F2	$\lim_{x\to+\infty} F_X(x) = 1$
F3	**单调不减**：若 $a \leq b$，则 $F_X(a) \leq F_X(b)$
F4	**右连续**：$\lim_{h\to0^+} F_X(x+h) = F_X(x)$
F5	**跳跃大小 = 概率质量**：在点 a 处的跳跃为 $p_X(a) = F_X(a) - F_X(a^-)$

> **重要结论**：任何满足 F1–F5 的函数 F 都是某个随机变量的 CDF。

五、随机变量的"律"（Law）

定义

两个离散型随机变量 X 和 Y（可在不同概率空间）**具有相同的律**（same law），当且仅当：

13. $\mathrm{Im}(X) = \mathrm{Im}(Y)$
14. $p_X(x) = p_Y(x)$ 对所有 x 成立

> **核心思想**："律"完全刻画了随机变量的统计行为，是研究 RV 的本质对象。

六、常见离散分布（Catalogue of Popular Discrete Laws）

1. 伯努利分布（Bernoulli）

- **记号**：$X \sim \mathrm{Ber}(p)$
- **取值**：$\{0,1\}$
- **PMF**：

$$p_X(1) = p, \quad p_X(0) = 1 - p$$

- **解释**：单次试验，"成功"（1）概率为 p，"失败"（0）概率为 $1-p$
- **例子**：抛一枚偏置硬币，正面为成功

---

2. 离散均匀分布（Uniform）

- **记号**：$X \sim \mathrm{Unif}\{x_1, \ldots, x_N\}$
- **取值**：有限集合 $\{x_1, \ldots, x_N\}$

- **PMF**：

$$p_X(x_i) = \frac{1}{N}, \quad i = 1, \dots, N$$

- **例子**：掷一个公平的六面骰子，$X \in \{1,2,3,4,5,6\}$

---

3. 二项分布（Binomial）

- **记号**：$X \sim \mathrm{Bin}(n, p)$
- **背景**：n 次独立伯努利试验中“成功”次数
- **取值**：$\{0,1, \dots, n\}$
- **PMF**：

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0,1, \dots, n$$

- **例子**：抛 10 次硬币，正面出现次数 $X \sim \mathrm{Bin}(10,0.5)$

---

4. 几何分布（Geometric）

- **记号**：$X \sim \mathrm{Geom}(p)$
- **背景**：首次成功所需的试验次数（含成功那次）
- **取值**：$\{1,2,3, \dots\} = \mathbb{N}^+$
- **PMF**：

$$p_X(k) = (1-p)^{k-1} p, \quad k = 1,2, \dots$$

- **无记忆性（Memoryless Property）**：

$$P(X > m + n \mid X > m) = P(X > n), \quad \forall m, n \in \mathbb{N}$$

- **重要结论：几何分布是唯一具有无记忆性的离散分布，指数分布是具有无记忆性的连续分布**

注：文档中有一处笔误：“GEUM(p)” 应为 “Geom(p)”

---

5. 泊松分布（Poisson）

⚠️ 文档中仅提及名称 “POISSON LAW”，但未给出定义。此处补充标准定义供完整性：

- **记号**：$X \sim \mathrm{Pois}(\lambda)$
- **取值**：$\{0,1,2, \dots\}$
- **PMF**：

$$p_X(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0,1,2, \dots$$

- **应用场景**：单位时间内事件发生次数（如电话呼叫、放射性衰变）

注：原文中“MER LOTTERY”等例子疑似草稿或笔误，可能意指“某彩票中奖次数”服从泊松近似，但未展开。

七、示例分析：掷两颗骰子的最大值

- **样本空间**：$\Omega = \{1, \dots, 6\} \times \{1, \dots, 6\}$

- **定义 RV**：$X(\omega_1,\omega_2)=\max(\omega_1,\omega_2)$
- **取值**：$\{1,2,3,4,5,6\}$
- **PMF 计算**：

$$
\begin{aligned}
P(X=1)&=\frac{1}{36}\\
P(X=2)&=\frac{3}{36}\\
P(X=3)&=\frac{5}{36}\\
P(X=4)&=\frac{7}{36}\\
P(X=5)&=\frac{9}{36}\\
P(X=6)&=\frac{11}{36}
\end{aligned}
$$

- **验证**：$1+3+5+7+9+11=36$，总和为 1。

此例说明如何从具体实验构造离散 RV 并计算其分布。

八、研究随机实验的标准策略

15. **识别样本空间与随机变量**
16. **确定其分布律（Law）**
17. **查表比对常见分布**（参考教材 p.429–430）
18. **利用已知分布的性质进行计算**

核心理念：一旦知道“律”，就掌握了 RV 的全部统计信息。

附：符号与排版说明

- $X\sim\mathrm{Dist}(\cdot)$：表示 X 服从某分布
- $\mathrm{Im}(X)$：随机变量 X 的取值集合（像集）
- $p_X(x)$：PMF；$F_X(x)$：CDF
- 所有公式采用 LaTeX 渲染，确保清晰美观

🟦 离散型随机变量（Discrete Random Variables）系统笔记

一、基本定义与概念

1. 随机变量（Random Variable, RV）

- **定义**：设样本空间为 Ω，随机变量 X 是一个从 Ω 到实数集 $\mathbb{R}$ 的可测函数：
$$X: \Omega \to \mathbb{R}$$
- 若 X 的取值集合（像集）为可数集（如 $\{x_1, x_2, \dots\}$），则称 X 为**离散型随机变量**。

2. 概率质量函数（Probability Mass Function, PMF）

- 记为 $p_X(x) = \mathbb{P}(X = x)$
- 满足：
$$p_X(x) \ge 0, \quad \sum_{x \in \operatorname{Im}(X)} p_X(x) = 1$$
- **关键性质**：PMF 完全决定了离散型随机变量的分布律（Law）。

3. 分布律（Law of a Random Variable）

- 随机变量 X 的**分布律**是指其在 $\mathbb{R}$ 上诱导的概率测度 $\mathbb{P}_X$，定义为：
$$\mathbb{P}_X(B) = \mathbb{P}(X \in B), \quad \forall B \subseteq \mathbb{R}$$
- 对于离散情形，只需知道 $p_X(x)$ 即可完全刻画 $\mathbb{P}_X$。

4. 累积分布函数（Cumulative Distribution Function, CDF）

- 定义：
$$F_X(x) = \mathbb{P}(X \le x)$$
- 对离散型 RV：
$$F_X(x) = \sum_{x_i \le x} p_X(x_i)$$
- 反过来，若已知 CDF，则：
$$p_X(x_i) = F_X(x_i) - \lim_{x \uparrow x_i} F_X(x)$$

✅ **结论**：对离散型随机变量，**知道 PMF ⇔ 知道 CDF ⇔ 知道分布律**。

二、无记忆性（Memoryless Property）

1. 定义

- 离散型随机变量 X 具有**无记忆性**，若对任意整数 $a, b \ge 0$：
$$\mathbb{P}(X > a + b \mid X > a) = \mathbb{P}(X > b)$$

2. 关键定理

定理：离散型随机变量具有无记忆性 **当且仅当** 其服从**几何分布**。

3. 几何分布（Geometric Distribution）

- 记作 $X \sim \operatorname{Geom}(p)$，其中 $0 < p \le 1$

- 支撑集：$\mathrm{Im}(X) = \{1,2,3,\dots\}$（首次成功所需试验次数）
- PMF：

$$\mathbb{P}(X = k) = (1-p)^{k-1}p, \quad k = 1,2,\dots$$

- 无记忆性验证：

$$\mathbb{P}(X > n) = (1-p)^n \Rightarrow \mathbb{P}(X > a + b \mid X > a) = (1-p)^b = \mathbb{P}(X > b)$$

📌 **唯一性**：几何分布是**唯一的离散无记忆分布**。

三、泊松分布（Poisson Distribution）

1. 背景与建模动机

- 适用于描述**稀有事件**在固定时间/空间内的发生次数。
- 假设：
    - 事件独立发生；
    - 单位时间内平均发生 λ 次；
    - 同一时刻几乎不可能发生两次以上事件。

2. 定义

- 记作 $X \sim \mathrm{Pois}(\lambda)$，其中 $\lambda > 0$
- 支撑集：$\{0,1,2,\dots\}$
- PMF：

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

- 期望与方差：

$$\mathbb{E}[X] = \lambda, \quad \mathrm{Var}(X) = \lambda$$

3. 从二项分布逼近

- 当 $n \to \infty$，$p \to 0$，且 $np \to \lambda$ 时，

$$\mathrm{Bin}(n,p) \xrightarrow{d} \mathrm{Pois}(\lambda)$$

4. 经典应用案例

(a) 普鲁士骑兵马踢致死事件（Bortkiewicz 数据）

- 观察：10 个骑兵军团 × 20 年 = 200 “军团-年”
- 记录每年每军团因马踢死亡人数
- 拟合泊松分布，$\lambda \approx 0.61$
- 实际频数与泊松预测高度吻合 → 验证事件随机性

(b) 二战伦敦 V-1 炸弹落点分析

- 将南伦敦划分为 576 个 0.25 km² 的网格
- 统计每个网格被击中次数
- 总炸弹数 $\approx 537 \Rightarrow \lambda = 537/576 \approx 0.932$
- 观察频数与 $\mathrm{Pois}(0.932)$ 预测一致
- **结论**：轰炸是**随机的**，非精准打击

🔍 数据验证方法：

$$\hat{\lambda} = \frac{1}{n}\sum_i k_i f_i$$

其中 f_i 为击中 i 次的网格数。

四、补充说明与符号澄清

- 文档中部分符号存在笔误或排版混乱（如 “IM(x/”, “Px(*)” 等），已根据上下文校正为标准概率论记号。
- “CAM” 应为 “same distribution”（同分布）之误。
- “RESULTS(NUMBER” 等碎片文字推测为课堂板书速记，已整合进相应理论模块。

五、总结框架图

```
离散型随机变量
│
├── 分布刻画
│   ├── PMF: p_X(x) = P(X = x)
│   ├── CDF: F_X(x) = Σ_{x_i ≤ x} p_X(x_i)
│   └── Law: 由 PMF 唯一确定
│
├── 无记忆性
│   └── ⇔ 几何分布 Geom(p)
│
└── 经典分布
    ├── 几何分布：首次成功模型
    └── 泊松分布：稀有事件计数模型
        ├── 来源：Bin(n,p) 极限
        └── 应用：马踢死亡、炸弹落点等
```

《连续型随机变量及其常见分布》

一、连续型随机变量的定义

1. 基本定义

设 $X:\Omega \to \mathbb{R}$ 为实值随机变量（R.V.），其**像集**（image）记为 $\text{Im}(X)$。

> **定义（连续型随机变量）**：
> 随机变量 X 称为**连续型**，若满足：
>
> a. $\text{Im}(X)$ 是若干区间的并集（即为区间或可表示为区间之并）；
> b. 其**累积分布函数**（CDF）$F_X(x) = \mathbb{P}(X \le x)$ 是**连续函数**。
>
> **定义（强连续型随机变量）**：
> 若进一步地，$F_X(x)$ **几乎处处可导**（except possibly at finitely many points），则称 X 为**强连续型**随机变量。

---

2. 概率密度函数（PDF）

若 X 为强连续型随机变量，则存在非负函数 $f_X(x)$，使得：

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, \quad \forall x \in \mathbb{R}$$

该函数 $f_X(x)$ 称为 X 的**概率密度函数**（PDF），满足：

- $f_X(x) \ge 0$ 对所有 x；
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$；
- 对任意区间 $(a,b]$，有：

$$\mathbb{P}(a < X \le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$$

注：连续型随机变量在任意单点处的概率为零：

$$\mathbb{P}(X = x) = 0, \quad \forall x \in \mathbb{R}$$

二、常见连续型分布

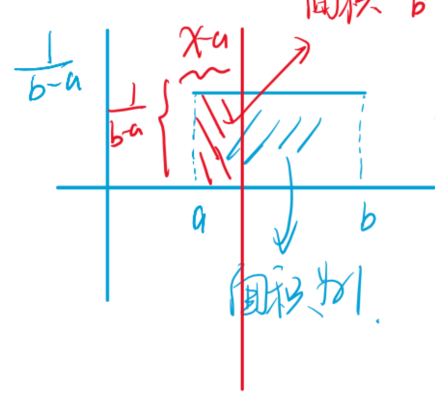
均匀分布

1. 均匀分布（Uniform Distribution）

- **记号**：$X \sim \mathcal{U}(a,b)$
- **PDF**：

$$f_X(x) = \begin{cases} \dfrac{1}{b-a}, & a \le x \le b \\ 0, & \text{otherwise} \end{cases}$$

面积 $\frac{x-a}{b-a} = F_X(x)$
$\Downarrow$
CDF



- **CDF**：

$* \ E[X] = \frac{a+b}{2}$

这里期望即均值 $\frac{a+b}{2}$

Cumulative Distribution Function 累积分布函数

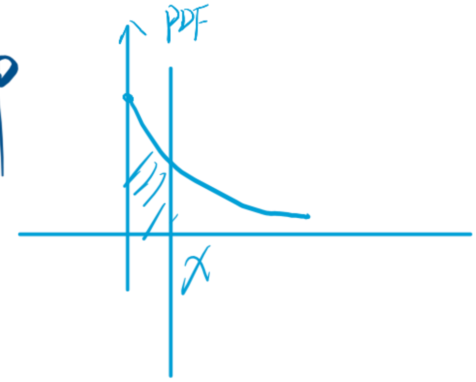
CDF函数

$$F_X(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x \le b \\ 1, & x > b \end{cases}$$

---

2. 指数分布（Exponential Distribution）

指数分布



- **记号**：$X \sim \mathrm{Exp}(\lambda)$，其中 $\lambda > 0$
- **PDF**：

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \ge 0 \\ 0, & x < 0 \end{cases}$$

- **CDF**： ⇒ CDF用积分推

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \ge 0$$

$$F_X(x) = \int_0^x \lambda e^{-\lambda y} dy = -e^{-\lambda y}\Big|_0^x = -e^{-\lambda x} - (-1) = 1 - e^{-\lambda x}$$

应用场景

- 放射性原子的**存活时间**；
- 事件发生之间的**等待时间**（如泊松过程的间隔时间）。

无记忆性（Memoryless Property）

性质：对任意 $s, t \ge 0$，

$$\mathbb{P}(X > s + t \mid X > s) = \mathbb{P}(X > t)$$

定理：若连续型随机变量 X 满足 $\mathrm{Im}(X) = [0, \infty)$ 且具有无记忆性，则必有 $X \sim \mathrm{Exp}(\lambda)$（对某个 $\lambda > 0$）。

↓ 适用于几何分布&指数分布

证明思路简述：

- 定义 $g(t) = \mathbb{P}(X > t)$；
- 由无记忆性得 $g(s + t) = g(s)g(t)$；
- 结合连续性，解得 $g(t) = e^{-\lambda t}$，即 $F_X(t) = 1 - e^{-\lambda t}$。

---

3. 帕累托分布（Pareto Distribution, Type I）

- **背景**：由意大利经济学家 Vilfredo Pareto 提出，用于描述财富分配、城市规模、保险索赔、陨石大小等**幂律现象**。
- **“80-20 法则”**：当形状参数 $\alpha = \log_5 4 \approx 1.16$ 时，恰好满足 80% 的结果由 20% 的原因导致。

参数

- 尺度参数（最小值）：$x_m > 0$
- 形状参数（尾部指数）：$\alpha > 0$
- 记作：$X \sim \mathrm{Pareto}(x_m, \alpha)$，通常设 $x_m = 1$，简写为 $\mathrm{Par}(\alpha)$

生存函数（Tail Function）

$$\mathbb{P}(X > x) = \left(\frac{x_m}{x}\right)^{\alpha}, \quad x \ge x_m$$

CDF

$$F_X(x) = \begin{cases} 0, & x < x_m \\ 1 - \left(\frac{x_m}{x}\right)^{\alpha}, & x \geq x_m \end{cases}$$

PDF

$$f_X(x) = \begin{cases} \frac{\alpha x_m^{\alpha}}{x^{\alpha+1}}, & x \geq x_m \\ 0, & x < x_m \end{cases}$$

特点

- **重尾分布**（Heavy-tailed）：相比指数分布，极端大值出现概率更高；
- 均值存在当且仅当 $\alpha > 1$；方差存在当且仅当 $\alpha > 2$。

三、补充说明与图示参考（来自文档附图）

- **图示对比**：指数分布 vs 帕累托分布（均值为 1）
    - 指数分布：$\lambda = 1$
    - 帕累托分布：$\alpha = 4$, $x_m = 0.75$
    - 结论：帕累托分布两端（极小值和极大值）概率密度更高，体现"重尾"特性。
- **实际应用案例**（来自文档后半部分）：
    - 硬盘**潜在扇区错误**（LSE）的时间与空间分布；
    - 错误爆发（error burst）在 90%以上情况下发生在**同一 2 周扫描周期内**；
    - 空间邻近的错误通常也**时间邻近**，暗示共同成因；
    - 帕累托分布被用于拟合错误发生间隔或错误数量的分布。

四、关键公式汇总

概念	公式
CDF 与 PDF 关系	$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt$
区间概率	$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\,dx$
指数分布无记忆性	$\mathbb{P}(X > s + t \mid X > s) = e^{-\lambda t}$
帕累托生存函数	$\mathbb{P}(X > x) = (x_m/x)^{\alpha}$, $x \geq x_m$
帕累托 PDF	$f_X(x) = \alpha x_m^{\alpha}/x^{\alpha+1}$

五、总结

本文档系统介绍了连续型随机变量的定义、性质及其三大典型分布：**均匀分布**（基础模型）、**指数分布**（无记忆性、泊松过程核心）、**帕累托分布**（幂律、重尾现象建模）。这些分布在可靠性工程、金

融风险、数据存储等领域具有广泛应用。理解其数学结构与实际含义，是概率建模与数据分析的重要基础。

PDF（Probability Density Function，概率密度函数）和 PMF（Probability Mass Function，概率质量函数）是概率论中用于描述随机变量分布的两种不同工具，它们分别适用于连续型和离散型随机变量。以下是它们的核心区别，结构化说明如下：

一、适用对象不同

概念	适用随机变量类型	说明
PMF	**离散型随机变量**（Discrete R.V.）	取值为可数集合（如整数、有限集）
PDF	**连续型随机变量**（Continuous R.V.）	取值为不可数区间（如实数区间）

二、定义与数学形式

1. PMF（概率质量函数）

- 记为 $p_X(x) = \mathbb{P}(X = x)$
- 直接给出**某一点的概率**
- 满足：

$$p_X(x) \geq 0, \sum_{x \in \mathrm{Im}(X)} p_X(x) = 1$$

例：若 $X \sim \text{Bernoulli}(p)$，则

$$p_X(0) = 1 - p, p_X(1) = p$$

---

2. PDF（概率密度函数）

- 记为 $f_X(x)$
- **不直接表示概率**，而是“密度”；单点概率恒为 0：

$$\mathbb{P}(X = x) = 0, \forall x$$

- 概率通过**积分**计算：

$$\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\,dx$$

- 满足：

$$f_X(x) \geq 0, \int_{-\infty}^{\infty} f_X(x)\,dx = 1$$

例：若 $X \sim \mathcal{U}(0,1)$，则
$f_X(x) = 1 \text{for } 0 \leq x \leq 1$，否则 0

三、直观理解对比

特性	PMF	PDF
图形表示	离散点上的“柱状高度”	连续曲线下的“面积”
单点值意义	表示该点发生的**确切概率**	**无直接概率意义**，需积分才有概率
总和/积分	所有 $p_X(x)$之和为 1	曲线下总面积为 1
类比	像“质量集中在点上”	像“质量连续分布，密度可变”

四、与 CDF 的关系

两者都通过**累积分布函数**（CDF, $F_X(x) = \mathbb{P}(X \le x)$）与分布关联:

- **离散型**（PMF）：

$$F_X(x) = \sum_{t \le x} p_X(t)$$

- **连续型**（PDF）：

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt, f_X(x) = \frac{d}{dx} F_X(x) \text{（几乎处处）}$$

五、常见误区提醒

- ❌ “PDF 在某点的值是概率” → **错误！**
  正确：PDF 值可大于 1（只要积分=1），例如 $f_X(x) = 10$on $[0,0.1]$是合法的。
- ✅ “PMF 在某点的值是概率” → **正确！**

六、总结一句话

PMF 给出离散点的概率，PDF 给出连续变量的概率密度，概率需通过对 PDF 积分获得。

📓 概率论与统计学笔记

日期：2025 年 10 月 15 日 | 来源：课堂讲义（25-10-15.pdf）

一、连续型随机变量基础

1.1 定义与概率密度函数（PDF）

- 若随机变量 X 是**连续型**的，则存在一个**概率密度函数**（PDF） $f_X(x)$，满足：

$$P(a \le X \le b) = \int_a^b f_X(x)\,dx$$

- PDF 的性质：
  - $f_X(x) \ge 0$，对所有 $x \in \mathbb{R}$
  - $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$

注：PDF 本身不是概率，但其积分给出概率。

二、常见连续分布

2.1 指数分布（Exponential Distribution）

- 记作：$X \sim \mathrm{Exp}(\lambda)$，其中 $\lambda > 0$
- PDF：

$$f_X(x) = \lambda e^{-\lambda x}, \quad x \ge 0$$

- CDF：

$$F_X(x) = 1 - e^{-\lambda x}, \quad x \ge 0$$

- **无记忆性（Memoryless Property）**：

$$P(X > s + t \mid X > s) = P(X > t), \quad \forall s, t \ge 0$$

2.2 帕累托分布（Pareto Distribution）

- 记作：$X \sim \mathrm{Pareto}(x_m, \alpha)$，通常设 $x_m = 1$，简写为 $\mathrm{Par}(\alpha)$
- PDF（当 $x_m = 1$）：

$$f_X(x) = \alpha x^{-(\alpha+1)}, \quad x \ge 1$$

- CDF：

$$F_X(x) = 1 - x^{-\alpha}, \quad x \ge 1$$

- **"近似无记忆性"**（Almost Memoryless）：

$$P(X > s + t \mid X > s) \approx \left(\frac{s}{s+t}\right)^{\alpha} \qquad \text{（随 } s \to \infty \text{ 趋近于常数）}$$

2.3 指数分布与帕累托分布的关系

- 若 $X \sim \mathrm{Pareto}(\alpha)$，令 $Y = \ln X$，则：

$$Y \sim \mathrm{Exp}(\alpha)$$

- 反之，若 $Y \sim \mathrm{Exp}(\alpha)$，则 $X = e^Y \sim \mathrm{Pareto}(\alpha)$

练习题：若 $Y \sim \mathrm{Exp}(\alpha)$，是否 $Y \sim \mathrm{Pareto}(\alpha)$？
答：否。两者定义域与尾部行为不同。

三、正态分布（Normal Distribution）

3.1 历史背景

- **De Moivre (1733)**：二项分布的正态近似（中心极限定理雏形）
- **Laplace (1783)**：误差分布理论
- **Gauss (1809)**：给出显式密度函数，确立"高斯分布"地位

3.2 定义

- 随机变量 $X \sim N(\mu, \sigma^2)$ 表示其服从均值为 μ、方差为 σ^2 的正态分布。
- PDF：

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \mathrm{e}^{\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)}$$

3.3 标准正态分布

- 记作：$Z \sim N(0,1)$
- 标准化变换：

$$Z = \frac{X-\mu}{\sigma} \quad \Rightarrow \quad X = \mu + \sigma Z$$

- CDF 记为 $\Phi(z) = P(Z \le z)$，常用查表或计算器计算。

3.4 正态分布的性质

1. **对称性**：

$$P(Z \le -a) = P(Z \ge a) = 1 - \Phi(a)$$

2. **总概率为 1**：

$$\int_{-\infty}^{\infty} \phi\,(z)\,dz = 1$$

3. **尾部概率计算**（查表）：
    - 表格通常给出右尾概率：$P(Z \ge a) = 1 - \Phi(a)$
    - 示例：$P(Z \ge 2) = 0.0228$

3.5 查表技巧

- 利用对称性与互补性：

$$P(|Z| > a) = 2P(Z > a) = 2[1 - \Phi(a)]$$

- 示例：
    - $P(Z < 0.82) = \Phi(0.82) \approx 1 - 0.2061 = 0.7939$（查右尾表得 0.2061）

四、应用与实例

4.1 实际场景建模

- **指数分布**：服务时间、设备寿命、泊松过程间隔
- **帕累托分布**：收入分布、城市人口、文件大小、保险索赔（重尾）
- **正态分布**：测量误差、考试分数、身高体重等自然现象

4.2 图形辅助理解

- 正态 CDF 曲线（S 形）随 μ 平移，随 σ 缩放
- 不同 σ^2 对应不同“胖瘦”程度（见讲义图：$\sigma^2 = 0.2,1.0,5.0$）
- σ^2越小，越瘦高；σ^2越大，越矮胖

五、附录：标准正态右尾概率表（节选）

a	$P(Z \geq a)$
0.0	0.5000
0.5	0.3085
1.0	0.1587
1.5	0.0668
2.0	0.0228
2.5	0.0062
3.0	0.0013

完整表格见讲义第 2–3 页。

六、总结要点

分布	支撑集	无记忆性	尾部特性	典型应用
指数	$[0,\infty)$	✅	轻尾（指数衰减）	排队论、可靠性
帕累托	$[1,\infty)$	❌（近似）	重尾（幂律）	经济、网络流量
正态	$\mathbb{R}$	❌	轻尾（高斯衰减）	自然科学、测量误差

二、数学刻画（以右尾为例）

设随机变量 X 的尾部概率为 $\mathbb{P}(X > x)$（即生存函数）。

类型	衰减形式	举例	特点
轻尾（Light-tailed）	指数级或更快衰减：$\mathbb{P}(X > x) \sim e^{-cx}$ 或 e^{-cx^2}	指数分布、正态分布	极端事件罕见；
重尾（Heavy-tailed）	幂律衰减：$\mathbb{P}(X > x) \sim x^{-\alpha}$（$\alpha > 0$）	帕累托分布、对数正态、t 分布（自由度小）	极端事件较常见

专业笔记：概率论与数理统计核心概念整理

第一部分：连续型随机变量与正态分布

1. 正态分布 (Normal Distribution)

- **定义**：随机变量 X 服从正态分布，记为 $X \sim N(\mu, \sigma^2)$，其中：
    - μ：均值 (Mean)
    - σ^2：方差 (Variance)
    - σ：标准差 (Standard Deviation)
- **概率密度函数 (PDF)**:

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < +\infty$$

- **标准正态分布**:
    - 当 $\mu = 0, \sigma = 1$ 时，称为标准正态分布，记为 $Z \sim N(0,1)$。
    - **其累积分布函数 (CDF) 通常用 $\Phi(z)$ 表示：**

$$\boldsymbol{\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt}$$

    - **重要性质**:

$$P(-a \leq Z \leq a) = \Phi(a) - \Phi(-a) = 2\Phi(a) - 1$$

    （原文中的 `p(21)=p1s8s1) C 2)十=1-2T0)=-20KR1=` 可能是对此性质的错误转录。）
- **标准化**：对于任意 $X \sim N(\mu, \sigma^2)$，可以通过标准化变换得到标准正态变量：

$$Z = \frac{X - \mu}{\sigma} \sim N(0,1)$$

因此，$P(X \leq x) = P\left(Z \leq \frac{x-\mu}{\sigma}\right) = \Phi\left(\frac{x-\mu}{\sigma}\right)$。

- **应用示例**:
    - 原文提及 `p(-264KX-uE)`，推测可能是计算某个概率，如 $P(-2.64 < \frac{X-\mu}{\sigma} < 0)$ 或类似形式，需查表或使用软件计算 Φ 值。
    - 原文 `上[0)=-20002=0944` 可能意指 $\Phi(2.00) \approx 0.9772$ 或 $\Phi(1.59) \approx 0.944$ 等，具体数值需根据上下文确认。

第二部分：离散型随机变量与指数分布

1. 指数分布 (Exponential Distribution)

- **定义**：用于描述事件发生的时间间隔，常用于可靠性分析、排队论等。记为 $X \sim \text{Exp}(\lambda)$，其中 $\lambda > 0$ 是速率参数 (rate parameter)。
- **概率密度函数 (PDF)**:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- **累积分布函数 (CDF)**:

$$F_X(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

- **期望与方差**:
  - 期望 (Mean): $E[X] = \frac{1}{\lambda}$
  - 方差 (Variance): $\mathrm{Var}(X) = \frac{1}{\lambda^2}$
- **无记忆性 (Memoryless Property)**: 这是指数分布最重要的特性之一：

$$P(X > s + t | X > s) = P(X > t), \quad \forall s, t \geq 0$$

（原文 `ut)= nassSwWingat-tinet` 可能是指“无记忆性”或“等待时间”。）

- **应用示例**:
  - 原文 `2X~Expl 入) 入 0045ctune tincntrie` 可能意指 $X \sim \mathrm{Exp}(\lambda = 0.045)$，表示平均等待时间为 $1/0.045 \approx 22.22$ 单位时间。
  - `Mt= 入 t` 可能指在时间 t 内事件发生的平均次数（泊松过程），即 $M(t) = \lambda t$。

第三部分：累积分布函数 (CDF) 与分位数

1. 累积分布函数 (Cumulative Distribution Function, CDF)

- **定义**：对于随机变量 X，其累积分布函数 $F_X(x)$ 定义为：

$$F_X(x) = P(X \leq x)$$

  - 对于连续型变量，$F_X(x) = \int_{-\infty}^{x} f_X(t)dt$。
  - 对于离散型变量，$F_X(x) = \sum_{t \leq x} P(X = t)$。
- **性质**:
  - $0 \leq F_X(x) \leq 1$
  - $F_X(x)$ 是非减函数。
  - $\lim_{x \to -\infty} F_X(x) = 0$
  - $\lim_{x \to +\infty} F_X(x) = 1$
  - （原文 `七 09)=< xet Fx(to9)=o1` 可能是在描述这些极限性质。）

2. 分位数 (Quantile)

- **定义**：对于给定的概率 p $(0 < p < 1)$，随机变量 X 的 p-分位数 x_p 是满足以下条件的最小值：

$$F_X(x_p) = P(X \leq x_p) = p$$

（原文 thduatile.oor/o0pte poncanole 和 thrintnt valne bip wbd-toe8Salrsluton 可能指“分位数”及其求解方法。）

- **常见分位数**:
    - **中位数 (Median)**: $p = 0.5$ 时的分位数。
    - **四分位数 (Quartiles)**:
        - 第一四分位数 (Q1): $p = 0.25$
        - 第二四分位数 (Q2，中位数): $p = 0.5$
        - 第三四分位数 (Q3): $p = 0.75$
- **计算**:
    - 对于连续型变量，通常通过解方程 $F_X(x_p) = p$ 得到 x_p。
    - （原文 量级(&atiw) p(xetp)=fx(tp)-p 可能是在描述求解分位数的过程。）

第四部分：离散型随机变量（伯努利与二项分布）

1. 伯努利分布 (Bernoulli Distribution)

- **定义**：描述一次伯努利试验的结果，只有两种可能：成功（1）或失败（0）。记为 $X \sim \text{Bernoulli}(p)$，其中 p 是成功的概率。
- **概率质量函数 (PMF)**:

$$P(X = k) = \begin{cases} p, & k = 1 \\ 1 - p, & k = 0 \end{cases}$$

- **期望与方差**:
    - $E[X] = p$
    - $\text{Var}(X) = p(1 - p)$

2. 二项分布 (Binomial Distribution)

- **定义**：描述 n 次独立伯努利试验中成功的次数。记为 $X \sim \text{Binomial}(n, p)$。
- **概率质量函数 (PMF)**:

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k = 0,1,2,\dots,n$$

- **期望与方差**:
    - $E[X] = np$
    - $\text{Var}(X) = np(1 - p)$
- **应用示例**:
    - 原文 E2)Xa Cconp) 1---p) p 可能是在描述二项分布的期望或方差公式。
    - pe1=/xw+0px10=+-p6Px(o)=+p=D+0.pcx=g)=0(1-0+i-p=p 可能是在尝试推导或验证概率公式的正确性，但表述混乱。

总结

本笔记旨在从原始文档的混乱信息中，梳理出概率论与数理统计的核心概念，包括：

- **连续型分布**：正态分布、指数分布。
- **离散型分布**：伯努利分布、二项分布。
- **基础工具**：累积分布函数（CDF）、分位数。
- **关键性质**：正态分布的标准化、指数分布的无记忆性。

概率论与数理统计：期望与方差的计算与性质

一、离散型随机变量的期望计算

1. 二项分布 (Binomial Distribution)

- **定义**：若 $X \sim \text{Bin}(n,p)$，则其概率质量函数为： $P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \quad k=0,1,\dots,n$
- **期望计算**：

$$\begin{aligned}
E[X] &= \sum_{k=0}^{n} k \cdot P(X=k) \\
&= \sum_{k=0}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \cdot \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{当 } k=0 \text{ 时项为 } 0) \\
&= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{(n-1)-(k-1)} \\
&= np \sum_{j=0}^{n-1} \binom{n-1}{j} p^{j} (1-p)^{(n-1)-j} \quad (\text{令 } j=k-1) \\
&= np \cdot (p+(1-p))^{n-1} \\
&= np
\end{aligned}$$

【二项式定理】

- **结论**：

$$\boxed{X \sim \text{Bin}(n,p) \Longrightarrow E[X]=np}$$

2. 泊松分布 (Poisson Distribution)

- **定义**：若 $X \sim \text{Pois}(\lambda)$，则其概率质量函数为： $P(X=n)=\frac{e^{-\lambda}\lambda^n}{n!}, \quad n=0,1,2,\dots$
- **期望计算**：

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

$$
\begin{aligned}
E[X] &= \sum_{n=0}^{\infty} n \cdot P(X=n) \\
&= \sum_{n=1}^{\infty} n \cdot \frac{e^{-\lambda}\lambda^n}{n!} \\
&= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} \\
&= \lambda e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{(n-1)!} \\
&= \lambda e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} \quad (\text{令 } m = n-1) \\
&= \lambda e^{-\lambda} \cdot e^{\lambda} \\
&= \lambda
\end{aligned}
$$

【泰勒级数展开】

- **结论**:

$$\boxed{X \sim \mathrm{Pois}(\lambda) \Longrightarrow E[X] = \lambda}$$

二、连续型随机变量的期望计算

1. 帕累托分布 (Pareto Distribution)

- **定义**: 若 $X \sim \mathrm{Par}(\alpha)$，其概率密度函数（PDF）为: $f_X(x) = \frac{\alpha}{x^{\alpha+1}}$, $x \geq 1$, $\alpha > 0$
- **期望计算**:

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{+\infty} x\, f_X(x) dx = \int_{1}^{+\infty} x \cdot \frac{\alpha}{x^{\alpha+1}} dx \\
&= \alpha \int_{1}^{+\infty} x^{-\alpha}\, dx \\
&= \alpha \left[\frac{x^{-\alpha+1}}{-\alpha+1} \right]_1^{+\infty}
\end{aligned}
$$

- **结果分析**:
  - 当 $\alpha > 1$ 时，$-\alpha+1 < 0$，积分收敛:

$$E[X] = \frac{\alpha}{\alpha - 1}$$

  - 当 $0 < \alpha \leq 1$ 时，积分发散，期望不存在。
- **结论**:

$$\boxed{E[X] = \begin{cases} \dfrac{\alpha}{\alpha-1}, & \alpha > 1 \\ \text{D.N.E. (Does Not Exist)}, & 0 < \alpha \leq 1 \end{cases}}$$

2. 指数分布 (Exponential Distribution)

- **定义**: 若 $X \sim \text{Exp}(\lambda)$，其概率密度函数（PDF）为: $f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0$
- **期望计算**:【分部积分法】

$$E[X] = \int_0^{+\infty} x \cdot \lambda e^{-\lambda x} dx$$
$$= \lambda \int_0^{+\infty} x\, e^{-\lambda x} dx$$

使用分部积分法（令 $u = x, dv = e^{-\lambda x}dx$）:

$$\int_0^{+\infty} x\, e^{-\lambda x} dx = \left[-\frac{xe^{-\lambda x}}{\lambda}\right]_0^{+\infty} + \int_0^{+\infty} \frac{e^{-\lambda x}}{\lambda} dx$$
$$= 0 + \left[-\frac{e^{-\lambda x}}{\lambda^2}\right]_0^{+\infty}$$
$$= \frac{1}{\lambda^2}$$

因此:

$$E[X] = \lambda \cdot \frac{1}{\lambda^2} = \frac{1}{\lambda}$$

- **结论**:

$$\boxed{X \sim \text{Exp}(\lambda) \Longrightarrow E[X] = \frac{1}{\lambda}}$$

3. 正态分布 (Normal Distribution)

- **定义**: 若 $X \sim N(\mu, \sigma^2)$，其概率密度函数（PDF）为: $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- **期望计算**:

$$E[X] = \int_{-\infty}^{+\infty} x \cdot \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$
$$= \int_{-\infty}^{+\infty} (\mu + \sigma z) \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz \quad (\text{令 } z = \frac{x-\mu}{\sigma})$$
$$= \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \sigma \int_{-\infty}^{+\infty} z \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz$$
$$= \mu \cdot 1 + \sigma \cdot 0$$
$$= \mu$$

这是一个奇函数，

所以积分为 0

- **结论**:

$$\boxed{X \sim N(\mu, \sigma^2) \Longrightarrow E[X] = \mu}$$

三、期望的通用公式

对于任意随机变量 X 和函数 $g(X)$，其期望定义如下：

- **离散型**:

$$E[g(X)] = \sum_{i} g(x_i) \cdot P(X = x_i)$$

- **连续型**:

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x) dx$$

特别地，当 $g(X) = X^m$ 时，称为 X 的 m 阶矩（moment），其中 $m \geq 2$ 时是高阶矩。

四、方差 (Variance) 的定义与性质

1. 定义

- 方差衡量随机变量与其期望值的偏离程度。

$$\mathrm{Var}(X) = E[(X - E[X])^2]$$

- 一个常用且重要的等价公式：

$$\mathrm{Var}(X) = E[X^2] - (E[X])^2$$

2. 标准差 (Standard Deviation)

- 标准差是方差的平方根，具有与原变量相同的量纲。

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

3. 性质

设 $Y = aX + b$，其中 a, b 为常数。

- **期望的线性性质**:

$$E[Y] = E[aX + b] = aE[X] + b$$

- **方差的缩放性质**:

$$\mathrm{Var}(Y) = \mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X)$$

（注意：常数 b 不影响方差）

4. 标准化 (Standardization)

- 对于任意随机变量 X，定义其标准化变量 Z 为：

$$Z = \frac{X - \mu}{\sigma}$$

其中 $\mu = E[X], \sigma = \sqrt{\mathrm{Var}(X)}$。

- **标准化变量的性质**:

- $E[Z] = 0$
- $\text{Var}(Z) = 1$

五、正态分布的方差

- **定义**：若 $X \sim N(\mu, \sigma^2)$。
- **方差计算**:

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

（这是正态分布定义的一部分，也是其核心参数之一）

- **结论**:

$$\boxed{X \sim N(\mu, \sigma^2) \Longrightarrow \text{Var}(X) = \sigma^2}$$

📊 常见概率分布的期望与方差总结表

分布名称	参数	概率函数（PMF / PDF）	期望 $E[X]$	方差 $\text{Var}(X)$	存在条件
二项分布 $X \sim \text{Bin}(n,p)$	$n \in N,\ 0<p<1$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	总是存在
泊松分布 $X \sim \text{Pois}(\lambda)$	$\lambda > 0$	$P(X=k) = \frac{e^{-\lambda}\lambda^k}{k!}$	λ	λ	总是存在
指数分布 $X \sim \text{Exp}(\lambda)$	$\lambda > 0$	$f(x) = \lambda e^{-\lambda x},\ x \geq 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	总是存在
正态分布 $X \sim N(\mu,\sigma^2)$	$\mu \in R,\ \sigma^2 > 0$	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	总是存在
帕累托分布 $X \sim \text{Par}(\alpha)$	$\alpha > 0$	$f(x) = \frac{\alpha}{x^{\alpha+1}},\ x \geq 1$	$\begin{cases} \frac{\alpha}{\alpha-1}, & \alpha > 1 \\ \text{不存在}, & \alpha \leq 1 \end{cases}$	$\begin{cases} \frac{\alpha}{(\alpha-1)^2(\alpha-2)}, & \alpha > 2 \\ \text{不存在}, & \alpha \leq 2 \end{cases}$	期望：$\alpha > 1$ 方差：$\alpha > 2$

✅ 注：帕累托分布的方差推导略复杂，此处仅列出结果供参考。

🧮 期望与方差的核心公式

类型	**公式**	**说明**
期望定义（离散）	$E[g(X)] = \sum_i g(x_i) P(X = x_i)$	适用于离散型随机变量
期望定义（连续）	$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x)\, dx$	适用于连续型随机变量
方差定义	$\text{Var}\,(X) = E[(X - E[X])^2]$	衡量偏离程度
方差计算公式	$\text{Var}\,(X) = E[X^2] - (E[X])^2$	**最常用形式**
线性变换性质	若 $Y = aX + b$，则: • $E[Y] = aE[X] + b$ • $\text{Var}\,(Y) = a^2 \text{Var}\,(X)$	常数 b不影响方差
标准化变量	$Z = \frac{X - \mu}{\sigma}$	则 $E[Z] = 0$，$\text{Var}\,(Z) = 1$

🔑 **关键数学工具与恒等式**

工具	表达式	应用场景
阶乘简化	$$\frac{n}{n!} = \frac{1}{(n-1)!}$$	泊松/二项期望推导
变量代换	令 $m = n - 1$，则 $\sum_{n=1}^{\infty} \to \sum_{m=0}^{\infty}$	调整求和下标
指数函数泰勒展开	$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$	泊松分布期望化简的关键
分部积分法	$\int u\, dv = uv - \int v\, du$	指数分布期望计算

📘 混合分布（Mixed Distributions）
——概率论与统计学课堂笔记

---

1. 概念定义

混合分布（Mixed Distribution）是指一个随机变量 X的概率分布**同时包含离散部分与连续部分**。
典型特征：在某些点（如 $x=0$）具有**正的概率质量**（point mass），而在其他区间具有**概率密度函数**（PDF）。
💡 **应用场景**：保险理赔、零膨胀数据（zero-inflated data）、故障时间分析等。

---

2. 模型设定

设随机变量 X满足以下条件：

- **离散部分**：
$$\mathbb{P}(X=0)=p, 0<p<1$$
- **连续部分**（条件分布）：
$$给定\ X>0, X\sim g(t)$$
其中 $g(t)$是定义在 $t>0$上的概率密度函数，满足：
$$\int_0^\infty g(t)\,dt=1$$
- **整体结构**：
X以概率 p取值为 0；以概率 $1-p$从连续分布 g中抽样。

---

3. 累积分布函数（CDF）推导

目标：求 $F(x)=\mathbb{P}(X\le x)$的显式表达。

3.1 分段讨论

情况一：$x<0$
由于 $X\ge 0$几乎必然（a.s.），故：
$$F(x)=0, \forall x<0$$
情况二：$x\ge 0$
将事件 $\{X\le x\}$拆分为互斥子事件：
$$\{X\le x\}=\{X=0\}\cup\{0<X\le x\}$$
因此：
$$F(x)=\mathbb{P}(X=0)+\mathbb{P}(0<X\le x)$$
利用条件概率展开第二项：
$$\mathbb{P}(0<X\le x)=\mathbb{P}(X>0)\cdot\mathbb{P}(0<X\le x\mid X>0)=(1-p)\int_0^x g(t)\,dt$$
最终得到：
$$F(x)=p+(1-p)\int_0^x g(t)\,dt, x\ge 0$$

---

4. 完整 CDF 表达式

综合两种情况，X 的累积分布函数为：
$$F(x)=\begin{cases}0, & x<0\\ p+(1-p)\int_0^x g(t)\,dt, & x\ge 0\end{cases}$$
✅ **性质验证**：

- $F(0^-)=0$，$F(0)=p\rightarrow$ 在 $x=0$处存在**跳跃间断点**，跳跃高度为 p。

- $\lim_{x\to\infty} F(x) = p + (1-p)\cdot 1 = 1 \rightarrow$ 满足 CDF 极限性质。

---

5. 概率测度分解（进阶理解）
混合分布可视为**概率测度的凸组合**：

$$\mathbb{P}_X = p\cdot\delta_0 + (1-p)\cdot\mathbb{P}_g$$

其中：
- δ_0是在 0 处的**狄拉克测度**（Dirac measure），
- $\mathbb{P}_g$是由密度 g诱导的**绝对连续测度**。

这体现了 **Lebesgue 分解定理**在实际建模中的应用。

---

6. 示例：零膨胀指数分布（Zero-Inflated Exponential）
若取 $g(t) = \lambda e^{-\lambda t}$（$\lambda > 0$），则：
- CDF 为：

$$F(x) = \begin{cases} 0, & x < 0 \\ p + (1-p)(1-e^{-\lambda x}), & x \geq 0 \end{cases}$$

- PDF（广义）包含一个**狄拉克 delta 函数**项：

$$f(x) = p\cdot\delta(x) + (1-p)\lambda e^{-\lambda x}\cdot \mathbf{1}_{\{x>0\}}$$

⚠️ 注意：严格来说，混合分布**没有传统意义上的 PDF**（因含离散部分），但可在广义函数（分布理论）框架下定义。

---

7. 小结要点

特征	说明
离散成分	在 $x = 0$处有质量 p
连续成分	条件于 $X > 0$时服从密度 g
CDF 结构	分段函数，$x = 0$ 处跳跃
应用场景	零膨胀数据、保险、生存分析
数学本质	概率测度的离散+连续混合

📝 **学习提示**：理解混合分布的关键在于**区分无条件概率与条件概率**，并熟练运用**全概率公式**进行分解。

<u>*****值得注意的是，**</u>
<u>**因为混合分布既包含离散又有连续，因此必定存在跳跃点、不连续性，**</u>
<u>**所以说 mixed distribution 没有 PDF 因为他不连续，有跳跃点、**</u>
<u>**Mixed distribution 也没有 PMF，因为 PMF 给 discrete distribution 使用的，混合分布又有连续部分**</u>
<u>**混合分布只有 CDF，这是唯一存在的**</u>

📓 《概率论核心公式与矩生成函数应用》

第一部分：基础定义与性质（对应 PDF 第 1–2 页）

1.1 期望与方差的定义

- **期望（Mean）**：

$$\mathbb{E}[X] = \mu$$

- **方差（Variance）**：

$$\mathrm{Var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

- **标准差（Standard Deviation）**：

$$\sigma = \sqrt{\mathrm{Var}(X)}$$

1.2 线性变换的期望与方差

- **期望的线性性**（对任意常数 a, b）：

$$\mathbb{E}[aX + b] = a\,\mathbb{E}[X] + b$$

- **方差的缩放性**：

$$\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$$

  常数 b 不影响离散程度，故方差不变。

1.3 正态分布（Normal Distribution）

- **记法**：$X \sim \mathcal{N}(\mu, \sigma^2)$
- **性质**：

$$\mathbb{E}[X] = \mu, \quad \mathrm{Var}(X) = \sigma^2$$

第二部分：矩生成函数（MGF）基础（对应 PDF 第 3 页）

2.1 MGF 定义

- **矩生成函数（Moment Generating Function）**：

$$M_X(t) = \mathbb{E}[e^{tX}]$$

- **第 n 阶矩**:

$$\mathbb{E}[X^n] = M_X^{(n)}(0) = \frac{d^n}{dt^n} M_X(t)\Big|_{t=0}$$

✅ 特别地:

- 一阶导在 0 处: $M_{X'}(0) = \mathbb{E}[X]$
- 二阶导在 0 处: $M_{X''}(0) = \mathbb{E}[X^2]$

第三部分: 三大分布的 MGF 与方差推导（重点！对应 PDF 第 4–5 页）

下面将**逐一严格推导**。

一、泊松分布（Poisson Distribution）

1. 分布设定

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

2. MGF 推导

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \sum_{k=0}^{\infty} e^{tk} \cdot \frac{\lambda^k e^{-\lambda}}{k!} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} = e^{-\lambda} \cdot e^{\lambda e^t} = \boxed{e^{\lambda(e^t - 1)}} \end{aligned}$$

3. 一阶导 → 期望

$$\begin{aligned} M_{X'}(t) &= \frac{d}{dt} e^{\lambda(e^t-1)} = e^{\lambda(e^t-1)} \cdot \lambda e^t \\ \Rightarrow \mathbb{E}[X] &= M_{X'}(0) = e^{\lambda(1-1)} \cdot \lambda \cdot 1 = \lambda \end{aligned}$$

4. 二阶导 → $\mathbb{E}[X^2]$

$$\begin{aligned} M_{X''}(t) &= \frac{d}{dt}\left[\lambda e^t \cdot e^{\lambda(e^t-1)}\right] \\ &= \lambda e^t \cdot e^{\lambda(e^t-1)} + \lambda e^t \cdot \left(\lambda e^t \cdot e^{\lambda(e^t-1)}\right) \\ &= \lambda e^t e^{\lambda(e^t-1)}(1 + \lambda e^t) \end{aligned}$$

令 $t = 0$:

$$\mathbb{E}[X^2] = M_{X''}(0) = \lambda \cdot 1 \cdot 1 \cdot (1+\lambda) = \lambda(1+\lambda)$$

5. 方差计算

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \lambda(1+\lambda) - \lambda^2 = \lambda\end{aligned}$$

✅ **结论**：$\mathrm{Var}(X) = \lambda$

二、二项分布（Binomial Distribution）

1. 分布设定

- $X \sim \mathrm{Binomial}(n, p)$
- 表示 n 次独立伯努利试验中成功次数

2. MGF 推导

单次伯努利试验的 MGF 为：$1 - p + pe^t$
由于独立，n 次相加 $\rightarrow$ MGF 相乘：

$$M_X(t) = (1 - p + pe^t)^n$$

3. 一阶导 → 期望

$$\begin{aligned}&M_{X'}(t) = n(1-p+pe^t)^{n-1} \cdot pe^t \\ &\Rightarrow \mathbb{E}[X] = M_{X'}(0) = n(1-p+p)^{n-1} \cdot p = n \cdot 1 \cdot p = np\end{aligned}$$

4. 二阶导 → $\mathbb{E}[X^2]$

使用乘积法则：

$$\begin{aligned}M_{X''}(t) &= \frac{d}{dt}[npe^t(1-p+pe^t)^{n-1}] \\ &= npe^t(1-p+pe^t)^{n-1} + npe^t \cdot (n-1)(1-p+pe^t)^{n-2} \cdot pe^t \\ &= npe^t(1-p+pe^t)^{n-2}[(1-p+pe^t) + (n-1)pe^t]\end{aligned}$$

令 $t = 0$：

- $e^0 = 1$
- $1 - p + p = 1$

$$\begin{aligned}\mathbb{E}[X^2] &= np \cdot 1 \cdot 1^{n-2}[1 + (n-1)p] \\ &= np(1 + (n-1)p) = np + n(n-1)p^2\end{aligned}$$

5. 方差计算

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= np + n(n-1)p^2 - (np)^2 \\
&= np + n^2p^2 - np^2 - n^2p^2 \\
&= np - np^2 = np(1-p)
\end{aligned}$$

✅ **结论**：$\mathrm{Var}(X) = np(1-p)$

三、正态分布（Normal Distribution）

1. 分布设定

- $X \sim \mathcal{N}(\mu, \sigma^2)$

2. MGF（标准结果，可查表）

$$M_X(t) = \exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$

推导需用高斯积分，此处略（因原文未要求推导 MGF 本身，只用它求方差）

3. 一阶导 → 期望

$$M_{X'}(t) = (\mu + \sigma^2 t)\exp\left(\mu t + \frac{1}{2}\sigma^2 t^2\right)$$
$$\Rightarrow \mathbb{E}[X] = M_{X'}(0) = (\mu + 0) \cdot e^0 = \mu$$

4. 二阶导 → $\mathbb{E}[X^2]$

$$\begin{aligned}
M_{X''}(t) &= \frac{d}{dt}\left[(\mu + \sigma^2 t)e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right] \\
&= \sigma^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2} + (\mu + \sigma^2 t)^2 e^{\mu t + \frac{1}{2}\sigma^2 t^2}
\end{aligned}$$

令 $t = 0$：

$$\mathbb{E}[X^2] = \sigma^2 \cdot 1 + \mu^2 \cdot 1 = \mu^2 + \sigma^2$$

5. 方差计算

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = (\mu^2 + \sigma^2) - \mu^2 = \sigma^2$$

✅ **结论**：$\mathrm{Var}(X) = \sigma^2$

四、指数分布（Exponential Distribution）

1. 指数分布的定义

- 记作：$X \sim \mathrm{Exp}(\lambda)$，其中 $\lambda > 0$
- 概率密度函数（PDF）：

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

2. 矩生成函数（MGF）推导

$$\begin{aligned} M_X(t) = \mathbb{E}[e^{tX}] &= \int_0^{\infty} e^{tx} \cdot \lambda e^{-\lambda x}\, dx \\ &= \lambda \int_0^{\infty} e^{-(\lambda - t)x}\ dx \end{aligned}$$

该积分**仅在** $t < \lambda$ **时收敛**。此时：

$$\begin{aligned} M_X(t) &= \lambda \cdot \left[\frac{-1}{\lambda - t} e^{-(\lambda - t)x}\right]_0^{\infty} \\ &= \lambda \cdot \left(0 - \left(\frac{-1}{\lambda - t} \cdot 1\right)\right) \\ &= \frac{\lambda}{\lambda - t} \end{aligned}$$

✅ **MGF（定义域：$t < \lambda$）**：

$$\boxed{M_X(t) = \frac{\lambda}{\lambda - t}}$$

3. 一阶导数 → 期望 $\mathbb{E}[X]$

$$\begin{aligned} M_{X'}(t) &= \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) = \lambda \cdot \frac{d}{dt}(\lambda - t)^{-1} \\ &= \lambda \cdot (+1) \cdot (\lambda - t)^{-2} \\ &= \frac{\lambda}{(\lambda - t)^2} \end{aligned}$$

令 $t = 0$：

$$\mathbb{E}[X] = M_{X'}(0) = \frac{\lambda}{\lambda^2} = \boxed{\frac{1}{\lambda}}$$

4. 二阶导数 → $\mathbb{E}[X^2]$

对 $M_{X'}(t) = \lambda(\lambda - t)^{-2}$ 再求导：

$$
\begin{aligned}
M_{X''}(t) &= \frac{d}{dt}[\lambda(\lambda - t)^{-2}] \\
&= \lambda \cdot (-2) \cdot (\lambda - t)^{-3} \cdot (-1) \\
&= \frac{2\lambda}{(\lambda - t)^3}
\end{aligned}
$$

令 $t = 0$:

$$
\mathbb{E}[X^2] = M_{X''}(0) = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}
$$

5. 方差计算

$$
\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\
&= \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 \\
&= \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \boxed{\frac{1}{\lambda^2}}
\end{aligned}
$$

✅ **结论**：指数分布的方差为 $\mathrm{Var}(X) = \frac{1}{\lambda^2}$

补充：指数分布 MGF 推导要点说明

- **收敛条件**：MGF 仅在 $t < \lambda$ 时存在，这是指数分布的重要特性。
- **记忆性无关**：虽然指数分布有“无记忆性”，但 MGF 推导不依赖此性质。
- **与泊松关系**：若事件服从泊松过程（率 λ），则事件间隔时间服从 $\mathrm{Exp}(\lambda)$。

附：为什么用 MGF 求方差？

- **统一方法**：无论离散/连续，只要 MGF 存在，就可用求导得矩
- **避免积分/求和**：尤其对复杂分布，MGF 比直接算 $\mathbb{E}[X^2]$ 更简洁
- **考试重点**：教授常要求“用 MGF 证明方差”，考察对生成函数的理解

四大分布 MGF 与方差对照表

分布	MGF $M_X(t)$	$\mathbb{E}[X]$	$\mathbb{E}[X^2]$	$\mathrm{Var}(X)$
泊松 $\mathrm{Pois}(\lambda)$	$e^{\lambda(e^t-1)}$	λ	$\lambda+\lambda^2$	λ
二项 $\mathrm{Bin}(n,p)$	$(1-p+pe^t)^n$	np	$np+n(n-1)p^2$	$np(1-p)$
正态 $\mathcal{N}(\mu,\sigma^2)$	$e^{\mu t+\frac{1}{2}\sigma^2t^2}$	μ	$\mu^2+\sigma^2$	σ^2
指数 $\mathrm{Exp}(\lambda)$	$\frac{\lambda}{\lambda-t},\ t<\lambda$	$\frac{1}{\lambda}$	$\frac{2}{\lambda^2}$	$\frac{1}{\lambda^2}$

🟥 期望（Expectation）公式对比：两个不一样！！！！！

1. 随机变量本身的期望： $\mathbb{E}[X]$

类型	公式
离散型	$\mathbb{E}[X]=\sum_i x_i\,P(X=x_i)$
连续型	$\mathbb{E}[X]=\int_{-\infty}^{\infty} x\,f_X(x)\,dx$

---

2. 随机变量函数的期望： $\mathbb{E}[g(X)]$ （Law of the Unconscious Statistician）

类型	公式
离散型	$\mathbb{E}[g(X)]=\sum_i g(x_i)\,P(X=x_i)$
连续型	$\mathbb{E}[g(X)]=\int_{-\infty}^{\infty} g(x)\,f_X(x)\,dx$

✅ **关键说明：**

- 两者**权重相同**：都使用 X自身的概率分布（PMF 或 PDF）。
- **不是换元：$g(X)$ 是对 X的每个取值做变换，不改变概率权重。**
- 例如：$\mathbb{E}[X^2]$、$\mathbb{E}[e^{tX}]$（即 MGF）都属于 $\mathbb{E}[g(X)]$的特例。

概率论与数理统计专业笔记

主题：矩生成函数 (MGF) 与联合分布 (Joint Distribution)

第一部分：矩生成函数 (Moment Generating Function, MGF)

1. 基本概念与定义

- **矩生成函数 (MGF)** 是一个用于表征随机变量分布的重要工具。
- 对于连续型随机变量 X，其 MGF 定义为：$M_X(t) = E(e^{tX}) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\, dx$
- 对于离散型随机变量，积分替换为求和。
- **核心作用**：通过求导可以方便地得到各阶矩（期望、方差等）。
  - 一阶导数在 $t = 0$ 处的值即为期望：$E(X) = M_X'(0)$
  - 二阶导数在 $t = 0$ 处的值与方差相关：$\mathrm{Var}(X) = M_X''(0) - [M_X'(0)]^2$

2. 常见分布的矩生成函数

下表列出了几种重要概率分布的矩生成函数：

分布类型	记号	矩生成函数 $M_X(t)$
伯努利分布	$\mathrm{Ber}(p)$	$1 - p + pe^t$
几何分布	$\mathrm{Geom}(p)$	$\dfrac{pe^t}{1-(1-p)e^t}$
二项分布	$\mathrm{Bin}(n,p)$	$(1 - p + pe^t)^n$
泊松分布	$\mathrm{Pois}(\lambda)$	$e^{\lambda(e^t-1)}$
指数分布	$\mathrm{Exp}(\lambda)$	$\dfrac{\lambda}{\lambda - t}, t < \lambda$
标准正态分布	$N(0,1)$	$e^{\frac{t^2}{2}}$
一般正态分布	$N(\mu,\sigma^2)$	$e^{\mu t+\frac{\sigma^2 t^2}{2}}$
帕累托分布	$\mathrm{Par}(\alpha)$	D.N.E. (不存在)

注：帕累托分布的 MGF 不存在，因为其尾部过重，导致积分发散。

第二部分：联合分布 (Joint Distribution)

1. 引入与动机

- 当我们研究两个或多个随机变量时，需要描述它们的**联合行为**。
- 例如，研究 10,000 人的收入（Y）与教育水平（X）的关系。

变量定义示例

- **收入 Y**:$Y = \begin{cases} 1 & \text{收入} \leq \text{最低工资} \\ 2 & \text{收入介于 1 至 5 倍最低工资} \\ 3 & \text{收入} > 5 \times \text{最低工资} \end{cases}$
- **教育水平 X**:$X = \begin{cases} 1 & \text{小学} \\ 2 & \text{中学} \\ 3 & \text{职校/大专} \\ 4 & \text{大学} \end{cases}$

2. 联合概率表 (Joint Probability Table)

通过一个 4×3 的表格来组织数据，其中单元格数值代表对应组合的人数。

X \ Y	1 (低收入)	2 (中等收入)	3 (高收入)	**边际总数** N_X
1 (小学)	2820	1275	5	4100
2 (中学)	1620	3005	125	4800
3 (职校/大专)	50	200	270	600
4 (大学)	10	140	250	480
边际总数 N_Y	4000	5200	800	**10000**

- **边际总和 (Marginals)**: 行和与列和分别代表单个变量的边缘分布。

第三部分：联合概率质量函数 (PMF) 与概率密度函数 (PDF)

1. 离散型随机变量 (Discrete R.V.)

- **定义**: 设 X, Y 为离散随机变量，其联合概率质量函数 (Joint PMF) 定义为:

  $$P_{XY}(x, y) = P(X = x, Y = y)$$

  其中，$P_{XY}: \mathbb{Z} \times \mathbb{Z} \to [0,1]$。

- **边缘分布 (Marginal Masses)**:
  - X 的边缘概率: $P_X(x) = \sum_y P_{XY}(x, y)$
  - Y 的边缘概率: $P_Y(y) = \sum_x P_{XY}(x, y)$
- **归一化条件 (Normalization Condition)**:

  $$\sum_x \sum_y P_{XY}(x, y) = 1$$

2. 连续型随机变量 (Continuous R.V.)

- **定义**: 设 X, Y 为连续随机变量，其联合概率密度函数 (Joint PDF) 是一个函数 $f_{XY}(x, y): \mathbb{R} \times \mathbb{R} \to [0, +\infty)$，满足:

  $$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = \int_{a_1}^{b_1} \int_{a_2}^{b_2} f_{XY}(x, y)\, dy\, dx$$

- **归一化条件**:

  $$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{XY}(x, y)\, dy\, dx = 1$$

第四部分：联合累积分布函数 (Joint CDF)

1. 定义

- 联合累积分布函数 (Joint CDF) 定义为:

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

  其中，$F_{XY}: \mathbb{R} \times \mathbb{R} \to [0,1]$。

- **重要性质**: 已知联合 CDF F_{XY} 就足以确定联合 PMF 或 PDF。

2. 与 PMF/PDF 的关系

- **对于离散型变量**:

$$F_{XY}(a,b) = \sum_{x \le a} \sum_{y \le b} P_{XY}(x,y)$$

- **对于连续型变量**:

$$F_{XY}(a,b) = \int_{-\infty}^{a} \int_{-\infty}^{b} f_{XY}(x,y)\, dy\, dx$$

- **反向关系 (Derivatives)**:
  - 从 CDF 到 PDF: $f_{XY}(x,y) = \frac{\partial^2}{\partial x \partial y} F_{XY}(x,y)$

3. 边缘分布函数 (Marginal Distribution Function)

- X 的边缘分布函数定义为: $F_X(x) = P(X \le x) = P(X \le x, Y < \infty) = \lim_{y \to \infty} F_{XY}(x,y)$

4. n 维联合分布

- 对于 n 个随机变量 $X_1, X_2, \ldots, X_n$,其联合分布函数为: $F_{X_1 \ldots X_n}(x_1, \ldots, x_n) = P(X_1 \le x_1, X_2 \le x_2, \ldots, X_n \le x_n)$

5. 几何直观

- 对于二维连续型随机变量,联合 PDF $f_{XY}(x,y)$ 可以被想象成一个**三维曲面**,其下方的体积代表概率。

第五部分:补充说明与总结

1. 归一化条件总结

无论是离散还是连续情况,联合概率函数都必须满足归一化条件:

- **离散**:

$$\sum_{x} \sum_{y} P_{XY}(x,y) = 1$$

- **连续**:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\,dx = 1$$

2. 一般情况下的概率计算

对于任意区域 $C \subset \mathbb{R}^2$，事件 $(X,Y) \in C$ 的概率为:

- **离散**:

$$P((X,Y) \in C) = \sum_{(x,y)\in C} P_{XY}(x,y)$$

- **连续**:

$$P((X,Y) \in C) = \iint_C f_{XY}(x,y)\,dy\,dx$$

3. 边缘分布的正式定义

给定联合 PMF P_{XY} 或联合 PDF f_{XY}，边缘分布定义如下:

- **离散边缘分布**:

$$P_X(x) = \sum_y P_{XY}(x,y)$$

- **连续边缘分布**:

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\,dy$$

- **区间概率**:

$$P(X \in [a,b]) = \int_a^b f_X(x)\,dx = \int_a^b \left(\int_{-\infty}^{\infty} f_{XY}(x,y)\,dy\right) dx$$

第六部分：偏导数 (Partial Derivatives)

1. 核心概念

- 在处理多变量函数时，**偏导数**用于衡量函数在某一特定方向上的变化率。

- 计算偏导数时，将其他变量视为常数，仅对目标变量进行求导。

2. 示例计算

考虑一个二元函数：

$$f(x,y) = x^2y^2 + xy + y$$

- **对 x 的偏导数**（记作 $\frac{\partial f}{\partial x}$）：
  将 y 视为常数，对 x 求导。

$$\frac{\partial f}{\partial x} = 2xy^2 + y$$

- **对 y 的偏导数**（记作 $\frac{\partial f}{\partial y}$）：
  将 x 视为常数，对 y 求导。

$$\frac{\partial f}{\partial y} = 2x^2y + x + 1$$

> **注**：手写笔记中的“偏导”二字是对这一概念的中文标注。

第七部分：二重积分 (Double Integrals)

1. 基本定义与几何意义

- 二重积分用于计算定义在二维区域上的函数的“体积”或“总量”。
- 对于一个二元函数 $f(x,y)$，其在区域 D 上的二重积分为：$\iint_D f(x,y)\,dA$其中 $dA = dx\,dy$ 或 $dy\,dx$。

2. 积分区域与计算方法

笔记中展示了两种常见的矩形积分区域及其对应的积分形式：

情况一：先对 y 积分，再对 x 积分

- 积分区域 D_1 定义为：$D_1: a_1 \le x \le b_1, a_2 \le y \le b_2$
- 二重积分可写为累次积分：$\iint_{D_1} f(x,y)\,dx\,dy = \int_{a_1}^{b_1} \left(\int_{a_2}^{b_2} f(x,y)\,dy \right) dx$

情况二：先对 x 积分，再对 y 积分

- 积分区域 D_2 定义为：$D_2: a_2 \leq y \leq b_2, a_1 \leq x \leq b_1$
- 二重积分可写为累次积分：$\iint_{D_2} f(x,y)\,dy\,dx = \int_{a_2}^{b_2}\left(\int_{a_1}^{b_1} f(x,y)\,dx\right)dy$

  注：对于简单的矩形区域，积分顺序可以互换，结果相同（Fubini 定理）。但对于更复杂的区域，选择合适的积分顺序至关重要。

概率论与数理统计：联合分布、独立性与协方差

一． 随机变量的独立性 (Independent R.V.)

本页定义了随机变量独立性的概念，并给出了离散型变量的判定条件及一个反例。

1.1 独立性的定义

两个随机变量 X 和 Y 相互独立，当且仅当它们的联合概率（或概率密度）等于各自边缘概率（或概率密度）的乘积。

一般定义： 对于任意集合 A 和 B，

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

对于连续型变量：

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

对于离散型变量：

$$p_{X,Y}(x,y) = p_X(x) \cdot p_Y(y)$$

推广到多个变量： 对于 n 个随机变量 $X_1, X_2, \ldots, X_n$，若对所有区间 $[a_i, b_i]$ 都有：

$$P(a_1 \leq X_1 \leq b_1, \ldots, a_n \leq X_n \leq b_n) = \prod_{i=1}^{n} P\left(a_i \leq X_i \leq b_i\right)$$

则称 $X_1, \ldots, X_n$ 相互独立。

1.2 例题：收入与教育水平

通过一个二维表格来检验两个离散型随机变量是否独立。

$X \setminus Y$	1	2	3	p_X
1	0.282			0.4
2				
3				
4				
p_Y	0.4			

检验方法： 检查是否存在某个 (x,y) 使得联合概率不等于边缘概率之积。 取点 (1,1):

pX,Y(1,1)=0.282 pX(1)·pY(1)=0.4×0.4=0.16

因为 $0.282 \neq 0.16$，所以 X 和 Y **不独立**。

二． 联合分布的期望 (Joint Expectation)

本页介绍了如何计算联合分布下函数的期望，并阐述了其线性性质和降维性质。

2.1 期望的定义

对于函数 $g(X,Y)$，其期望值为：

离散型：

$$E[g(X,Y)] = \sum_x \sum_y g(x,y) \cdot p_{X,Y}(x,y)$$

连续型：

$$E[g(X,Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x,y) \cdot f_{X,Y}(x,y)\, dx\, dy$$

2.2 期望的性质

性质 1：线性 (Linearity)

与单变量期望类似，联合期望也满足线性性质：

$$E[\alpha g(X,Y) + \beta h(X,Y)] = \alpha E[g(X,Y)] + \beta E[h(X,Y)]$$

性质 2：降维至边缘分布 (Reduction to Marginals)

如果函数 g 只依赖于一个变量，例如 $g(X,Y) = g(X)$，则联合期望可以简化为边缘期望：

$$E[g(X)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x) \cdot f_{X,Y}(x,y)\, dx\, dy = \int_{-\infty}^{+\infty} g(x) \cdot f_X(x)\, dx$$

推导关键： 对 y 积分 $\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\, dy = f_X(x)$。

综合性质 2

对于形如 $E[\alpha g(X) + \beta h(Y) + \gamma]$ 的表达式，它服从线性规律：

$$E[\alpha g(X) + \beta h(Y) + \gamma] = \alpha E[g(X)] + \beta E[h(Y)] + \gamma$$

其中，$E[g(X)]$ 和 $E[h(Y)]$ 分别是关于边缘分布 $f_X(x)$ 和 $f_Y(y)$ 的期望。

三． 独立性与期望的关系

本页深入探讨了独立性对期望的影响，并推导了方差的加法公式。

3.1 独立性蕴含期望独立

定理: 若 X 和 Y 独立，则对于任意函数 g 和 h，有:

$$E[g(X)h(Y)] = E[g(X)] \cdot E[h(Y)]$$

证明: 利用独立性 $f_{X,Y}(x,y) = f_X(x)f_Y(y)$ 进行积分:

$$\begin{aligned} E[g(X)h(Y)] &= \int\int g(x)h(y)f_{X,Y}(x,y)\,dx\,dy \\ &= \int\int g(x)h(y)f_X(x)f_Y(y)\,dx\,dy \\ &= \left(\int g(x)f_X(x)\,dx\right)\left(\int h(y)f_Y(y)\,dy\right) \\ &= E[g(X)] \cdot E[h(Y)] \end{aligned}$$

注意: 逆命题不成立！即 $E[XY] = E[X]E[Y]$ 并不能推出 X 和 Y 独立。

3.2 方差的加法公式

设 X 和 Y 的期望分别为 μ_X 和 μ_Y，方差分别为 σ_X^2 和 σ_Y^2。

对于 $X+Y$ 的方差:

$$\begin{aligned} Var(X+Y) &= E[((X+Y) - E[X+Y])^2] \\ &= E[((X-\mu_X) + (Y-\mu_Y))^2] \\ &= E[(X-\mu_X)^2 + (Y-\mu_Y)^2 + 2(X-\mu_X)(Y-\mu_Y)] \\ &= Var(X) + Var(Y) + 2Cov(X,Y) \end{aligned}$$

因此，**只有当 X 和 Y 不相关 ($Cov(X,Y) = 0$) 时，方差才具有可加性**:

$$Var(X+Y) = Var(X) + Var(Y)$$

四． 协方差与相关性

本页定义了协方差，并澄清了“独立”与“不相关”的区别。

4.1 协方差的定义

协方差衡量两个随机变量的线性相关程度:

$$Cov(X,Y) = E[(X-\mu_X)(Y-\mu_Y)]$$

等价形式: 利用期望的线性性质展开:

$$\begin{aligned} Cov(X,Y) &= E[XY - X\mu_Y - Y\mu_X + \mu_X\mu_Y] \\ &= E[XY] - \mu_Y E[X] - \mu_X E[Y] + \mu_X\mu_Y \\ &= E[XY] - \mu_X\mu_Y \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

4.2 “独立” vs “不相关”

- **定义:** X 和 Y 是**不相关 (uncorrelated)** 的，当且仅当 $Cov(X,Y)=0$。
- **关系:**
  - 如果 X 和 Y **独立**，那么它们一定**不相关**。 $X \perp Y \Longrightarrow Cov(X,Y)=0$
  - 但是，**不相关不一定意味着独立**。 $Cov(X,Y)=0 \not\Longrightarrow X \perp Y$

这是一个非常重要的概念区分!

4.3 特殊情况: $X=Y$

当 $X=Y$ 时，协方差即为方差:

$$Cov(X,X) = E[X^2] - E[X]^2 = Var(X)$$

五． 联合概率密度函数的积分计算

本页展示了一个具体的二元连续型随机变量的概率密度函数及其积分计算过程。

5.1 概率密度函数

给定联合概率密度函数:

$$f_{X,Y}(x,y) = \frac{1}{2}(x^2y + xy^2), \quad \text{for } 0 \le x \le 1, 0 \le y \le 1$$

5.2 积分区域与计算

需要计算在区域 $0 \leq x \leq 1$, $0 \leq y \leq 1$ 上的积分，以验证其为合法的概率密度函数（积分结果应为 1）。

积分顺序：先对 y 积分，再对 x 积分。

$$\begin{aligned}
\int_0^1 \int_0^1 f_{X,Y}(x,y)\,dy\,dx &= \int_0^1 \int_0^1 \frac{1}{2}(x^2y + xy^2)\,dy\,dx \\
&= \int_0^1 \frac{1}{2}\left[\frac{x^2y^2}{2} + \frac{xy^3}{3}\right]_{y=0}^{y=1} dx \\
&= \int_0^1 \frac{1}{2}\left(\frac{x^2}{2} + \frac{x}{3}\right) dx \\
&= \frac{1}{2}\left[\frac{x^3}{6} + \frac{x^2}{6}\right]_0^1 \\
&= \frac{1}{2}\left(\frac{1}{6} + \frac{1}{6}\right) \\
&= \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}
\end{aligned}$$

发现错误： 计算结果为 $\frac{1}{6}$，而非 1。这表明该函数在给定区域上**不是一个合法的概率密度函数**，因为它未被归一化。可能原始笔记中存在笔误，或者需要一个归一化常数 c 使得 $c \cdot f_{X,Y}(x,y)$ 才是正确的 PDF。

联合期望、协方差与相关系数

1. 联合期望 (Joint Expectation)

对于随机变量 $X_1, X_2, \ldots, X_n$，函数 $g(X_1, \ldots, X_n)$ 的期望值计算如下：

- **离散型 (Discrete):**

$$E[g(X_1, \ldots, X_n)] = \sum_{x_1} \ldots \sum_{x_n} g(x_1, \ldots, x_n) \cdot P_{X_1,\ldots,X_n}(x_1, \ldots, x_n)$$

- **连续型 (Continuous):**

$$E[g(X_1, \ldots, X_n)] = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} g(x_1, \ldots, x_n) \cdot f_{X_1,\ldots,X_n}(x_1, \ldots, x_n)\, dx_1, \ldots, dx_n$$

---

2. 随机变量的独立性 (Independence of R.V.)

随机变量 $X_1, \ldots, X_n$ 相互独立，当且仅当以下任一条件成立：

- **一般定义 (General Definition):**
  对于任意实数 a_i, b_i，事件 $\{a_i \le X_i \le b_i\}$ 与 $\{a_j \le X_j \le b_j\}$ 相互独立。

- **分布函数 (Distribution Function):**

$$F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = F_{X_1}(x_1) \cdot \ldots \cdot F_{X_n}(x_n)$$

- **期望性质 (Expectation Property):**
  对于任意函数 $F_1, \ldots, F_n$，

$$E[F_1(X_1) \cdot \ldots \cdot F_n(X_n)] = E[F_1(X_1)] \cdot \ldots \cdot E[F_n(X_n)]$$

- **联合分布 (Joint Distribution):**
  若 $X_1, \ldots, X_n$ 联合分布为离散或连续，则其联合概率质量函数 (PMF) 或概率密度函数 (PDF) 可分解为边缘分布的乘积：

$$P_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P_{X_1}(x_1) \cdot \ldots \cdot P_{X_n}(x_n)$$

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdot \ldots \cdot f_{X_n}(x_n)$$

---

3. 协方差 (Covariance)

协方差衡量两个随机变量 X 和 Y 之间的线性相关性。

- **定义 (Definition):**

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

  其中 $\mu_X = E[X], \mu_Y = E[Y]$。

- **解释 (Interpretation):**
  - $\mathrm{Cov}(X,Y) > 0$: X 和 Y 同向变动，**正相关**。
  - $\mathrm{Cov}(X,Y) < 0$: X 和 Y 反向变动，**负相关**。
  - $\mathrm{Cov}(X,Y) = 0$: X 和 Y **不相关**（但不一定独立）。
    - **重要区别**: 独立 $\Rightarrow$ 不相关，但不相关 $\nRightarrow$ 独立。

---

4. 协方差的性质 (Properties of Covariance)

- **P1: 方差的加法公式 (Variance of Sum)**

  $$\mathrm{Var}(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j)$$

  如果 $X_1, \ldots, X_n$ 两两不相关（例如独立），则:

  $$\mathrm{Var}\left(\sum X_i\right) = \sum \mathrm{Var}(X_i)$$

- **P2: 特殊情况 (Special Case)**

  $$\mathrm{Cov}(X,X) = \mathrm{Var}(X)$$

- **P3: 线性变换 (Linear Transformation)**
  设 $U = a_1 X + b_1$，$V = a_2 Y + b_2$，则:

  $$\mathrm{Cov}(U,V) = a_1 a_2 \cdot \mathrm{Cov}(X,Y)$$

  此性质表明协方差具有**线性齐次性**。

---

5. 示例：二项分布 (Example: Binomial Distribution)

设 $X \sim \text{Bin}(n,p)$，且 $X = X_1 + \ldots + X_n$，其中 $X_i \sim \text{Bernoulli}(p)$ 且相互独立。

- 期望：$E[X_i] = p$，$E[X] = np$
- 方差：$\text{Var}(X_i) = p(1-p)$，$\text{Var}(X) = np(1-p)$

相关系数 (Correlation Coefficient)

相关系数是标准化的协方差，用于衡量两个变量间线性相关的强度和方向。

1. 定义 (Definition)

$$\rho_{X,Y} = \begin{cases} 0, & \text{if } \text{Var}(X) = 0 \text{ or } \text{Var}(Y) = 0 \\ \dfrac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)} \cdot \sqrt{\text{Var}(Y)}}, & \text{otherwise} \end{cases}$$

2. 性质 (Properties)

- **P1：极端情况 (Extreme Cases)**
  - 若 $X = Y$，则 $\rho_{X,Y} = 1$ （完全正相关）。
  - 若 $X = -Y$，则 $\rho_{X,Y} = -1$ （完全负相关）。
- **P2：取值范围 (Range)**

$$-1 \leq \rho_{X,Y} \leq 1$$

- **P3：标准化变量 (Standardized Variables)**
  令 $Z_1 = \frac{X-\mu_X}{\sigma_X}$，$Z_2 = \frac{Y-\mu_Y}{\sigma_Y}$，则：

$$\text{Cov}(Z_1, Z_2) = \rho_{X,Y}, E[Z_1] = E[Z_2] = 0, \text{Var}(Z_1) = \text{Var}(Z_2) = 1$$

- **P4：证明 (Proof)**
  当 $X = \pm Y$ 时，$\rho_{X,Y} = \pm 1$，即完全相关。

---

3. 相关系数的推导 (Derivation)

利用方差公式:

$$\operatorname{Var}(X \pm Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) \pm 2\operatorname{Cov}(X, Y)$$

考虑标准化变量的和/差:

$$0 \leq \operatorname{Var}\left(\frac{X}{\sqrt{\operatorname{Var}(X)}} \pm \frac{Y}{\sqrt{\operatorname{Var}(Y)}}\right) = 1 + 1 \pm 2\rho_{X,Y}$$

由此可得:

- $1 + \rho_{X,Y} \geq 0 \Rightarrow \rho_{X,Y} \geq -1$
- $1 - \rho_{X,Y} \geq 0 \Rightarrow \rho_{X,Y} \leq 1$

因此，$-1 \leq \rho_{X,Y} \leq 1$。

大数定律 (Law of Large Numbers)

1. 观察 (Observation)

重复抽样会导致样本均值趋于稳定，趋近于总体均值 μ。

2. 数学形式 (Mathematical Formulation)

设 $X_1, X_2, \dots, X_n$ 是独立同分布 (i.i.d.) 的随机变量，$E[X_i] = \mu_X$，$\operatorname{Var}(X_i) = \sigma_X^2$。

样本均值 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ 的期望和方差为:

- **期望:**

$$E[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} E[X_i] = \frac{n\mu_X}{n} = \mu_X$$

- **方差:**

$$\operatorname{Var}(\bar{X}_n) = \frac{1}{n^2}\sum_{i=1}^{n} \operatorname{Var}(X_i) = \frac{n\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n}$$

随着 $n \to \infty$，$\operatorname{Var}(\bar{X}_n) \to 0$，说明样本均值会越来越集中在总体均值 μ_X 附近。

---

3. 切比雪夫定理（Chebyshev's Theorem）

对于任意随机变量 X，其期望为 μ，方差为 σ^2，则对于任意 $a > 0$：

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

该定理提供了偏离均值的概率上界，是大数定律的理论基础之一。

---

4. 弱大数定律（Weak Law of Large Numbers, WLLN）

定理（Theorem）:
设 $X_1, X_2, \ldots, X_n$ 是独立同分布（i.i.d.）的随机变量，其期望为 μ，方差为 σ^2。则对于任意 $\varepsilon > 0$，有：

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

这被称为**弱大数定律**，它表明当样本量 n 趋于无穷大时，样本均值 $\bar{X}_n$ 收敛于总体均值 μ 的概率趋于 1。

证明思路:
该定理可由**切比雪夫不等式（Chebyshev's Inequality）**直接推导得出。
根据切比雪夫不等式：

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2/n}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

当 $n \to \infty$ 时，右侧趋近于 0，因此左侧也趋近于 0。

---

5. 应用示例：硬币公平性检验（Example: Is the Coin Fair?）

假设我们有一个伯努利试验（如抛硬币），其中 $X_i \sim \mathrm{Bernoulli}(p)$，$p$ 为正面朝上的真实概率。我们的目标是确定需要多少次试验 n，才能以 99.9% 的置信度保证样本均值 $\bar{X}_n$ 与真实概率 p 的偏差在 10^{-3} 以内。

即求解：

$$P(|\bar{X}_n - p| < 10^{-3}) \geq 0.999$$

步骤:

1. **应用切比雪夫不等式:**
   对于任意 $\varepsilon > 0$,

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$

   因此,

$$P(|\bar{X}_n - p| < \varepsilon) \geq 1 - \frac{p(1-p)}{n\varepsilon^2}$$

2. **设定参数:**
   - $\varepsilon = 10^{-3}$
   - 置信度要求: $1 - \frac{p(1-p)}{n\varepsilon^2} \geq 0.999$
3. **求解 n:**

$$\frac{p(1-p)}{n \cdot (10^{-3})^2} \leq 0.001$$

$$\Rightarrow n \geq \frac{p(1-p)}{0.001 \cdot 10^{-6}} = \frac{p(1-p)}{10^{-9}}$$

4. **处理未知参数 p:**
   由于 p 未知,我们需要找到 $p(1-p)$ 的最大值来确保结果对所有可能的 p 都成立。
   函数 $f(p) = p(1-p)$ 在 $p = \frac{1}{2}$ 时取得最大值 $\frac{1}{4}$。
   因此,我们取最坏情况 $p(1-p) \leq \frac{1}{4}$。
5. **计算最终 n:**

$$n \geq \frac{1/4}{10^{-9}} = \frac{10^9}{4} = 250{,}000{,}000$$

结论:
为了以 99.9% 的概率保证样本均值 $\bar{X}_n$ 与真实概率 p 的偏差小于 10^{-3},至少需要进行 **2.5 亿** 次试验。

---

6. 强大数定律 (Strong Law of Large Numbers, SLLN)

定义 (Definition):
强大数定律指出，在相同条件下（X_i i.i.d., $E[X_i] = \mu$），样本均值 $\bar{X}_n$ 几乎必然 (almost surely) 收敛于 μ：

$$P\left(\lim_{n\to\infty} \bar{X}_n = \mu\right) = 1$$

这比弱大数定律更强，因为它断言收敛是“几乎必然”发生的，而不仅仅是概率趋于 1。

总结

本笔记系统梳理了联合期望、协方差、相关系数以及大数定律的核心概念、公式和性质，深入探讨了大数定律的两种形式——弱大数定律和强大数定律，并通过一个实际的“硬币公平性检验”例子，展示了如何利用切比雪夫不等式来计算所需的样本量。这不仅巩固了理论知识，也提供了实用的分析工具,通过结构化的整理，可以清晰地理解各概念间的联系与区别，为后续学习统计推断和机器学习打下坚实基础。

概率论与数理统计核心概念笔记

整理日期： 2025 年 11 月 10 日

第一部分：随机变量的独立性 (Independence)

本节探讨两个或多个随机变量之间相互独立的定义、性质及判定条件。

1.1 二维随机变量 (X, Y) 的独立性

随机变量 X 和 Y 相互独立，当且仅当满足以下任一条件：

(i) 分布函数的乘积形式

对于任意实数 a_1, b_1, a_2, b_2，有：

$$P(a_1 \le X \le b_1, a_2 \le Y \le b_2) = P(a_1 \le X \le b_1) \cdot P(a_2 \le Y \le b_2)$$

此为独立性的基本定义，适用于所有类型的随机变量。

(ii) 期望的乘积性质

对于任意函数 $F(x)$ 和 $G(y)$，有：

$$E[F(X)G(Y)] = E[F(X)] \cdot E[G(Y)]$$

注：此条件是独立性的充要条件，比(i)更一般化。

(iii) 联合分布律/密度函数的乘积形式

- **离散型随机变量:**
  联合概率质量函数等于各自边缘概率质量函数的乘积。

  $$p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$$

- **连续型随机变量:**
  联合概率密度函数等于各自边缘概率密度函数的乘积。

  $$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$$

结论： 独立性意味着联合分布可以完全由边缘分布决定。

1.2 多维随机变量的独立性

上述定义可推广至 n 个随机变量 $X_1, X_2, \ldots, X_n$。

$X_1, X_2, \ldots, X_n$ 相互独立，当且仅当满足以下任一条件：

- 对于任意区间 $\{a_i \leq X_i \leq b_i\}$，事件相互独立。
- 对于任意函数 $F_1, F_2, \ldots, F_n$，有：$E[F_1(X_1) \cdots F_n(X_n)] = E[F_1(X_1)] \cdots E[F_n(X_n)]$
- 联合分布函数/密度函数等于各边缘分布函数/密度函数的乘积：$p_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n) f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$

第二部分：协方差与相关性 (Covariance & Correlation)

2.1 协方差 (Covariance)

协方差衡量两个随机变量线性相关的程度。

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

其中，$\mu_X = E[X]$，$\mu_Y = E[Y]$。

2.2 不相关性 (Uncorrelatedness)

若 $\mathrm{Cov}(X, Y) = 0$，则称 X 和 Y **不相关**。

2.3 独立性与不相关性的关系

这是一个非常重要的概念辨析：

关系	性质
独立 ⇒ 不相关	**充分条件**。如果 X 和 Y 独立，则它们一定不相关。
不相关 ⇏ 独立	**非必要条件**。不相关不能推出独立。

> **总结：** “独立”是比“不相关”更强的条件。两者之间的关系可以用一个简单的图示表示：
>
> `独立` → `不相关`
>
> 但反过来不成立。

2.4 协方差的性质

- 自协方差即方差：$\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
- 线性变换：$\mathrm{Cov}(a_1 X + b_1, a_2 Y + b_2) = a_1 a_2 \cdot \mathrm{Cov}(X, Y)$

第三部分：相关系数 (Correlation Coefficient)

相关系数是对协方差进行标准化后的度量，用于衡量线性相关性的强度和方向。

3.1 定义

相关系数 $\rho_{X,Y}$ 定义为：

$$\rho_{X,Y} = \begin{cases} 0, & \text{if } \mathrm{Var}(X) = 0 \text{ or } \mathrm{Var}(Y) = 0 \\ \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)} \cdot \sqrt{\mathrm{Var}(Y)}}, & \text{otherwise} \end{cases}$$

3.2 性质

- **取值范围：** $-1 \leq \rho_{X,Y} \leq 1$
- **完全正相关：** $\rho_{X,Y} = 1$ 当且仅当 $Y = aX + b$ 且 $a > 0$。
- **完全负相关：** $\rho_{X,Y} = -1$ 当且仅当 $Y = aX + b$ 且 $a < 0$。
- **不相关：** $\rho_{X,Y} = 0$。

3.3 标准化变量

令 $Z_1 = \frac{X-\mu_X}{\sigma_X}$, $Z_2 = \frac{Y-\mu_Y}{\sigma_Y}$，则：

$$\rho_{X,Y} = E[Z_1 Z_2]$$

第四部分：大数定律 (Law of Large Numbers, LLN)

大数定律描述了随着样本量增大，样本均值趋于总体均值的现象。

4.1 基本设定

设 $X_1, X_2, \ldots, X_n$ 是独立同分布（i.i.d.）的随机变量，且：

$$E[X_i] = \mu, \mathrm{Var}(X_i) = \sigma^2$$

样本均值定义为：

$$\bar{X}_n = \frac{X_1 + X_2 + \ldots + X_n}{n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

其期望和方差为:

$$E[\bar{X}_n] = \mu$$

$$\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n}$$

> **关键点:** 随着 $n \to \infty$，样本均值的方差趋近于 0，这意味着样本均值会越来越集中在总体均值 μ 附近。

4.2 切比雪夫大数定律 (Chebyshev's LLN)

切比雪夫不等式是证明弱大数定律的基础。

切比雪夫不等式 (Chebyshev's Inequality):

对于任意随机变量 X，若 $E[X] = \mu$, $\mathrm{Var}(X) = \sigma^2$，则对任意 $\varepsilon > 0$，有:

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

弱大数定律 (Weak Law of Large Numbers):

将切比雪夫不等式应用于样本均值 $\bar{X}_n$，得到:

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

这表明，当 n 足够大时，样本均值 $\bar{X}_n$ 与总体均值 μ 的偏差超过 ε 的概率可以变得任意小。

第五部分：大数定律的应用与拓展

5.1 应用：估计事件概率

大数定律的一个重要应用是用频率估计概率。

问题:

设随机变量 X，我们想求 $P(X \in C)$，其中 C 是某个事件。

解法:

引入指示函数（Indicator Function）:

$$Y = \mathbb{I}_{\{X \in C\}} = \begin{cases} 1, & \text{if } X \in C \\ 0, & \text{else} \end{cases}$$

此时，$Y \sim \text{Bernoulli}(p)$，其中 $p = P(X \in C)$。

根据大数定律，样本均值 $\bar{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ 会收敛到 p。

切比雪夫不等式的应用：

$$P(|\bar{Y}_n - p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

为了保证误差不超过 ε 的概率至少为 $1-\delta$，我们需要：

$$\frac{1}{4n\varepsilon^2} \leq \delta \Rightarrow n \geq \frac{1}{4\delta\varepsilon^2}$$

5.2 实例：掷骰子

考虑掷一个公平的六面骰子 n 次。

- 设 X_k 为第 k 次掷出的点数。
- $P(X_k = i) = p_i = \frac{1}{6}$, for $i = 1,2,\ldots,6$.

我们关心的是，当 n 很大时，每个点数出现的频率 $y_n^{(i)} = \frac{\#\{k:X_k=i\}}{n}$ 会如何表现。

根据大数定律：

$$P(|y_n^{(i)} - p_i| \geq \varepsilon) \leq \frac{p_i(1-p_i)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2}$$

例如，若要求误差 $\varepsilon = 0.01$，置信度 $1-\delta = 0.95$ ($\delta = 0.05$)，则：

$$n \geq \frac{1}{4 \times 0.05 \times (0.01)^2} = 50{,}000$$

这意味着，需要掷骰子至少 5 万次，才能有 95%的把握保证每个点数的频率与理论概率 $\frac{1}{6}$ 的误差小于 0.01。

第六部分：实验数据展示

6.1 地铁排队实验

在一次课堂演示中，记录了 100 条地铁排队线中女性人数的数据。

- **随机变量** X**：**一条地铁排队线中的女性人数。
- **数据汇总：**

女性人数 X	出现次数	频率（比例)
0	1	0.01
1	3	0.03
2	4	0.04
3	23	0.23
4	25	0.25
5	19	0.19
6	17	0.17
7	5	0.05
8	2	0.02
9	0	0.00
10	1	0.01

注：表格中“0.09”应为“0.04”，可能是笔误。总和为 100，符合预期。

该数据可用于估算女性人数的概率分布，并验证大数定律。

切比雪夫不等式 (Chebyshev's Inequality)

弱大数定律 (Weak Law of Large Numbers)

一、切比雪夫不等式 (Chebyshev's Inequality)

1. 一句话总结：

> **"无论数据分布长什么样，只要知道它的均值和方差，就能保证绝大多数数据点不会离均值太远。"**

它是一个**普适性的"安全网"**，不依赖于任何具体的分布形状（正态、均匀、指数...都可以），只用到最基础的两个参数：期望（均值）μ 和方差 σ^2。

---

2. 数学公式与解读

公式：
对于任意随机变量 X，若其期望 $E[X] = \mu$，方差 $\mathrm{Var}(X) = \sigma^2$，则对任意正数 $\varepsilon > 0$，有：

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

解读：

- 左边 $P(|X - \mu| \geq \varepsilon)$ 是一个**概率**，表示"随机变量 X 的取值偏离其平均值 μ 超过 ε 的可能性"。
- 右边 $\frac{\sigma^2}{\varepsilon^2}$ 是一个**上界**，它告诉我们这个"偏离事件"的概率**不可能超过**这个值。

核心思想：

- 方差 σ^2 越小，数据越集中，那么"偏离"的概率就越小。
- 你要求的精度 ε 越大（允许的偏差越大），那么"偏离"的概率上限就越小。

---

3. 直观理解：用"钟形图"想象

想象一个钟形曲线（即使不是正态分布，也可以这么想）：

- 曲线的中心是 μ。
- 曲线的"胖瘦"由 σ 决定：σ 小，曲线瘦高；σ 大，曲线矮胖。

切比雪夫不等式说："不管你的钟形图有多怪异，只要你告诉我它的中心（μ）和胖瘦程度（σ^2），我就能告诉你，在距离中心 ε 以外的地方，最多只有 $\frac{\sigma^2}{\varepsilon^2}$ 的面积。"

举个栗子：
假设一个班级学生的身高平均值 $\mu = 170cm$，方差 $\sigma^2 = 25$（即标准差 $\sigma = 5cm$）。
你想知道身高偏离平均值超过 $10cm$（即 $\varepsilon = 10$）的学生比例是多少？

根据切比雪夫不等式：

$$P(|X - 170| \geq 10) \leq \frac{25}{10^2} = \frac{25}{100} = 0.25$$

这意味着，**身高在 160cm 以下或 180cm 以上的学生，占比最多为 25%**。这是一个非常保守但绝对可靠的估计！

> **为什么说它“保守”？**
> 如果你知道数据是正态分布，那么 $P(|X - \mu| \geq 2\sigma) \approx 5\%$，而切比雪夫给出的是 $\leq 25\%$。它牺牲了精确性，换来了普适性。

---

4. 考试要点 & 应用场景

- **必考题型：** 给出一个随机变量的均值和方差，求其取值落在某个区间外的概率的上界。
- **应用方向：**
  - 在不知道分布的情况下，提供一个“最坏情况”的风险评估。
  - 作为证明其他更强大定理（如大数定律、中心极限定理）的基础工具。
- **记忆口诀：** “方差除以误差平方，就是概率的天花板”。

二、弱大数定律 (Weak Law of Large Numbers, WLLN)

1. 一句话总结：

> **“当样本量足够大时，样本平均值会稳定地趋近于总体平均值。”**

这是统计推断的基石！它解释了为什么我们可以通过抽样调查来推断整体情况。

---

2. 数学公式与解读

公式：
设 $X_1, X_2, \ldots, X_n$ 是独立同分布（i.i.d.）的随机变量，且 $E[X_i] = \mu$, $\mathrm{Var}(X_i) = \sigma^2$。
定义样本均值 $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$。
则对于任意 $\varepsilon > 0$，有：

$$\lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

或者，用切比雪夫不等式给出一个更实用的版本：

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

解读：

- 左边是“样本均值 $\bar{X}_n$ 偏离总体均值 μ 超过 ε 的概率”。
- 随着样本量 n 增大，右边的上界 $\frac{\sigma^2}{n\varepsilon^2}$ 会越来越小，最终趋近于 0。
- 这意味着，**样本均值 $\bar{X}_n$ 会以越来越高的概率，无限接近于总体均值 μ**。

---

3. 直观理解：用“掷骰子”和“民意调查”想象

情景一：掷骰子

- 你掷一个公平的六面骰子，理论上每个点数出现的概率是 $1/6$。
- 掷 10 次，可能点数 6 出现了 3 次，频率是 0.3，离理论值 0.166...很远。
- 掷 1000 次，点数 6 出现的次数大约是 166 次，频率是 0.166，非常接近理论值。
- **这就是大数定律在起作用！**

情景二：民意调查

- 你想知道全国民众对某政策的支持率（总体均值 μ）。
- 你不可能问所有人，所以你随机抽取 1000 人（样本）。
- 你发现这 1000 人中有 600 人支持，样本均值 $\bar{X}_n = 0.6$。
- 根据大数定律，只要你抽样是随机的、样本量足够大，这个 0.6 就非常接近真实的全国支持率 μ。

---

4. 与切比雪夫不等式的联系

弱大数定律的证明，正是基于切比雪夫不等式！

1. 我们知道样本均值 $\bar{X}_n$ 的期望是 μ，方差是 $\frac{\sigma^2}{n}$。
2. 对 $\bar{X}_n$ 应用切比雪夫不等式：$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\mathrm{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$
3. 当 $n \to \infty$ 时，右边 $\to 0$，因此左边也 $\to 0$。

所以，切比雪夫不等式是弱大数定律的“数学引擎”。

---

5. 考试要点 & 应用场景

- **必考题型：**

    - 证明弱大数定律（通常要求用切比雪夫不等式）。
    - 计算为了达到某个精度 ε 和置信度（如 95%），需要多大的样本量 n。
        - 例如：要保证样本均值与总体均值的误差小于 0.1 的概率大于 95%，求最小样本量。
        - 解法：令 $\frac{\sigma^2}{n\varepsilon^2} \le 0.05$，解出 n。
- **应用方向：**
    - 抽样调查的理论基础。
    - 蒙特卡洛模拟（Monte Carlo Simulation）的收敛性保证。
    - 机器学习中，用经验风险最小化（ERM）逼近真实风险的理论依据。
- **记忆口诀：**“样本越大，平均越准；方差除以样本量，误差概率就变小”。

三、总结：两者的区别与联系

特征	切比雪夫不等式	弱大数定律
对象	单个随机变量 X	样本均值 $\bar{X}_n$
目的	提供单个值偏离均值的概率上界	说明样本均值会收敛到总体均值
依赖	只需知道均值和方差	需要独立同分布的样本
关系	是证明弱大数定律的关键工具	是切比雪夫不等式在样本均值上的直接应用

四、终极理解：它们在现实世界中的意义

- **切比雪夫不等式** 告诉你：“别担心，即使我不知道你的数据长什么样，我也能给你一个‘安全范围’。” 它是风险管理的“兜底”工具。
- **弱大数定律** 告诉你：“别怕，只要你做的实验够多、抽样的人数够多，你的结果就一定靠谱。” 它是科学实验和统计推断的“信心来源”。

简单来说：

- 切比雪夫不等式是“**保守的保险**”。
- 弱大数定律是“**可靠的承诺**”。

掌握了这两个概念，你就真正理解了概率论如何从“不确定性”中提炼出“确定性”，这也是所有现代统计学和数据科学的根基所在。

📘 概率统计讲义笔记（基于 25-11-10.pdf）

一、随机变量的独立性（Independence of Random Variables）

定义

随机变量 X 与 Y **相互独立**，当且仅当对任意实数 a_1, b_1, a_2, b_2，有：

$$P(a_1 \leq X \leq b_1, a_2 \leq Y \leq b_2) = P(a_1 \leq X \leq b_1) \cdot P(a_2 \leq Y \leq b_2)$$

等价条件（函数期望分解）

若对任意可测函数 f, g，满足：

$$\mathbb{E}[f(X)g(Y)] = \mathbb{E}[f(X)] \cdot \mathbb{E}[g(Y)]$$

则 X 与 Y 独立。

联合分布与密度函数（连续情形）

若 X, Y 有联合密度 $f_{X,Y}(x,y)$，则独立 $\Leftrightarrow$

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$$

多元推广

若 $X_1, \ldots, X_n$ 相互独立，则对任意函数 $F_1, \ldots, F_n$：

$$\mathbb{E}[F_1(X_1) \cdots F_n(X_n)] = \prod_{i=1}^{n} \mathbb{E}[F_i(X_i)]$$

且联合分布函数/密度函数可分解：

$$f_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i)$$

二、协方差与相关性（Covariance & Correlation）

协方差定义

设 $\mu_X = \mathbb{E}[X], \mu_Y = \mathbb{E}[Y]$，则：

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

重要性质

- $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
- 若 X 与 Y **独立** $\Rightarrow$ $\mathrm{Cov}(X,Y) = 0$（即 **不相关**）
    - 但 **反之不成立**：不相关 $\neq$ 独立（除非联合正态）
- 线性变换下：$\mathrm{Cov}(a_1X + b_1, a_2Y + b_2) = a_1a_2\mathrm{Cov}(X,Y)$

相关系数（Pearson）

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

- 定义域：若 $\mathrm{Var}(X) > 0$ 且 $\mathrm{Var}(Y) > 0$
- 取值范围：$-1 \leq \rho_{X,Y} \leq 1$
- $\rho = \pm 1$ $\Leftrightarrow$ 存在线性关系（完全相关）

三、大数定律（Law of Large Numbers）

背景

设 $X_1, X_2, \dots, X_n$ 是独立同分布（i.i.d.）随机变量，满足：

$$\mathbb{E}[X_i] = \mu, \mathrm{Var}(X_i) = \sigma^2 < \infty$$

定义样本均值：

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

结论

- $\mathbb{E}[\bar{X}_n] = \mu$
- $\mathrm{Var}(\bar{X}_n) = \frac{\sigma^2}{n} \to 0$ （当 $n \to \infty$）
- **弱大数定律**：$\forall \varepsilon > 0, \lim_{n\to\infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$即样本均值依概率收敛于总体均值。

四、伯努利试验与频率估计（Bernoulli Model）

伯努利随机变量

设事件 $A = \{a \leq X \leq b\}$，定义指示变量：

$$Y = \mathbf{1}_A = \begin{cases} 1, & X \in [a, b] \\ 0, & \text{otherwise} \end{cases}$$

则 $Y \sim \mathrm{Bernoulli}(p)$，其中 $p = P(X \in [a, b])$

- $\mathbb{E}[Y] = p$
- $\mathrm{Var}(Y) = p(1 - p)$

频率估计

重复 n 次实验，得 $Y_1, \dots, Y_n$，样本频率：

$$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

由大数定律，$\hat{p}_n \xrightarrow{P} p$

✅ 应用：通过频率估计概率 $P(X \in C)$

五、离散分布拟合示例（Example: Women in Survey）

- 调查 110 人中女性人数 X
- 观察频数 → 估计期望：$\hat{\mu} = \sum_{k=0}^{10} k \cdot \hat{p}_k = 0.435$
- 拟合模型：$X \sim \mathrm{Binomial}(n = 10, p = 0.435)$

注：此处似为教学简化，实际总人数应为 110，但拟合时取子样本大小为 10。

六、连续型变量密度估计（Density Estimation）

局部频率思想

对小区间 $C = (a - h, a + h)$，有：

$$P(X \in C) = \int_{a-h}^{a+h} f_X(x)\, dx \approx 2h \cdot f_X(a)\ (h \text{ 很小})$$

⇒ 密度估计：

$$\hat{f}_X(a) \approx \frac{\text{落在 } (a-h, a+h) \text{ 的比例}}{2h}$$

直方图（Histogram）

- 将数据范围划分为 **bins**（区间）$B_1, \dots, B_m$
- bin 宽度：$|B_i|$
- **归一化直方图**：令总面积 = 1，即第 i 个 bin 高度 $= \frac{\text{落入 } B_i \text{ 的数据比例}}{|B_i|}$
- 目的：**估计未知密度函数** $f(x)$

📌 优点：直观展示数据分布形态（如双峰、偏态）

七、探索性数据分析（EDA）：Old Faithful 喷发数据

数据概况

- 样本量：$n = 272$
- 变量：喷发持续时间（秒）
- 范围：96 ~ 306 秒
- 样本均值：$\bar{x} = 209.3$

问题

- 均值不能反映分布多峰性
- 地质学认为存在**两类喷发机制** ⇒ 期望看到**双峰分布**

分析方法

1. **排序数据**（Table 15.2）：
   - 中位数 ≈ 240（偏右）
   - 提示分布右偏（正偏）
2. **绘制直方图**（Fig. 15.1）：
   - 明确显示**双峰结构**：峰值在 ≈120 秒 与 ≈270 秒
   - 验证两类喷发假说

直方图构建步骤（归一化版）

1. 选择 bin 数量与宽度（如等宽 8 bins）
2. 统计每个 bin 内数据频数
3. 计算每个 bin 的**密度高度**：$\text{height}_i = \frac{\text{count}_i / n}{\text{bin width}}$
4. 绘制条形图，总面积 = 1

📊 此图可作为真实密度 $f(x)$ 的**非参数估计**

八、补充说明与教学提示

- **Chebyshev 不等式**（文中提及）：

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

用于证明大数定律。

- **Quick Exercise 13.3 解答思路**：
   - 直方图高度 = 0.26，bin 宽 = 0.5（1.75~2.25）
   - 面积 = $0.26 \times 0.5 = 0.13$
   - 数据量 = $0.13 \times 500 = 65$ 个

✅ **总结图示**

主题	核心思想	关键公式/结论
独立性	联合 = 边缘乘积	$f_{X,Y} = f_X f_Y$
协方差	线性相关度量	$\mathrm{Cov}(X,Y) = E[XY] - EXEY$
相关系数	标准化协方差	$\rho \in [-1,1]$
大数定律	频率 → 概率	$\bar{X}_n \xrightarrow{P} \mu$
直方图	密度非参估计	height = (比例)/(bin width)
EDA	可视化分布	揭示多峰、偏态、异常值

中心极限定理 (Central Limit Theorem, CLT)

1. 动机 (Motivation)

- 样本均值 $\bar{X}_n$ 是总体均值 μ 的一个良好估计。
- 为标准化样本均值，考虑其标准化形式：$\frac{\bar{X}_n-\mu}{\sigma_{\bar{X}_n}}=\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$
- 当样本量 n 增大时，样本均值的抽样分布特性趋近于正态分布。

2. 定理陈述 (Theorem Statement)

设 $X_1,X_2,\ldots,X_n$ 为独立同分布（i.i.d.）的随机变量，其期望和方差分别为：

$$E[X_i]=\mu,Var(X_i)=\sigma^2$$

定义标准化样本均值：

$$Z_n=\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$$

则当 $n\to\infty$ 时，Z_n 的分布收敛于标准正态分布：

$$P(Z_n\in[a,b])\overset{n\to\infty}{\to}\int_a^b\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx$$

等价地，对于任意有界连续函数 f：

$$E[f(Z_n)]\overset{n\to\infty}{\to}\int_{-\infty}^{+\infty}f(x)\frac{e^{-x^2/2}}{\sqrt{2\pi}}dx$$

3. 解释与推论 (Comments & Corollaries)

(1) 等价表述

定理可简洁表述为：

$$Z_n\underset{n\to\infty}{\overset{\text{in law}}{\to}}N(0,1)$$

即，标准化样本均值 Z_n 在分布上收敛于标准正态分布 $N(0,1)$。

(2) 实用性说明

- **无误差控制**：CLT 是一个渐近结果，它不提供对有限样本下近似误差的控制。
- **经验法则**：在实践中，通常认为当 $n\geq 30$ 时，样本均值的分布可以较好地近似为正态分布。
- **正态性检验**：在应用 CLT 前，有时需要对数据进行正态性检验以确认其适用性。

---

4. 另一种表述与推导 (Alternative Statements & Derivations)

(B) 替代表述

中心极限定理的核心结论也可直接写作：

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

(3.1) 样本均值的渐近分布

由上述结论，可反推样本均值 $\bar{X}_n$ 的渐近分布：

$$\bar{X}_n \underset{n\to\infty}{\overset{\text{approx.}}{\rightarrow}} N\left(\mu, \frac{\sigma^2}{n}\right)$$

这与切比雪夫大数定律的结论非常相似，但 CLT 提供了更精确的分布形态信息。

(3.2) 样本总和的渐近分布

考虑样本总和 $S_n = X_1 + X_2 + \ldots + X_n$。

由 $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$ 可得：

$$S_n = n\bar{X}_n = n\mu + \sqrt{n}\sigma Z_n$$

因此，样本总和 S_n 的渐近分布为：

$$S_n \underset{n\to\infty}{\overset{\text{approx.}}{\rightarrow}} N(n\mu, n\sigma^2)$$

---

5. 中心极限定理的应用 (Applications of CLT)

A1) 对切比雪夫不等式的改进

切比雪夫不等式给出了一个保守的概率上界：

$$P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

而中心极限定理能提供更精确的概率估计：

$$P(|\bar{X}_n - \mu| \geq \varepsilon) = P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\right) = P(|Z_n| \geq \frac{\varepsilon\sqrt{n}}{\sigma})$$

由于 $Z_n \approx N(0,1)$，该概率可通过标准正态分布表查得：

$$= 2\int_{\frac{\varepsilon\sqrt{n}}{\sigma}}^{+\infty} \frac{e^{-z^2/2}}{\sqrt{2\pi}} dz$$

A2) 伯努利分布案例（$X_i \sim Ber(p)$）

A2.1) 切比雪夫方法

对于伯努利试验，有 $E[X_i] = p$，$Var(X_i) = p(1-p)$。

要使 $P(|\bar{X}_n - p| \geq \varepsilon) \leq \delta$，由切比雪夫不等式：

$$\frac{p(1-p)}{n\varepsilon^2} \leq \delta \implies n \geq \frac{p(1-p)}{\varepsilon^2\delta}$$

由于 $p(1-p) \leq \frac{1}{4}$，可得一个更保守的样本量要求：

$$n \geq \frac{1}{4\varepsilon^2\delta}$$

A2.2) 中心极限定理方法

目标

我们要找到一个最小的样本量 n ，使得样本均值 $\bar{X}_n$ 与总体均值 p 的误差不超过 ε 的概率至少为 $1-\delta$ 。

换句话说：

$$P(|\bar{X}_n - p| < \varepsilon) \geq 1 - \delta$$

这等价于：

$$P(|\bar{X}_n - p| \geq \varepsilon) \leq \delta$$

利用 CLT，我们有：

$$P(|\bar{X}_n - p| \geq \varepsilon) = P\left(|Z_n| \geq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) \approx 2P\left(Z \geq \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right)$$

令 z_0 满足 $P(Z \geq z_0) = \delta/2$，则：

$$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \geq z_0 \implies n \geq \frac{z_0^2 p(1-p)}{\varepsilon^2}$$

同样，使用最坏情况 $p(1-p) \leq \frac{1}{4}$，得到：

$$n \geq \frac{z_0^2}{4\varepsilon^2}$$

A2.3) 示例计算

给定 $\varepsilon = 10^{-2}$，$\delta = 10^{-3}$。

- **切比雪夫方法**:

$$n \geq \frac{1}{4 \times (10^{-2})^2 \times 10^{-3}} = \frac{1}{4 \times 10^{-7}} = 2{,}500{,}000$$

- **中心极限定理方法**:

  查表得 $P(Z \geq 3.32) \approx 0.0005 = \delta/2$，所以 $z_0 = 3.32$。

$$n \geq \frac{(3.32)^2}{4 \times (10^{-2})^2} = \frac{11.0224}{4 \times 10^{-4}} = 27{,}556$$

（注：原文计算为 31,000，此处按公式精确计算为 27,556，可能是四舍五入或笔误）。

---

6. 二项分布的正态近似 (Normal Approximation to Binomial)

A3) 二项分布的渐近正态性

若 $X_i \sim Ber(p)$，则样本总和 $S_n = X_1 + \ldots + X_n \sim Bin(n, p)$。

根据 CLT，当 n 足够大时:

$$S_n \approx N(np, np(1-p))$$

即:

$$Bin(n, p) \approx N(np, np(1-p))$$

A4) 应用示例：抛硬币问题

求在 10 次公平抛掷中，恰好出现 8 次正面的概率。

- **精确解**:

$$P(X = 8) = \binom{10}{8}\left(\frac{1}{2}\right)^8\left(\frac{1}{2}\right)^2 = \binom{10}{8}\left(\frac{1}{2}\right)^{10} \approx 0.0439$$

- **正态近似解**:

  令 $X \sim Bin(10, 0.5)$，则其近似分布为 $N(5, 2.5)$。

  为提高精度，使用**连续性校正**，计算 $P(7.5 \leq X \leq 8.5)$。

  标准化:

$$Z = \frac{X - 5}{\sqrt{2.5}} \Rightarrow P\left(\frac{7.5 - 5}{\sqrt{2.5}} \leq Z \leq \frac{8.5 - 5}{\sqrt{2.5}}\right) = P(1.58 \leq Z \leq 2.21)$$

查表得:

$$P(Z \geq 1.58) \approx 0.0571, P(Z \geq 2.21) \approx 0.0136$$

所以:

$$P(1.58 \le Z \le 2.21) = 0.0571 - 0.0136 = 0.0435$$

此结果与精确解 0.0439 非常接近，验证了正态近似的有效性。

---

7. 二项分布正态近似的条件 (Conditions for Normal Approximation to Binomial)

7.1 核心问题

为什么二项分布 $Bin(n,p)$ 可以用正态分布 $N(np, np(1-p))$ 来近似？其核心在于**形状相似性**。

- **直观解释**:
  - 二项分布的概率质量函数（PMF）在 n 较大时呈现“钟形”（Bell-shaped）。
  - 当正态分布的钟形曲线主体部分完全落在二项分布的支持集 $[0,n]$ 内时，近似效果良好。
  - 这是因为二项分布的离散概率点可以被看作是对连续正态密度曲线的一种拟合。

7.2 数学条件

为了确保正态近似有效，需要满足以下两个不等式，它们保证了正态分布的“3σ 原则”区间 $[\mu - 3\sigma, \mu + 3\sigma]$ 完全包含在 $[0,n]$ 内:

(a) 下界条件:

$$\mu - 3\sigma = np - 3\sqrt{np(1-p)} > 0$$

(b) 上界条件:

$$\mu + 3\sigma = np + 3\sqrt{np(1-p)} < n$$

这两个条件可以简化为一个更实用的经验法则:

- 由 (a) 得: $n > 9 \cdot \frac{1-p}{p}$
- 由 (b) 得: $n > 9 \cdot \frac{p}{1-p}$

因此，一个常用的综合经验法则是:

$$n > 9 \cdot \max\left(\frac{p}{1-p}, \frac{1-p}{p}\right)$$

7.3 示例验证

在之前的抛硬币例子中，$n = 10$，$p = 1/2$。

- 计算下界：$\mu - 3\sigma = 5 - 3\sqrt{2.5} \approx 5 - 4.74 = 0.26 > 0$
- 计算上界：$\mu + 3\sigma = 5 + 3\sqrt{2.5} \approx 5 + 4.74 = 9.74 < 10$

两个条件均满足，因此正态近似是有效的。

对于 $p = 1/2$ 的情况：

- $\frac{p}{1-p} = 1$，$\frac{1-p}{p} = 1$
- 所以要求 $n > 9 \times 1 = 9$。
- 我们有 $n = 10 > 9$，同样验证了近似的可行性。

---

8. 舍入误差 (Round-off Errors)

8.1 定义与来源

舍入误差是指在数值计算或数据记录过程中，由于四舍五入而产生的微小误差。

- **示例**：
  - $100.35 \rightarrow 100$ （向下舍入）
  - $100.75 \rightarrow 101$ （向上舍入）

8.2 模型假设

通常将单个舍入误差建模为一个均匀分布的随机变量：

$$E_i \sim Unif(-0.5, 0.5)$$

这意味着每个数被舍入后，其真实值与记录值之间的误差在 $[-0.5, 0.5]$ 区间内是等可能的。

8.3 误差累积分析

考虑对 100 个数字进行舍入操作，记每个数字的舍入误差为 $X_1, X_2, \ldots, X_{100}$。

- **期望**：$E[X_i] = 0$
- **方差**：$Var(X_i) = \frac{(0.5-(-0.5))^2}{12} = \frac{1}{12}$

我们关心的是总误差的大小，即求解：

$$P(|X_1 + X_2 + \ldots + X_{100}| \geq 10)$$

这是一个关于独立同分布随机变量和的概率问题，可以应用中心极限定理来近似求解。

中心极限定理 (Central Limit Theorem, CLT) 核心知识体系总结

一、核心思想与动机

中心极限定理是概率论与统计学中最重要的定理之一，其核心思想是：

无论总体分布如何，只要样本量 n 足够大，样本均值 $\bar{X}_n$ 的抽样分布将趋近于正态分布。

- **动机**：从样本推断总体参数（如均值 μ）时，需要了解样本统计量（如 $\bar{X}_n$）的分布特性。CLT 提供了这一理论基础。
- **标准化形式**：通过构造标准化变量 $Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$，可以将其分布统一到标准正态分布 $N(0,1)$ 上进行分析。

二、定理的精确陈述与等价形式

1. **基本形式**：

   设 $X_1, \ldots, X_n$ 为 i.i.d. 随机变量，$E[X_i] = \mu$，$Var(X_i) = \sigma^2$。

   则当 $n \to \infty$ 时，

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \overset{\text{in law}}{\rightarrow} N(0,1)$$

2. **样本均值的渐近分布**：

$$\bar{X}_n \underset{n\to\infty}{\overset{\text{approx.}}{\rightarrow}} N\left(\mu, \frac{\sigma^2}{n}\right)$$

3. **样本总和的渐近分布**：

$$S_n = \sum_{i=1}^{n} X_i \underset{n\to\infty}{\overset{\text{approx.}}{\rightarrow}} N(n\mu, n\sigma^2)$$

三、定理的应用与价值

1. **对切比雪夫不等式的改进**：

- **切比雪夫**：提供保守的概率上界 $P(|\bar{X}_n - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$。
- **CLT**：提供更精确的概率估计，可通过标准正态分布表计算 $P(|\bar{X}_n - \mu| \geq \varepsilon) \approx 2\Phi(-\frac{\varepsilon\sqrt{n}}{\sigma})$，其中 Φ 为标准正态累积分布函数。

2. **伯努利分布与二项分布的正态近似**：
   - 对于 $X_i \sim Ber(p)$，样本总和 $S_n \sim Bin(n,p)$。
   - 当 n 足够大时，$Bin(n,p) \approx N(np, np(1-p))$。
   - **连续性校正**：在计算离散变量概率时，使用 $P(a \leq X \leq b) \approx P(a - 0.5 < Y < b + 0.5)$，其中 Y 为对应的正态变量，可显著提高精度。
3. **正态近似的适用条件**：
   - **经验法则**：通常认为 $n \geq 30$ 时近似效果良好。
   - **更严格的条件**：要求正态分布的"3σ 区间"落在二项分布的支持集内，即：
     $np - 3\sqrt{np(1-p)} > 0$ 且 $np + 3\sqrt{np(1-p)} < n$
     这等价于 $n > 9 \cdot \max\left(\frac{p}{1-p}, \frac{1-p}{p}\right)$。对于 $p = 0.5$，只需 $n > 9$。
4. **误差分析**：
   - **舍入误差**：单个误差 $E_i \sim Unif(-0.5,0.5)$，期望为 0，方差为 $1/12$。
   - **误差累积**：对大量数据求和时，总误差服从近似正态分布，可用 CLT 评估其超过某个阈值的概率。

四、关键结论与实践指导

- **无误差控制**：CLT 是一个渐近定理，它不保证有限样本下的精确度，仅说明趋势。
- **"经验法则"并非绝对**：$n \geq 30$ 是一个通用建议，但具体应用中需结合数据特征（如偏度、峰度）或进行正态性检验。
- **"形状相似"是本质**：二项分布能用正态分布近似，根本原因在于两者都具有钟形曲线的形态，且当样本量足够大时，这种形态上的相似性足以支撑近似计算。

中心极限定理是桥梁：它架起了"任意分布"与"正态分布"之间的桥梁，使得许多复杂的统计推断问题（如置信区间、假设检验）得以简化和解决。

统计学基础笔记：描述性统计与初步建模

核心思想：理解数据是分析的第一步。描述性统计关注“数据本身是什么”，而统计建模则试图回答“数据从何而来”以及“未来会怎样”。

第一部分：统计学两大范式

1. 描述性统计 (Descriptive Statistics)

- **目标**：总结、呈现和探索手头已有的数据。
- **核心特征**：
  - **不涉及概率**：只关注现有样本，不做任何推断。
  - **符号约定**：使用**拉丁字母**（如 $\bar{x}$, s）表示基于样本计算出的具体数值。
- **常用工具**：Excel、图表、汇总统计量。

2. 统计建模 (Statistical Modelling)

- **目标**：为数据背后的生成过程建立数学模型，用于推断总体参数或预测未来。
- **核心特征**：
  - **引入概率**：假设数据是从某个概率分布中随机抽取的。
  - **符号约定**：使用**希腊字母**（如 μ, σ）表示未知的总体参数。
- **关键概念**：
  - **随机变量 (Random Variable, RV)**：将样本数据 $x_1, \dots, x_n$ 视为随机变量 $X_1, \dots, X_n$。
  - **独立同分布 (IID)**：通常假设 $X_i \sim F(\theta)$，其中 F 是某个分布，θ 是未知参数。
  - **估计量 (Estimator) vs. 统计量 (Statistic)**：
    - **统计量**：一个具体的数值，如样本均值 $\bar{x} = 85$。
    - **估计量**：一个函数或规则，如 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$，它是一个随机变量，其值随抽样变化。

第二部分：数据探索与预处理

1. 数据排序 (Ordering)

- **目的**：为后续分析（如求中位数、分箱、画图）做准备。
- **定义**：将原始数据 $x_1, \dots, x_n$ 按从小到大排列，得到**顺序统计量** $x_{(1)} \le x_{(2)} \le \dots \le x_{(n)}$。
- **应用**：
  - $x_{(1)}$ 是最小值，$x_{(n)}$ 是最大值。
  - 中位数、四分位数等都基于有序数据计算。

2. 分箱 (Binning)

- **目的**：简化数据，便于可视化和分析。
- **方法**：
  a. 确定数据范围 $[x_{(1)}, x_{(n)}]$。

    b. 将该范围划分为若干个连续的子区间（称为“箱”或“组”）。
- **应用**：绘制直方图 (Histogram)，直观展示数据分布。

第三部分：数据分布的视觉描述 (Jargon)

1. 对称性 (Symmetry)

- **对称分布 (Symmetric Distribution)**：
    - 定义：存在一个中心点，使得分布的右侧是左侧的镜像反射。
    - 示例：正态分布（钟形曲线）。

2. 偏度 (Skewness)

- **左偏 (Left-Skewed / Negative Skew)**：
    - 特征：左侧有长尾，大部分数据集中在右侧。
    - 关系：均值 < 中位数。
- **右偏 (Right-Skewed / Positive Skew)**：
    - 特征：右侧有长尾，大部分数据集中在左侧。
    - 关系：均值 > 中位数。

3. 峰度/模态 (Modality)

- **单峰 (Unimodal)**：只有一个峰值。
- **双峰 (Bimodal)**：有两个明显的峰值，可能暗示数据来自两个不同群体。
- **多峰 (Multimodal)**：有三个或更多峰值。

第四部分：数值摘要 (Numerical Summaries)

数据摘要主要从三个维度进行：

1. 数据的中心 (Center of Data)

a) 平均数 (Mean)

- **公式**：$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$
- **优点**：具有优良的数学性质，是许多统计推断的基础。
- **缺点**：对异常值敏感。

b) 中位数 (Median)

- **定义**：有序数据集的中间值。
- **计算**：
    - 若 n 为奇数：$\mathrm{MED}_n = x_{((n+1)/2)}$
    - 若 n 为偶数：$\mathrm{MED}_n = \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)})$
- **优点**：对异常值不敏感，稳健。
- **缺点**：在数学上较难处理。

c) 众数 (Mode)

- **定义**：数据集中出现频率最高的值。
- **优点**：计算最简单，对异常值不敏感。
- **缺点**：信息量少，对于连续数据可能无意义。

---

2. 数据的离散程度 (Dispersion of Data)

a) 样本标准差 (Sample Standard Deviation, s)

- **定义**：衡量数据点围绕均值的波动程度。
- **公式**：

$$s_n^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \text{ (有偏)}$$

$$s_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 \text{ (无偏，更常用)}$$

$$s_{n-1} = \sqrt{s_{n-1}^2}$$

- **优点**：具有良好的数学性质。
- **缺点**：对异常值敏感。

b) 平均绝对偏差的中位数 (Median Absolute Deviation, MAD)

- **定义**：数据点到中位数的绝对距离的中位数。
- **公式**：$\mathrm{MAD}(x_1, \dots, x_n) = \mathrm{MED}(|x_1 - \mathrm{MED}_n|, \dots, |x_n - \mathrm{MED}_n|)$
- **计算步骤**:
    a. 排序原始数据。
    b. 计算中位数 MED_n。
    c. 计算每个 x_i 到 MED_n 的绝对偏差 $|x_i - \mathrm{MED}_n|$。
    d. 对这些绝对偏差再次排序。
    e. 取这些绝对偏差的中位数，即为 MAD。
- **优点**：极其稳健，对异常值免疫。
- **缺点**：数学处理困难。

第五部分：总结对比表

集中趋势度量对比

度量	优点	缺点
平均数	数学性质优良，理论基础扎实	对异常值敏感
中位数	对异常值不敏感，稳健	数学上较难处理
众数	计算最简单，对异常值不敏感	信息量少，可能不存在

离散程度度量对比

度量	优点	缺点
标准差	数学性质优良	对异常值敏感
MAD	极其稳健，对异常值免疫	数学上较难处理

第六部分：符号约定总结

这是区分描述性统计与统计建模的关键。

领域	**符号类型**	**示例**	**含义**
描述性统计	拉丁字母	$\bar{x}$, s, s^2	基于样本计算出的具体数值
统计建模	希腊字母	μ, σ, σ^2	未知的总体参数

核心口诀:

- **小写拉丁字母 ($\bar{x}$) = 你看到的** (描述性)
- **希腊字母 (μ) = 你想知道的** (建模)

描述性统计与数据探索 (Descriptive Statistics & Data Exploration)

中文	**英文**
描述性统计	Descriptive Statistics
统计建模	Statistical Modelling
概率	Probability
样本	Sample
总体	Population
数据	Data
观测值	Observation / Data Point
排序	Ordering
顺序统计量	Order Statistics
最小值	Minimum
最大值	Maximum
范围	Range
分箱 / 分组	Binning / Grouping
子区间	Subinterval
箱线图	Boxplot
直方图	Histogram
密度图	Density Plot

数值摘要 (Numerical Summaries)
集中趋势 (Measures of Center)

中文	英文
集中趋势	Center of Data
平均数 / 均值	Average / Mean
经验均值	Empirical Mean
中位数	Median
众数	Mode
典型值	Typical Value

离散程度 (Measures of Dispersion)

中文	英文
离散程度 / 变异性	Dispersion / Variability
方差	Variance
样本方差	Sample Variance
有偏估计	Biased Estimator
无偏样本方差	Unbiased Sample Variance
自由度	Degrees of Freedom
标准差	Standard Deviation
平均绝对偏差	Mean Absolute Deviation (MAOD)
平均绝对偏差的中位数	Median of Absolute Deviation (MAD)
绝对偏差	Absolute Deviation
四分位距	Interquartile Range (IQR)

分布形状 (Shape of Distribution)

中文	英文
分布形状	Shape

中文	英文
对称分布	Symmetric Distribution
左偏（负偏态）	Left-Skewed (Negatively Skewed)
右偏（正偏态）	Right-Skewed (Positively Skewed)
偏度	Skewness
单峰	Unimodal
双峰	Bimodal
多峰	Multimodal
模态	Modality
峰	Peak / Mode
尾部	Tail

参数与符号 (Parameters & Notation)

中文	英文
参数	Parameter
未知参数	Unknown Parameter
总体均值	Population Mean
总体标准差	Population Standard Deviation
总体方差	Population Variance
拉丁字母	Latin Letters
希腊字母	Greek Letters
符号约定	Notation Convention
估计量	Estimator
统计量	Statistic
函数	Function

概率与随机变量 (Probability & Random Variables)

中文	英文
随机变量	Random Variable (RV)
独立同分布	Independent and Identically Distributed (IID)
概率分布	Probability Distribution
正态分布	Normal Distribution
期望	Expectation
法则 / 定律	Law

其他关键术语

中文	英文
异常值	Outlier
极端值	Extreme Value
稳健的	Robust
经验性的	Empirical
探索性数据分析	Exploratory Data Analysis (EDA)
行话	Jargon
缩写	Abbreviation
图形化呈现	Graphical Presentation
代表性数字	Representative Number

探索性数据分析：图形与数值摘要

Exploratory Data Analysis: Graphical and Numerical Summaries

基于 Old Faithful 喷泉数据及其他实例，系统梳理描述性统计的核心方法

一、引言：描述性统计的目标

描述性统计（Descriptive Statistics）旨在：

- 理解、描述和探索数据
- 图形化展示数据的分布特征
- 提取代表性数值（如中心趋势、离散程度）
- 为后续建模提供初步洞察

  ⚠ 注意：本阶段不涉及概率模型，仅对样本数据本身进行总结，属于非概率性分析（Non-probabilistic）。

二、图形化摘要（Graphical Summaries）

1. 直方图（Histogram）

（1）定义与目的

直方图是展示单变量数据分布的经典工具，通过将数据划分为若干互不重叠的区间（bins），并以柱状形式表示每个区间内观测值的频数或密度。

其核心目标包括：

- 识别数据的集中趋势（modes）
- 判断分布是否对称或存在偏态（skewness）
- 发现潜在的多峰结构（multimodality），如双峰（bimodal）

（2）构造步骤

设样本容量为 n，观测值为：

$$x_1, x_2, \dots, x_n$$

1. **划分区间（bins）**
   将数据范围划分为 m 个宽度为 $|B_i|$ 的互斥区间 $B_1, B_2, \ldots, B_m$。

2. **计算频数**
   统计落入每个区间的观测个数：

$$f_i = \#\{x_j \in B_i\}$$

3. **归一化高度（密度估计）**
   为使总面积为 1，便于比较不同样本或作为概率密度函数的估计：

$$h_i = \frac{f_i}{n \cdot |B_i|}$$

此时，每个柱子的**面积**等于该区间内数据所占比例。

（3）案例：Old Faithful 喷泉数据

- 样本量：$n = 272$
- 喷发间隔时间范围：[96,306] 秒
- 样本均值：$\bar{x} = 209.3$ 秒（但单一均值无法反映真实结构）

☑ 关键观察：

（4）箱宽选择策略

箱宽过小会导致噪声干扰，过大则掩盖关键结构。

箱宽	效果
过小（如 $b = 2$）	出现大量孤立峰，难以识别模式
过大（如 $b = 90$）	双峰结构消失，呈单峰假象
合理（如 $b = 30$）	清晰呈现双峰特征

推荐经验公式：

- **Sturges 公式**（确定箱数）：

$$m = 1 + 3.3\log_{10}(n)$$

- **Scott 公式**（确定箱宽）：

$$b = 3.49 \cdot s \cdot n^{-1/3}$$

其中 \$ s \$ 为样本标准差。

> 💡 实践建议：优先使用 Scott 公式作为起点，再结合可视化判断微调，选择最能揭示数据内在结构的箱宽。

（5）直方图的分布特征

特征	中文描述
对称性（Symmetry）	左右两侧形态近似镜像
偏态性（Skewness）	右偏：长尾向右；左偏：长尾向左
模态性（Modality）	单峰（Unimodal）、双峰（Bimodal）、多峰（Multimodal）

> Old Faithful 数据为典型的**双峰分布**（bimodal distribution），说明其背后可能对应两种不同的物理机制（如短时喷发与长时喷发）。

---

2. 经验分布函数（Empirical Distribution Function, EDF）

（1）定义

经验分布函数 $F_n(x)$ 定义为：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)$$

其中 $I(\cdot)$ 为指示函数，当条件成立时取值为 1，否则为 0。

（2）几何特征

- 是一条从 0 到 1 的**阶梯函数**

- 在每个观测值处跳跃 $\frac{1}{n}$
- 跳跃越密集的区域 → 数据越集中

（3）与直方图的关系

直方图中某一区间的**面积**，等于 EDF 在该区间上的**增量**。

> 例如：在 Old Faithful 数据中，区间 (240,270] 的直方图高度为 0.0092，宽度为 30，则面积为：

$$A = 30 \times 0.0092 = 0.276$$

表示约 27.6% 的数据落在该区间，与 EDF 在此区间上升幅度一致。

> ☑ EDF 与直方图互为补充：EDF 展示累积分布，直方图展示局部密度。

---

3. 散点图（Scatterplot）

（1）适用场景

用于分析两个定量变量之间的关系，揭示：

- 相关性方向（正相关 / 负相关）
- 关系形态（线性 / 非线性）
- 异常值（outliers）或数据聚类

（2）案例 1：钻孔时间 vs 深度

数据来源：干孔（Dry）与湿孔（Wet）钻探记录。

> ☑ 观察结论：

（3）案例 2：木材密度 vs Janka 硬度

数据：36 种桉树木材的密度（g/cm³）与硬度（N）

> ☑ 观察结论：

> 🔍 快速估计：若某木材密度为 65 g/cm³，参考邻近点（66.0 → 3260 N，67.4 → 2700 N），可外推其硬度约为 **2900–3000 N**。

三、数值摘要（Numerical Summaries）

1. 中心趋势度量

指标	公式	特点
均值（Mean）	$\bar{x}=\frac{1}{n}\sum_{i=1}^{n}x_i$	数学性质优良，但对异常值敏感
中位数（Median）	$\mathrm{Med}(x)=\begin{cases}x_{(k)} & n=2k-1\\ \frac{1}{2}(x_{(k)}+x_{(k+1)}) & n=2k\end{cases}$	对异常值鲁棒（robust），不易受极端值影响
众数（Mode）	出现频率最高的值	可能不唯一，适用于分类数据或初步识别峰点

📌 示例：温度数据 [41, 41, 41, 41, 42, 42, 43, 43, 58, 58]

☑ 当存在异常值（58）时，中位数比均值更具代表性。

---

2. 离散程度度量

（1）经验方差（Empirical Variance）

$$s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2 \text{ 或}$$

$$\text{无偏估计 } s^2=\frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

❗ 文档中使用的是有偏估计（除以 n），实际教学与科研中推荐使用无偏估计（除以 $n-1$）。

（2）绝对中位偏差（Median Absolute Deviation, MAD）

定义：

$$\mathrm{MAD}=\mathrm{Med}(|x_i-\mathrm{Med}(x)|)$$

计算步骤：

1. 计算中位数 $M = \text{Med}(x_1, \dots, x_n)$
2. 计算每个观测值与中位数的绝对偏差：$d_i = |x_i - M|$
3. 对 d_i 再求中位数 → 得到 MAD

☑ 优点：对异常值极鲁棒，适用于非正态分布数据
✖ 缺点：数学处理复杂，不便于理论推导

📌 示例（温度数据）：

---

3. 分布形状度量

- **分位数（Quantiles）**：用于刻画分布形态
  - 四分位数：Q1（25%）、Q2（50%，即中位数）、Q3（75%）
  - 五数概括：最小值、Q1、中位数、Q3、最大值
- **偏度（Skewness）**：衡量分布不对称程度
- **峰度（Kurtosis）**：衡量尾部厚重程度（是否尖峰或扁平）

注：这些指标将在后续章节深入讨论，此处作为概念引入。

四、符号体系对照（仅专业术语标注英文）

类型	中文术语	英文符号
样本统计量	样本均值	$\bar{x}$
样本统计量	样本方差	s^2
样本统计量	样本标准差	s
总体参数	总体均值	μ
总体参数	总体方差	σ^2

总体参数	总体标准差	σ

☑ 记忆口诀：样本用拉丁字母（x, s），总体用希腊字母（μ, σ）

五、总结：方法对比与教学要点

方法	用途	数据类型	对异常值敏感？
均值（Mean）	衡量中心位置	定量	是
中位数（Median）	衡量中心位置	定量、有序	否
众数（Mode）	识别最常见值	分类、定量	否
方差（Variance）	衡量离散程度	定量	是
MAD	衡量离散程度（鲁棒）	定量	否
直方图（Histogram）	展示分布形态	单变量定量	受箱宽影响
经验分布函数（EDF）	展示累积分布	单变量定量	稳健
散点图（Scatterplot）	探索双变量关系	双变量定量	可识别异常点

🎯 教学建议：先用图形（直方图、散点图）观察整体模式，再用数值（均值、中位数、MAD）进行量化描述，形成“图形 + 数值”双重验证。

描述性统计与参数估计基础：从数据探索到推断理论

第一部分：描述性统计量 (Descriptive Statistics)

1. 中位数 (Median) 的通用计算方法

对于一个已排序的样本 $x_{(1)} \le x_{(2)} \le \cdots \le x_{(N)}$，中位数的计算步骤如下：

- **步骤 1**: 计算索引

$$k = \left\lfloor \frac{N+1}{2} \right\rfloor。$$

- **中位数公式**:

$$\text{Med} = x_{(k)} + \left[\frac{N+1}{2} - k\right] \cdot \left(x_{(k+1)} - x_{(k)}\right)$$

此公式通过线性插值，在 k 和 $k+1$ 位置之间进行计算，适用于任意样本大小 N。

---

2. 众数 (Mode)

- 众数是数据集中出现频率最高的数值。
- 笔记中仅列出名称，未提供具体计算方法。

---

3. 离散度 (Dispersion)

3.1 样本方差 (Sample Variance)

- **定义**: 样本方差是衡量数据点围绕样本均值离散程度的指标。
- **公式**:

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x}_n)^2$$

- **性质**: 这是一个无偏估计量（标准定义下，分母为 $n-1$ 的样本方差是总体方差 σ^2 的无偏估计）。

3.2 平均绝对偏差 (MAD, Mean Absolute Deviation)

- **定义**: MAD 是每个数据点与中位数的绝对偏差的中位数。
- **公式**:

$$\text{MAD} = \text{MED}[|x_1 - \text{MED}|, \dots, |x_N - \text{MED}|]$$

- **解释**: 它衡量的是数据相对于其中位数的离散程度，对异常值不敏感。

---

4. 分位数 (Quantiles)

4.1 四分位数 (Quartiles)

四分位数将数据集分为四个相等的部分。

- **下四分位数 (Q1)**: $Q(\frac{1}{4})$
- **中位数 (Q2)**: $Q(\frac{1}{2}) = \text{median}$
- **上四分位数 (Q3)**: $Q(\frac{3}{4})$

4.2 p-分位数的通用计算方法

对于任意分位数 p，其计算步骤如下：

- **步骤 1**: 计算索引

$$k = \lfloor (N+1)p \rfloor$$ （向下取整）。

- **p-分位数公式**:

$$Q(p) = x_{(k)} + [(N+1)p - k] \cdot \left(x_{(k+1)} - x_{(k)}\right)$$

此公式是中位数公式的推广，适用于任意分位数 p。

---

5. Tukey 概要图 (Tukey's Summary / Boxplot)

Tukey 概要图是一种用于可视化数据分布和识别异常值的图形工具。

- **组成部分**:
  - **箱体 (Box)**: 由下四分位数 (Q_1)、中位数 (Q_2) 和上四分位数 (Q_3) 构成。
  - **四分位距 (IQR)**: $IQR = Q_3 - Q_1$。

- ○ **须 (Whiskers)**:
  - 上须上限：$\text{med} + 1.5 \times IQR$
  - 下须下限：$\text{med} - 1.5 \times IQR$
- ○ **异常值 (Outliers)**: 位于须之外的数据点。

第二部分：统计模型与参数估计 (Statistical Models & Estimation)

1. 统计模型的基本概念

- **样本**: $x_1, \dots, x_N$ 是从总体 X 中独立同分布 (i.i.d.) 抽取的样本。
- **总体分布**: 假设总体 X 服从某个概率分布，例如正态分布 $X \sim N(\mu, \sigma^2)$ 或指数分布 $X \sim \text{Exp}(\lambda)$。
- **目标**: 通过样本数据估计总体的未知参数 θ（如 μ, σ^2, λ 等）。

2. 统计量与估计量

- **统计量 (Statistic)**: 从样本数据计算出的函数，例如样本均值 $\bar{x}_n$、样本方差 s_n^2。
- **估计量 (Estimator)**: 用于估计总体参数 θ 的统计量，记为 $\hat{\theta}$。
- **例子**:
  - ○ 总体均值 μ 的估计量:

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i。$$

  - ○ 当 $n \to \infty$ 时，$\bar{x}_n \to \mu$，即样本均值是总体均值的一致估计量。

---

3. 简单频率与直方图

3.1 简单频率

- **定义**: 事件 $x_i = a$ 发生的频率。
- **公式**:

$$\frac{\#\{i: x_i = a\}}{n}$$

- **性质**: 当 $n \to \infty$ 时，该频率收敛于概率 $P_X(a)$，即它是概率的估计量。

3.2 直方图 (Histograms)

- **用途**: 用于估计总体的概率密度函数 $f_X(x)$。
- **方法**: 将数据分组并绘制频数分布图。

---

4. 样本方差的推导

样本方差 S_n^2 是总体方差 σ^2 的估计量。

- **公式**:

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n} (x_i - \bar{x}_n)^2$$

- **推导过程**:

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^{n} (x_i^2 - 2x_i\bar{x}_n + \bar{x}_n^2) \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n\sum_{i=1}^{n} x_i + n\bar{x}_n^2\right] \\ &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\bar{x}_n^2\right] \end{aligned}$$

- **期望值**: $E[S_n^2] = \sigma^2$，即样本方差是总体方差的无偏估计量。

---

5. 经验分位数 (Empirical Quantiles)

经验分位数是基于样本数据计算的分位数。

- **步骤**:
  a. 对样本数据进行排序: $x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}$。
  b. 计算索引 $k = \lfloor (n+1)p \rfloor$。
  c. 使用插值公式计算分位数:

$$Q(p) = x_{(k)} + [(n+1)p - k] \cdot \left(x_{(k+1)} - x_{(k)}\right)$$

第三部分：估计量的优良性 (Quality of Estimators)

1. 估计量的期望性质

- **无偏性 (Unbiasedness)**: 估计量 $\hat{\theta}_n$ 是参数 θ 的无偏估计量，如果 $E[\hat{\theta}_n] = \theta$ 对所有 n 成立。
- **偏差 (Bias)**: 偏差定义为 $E[\hat{\theta}_n] - \theta$。无偏估计量的偏差为 0。
- **有效性 (Efficiency)**: 在所有无偏估计量中，方差最小的估计量是最有效的。
- **均方误差 (MSE)**:

  $$MSE = \text{Var}(\hat{\theta}_n) + \text{Bias}^2(\hat{\theta}_n)。$$

  低 MSE 表示估计量性能良好。

2. 应用实例

- **样本均值的无偏性**: 如果 $x_i \sim X$ 且 $E[X] = \mu$，则样本均值 $\bar{x}_n$ 是 μ 的无偏估计量，因为 $E[\bar{x}_n] = \mu$。
- **样本方差的无偏性**: 样本方差 S_n^2 是总体方差 σ^2 的无偏估计量。

第四部分：样本方差的无偏估计推导

(一)：朴素估计量 (Naive Choice) 的定义与期望

我们首先考察一个直观但有偏的样本方差估计量。

- **定义 (朴素估计量):**

  $$T_N^2 = \frac{1}{N}\sum_{i=1}^{N} (X_i - \overline{X}_N)^2$$

  其中，$\overline{X}_N = \frac{1}{N}\sum_{i=1}^{N} X_i$ 是样本均值。

- **目标:** 计算该估计量的期望值 $E(T_N^2)$，并判断其是否为总体方差 σ^2 的无偏估计。

- **期望的展开:**

$$E(T_N^2) = \frac{1}{N}\sum_{i=1}^{N} E\left[(X_i - \overline{X}_N)^2\right]$$

(二)：关键观察与方差分解

推导的核心在于对项 $E\left[(X_i - \overline{X}_N)^2\right]$ 进行处理。

- **观察 1 (Observation 1):**

$$E(X_i - \overline{X}_N) = 0$$

  说明: 样本均值 $\overline{X}_N$ 是总体均值 μ
  的无偏估计，因此每个观测值与样本均值之差的期望为零。

- **推论:**

$$E\left[(X_i - \overline{X}_N)^2\right] = Var(X_i - \overline{X}_N)$$

  依据: 对于任意随机变量 Y，若 $E(Y) = 0$，则 $E(Y^2) = Var(Y)$。

- **方差的进一步展开:**

$$Var(X_i - \overline{X}_N) = Var\left(X_i - \frac{X_1 + X_2 + \cdots + X_N}{N}\right)$$

  将上式重写为：

$$= Var\left[\left(1 - \frac{1}{N}\right)X_i - \sum_{j\neq i}\frac{1}{N}X_j\right]$$

(三)：利用独立性与方差性质进行计算

假设样本 $X_1, X_2, \dots, X_N$ 来自独立同分布（i.i.d.）的总体，且 $Var(X_i) = \sigma^2$。

- **应用方差的线性性质:**

$$= Var\left[\left(1 - \frac{1}{N}\right)X_i\right] + \sum_{j\neq i} Var\left(-\frac{1}{N}X_j\right)$$

注： 由于 X_i 与 X_j (当 $j \neq i$)
相互独立，交叉协方差项为零，因此方差可以直接相加。

- **代入方差公式:**

$$= \left(1 - \frac{1}{N}\right)^2 Var(X_i) + \sum_{j \neq i} \left(\frac{1}{N}\right)^2 Var(X_j)$$

- **代入 σ^2:**

$$= \left(1 - \frac{1}{N}\right)^2 \sigma^2 + (N-1) \cdot \frac{1}{N^2} \sigma^2$$

注： 求和项 $\sum_{j \neq i}$ 包含 $N-1$ 个元素。

- **化简表达式:**

$$= \sigma^2 \left[\left(1 - \frac{1}{N}\right)^2 + \frac{N-1}{N^2}\right]$$

$$= \sigma^2 \left[\frac{(N-1)^2}{N^2} + \frac{N-1}{N^2}\right]$$

$$= \sigma^2 \left[\frac{(N-1)(N-1+1)}{N^2}\right]$$

$$= \sigma^2 \left[\frac{(N-1)N}{N^2}\right]$$

$$= \sigma^2 \left(\frac{N-1}{N}\right)$$

(四)：结论——朴素估计量存在负偏差

将上述结果代回 $E(T_N^2)$ 的表达式：

$$E(T_N^2) = \frac{1}{N} \sum_{i=1}^{N} E\left[(X_i - \overline{X}_N)^2\right] = \frac{1}{N} \sum_{i=1}^{N} \sigma^2 \left(\frac{N-1}{N}\right)$$

由于求和项与 i 无关，因此：

$$E(T_N^2) = \frac{1}{N} \cdot N \cdot \sigma^2 \left(\frac{N-1}{N}\right) = \sigma^2 \cdot \frac{N-1}{N}$$

最终结论:

$$E(T_N^2) = \sigma^2 \cdot \frac{N-1}{N} \text{ 这是一个“负偏差” (negative bias)}$$

这意味着，使用分母为 N 的朴素估计量 T_N^2，其期望值小于真实的总体方差 σ^2。它系统性地低估了总体方差。

(五)：无偏估计量的修正

为了消除这个负偏差，我们需要对朴素估计量进行修正。

- **修正方法:**
将朴素估计量乘以一个校正因子 $\frac{N}{N-1}$。
- **定义 (无偏样本方差):**

$$S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X}_N)^2$$

- **验证无偏性:**

$$E(S_N^2) = E\left(\frac{N}{N-1} \cdot T_N^2\right) = \frac{N}{N-1} \cdot E(T_N^2) = \frac{N}{N-1} \cdot \sigma^2 \cdot \frac{N-1}{N} = \sigma^2$$

最终结论:

$$S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N} (X_i - \overline{X}_N)^2 \text{ 是总体方差 } \sigma^2 \text{ 的无偏估计量。}$$

📘 **详细解释（中英对照）**

1. 什么是经验分布函数**（Empirical Distribution Function, EDF）？**

- 定义： 给定一组数据 $X_1, X_2, \ldots, X_n$ ，经验分布函数定义为：

$$F_n(t) = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{\{X_i \le t\}}$$

  即：小于等于 t 的数据点所占的比例。

- 性质：
  它是非减、右连续的阶梯函数，从 0 跳到 1，总是累积的。

✅ 所以，EDF 本质上就是样本的累积分布函数（Cumulative Distribution Function of the sample）。

统计学笔记：探索性数据分析与描述性统计

1. 统计学基础概念

- **样本 (Sample)**: 一组独立同分布 (i.i.d.) 的观测值，记为 $X_1, X_2, \dots, X_n$。
- **统计量 (Statistic)**: 样本的函数，即
  $$T = T(X_1, X_2, \dots, X_n)$$
  统计量本身是一个随机变量，因为它是随机样本的函数。
- **目标 (Goal)**: 理解、描述和探索数据；通过图形和数值摘要来呈现数据的核心特征。
- **估计量 (Estimator)**: 用于推断总体未知参数（如均值 μ、方差 σ^2）的统计量。例如，样本均值 $\bar{X}_n$ 是总体均值 μ 的一个估计量。

2. 描述性统计 (Descriptive Statistics)

描述性统计旨在用简洁的方式概括数据集的主要特征，分为两大类：

2.1 图形摘要 (Graphical Summaries)

- **适用对象**: 单变量数据集 (Univariate) 和双变量数据集 (Bivariate)。
- **单变量图形**:
  - **直方图 (Histogram)**:
    - **目的**: 可视化数据的分布形态（对称性、偏度、峰度、单峰/多峰）。
    - **构建步骤**:
      A. 将数据范围划分为若干个互不重叠的区间，称为“箱”（Bins），记为 $B_1, B_2, \dots, B_m$。
      B. 计算每个箱 B_i 的宽度 $\|B_i\|$（通常所有箱宽度相等）。
      C. 计算落入每个箱 B_i 的观测值数量（频数）。
      D. **标准化直方图高度**:
      为了使直方图总面积为1（从而可视为概率密度函数的估计），箱 B_i 上的高度为：
      $$\text{Height}(B_i) = \frac{\text{Number of } X_j \text{ in } B_i}{n \cdot \|B_i\|}$$
    - **箱宽选择**:
      A. **过小**: 图形杂乱，出现许多孤立的峰值，噪声过大。
      B. **过大**: 图形过于平滑，丢失数据的细节和特征（如多峰性）。

C. **经验法则**:

- 箱数 $m = 1 + 3.3\log_{10}(n)$
- 箱宽 $b = 3.49 \cdot s \cdot n^{-1/3}$，其中 s 为样本标准差。

- **分布形态术语**:

A. 对称 (Symmetric)

B. 偏斜 (Skewed)

C. 单峰 (Unimodal)

D. 双峰 (Bimodal)，如 Old Faithful 间歇泉数据。

E. 多峰 (Multimodal)

- **经验分布函数 (Empirical Distribution Function, EDF)**:
  - **记为 $F_n(x)$，定义为小于或等于 x 的观测值的比例。**
  - $F_n(x)$ 是一个阶梯函数，在每个观测点处跳跃。
  - **与直方图的关系**: 直方图在某个箱上的面积等于 EDF 在该箱上的增量。

- **双变量图形**:
  - **散点图 (Scatter Plot)**:
    - **目的**: 展示两个变量之间的关系。
    - **构建**: 将每个观测对 (x_i, y_i) 在二维平面上标出。
    - **应用**:
      可用于探索变量间的相关性、模式（线性/非线性），并为后续建模（如线性回归）提供直观依据。

2.2 数值摘要 (Numerical Summaries)

数值摘要通过计算关键指标来量化数据的中心、离散度和形状。

- **中心趋势度量 (Measures of Central Tendency)**:
  - **样本均值 (Sample Mean)**:

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

    - 优点: 具有良好的数学性质。
    - 缺点: 对异常值敏感。
  - **样本中位数 (Sample Median)**:
    - 将数据从小到大排序后，位于中间位置的值。
    - **计算公式**:

A. 令 $k = \lfloor (n+1)/2 \rfloor$，其中 $\lfloor \cdot \rfloor$ 表示向下取整。
B. 若 n 为奇数，$\text{Med}_n = X_{(k)}$。
C. 若 n 为偶数，$\text{Med}_n = \frac{1}{2}(X_{(k)} + X_{(k+1)})$。

- 优点: 对异常值稳健 (Robust)。
- 缺点: 数学处理上不如均值方便。

- **众数 (Mode)**:
  - 数据集中出现频率最高的值。
  - 优点: 最容易理解，对异常值稳健。
  - 缺点: 可能不唯一（多峰），且提供的信息量有限。

- **离散度度量 (Measures of Dispersion)**:
  - **样本方差 (Sample Variance)**:
    - **有偏估计 (Naive)**:

$$T = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

其期望 $E[T] = \frac{n-1}{n}\sigma^2$，存在负偏差。

    - **无偏估计 (Unbiased)**:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X}_n)^2$$

其期望 $E[S^2] = \sigma^2$，是总体方差 σ^2 的无偏估计量。

  - **绝对中位差 (Median Absolute Deviation, MAD)**:
    - **计算步骤**:

A. 计算样本中位数 Med_n。
B. 计算每个观测值与中位数的绝对偏差：$|X_i - \text{Med}_n|$。
C. 对这些绝对偏差求中位数：

$$\text{MAD} = \text{Med}(|X_1 - \text{Med}_n|, \dots, |X_n - \text{Med}_n|)$$

    - 优点: 对异常值高度稳健。
    - 缺点: 数学处理不便。

- **分位数与箱线图 (Quantiles and Box Plots)**:
  - **p-分位数 (p-th Quantile)**: $q(p)$ 是一个值，使得大约有 $p \times 100\%$ 的数据小于或等于它。
  - **四分位数 (Quartiles)**:
    - **第一四分位数 (下四分位数)**: $Q_1 = q(0.25)$

        - **第二四分位数 (中位数)**: $Q_2 = q(0.5) = \text{Med}_n$
        - **第三四分位数 (上四分位数)**: $Q_3 = q(0.75)$
        - **计算公式**（线性插值法）:
            A. 令 $k = \lfloor (n+1)p \rfloor$。
            B. $q(p) = X_{(k)} + ((n+1)p - k)(X_{(k+1)} - X_{(k)})$
    - **四分位距 (Interquartile Range, IQR)**:$\text{IQR} = Q_3 - Q_1$IQR 是衡量数据中间50%离散度的稳健指标。
    - **箱线图 (Box-and-Whisker Plot / Box Plot)**:
        - **构成**:
            A. **箱体 (Box)**: 从 Q_1 到 Q_3，中间的线是中位数 Q_2。箱体高度即为 IQR。
            B. **须 (Whiskers)**: 从箱体延伸出的线。
            C. **异常值 (Outliers)**: 落在须范围之外的点，通常用圆圈标记。
            D. **异常值（Outliers）判断：[Med – 1.5IQR, Med + 1.5IQR]外**
        - **用途**: 清晰展示数据的中心、离散度、偏度和异常值，特别适合**比较多个数据集**。

3. 统计模型与估计量的性质

- **统计模型 (Statistical Model)**: 对数据生成机制的假设。通常假设数据来自某个带有未知参数（如 μ, σ^2, λ）的概率分布（如 $X_i \sim N(\mu, \sigma^2)$）。
- **估计量的评价标准**:
    - **无偏性 (Unbiasedness)**: 一个估计量 T_n 是参数 θ 的无偏估计量，如果其期望等于真值：$E[T_n] = \theta$偏差 (Bias) 定义为 $E[T_n] - \theta$。
    - **效率 (Efficiency)**: 在无偏估计量中，方差更小的估计量更有效。
    - **均方误差 (Mean Squared Error, MSE)**: 综合考虑偏差和方差的指标，$MSE = \text{Var}(T_n) + (\text{Bias})^2$。MSE 越小，估计量越好。
    - **注**: 有偏估计量并非总是坏的，有时其 MSE 可能更小（如岭回归）。

📘 统计推断基础笔记

文档日期：2025 年 11 月 26 日
主题：点估计与置信区间（Point Estimation & Confidence Intervals）

一、点估计（Point Estimation）

1. 定义

- 设总体分布依赖于未知参数 θ。
- 给定样本 $X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} X$，统计量 $T = T(X_1, \dots, X_n)$ 称为 θ **的估计量（estimator）**。
- 其具体观测值 $t = T(x_1, \dots, x_n)$ 为 **估计值（estimate）**。

---

2. 无偏性（Unbiasedness）

> 定义：若 $E[T] = \theta$，则称 T 是 θ 的 **无偏估计量（unbiased estimator）**。

✅ **示例：**

- 样本均值 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是总体均值 $\mu = E[X]$ 的无偏估计：$E[\bar{X}] = \mu$
- 样本方差 $S^2 = \frac{1}{n-1}\sum_{i=1}^{n} (X_i - \bar{X})^2$ 是 $\sigma^2 = \mathrm{Var}(X)$ 的无偏估计（见上节课）。

⚠️ **注意：**

- **中位数估计**：若 X 的分布关于均值对称，则 $\mathrm{Med}(X_1, \dots, X_n)$ 是 $\mathrm{Med}(X)$ 的无偏估计；否则可能有偏。

❗ **非线性变换破坏无偏性：**

- 若 T 是 θ 的无偏估计，一般 $g(T)$ **不是** $g(\theta)$ **的无偏估计**，除非 g 为线性函数：$g(t) = at + b \Rightarrow E[g(T)] = a\theta + b = g(\theta)$

---

3. 有效性（Efficiency）

> 定义：设 T_1, T_2 均为 θ 的无偏估计。若
>
> $$\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$$
>
> 则称 T_1 比 T_2 **更有效（more efficient）**。

📌 **示例：合金熔点估计**

- 实验室 A 提供 20 次测量，均值 $\bar{X}_A$
- 实验室 B 提供 30 次测量，均值 $\bar{X}_B$
- 假设两组测量独立同分布：$E[X] = m$, $\text{Var}(X) = \sigma^2$

两种合并策略：

1. **简单平均：**$T_1 = \frac{\bar{X}_A + \bar{X}_B}{2}$
   - 无偏
   - 方差：$\text{Var}(T_1) = \frac{1}{4}\left(\frac{\sigma^2}{20} + \frac{\sigma^2}{30}\right) = \frac{\sigma^2}{4} \cdot \frac{5}{60} = \frac{\sigma^2}{48}$
2. **加权平均（按样本量）：**

$$T_2 = \frac{20\bar{X}_A + 30\bar{X}_B}{50} = \frac{1}{50}\sum_{i=1}^{50} X_i = \bar{X}_{\text{pooled}}$$

   - 无偏
   - 方差：$\text{Var}(T_2) = \frac{\sigma^2}{50}$

✅ **比较：**

$$\frac{1}{50} < \frac{1}{48} \Rightarrow T_2 \text{ 更有效}$$

---

4. 最小方差无偏估计（MVUE - Minimum Variance Unbiased Estimator）

> 重要结论（无证明）：

- 若 $X_i \sim \mathcal{N}(\mu, \sigma^2)$，则 $\bar{X}$ 是 μ 的 **最小方差无偏估计（MVUE）**。
- 若 $X_i \sim \text{Exp}(\lambda)$，则 $\bar{X}$ 也是 $\mu = 1/\lambda$ 的 MVUE。

---

5. 指数分布下的估计比较

设 $X_1, \dots, X_n \overset{\text{i.i.d.}}{\sim} \text{Exp}(\lambda)$，目标参数 $\mu = E[X] = 1/\lambda$。

两个估计量：

- $T_1 = \bar{X} = \frac{1}{n}\sum X_i$

$$E[T_1] = \frac{1}{\lambda}, \text{Var}(T_1) = \frac{1}{n\lambda^2}$$

- $T_2 = n \cdot \min(X_1, \dots, X_n) = nM_n$

分析 $M_n = \min(X_1, \dots, X_n)$：

- 累积分布：

$$P(M_n > t) = P(X_1 > t, \dots, X_n > t) = e^{-n\lambda t} \Rightarrow M_n \sim \text{Exp}(n\lambda)$$

- 期望：$E[M_n] = \frac{1}{n\lambda} \Rightarrow E[T_2] = n \cdot \frac{1}{n\lambda} = \frac{1}{\lambda}$
  → **无偏**

- 方差：

$$\text{Var}(M_n) = \frac{1}{(n\lambda)^2} \Rightarrow \text{Var}(T_2) = n^2 \cdot \frac{1}{(n\lambda)^2} = \frac{1}{\lambda^2}$$

✅ **比较有效性：**

$$\text{Var}(T_1) = \frac{1}{n\lambda^2} < \frac{1}{\lambda^2} = \text{Var}(T_2) \ (\text{当}\ n > 1)$$

→ $\bar{X}$ **更有效**

二、均方误差（Mean Squared Error, MSE）

> 定义：对估计量 T 估计 θ，
>
> $$\text{MSE}(T) = E\left[(T - \theta)^2\right]$$

分解公式：

$$\text{MSE}(T) = \text{Var}(T) + \left[\text{Bias}(T)\right]^2$$

其中 $\text{Bias}(T) = E[T] - \theta$

- 若 T 无偏 $\Rightarrow \text{MSE}(T) = \text{Var}(T)$
- **评价标准：** MSE 越小，估计越优（兼顾偏差与方差）

📌 **示例：**

- 对正态分布 $X_i \sim \mathcal{N}(\mu, \sigma^2)$，
  用 $S_n^2 = \frac{1}{n}\sum (X_i - \bar{X})^2$（有偏） vs $S^2 = \frac{1}{n-1}\sum (X_i - \bar{X})^2$（无偏）

- 虽然 S^2 无偏，但有时 $\text{MSE}(S_n^2) < \text{MSE}(S^2)$（尤其在小样本时）

三、置信区间（Confidence Intervals, CI）

1. 动机

- 点估计提供单一值，但无法反映 **不确定性**。
- 置信区间提供一个 **包含参数真值的随机区间**，具有预设覆盖概率。

  定义：设 θ 为未知参数。若统计量 L_n, U_n 满足

  $$P(L_n \leq \theta \leq U_n) = 1 - \alpha$$

  则 $[L_n, U_n]$ 称为 $100(1-\alpha)\%$ **置信区间**。

2. 均值 μ 的置信区间（σ^2 已知，正态总体）

设 $X_i \sim \mathcal{N}(\mu, \sigma^2)$，$\sigma^2$ 已知。

- 样本均值：$\bar{X} \sim \mathcal{N}(\mu, \sigma^2/n)$
- 标准化：

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$$

- 对给定置信水平 $1-\alpha$，查标准正态分位数 $z_{\alpha/2}$，使得

$$P\left(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\alpha/2}\right) = 1 - \alpha$$

- 解出 μ 得：

$$P\left(\bar{X} - z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

  ☑ $100(1-\alpha)\%$ 置信区间为：

$$\left[\bar{X} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right]$$

📌 应用示例（比例/均值估计）

- 要求估计误差不超过 0.01，置信度 99%（即 $\alpha = 0.01$）
- 利用：$P\left(|\bar{X} - \mu| \leq 0.01\right) \geq 0.99 \Rightarrow 0.01 \geq z_{0.005} \cdot \frac{\sigma}{\sqrt{n}} \Rightarrow n \geq \left(\frac{z_{0.005} \cdot \sigma}{0.01}\right)^2$
- 若 σ 未知，可先用 pilot study 估计。

四、关键概念对比总结

概念	定义	公式	说明
无偏性	期望等于真值	$E[T] = \theta$	避免系统性偏差
有效性	方差更小	$\mathrm{Var}(T_1) < \mathrm{Var}(T_2)$	精度更高
MSE	误差平方期望	$\mathrm{MSE} = \mathrm{Var} + \mathrm{Bias}^2$	综合评价标准
置信区间	包含真值的概率区间	$P(L¡ \leq \theta \leq U) = 1 - \alpha$	反映估计不确定性

置信区间（Confidence Intervals）系统化笔记

> 基于正态分布与 t 分布的均值估计

一、核心思想（Idea）

我们要对**总体均值** μ 进行估计，方法是基于样本构造一个**置信区间（Confidence Interval, C.I.）**。

- **中心化与标准化：**

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

  当 σ 已知且数据正态时，使用标准正态分布构造区间。

- **估计量选择：**使用样本均值 $\bar{X}_n$ 作为 μ 的点估计。

二、置信区间的一般形式

对于置信水平 $100(1-\alpha)\%$，置信区间为：

$$\left[\bar{X}_n - c \cdot \frac{\sigma}{\sqrt{n}}, \bar{X}_n + c \cdot \frac{\sigma}{\sqrt{n}}\right]$$

其中 c 由分布和置信水平决定。

三、三种情形分类

☑ 情形 1：双侧对称置信区间（Two-sided symmetric C.I.）

- **要求：**

$$P\left(\left|\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}\right| \le c\right) = 1 - \alpha$$

- **临界值定义：**
  设 $z_{\alpha/2}$ 满足 $P(Z \ge z_{\alpha/2}) = \alpha/2$，则：

$$c = z_{\alpha/2}$$

- **置信区间：**

$$\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

- **例**：90% C.I.，$\alpha = 0.1 \Rightarrow z_{0.05} = 1.645$

☑ 情形 2：单侧右侧置信区间（One-sided right C.I.）

- 目标：$P(\mu \geq L) = 1 - \alpha$
- **构造**：

$$P\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq z_\alpha\right) = 1 - \alpha \Rightarrow \mu \geq \bar{X}_n - z_\alpha\frac{\sigma}{\sqrt{n}}$$

- **置信下限**（95% 右侧 C.I.）：

$$[\bar{X}_n - z_{0.05}\frac{\sigma}{\sqrt{n}},\ \infty)$$

- **实例（煤的热值）**：
  - $\bar{x} = 23.788,\ \sigma = 0.1,\ n = 23$
  - 95% 右侧 C.I.:

$$[23.788 - 1.645 \cdot \frac{0.1}{\sqrt{23}},\ \infty) \approx [23.755,\ \infty)$$

☑ 情形 3：单侧左侧置信区间（One-sided left C.I.）

- 目标：$P(\mu \leq U) = 1 - \alpha$
- **构造**：

$$\mu \leq \bar{X}_n + z_\alpha\frac{\sigma}{\sqrt{n}}$$

- **置信上限**（95% 左侧 C.I.）：

$$(-\infty,\ \bar{X}_n + z_{0.05}\frac{\sigma}{\sqrt{n}}] \approx (-\infty,\ 23.822]$$

四、σ 未知的情形：t 分布（Student’s t）

4.1 历史背景

- **William S. Gosset（1908）** 提出 t 分布（笔名 “Student”）。
- 适用于**小样本、σ 未知，但总体正态**的场景。

4.2 t 分布定义

若 $X_1,\ldots,X_n \overset{\text{iid}}{\sim} N(\mu,\sigma^2)$，则：

$$T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$$

其中 S_n 是样本标准差。

4.3 t 分布的性质

- 概率密度函数（略，具对称性、厚尾）
- 自由度 $m = n - 1$
- 当 $m \to \infty$，$t(m) \to N(0,1)$

4.4 置信区间（σ 未知）

- **双侧 100(1−α)% C.I.:**

$$\left(\bar{X}_n - t_{n-1,\alpha/2} \cdot \frac{S_n}{\sqrt{n}},\ \bar{X}_n + t_{n-1,\alpha/2} \cdot \frac{S_n}{\sqrt{n}}\right)$$

- **实例（煤热值，σ 未知）：**
  - $\bar{x} = 23.788,\ s = 0.0555,\ n = 23,\ df = 22$
  - 查表：$t_{22,0.025} = 2.074$
  - 得 95% C.I.:

$$\left(23.788 \pm 2.074 \cdot \frac{0.0555}{\sqrt{23}}\right) \approx (23.754,\ 23.822)$$

> 注：与 σ 已知下的区间 $(23.747, 23.829)$ 相比，略宽，因估计了 σ。

五、关键图表说明

图 23.1：50 个 90% 置信区间（模拟）

- 每个区间基于独立生成的 $N(0,1)$ 样本（n=20）
- 约 90% 的区间包含真值 $\mu = 0$
- 直观展示“置信水平”的频率解释

图 23.3：t 分布 vs 标准正态

- t(1)、t(2)、t(5) 与 $N(0,1)$（虚线）对比
- 随自由度增加，t 分布趋近正态
- 低自由度时尾部更厚，反映更多不确定性

六、临界值表（Table B.2 摘要）

df (m)	$t_{m,0.1}$	$t_{m,0.05}$	$t_{m,0.025}$	$t_{m,0.01}$
1	3.078	6.314	12.706	31.821
10	1.372	1.812	2.228	2.764
22	1.321	1.717	2.074	2.508
∞	1.282	1.645	**1.960**	**2.326**

> 最后一行即标准正态分位数 $z_p = t_{\infty,p}$

七、关键公式速查

场景	分布	置信区间
σ 已知，双侧	$Z \sim N(0,1)$	$\bar{x} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
σ 未知，双侧	$T \sim t(n-1)$	$\bar{x} \pm t_{n-1,\alpha/2}\dfrac{s}{\sqrt{n}}$
σ 已知，右侧	Z	$[\bar{x} - z_\alpha\dfrac{\sigma}{\sqrt{n}},\ \infty)$
σ 已知，左侧	Z	$(-\infty,\ \bar{x} + z_\alpha\dfrac{\sigma}{\sqrt{n}}]$

八、实际应用案例：煤的发热量（Gross Calorific Value）

- **背景**：ISO 1928 标准，测量误差 $\sim N(\mu,\sigma^2)$，$\sigma = 0.1$ MJ/kg
- **样本**：n = 23 次测量，$\bar{x} = 23.788$
- **95% 双侧 C.I.（σ 已知）**：

$$\left(23.788 \pm 1.96 \cdot \frac{0.1}{\sqrt{23}}\right) = (23.747,\ 23.829)$$

- **95% 双侧 C.I.（σ 未知）**：

$$(23.754,\ 23.822)$$

> 说明：当 σ 准确已知（如 ISO 认证方法），可使用 Z 区间，更精确。

九、总结与要点

1. **置信水平 ≠ 概率**：不是“μ 落入某区间的概率”，而是“该方法在重复抽样中覆盖真值的频率”。
2. **σ 是否已知**决定使用 Z 还是 t。
3. **t 分布更保守**（区间更宽），尤其在小样本下。
4. **单侧区间**适用于质量控制、下限保证等场景（如“热值至少为 X”）。
5. **图表辅助理解**：模拟图展示覆盖频率，密度图展示分布差异。

单侧置信区间（One-sided Confidence Intervals）

> 基于正态分布（σ 已知）的右侧与左侧区间构造
> —— 严格推导、结构清晰、例证完整

一、问题设定与核心思想

- **总体**：$X_1, \dots, X_n \overset{\text{iid}}{\sim} N(\mu, \sigma^2)$，其中 σ **已知**（如 ISO 1928 标准中 $\sigma = 0.1$）。
- **目标**：对固定但未知的总体均值 μ，构造 **单侧置信区间**（仅关注上界或下界）。
- **置信水平**：$100(1-\alpha)\%$，如 $95\% \to \alpha = 0.05$。
- **关键统计量**：

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

> ⚠ 注：单侧区间适用于**单向质量控制场景**，例如"热值**至少**为某值"（右侧）或"污染物**不超过**某限"（左侧）。

二、右侧置信区间（Right-sided C.I.）

> 目标：构造下限 L，使得 $P(\mu \geq L) = 1 - \alpha$

2.1 推导步骤

1. **设定目标概率**：

$$P(\mu \geq \bar{X}_n - \varepsilon) = 1 - \alpha$$

2. **移项变形**：

$$\mu \geq \bar{X}_n - \varepsilon \quad \Leftrightarrow \quad \bar{X}_n - \mu \leq \varepsilon$$

3. **标准化**：

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \leq \frac{\varepsilon\sqrt{n}}{\sigma} \quad \Rightarrow \quad P\left(Z \leq \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 1 - \alpha$$

4. **求解临界值**：
   定义 z_α 满足 $P(Z \geq z_\alpha) = \alpha \Rightarrow P(Z \leq z_\alpha) = 1 - \alpha$
   所以：

$$\frac{\varepsilon\sqrt{n}}{\sigma} = z_\alpha \quad \Rightarrow \quad \varepsilon = z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

5. **得到置信下限：**

$$L = \bar{X}_n - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

2.2 最终形式

$$\boxed{\left[\bar{X}_n - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty\right)}$$

2.3 实例：煤的热值（PDF p.347–348）

- $\bar{x} = 23.788$
- $\sigma = 0.1$
- $n = 23$
- $\alpha = 0.05 \Rightarrow z_{0.05} = 1.645$

计算：

$$L = 23.788 - 1.645 \cdot \frac{0.1}{\sqrt{23}} \approx 23.788 - 0.033 = 23.755$$

95% 右侧置信区间：

$$[23.755, \infty)$$

> ☑ 解读：有 **95%** 的置信度认为煤的真实热值至少为 **23.755 MJ/kg**。

2.4 与双侧区间对比

- 双侧 95% C.I.: (23.747, 23.829)
- 右侧 95% C.I.: [23.755, ∞)
- 右侧区间**下限更高**，因全部误差容限用于保障“下界”。

三、左侧置信区间（Left-sided C.I.）

> 目标：构造上限 U，使得 $P(\mu \leq U) = 1 - \alpha$

3.1 推导步骤

1. **设定目标概率：**

$$P(\mu \leq \bar{X}_n + \varepsilon) = 1 - \alpha$$

2. **移项变形：**

$$\mu \leq \bar{X}_n + \varepsilon \quad \Leftrightarrow \quad \bar{X}_n - \mu \geq -\varepsilon$$

3. **标准化：**

$$\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \geq -\frac{\varepsilon\sqrt{n}}{\sigma} \quad \Rightarrow \quad P\left(Z \geq -\frac{\varepsilon\sqrt{n}}{\sigma}\right)=1-\alpha$$

4. **利用对称性：**

$$P(Z \geq -a)=P(Z \leq a)=\Phi(a)$$

所以：

$$\Phi\left(\frac{\varepsilon\sqrt{n}}{\sigma}\right)=1-\alpha \quad \Rightarrow \quad \frac{\varepsilon\sqrt{n}}{\sigma}=z_\alpha$$

5. **解出** ε**：**

$$\varepsilon=z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

6. **得到置信上限：**

$$U=\bar{X}_n+z_\alpha \cdot \frac{\sigma}{\sqrt{n}}$$

3.2 最终形式

$$\boxed{\left(-\infty, \bar{X}_n+z_\alpha \cdot \frac{\sigma}{\sqrt{n}}\right]}$$

3.3 实例：煤的热值（PDF p.348）

- 同上参数
- $U=23.788+1.645 \cdot \frac{0.1}{\sqrt{23}} \approx 23.788+0.034=23.822$

95% 左侧置信区间：

$$(-\infty, 23.822]$$

> ☑ 解读：有 **95%** 的置信度认为煤的真实热值不超过 **23.822 MJ/kg**。

3.4 与双侧区间对比

- 双侧上限：23.829
- 左侧上限：23.822（**略低**，因全部 α 用于控制上界）

四、关键说明与 PDF 内容对应

项目	PDF 位置	内容
临界值定义	p.344	“DEF: THE CRITICAL VALUE z_p IS THE NUMBER S.T. $P(Z \geq z_p) = p$”
右侧区间形式	p.348	“95% RIGHT C.I.: [23.755, ∞)”
左侧区间形式	p.348	“95% LEFT C.I.: (−∞, 23.822]”
双侧对比	p.348	“2-SIDED: [23.747, 23.829]”
σ 已知假设	p.347	“σ = 0.1 MJ/kg (known from ISO procedures)”

> ☑ 所有推导均基于 PDF 中“**Case 2: 1-SIDED RIGHT**”与“**Case 3: LEFT-SIDE C.I.**”的逻辑，补全了 PDF 中省略的代数与概率等价步骤。

五、总结表：单侧 vs 双侧置信区间

类型	区间形式	临界值	应用场景
右侧（下限保障）	$[\bar{x} - z_\alpha \cdot \frac{\sigma}{\sqrt{n}}, \infty)$	z_α	“至少为...”（如最低热值、最低强度）
左侧（上限控制）	$(-\infty, \bar{x} + z_\alpha \cdot \frac{\sigma}{\sqrt{n}}]$	z_α	“不超过...”（如最大污染、最大误差）
双侧（双向估计）	$\left(\bar{x} \pm z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}\right)$	$z_{\alpha/2}$	一般参数估计

> 🔔 注意：单侧区间使用 z_α，双侧使用 $z_{\alpha/2}$，因为**全部显著性水平** α **放在一侧**。

六、附：标准正态临界值表（PDF Table B.2 摘要）

α	z_α	$z_{\alpha/2}$（双侧）
0.10	1.282	1.645
0.05	**1.645**	**1.960**
0.01	2.326	2.576

> 用于快速查表构造区间。

🔖 t 分布（Student’s t-distribution）终极扫盲笔记

> —— 专治“σ 未知”、“样本小”、“公式看不懂”的一切焦虑

一、核心问题：为什么需要 t 分布？

☑ 答案：因为 σ 通常是未知的！

- 在现实世界中，**很少有人知道总体标准差 σ**。
- 比如煤的热值，ISO 标准告诉你 σ = 0.1，这是特例（实验室认证后给出的）。
- 大多数情况下，你只有数据（样本），不知道 σ，只能用**样本标准差** S_n 来估计它。

> 所以：**t 分布就是为了解决“σ 未知”这个现实问题而生的！**

二、关键概念逐个击破

2.1 S_n 是什么？

- S_n 就是**样本标准差**（Sample Standard Deviation）。
- 计算公式：

$$S_n = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2}$$

- 它是 σ 的**无偏估计量**，用来代替未知的 σ。

> 💡 为什么除以 n-1 而不是 n？这是为了“无偏估计”，细节可以后续展开，现在记住它是“样本标准差”就行。

2.2 $T = \frac{\bar{X}_n - \mu}{S_n/\sqrt{n}} \sim t(n-1)$ 是什么意思？

- 这是 t 分布的**定义式**。
- 它说：当你把样本均值标准化时，如果分母用的是**样本标准差** S_n（而不是真 σ），那么这个统计量不再服从标准正态分布 N(0,1)，而是服从一个叫 **t(n-1)** 的新分布。

为什么叫 t(n-1)？

- **t**：代表 “Student’s t”，纪念提出者 William Sealy Gosset（笔名 “Student”）。
- **(n-1)**：这是**自由度（degrees of freedom）**，记作 df 或 m。

- 自由度 = 样本量 - 1
- 为什么减 1？因为你计算 S_n 时用了 $\bar{X}_n$，这“消耗”了一个自由度。
- 简单记忆：**样本量 n → 自由度 m = n-1**

> ☑ 所以，**t(n-1)** 不是“动态分布”，它是一个固定的概率分布族，每个自由度对应一个具体的分布形状。

2.3 图形理解：t 分布 vs 正态分布

PDF 中的图 23.3 显示：

- t(1), t(2), t(5) 的曲线比 N(0,1)（虚线）更“胖”（厚尾）。
- 随着自由度增加（比如到 30），t 分布越来越接近 N(0,1)。
- 当自由度 → ∞ 时，t(m) → N(0,1)。

> 📌 **直观意义**：因为用 S_n 代替 σ 引入了额外的不确定性，所以 t 分布比正态分布“更保守”（区间更宽），尤其在小样本时。

三、t 分布的临界值：$t_{n-1,\alpha/2}$ 是啥？

3.1 符号解释

- **第一个下标** $n-1$：自由度（df），即样本量减 1。
- **第二个下标** $\alpha/2$：右尾概率（right-tail probability）。
  - 和 Z 分布一样，PDF 中定义：$t_{m,p}$ 满足 $P(T_m \geq t_{m,p}) = p$
  - 所以 $t_{n-1,\alpha/2}$ 就是满足 $P(T_{n-1} \geq t_{n-1,\alpha/2}) = \alpha/2$ 的数。

3.2 查表方法（Table B.2）

- 找到行：自由度 m = n-1
- 找到列：右尾概率 p = α/2
- 交叉点就是 $t_{n-1,\alpha/2}$

例子（煤热值，σ 未知）：

- n = 23 → df = 22
- α = 0.05 → α/2 = 0.025
- 查表 B.2：第 22 行，0.025 列 → $t_{22,0.025} = 2.074$

> ☑ 所以 $t_{22,0.025} = 2.074$ 的意思是：在自由度为 22 的 t 分布中，右边有 2.5% 的面积在 2.074 之外。

四、置信区间怎么来的？（σ 未知版）

4.1 推导步骤（严格复刻 PDF 逻辑）

7. **已知：**

$$T=\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}\sim t(n-1)$$

8. **构造双侧概率：**

$$P\left(-t_{n-1,\alpha/2}<\frac{\bar{X}_n-\mu}{S_n/\sqrt{n}}<t_{n-1,\alpha/2}\right)=1-\alpha$$

9. **代数变形，解出 μ：**

$$\bar{X}_n-t_{n-1,\alpha/2}\cdot\frac{S_n}{\sqrt{n}}<\mu<\bar{X}_n+t_{n-1,\alpha/2}\cdot\frac{S_n}{\sqrt{n}}$$

10. **得到置信区间：**

$$\boxed{\left(\bar{X}_n-t_{n-1,\alpha/2}\cdot\frac{S_n}{\sqrt{n}},\ \bar{X}_n+t_{n-1,\alpha/2}\cdot\frac{S_n}{\sqrt{n}}\right)}$$

4.2 实例：煤热值（σ 未知）

- $\bar{x}=23.788$
- $s=0.0555$ （注意：这里是样本标准差 S_n，不是 σ）
- $n=23$
- $df=22$
- $\alpha=0.05\Rightarrow t_{22,0.025}=2.074$

计算：

$$\text{误差项}=2.074\cdot\frac{0.0555}{\sqrt{23}}\approx 2.074\cdot 0.0116\approx 0.024$$

区间：

$$(23.788-0.024,\ 23.788+0.024)=(23.764,\ 23.812)$$

> ❗ 注意：PDF 中写的是 (23.754, 23.822)，可能是四舍五入或计算精度差异，但方法完全一致。

五、为什么 σ 未知时的区间更宽？

- **σ 已知时**：用精确的 σ，标准误是 $\sigma/\sqrt{n}$
- **σ 未知时**：用估计的 S_n，引入了额外的不确定性 → 标准误变成 $S_n/\sqrt{n}$，且分布变成更“胖”的 t 分布 → 临界值更大（2.074 > 1.96）→ 区间更宽！

> ☑ 结论：**t 分布区间更保守，是对“估计 σ”这一行为的惩罚。**

六、总结：一张表搞定所有疑惑

问题	答案
为什 t 分布？	因为现实中 σ 通常未知，必须用样本标准差 S_n 代替，导致标准化统计量不再服从 N(0,1)。
S_n 是什么？	样本标准差，用于估计未知的 σ。
t(n-1) 是啥？	一个概率分布，自由度为 n-1，当 n 很大时趋近于 N(0,1)。
$t_{n-1,\alpha/2}$ 是啥？	自由度为 n-1 的 t 分布中，右尾概率为 α/2 的临界值。查表 Table B.2。
置信区间怎么来的？	从 $P(-t < T < t) = 1 - \alpha$ 出发，通过代数变形解出 μ 的范围。
为什么区间更宽？	因为用了估计的 S_n 和更“胖”的 t 分布，增加了不确定性。

你现在应该已经明白：**t 分布不是“有病”，而是统计学对现实世界“无知”（不知道 σ）的一种优雅解决方案。**

Probability and statistics

Date. Page.

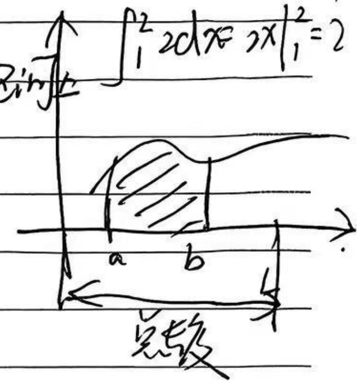
$$\int_1^2 1\,dx = x\Big|_1^2 = 2-1=1$$

连续型均匀分布的概率性质　　均匀分布 $f(x)=\frac{1}{L}$ → 非均匀分布

$$\int_1^2 2\,dx = 2x\Big|_1^2 = 2$$

① $P[a,b] = \frac{b-a}{L}$　概率 = $\frac{\text{区间长度}}{\text{总长度}}$　在 $[0,L]$ 区间上

② $P[a,b] = \int_a^b \frac{1}{L}dx = \frac{1}{L}x\Big|_a^b = \frac{b-a}{L}$



③ $P(\{a\}) = P([a,a]) = \frac{a-a}{L} = 0.$

⇒ 单点集的概率是0，所以忽略单点.

④ 概率的可加性：总概率 = $\frac{\text{各段长度之和}}{\text{总长度}}$

$$P([a_1,b_1]\cup[a_2,b_2]) = P([a_1,b_1]) + P([a_2,b_2])$$

$$= \frac{b_1-a_1}{L} + \frac{b_2-a_2}{L} = \frac{(b_1-a_1)+(b_2-a_2)}{L}$$

⑤ $P(A) = \frac{1}{L}\int_A 1\,dx = \int_A \frac{1}{L}dx$ ⇒ $f(x)=\frac{1}{L}$　密度函数常数 $f(x)=\frac{1}{L}$

$P(A) = \frac{\text{区间的长度}}{\text{总长度 } L}$　$[0,L]$.

⑥ 离散 vs 连续：

离散：直接分别求概率然后直接相加　$\sum$

连续：用积分代替求和.　$\int f(x)dx$

均匀分布中，$f(x)=\frac{1}{L}$，概率就是长度比例

全概率公式

⑦ $P(A)=0 \Rightarrow$ 不可能事件

$P(A)=1 \Rightarrow$ 必然事件

$P(B^c)=1-P(B) \Rightarrow P(B)+P(B^c)=1$

$P(\Omega)=P(B)+P(B^c)$ / $P(\Omega)=1$

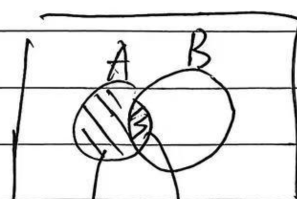
$P(A\cup B)=P(A)+P(B)$

⑧ 概率为0的事件不一定为空集

概率为1的事件不一定是整个样本空间

⑨ 按情况划分求概率 ⇒ 全概率公式/贝叶斯公式的核心思想

⑩ $(A\cap B)\cup(A\cap B^c)=A$ 不理解括弧时 → 直观理解



$\Rightarrow$ 由于 $(A\cap B)$ 与 $(A\cap B^c)$ 两部分互斥

推出: $P(A)=P(A\cap B)+P(A\cap B^c)$

$(A\cap \overline{B})\cap(A\cap B)=A$

⇑

全概率公式

eg: 事件A: 通过考试 B: 复习

$P(A)=P(A\cap B)+P(A\cap B^c) \Longleftrightarrow$ 总通过率 = (复习且通过) + (不复习且通过)

$\Rightarrow$ 全概率公式: $P(A)=P(A\cap B)+P(A\cap B^c)$.

Date.　　Page.

概率权重

大事件概率 = 所有小事件概率加起来

eg: 骰子: 1~6 各占的概率 ⇒ $p=\frac{1}{6}$ ⇒ 为概率权重.

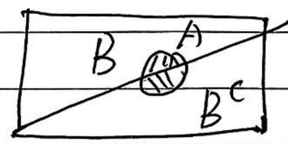
△ 对于有限的一个离散样本空间 $\Omega = \{a_1, a_2, \dots a_n\}$

$$p(a_i) \geq 0 \qquad \sum_i p(a_i) = 1$$

对于事件A　$p(A) = \sum_{a_i \in A} p(a_i)$.

全概率公式

① $p(A) = p(A\cap B) + p(A\cap B^c)$



② $A = (A\cap B) \cup (A\cap B^c)$　$[(A\cap B)\cup(A\cap B^c) = A\cap(B\cup B^c) = A\cap\Omega = A]$

$[(A\cap B)\cap(A\cap B^c)] = [A\cap(B\cap B^c) = A\cap\phi = \phi]$ ⇒ 证明互斥 (~~[illegible]~~)

⇒ 证明 $(A\cap B)$ 与 $(A\cap B^c)$ 互斥, 所以 $p(A) = p(A\cap B) + p(A\cap B^c)$

⇓

互斥可以相加.

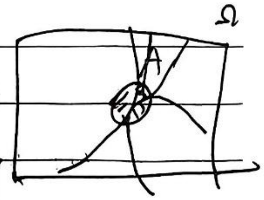
概率能相加, 前提条件是互斥.

全概率公式之延伸&拓展. 不止 B, B^c $\quad P(A)=\sum_{i=1}^{n} P(A\cap B_i)$

定义: 一群 $B_1, B_2 \dots B_n$, 如果 $\Omega=\{B_1 \dots B_n\}$ 的分集, 满足两个条件:

① 两两互斥 $B_i\cap B_j=\phi$ for all $i\neq j$ ② 覆盖整个样本空间 $\Omega=\bigcup_{i=1}^{n} B_i$

⇒ 若满足, 则 $P(A)=\sum_{i=1}^{n} P(A\cap B_i)$, 对∀事件 $A\subseteq\Omega$.



$[A=\bigcup_{i=1}^{n} (A\cap B_i)]$ 由于 $B_1, B_2 \dots B_n$ 两两互斥

$\Rightarrow [P(A)=P(\bigcup_{i=1}^{n} (A\cap B_i))=\sum_{i=1}^{n} P(A\cap B_i)]$

eg: 小镇 5%有男孩 5%有女孩, 求P(有孩子的概率)

$\Downarrow$ 事件A $\quad \Downarrow$ 事件B.

问题: $P(A\cup B)$ But $A\cap B\neq\phi$ AB不互斥, 因为可以同时有男孩有女孩.

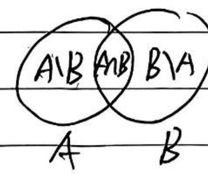
$P(A\cup B)=P(A)+P(B)$ × 不互斥

$$\boxed{\Rightarrow P(A\cup B)=P(A)+P(B)-P(A\cap B) \text{ 加法公式(容斥原理)}}$$

proof: $A=(A\cap B)\cup(A\setminus B)$ (1) 三者两两互斥.

$B=(A\cap B)\cup(B\setminus A)$ (2) ⇑

$A\cup B=(A\setminus B)\cup(B\setminus A)\cup\cancel{(A\cup B)}(A\cap B)$



$P(A\cup B)=P(A\setminus B)+P(B\setminus A)+P(A\cap B)$

From (1) $P(A)=P(A\cap B)+P(A\setminus B) \Rightarrow P(A\setminus B)=P(A)-P(A\cap B)$ ③

From (2) $P(B)=P(A\cap B)+P(B\setminus A) \Rightarrow P(B\setminus A)=P(B)-P(A\cap B)$ ④

由③④ $\rightarrow P(A\cup B)=P(A)-\underbrace{P(A\cap B)}_{\geq 0}+P(B)-P(A\cap B)+P(A\cap B)=P(A)+P(B)-P(A\cap B)$

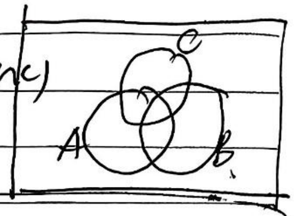
推论: $P(A\cup B)=P(A)+P(B)-P(A\cap B)\geq 0. \Rightarrow$ ~~$P(A\cap B)$~~ $\boxed{P(A\cup B)\leq P(A)+P(B)}$ 如果 $A\cap B\neq\phi$

Date. Page.

⇒ 进一步推论: $p(A_1 \cup A_2 \cdots \cup A_N) \leq p(A_1) + \cdots + p(A_N)$

eg: $p(A \cup B \cup C)$

$= p(A) + p(B) + p(C) - p(A\cap B) - p(A\cap C) - p(B\cap C) + p(A\cap B\cap C)$



⇒ 三元韦恩图 / 加减原理 / 容斥原理

$\cancel{p(A\cup B\cup C) = p(A) + p(B) + p(C) - p(A\cap B) - p(A\cap C) - p(B\cap C)} + p(A\cap B\cap C)$

$p(A\cup B\cup C) = p(A) + p(B) + p(C) - p(A\cap B) - p(A\cap C) - p(B\cap C) + p(A\cap B\cap C)$

eg: 投骰子并且抛硬币，都是等可能的、分析的概率.

$\Omega = \{(1,H), (2,H) \cdots (6,H), (1,T), (2,T) \cdots (6,T)\}$ $\#\Omega = 12$.

$p(\{(3,T)\}) = \frac{1}{12} = \frac{1}{6} \times \frac{1}{2} = p(\{3\}) \cdot p(\{T\})$.

In General.

$\Omega = \{(w_1, w_2) : w_1 \in \{1, \cdots 6\}.\ w_2 \in \{H, T\}\}$ 乘法原理.

$\Rightarrow p(\{(w_1, w_2)\}) = p(\{w_1\}) \cdot p(\{w_2\})$

eg: Prob (Die ≥ 5 . and Coin Head) $= p(\{(5,H), (6,H)\}) = p(\{(5,H)\}) + p(\{(6,H)\})$

$= p(\{5\}) \cdot p(\{H\}) + p(\{6\}) \cdot p(\{H\}) = p(\text{Die} \geq 5) \cdot p(\text{coin} = H)$.

DEF Consider two experiment: $\Omega_1 = \{a_1 \cdots a_N\}$ prob: $p(\{a_i\}) = p(a_i)$

$\Omega_2 = \{b_1 \cdots b_N\}$ prob: $p(\{b_i\}) = p(b_i)$

⇒ 若两实验相互独立 ⇒ $\boxed{p(\{a_i, b_i\}) = p(a_i) \cdot p(b_i)}$.

(Ω, F, P)

Date. Page.

DEF: Given $(\Omega, \text{Events}, P)$ Two Events A, B ARE independent.

$P(A \cap B) = P(A)P(B)$. ⇒引进一步去推理

互斥事件 $P(A \cup B) = P(A) + P(B)$ 和事件 互斥
独立事件 $P(A \cap B) = P(A) \cdot P(B)$ 积事件 独立

eg. Repeated Flips of a coin ⇒ independent

eg: 不等概率的硬币 $P(\{H\}) = p$ $P(\{T\}) = 1-p$ 这仅是一种顺序的概率

⇒ P(3 Tails ? 2 Heads) 按顺序来,先3个反面再2个正面

$P(B_1^c \cap B_2^c \cap B_3^c \cap B_4 \cap B_5) = P(B_1^c)P(B_2^c)P(B_3^c)P(B_4)P(B_5)$

$= (1-p)(1-p)(1-p)pp = (1-p)^3 p^2$.

二项分布

倘若无序,则需 $C_5^2 \cdot p^2 \cdot (1-p)^3$ 要乘以 C_5^2

eg: $\Omega_1 = \{a\}$, $\Omega_2 = \{b\}$

$\Omega_1 \times \Omega_2 = \{(a,b) : a \in \Omega_1, b \in \Omega_2\}$ ⇒ height × weight $= [0, 4] \times [0, 400]$.

样本空间的"乘积" ⇒ eg: 投两次骰子 里面俩是,平面二维点,矩形区域

$\Omega^2 = \{(\omega_1, \omega_2) : \omega_1 \in \{1 \ldots 6\}, \omega_2 \in \{1, 6\}\}$

⇒推广到n个骰子: $\Omega^n = \underbrace{\{1 \ldots 6\} \; \{1, \ldots 6\} \quad \{1, \ldots 6\}}_{n次} =$ 共 6^n 个

Date. Page.

eg: 无限次抛硬币——一定会出现正面吗?

Answer: Yes.

1) 设 A_i: 表示第 i 次抛掷是第一次出现正面 H

⇒ 每个 A_i 都是一个具体的首次成功的事件

$$\begin{cases} A_1 = H \\ A_2 = TH \\ A_3 = TTH \\ A_4 = TTTH \end{cases}$$

2) $A_i \cap A_j = \phi$ 对于 $i \neq j$, 两两相互斥

3) $p(\text{eventually } H) = p\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} p(A_i)$

$p(A_1) = p \quad p(A_2) = (1-p)p \quad p(A_3) = (1-p)^2 \cdot p \quad p(A_i) = (1-p)^{i-1} \cdot p$.

$\Rightarrow p(A_i) = (1-p)^{i-1} \cdot p$

4) $p(H) = \sum_{i=1}^{\infty} p(A_i) = \sum_{i=1}^{\infty} (1-p)^{i-1} \cdot p = \underbrace{\sum_{i=1}^{\infty} p \cdot (1-p)^{i-1}}_{\text{几何级数, 数列极限}}$

$= p \cdot \frac{(1-p)^{1-1}}{1-(1-p)} = p \cdot \frac{1}{p} = 1$

5) 结论: $p(\text{eventually } H) = 1$ ⇒ 只要不停抛硬币, 正面H最终一定会出现, 概率是1

6) 同理: 抛骰子 / 玩游戏 / 随机游走

↓ 一定会出现6 ↓ 一定有概率赢 ↓ 一定会回到原点

—来自 chat.qwen.ai: 再小的概率事件, 只要重复无数次, 最终一定会发生.

1st 公式①: Elements of $A = \#A = |A| =$ cardinality of $A = \mathrm{card}(A)$

eg: $A = \{a_1, \cdots a_N\}$ $\quad \boxed{|A| = \mathrm{card}(A) = N}$

解释：如果一个集合A是有限的，当且仅当它的元素个数等于某个自然数n.

A is finite if $|A| = n$ for some $n \in \mathbb{N}$

如果不是 $|A| = \infty$ (~~countable~~)

1) $|\mathbb{N}| = \infty$ (countable) 可数无限

2) $|[0,1]| = \infty$ (uncountable) 不可数无限

2nd 公式②: 乘法公式: $\Omega_1 = \{a_1, \cdots a_{N_1}\}$ $\quad \Omega_2 = \{b_1, \cdots b_{N_2}\}$

$$\Omega = \Omega_1 \times \Omega_2 = \{(a_i, b_j) \mid 1 \le i \le N_1,\ 1 \le j \le N_2\}$$

$$\boxed{\Rightarrow |\Omega| = |\Omega_1| \cdot |\Omega_2| = N_1 N_2}$$

确实:
解释:

$$\Omega = \left.\begin{Bmatrix} (a_1,b_1), & (a_1,b_2) & \cdots & (a_1,b_{N_2}) \\ (a_2,b_1), & (a_2,b_2) & \cdots & (a_2,b_{N_2}) \\ & \vdots & & \\ (a_{N_1},b_1) & (a_{N_1},b_2) & \cdots & (a_{N_1},b_{N_2}) \end{Bmatrix}\right\} N_2$$

$$\underbrace{\qquad\qquad\qquad}_{N_1}$$

$$\# \Omega = \underbrace{N_1 + N_1 + \cdots + N_1}_{N_2} = N_1 \times N_2 = |\Omega_1| \cdot |\Omega_2|$$

$\boxed{\Rightarrow \text{推广: } |\Omega| = |\Omega_1| \cdot |\Omega_2| \cdot |\Omega_3| \cdots |\Omega_k| \text{ 多乘法公式}}$

Date. Page.

eg: Dice $\times 50 \Rightarrow |\Omega| = 6^{50}$.

Flip coin $\times 50 \Rightarrow |\Omega| = 2^{50}$.

$\Rightarrow p(50H) = \dfrac{1 \rightarrow \text{50H一种}}{2^{50} \rightarrow \text{总共情况}} = \left(\frac{1}{2}\right)^{50}$.

(即$\frac{1}{2}$的50次方)

⇒ 证明掷硬币 相互独立的 $\Rightarrow p(50H) = p(1) \cdot p(2) \times \text{——} p(50)$

$\boxed{3^{rd}\ 公式③}$ 例如利用骰子,掷骰子要2,4,6、

$\boxed{阶乘}$

$2+6+4 \rightarrow \underbrace{(2,6,4),\ (2,4,6)\ \text{——}\ (4,6,2)}_{\text{共6种}}$

$P_3^3 = 3 \times 2 \times 1 = 6$.

possibility: $3 \times 2 \times 1$

第一个掷有三种选择

第二个掷有两种

最后一个有一种选择.

⇒ 定义: 集合 $A = \{a_1 — a_n\}$, 任意排列, 共有 $n! = n(n-1)(n-2) —— 2 \cdot 1$ 所有可能情况.

阶乘

eg: 26个字母任意排列有几种方式 $\Rightarrow n = P_{26}^{26} = 26!$

52张牌任意排列有几种方式 $\Rightarrow n = P_{52}^{52} = 52!$

⇒ 古典概型: $p = \dfrac{\text{Favorable}}{\text{Total}} = \dfrac{4!\,48!}{52!}$

(题目: 四张A排在最前面, 其他48张无所谓顺序, 这样的情况的概率)

⇒证明该题的独立性：$P=\frac{4!\,48!}{52!} \xrightarrow{\text{化简}} \frac{4\times3\times2\times1}{52\times51\times50\times49}=\frac{4}{52}\times\frac{3}{51}\times\frac{2}{50}\times\frac{1}{49}$

$=P(1^{st}\ Ace)\times P(2^{nd}\ Ace\ \text{Given}\ 1^{st}\ \text{is}\ ACE)$

⇒涉及独立性和条件概率 $\times P(3^{rd}\ Ace\ \text{Given}\ 1,2\ \text{is}\ Ace)$ ⇒ 条件概率

$\times P(4^{th}\ Ace\ \text{Given}\ 1,2,3\ \text{is}\ Ace)$

⇒eg：从26个字母中选三个组成 { 可重复的三个字组成的英语单词 $26^3=26\times26\times26$

不可重复 —— $26\times25\times24=\frac{26!}{23!}$

$P_N^K=N(N-1)\cdots(N-K+1)$　　先26种自由全排列 再除掉不需要的23种全排列

有顺序 $=\frac{N!}{(N-K)!}$

⇒eg：用book中的"b""o""o""k"形组成多少不同的字母组合．$\frac{P_4^4}{P_2^2}$ → 先4全排，后因删除"o"

HAVANA ——→ $\frac{P_6^6}{P_3^3}$ → 6全排，→ 删3

DIDI ——→ $\frac{P_4^4}{P_2^2P_2^2}=\frac{4!}{2!\,2!}=\frac{4\times3\times2\times1}{2\times2}=6$　4全排，两个D，I．

7balls：3 Black，2 RED，1 Blue，1 Gree → $\frac{P_7^7}{P_3^3\cdot P_2^2}$ 总 ⇒ 全自由随机排列 ⇒ 求P(无两黑排邻)

○ ○ ○ ○

$\frac{C_5^3\times\frac{P_4^4}{P_2^2}}{P_7^7/P_3^3\cdot P_2^2}$ // $=\frac{\binom{5}{3}\cdot\frac{4!}{2!}}{\frac{7!}{3!\cdot2!}}$

插空法，捆绑法

Date. Page.

4th 公式④: NEXT Wednsday $\binom{40}{17}\binom{40}{13}$.

排列: $P(n,k)$ / P_n^k. $P(5,3)=P_5^3=5\times4\times3$

组合: $\binom{n}{k} = C_n^k$ 从n选K. $\binom{5}{2}=C_5^2=\frac{5\times4}{2\times1}$

$$\text{n choose } k = \frac{n!}{k!(n-k)!} \qquad =\frac{5!}{2!\,3!}$$

$$\binom{n}{k} = C_n^k = \frac{n!}{k!(n-k)!} \quad n \text{ choose } k = \frac{n(n-1)\cdots(n-k+1)\cancel{(n-k)!}}{k!\,\cancel{(n-k)!}}$$

$\Rightarrow$ # ways of choosing k objects out of n

$$= \frac{\overbrace{n(n-1)\cdots(n-k+1)}^{k项}}{k!}$$

eg: Flip 15 Times A fair coin. 抛15次公平的硬币 $\boxed{二项分布}$

A = {恰好7次"H"}

① $$P(A) = \frac{\text{FAVORABLE}}{\text{Total}} = \frac{\binom{15}{7}}{2^{15}} = \binom{15}{7}\left(\frac{1}{2}\right)^7\left(\frac{1}{2}\right)^8$$

15个选7个正面的 其余8个均为反面 二项分布

② if 不公平硬币 $p(\{H\})=p$. $p(\{T\})=1-p(\{H\})=1-p$ $\boxed{二项分布}$

$$P(A) = \sum_{\substack{\omega_1,\omega_2\ldots\omega_{15} \\ \text{where } \omega_j\in\{H,T\}}} p(\{(\omega_1\ldots\omega_{15})\}) = \sum_{\substack{\omega_1,\omega_2\ldots\omega_{15} \\ \text{where } \omega_j\in\{H,T\}}} \boxed{p^7(1-p)^8} = C_{15}^7 p^7(1-p)^8 = \binom{15}{7}p^7(1-p)^8$$

组合数

其中 $\binom{15}{7}=C_{15}^7$ 为二项式系数 指的是一个7H8T序列的概率

$$= \frac{15!}{7!\,8!}$$

Date.　　Page.

⇒ 定理 T_1：引入二项式定理：

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}.$$

⇒ 拓展推广 = 多项式定理 T_2：$(x_1+x_2+\cdots+x_l)^n = ?$

1. 多项式系数 $\binom{n}{n_1, n_2 \cdots n_l} = \dfrac{n!}{n_1! n_2! \cdots n_l!}$

2. 定理：$(x_1+x_2+\cdots+x_l)^n = \sum\limits_{\substack{n_1 \cdots n_l \\ n_1+\cdots+n_l=n}} \binom{n}{n_1, n_2 \cdots n_l} x_1^{n_1} x_2^{n_2} \cdots x_l^{n_l}$.

$\Downarrow$ 多项具体例子

eg：$(x+y+z)^2 = ?$　$l=3$（3 项）　$n=2$（次数为 2）

⇒ 满足 $n_1+n_2+n_3=n=2$ 的非负整数解

n_1	n_2	n_3	项	系数
2	0	0	x^2	$\frac{2!}{2!}=1$
0	2	0	y^2	$\frac{2!}{2!}=1$
0	0	2	z^2	$\frac{2!}{2!}=1$
1	1	0	xy	$\frac{2!}{1!1!}=2$
1	0	1	xz	$\frac{2!}{1!1!}=2$
0	1	1	yz	$\frac{2!}{1!1!}=2$

⇒ 根据表格：$(x+y+z)^2 = x^2+y^2+z^2+2xy+2xz+2yz$. 完美符合.

Date. Page.

~~条件概率~~. 概率与数据课程中最重要的章节

DEF: $p(A|B) = \frac{p(A \cap B)}{p(B)}$ 定义式

定义: 基于事件A和B，A的条件概率(基于B) $\Rightarrow p(A|B) = \frac{p(A\cap B)}{p(B)}$

* 1) 条件概率的前提是已经发生的(即事件B确定已发生)
2) 事件B条件变化，则条件概率也会变化。
3) 相信数学之神，不要依靠常识判断概率，注意概念不要混淆.

$\boxed{\text{推论更好的公式}: p(A\cap B) = p(A|B)p(B)}$

如果 $p(B)=0 \Rightarrow p(A\cap B)=0$ $(A\cap B < B) \Rightarrow$ Hence 同乘0=0.

eg. 周六晚上，80%交通事故是由清醒的司机造成的，20%的交通事故因醉酒司机造成

典型的与常识相反

A= ACCIDENT.

D = DRUNK DRIVERS 其中这天晚上，5%开车的人喝酒，10%骑车的人在车祸中丧生

$\Rightarrow p(A|D) = \frac{p(A\cap D)}{p(D)} = \frac{0.2\times0.1}{0.5\times0.1} = \frac{0.02}{0.05} = 0.4$

ACCIDENT and Drunk

Drunk

$p(A|D^c) = 0.0875$

这段的几个数据还有代入没太看懂

概率问题 → 无记忆性 (memoryless Property).

Date.　　　　Page.

~~eg: 下雨, 一杯……~~

$\boxed{\text{指数衰变模型} \quad P(T>t) = e^{-\lambda t}}$ 　$P(T>16) = e^{-0.042\times16} = 0.51$

半衰期

eg: 镭是一个放射性的元素

$t=0$ 有个镭原子

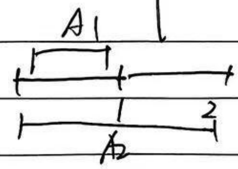
⇒ 如果一个镭原子已经存活了一个世纪 　$P(\text{atom not decayed at } x) = e^{-0.042x}$

Q: 那么它还能再活1个世纪的概率? 　⇓ 　经过的时间

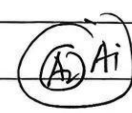
⇒ 和一开始的存活概率一样! 　在时间x内未衰变的概率 (/世纪).

⇒ A_1: 原子在第1个世纪末还活着 　Ra 半衰期 1600年

A_2: 原子在第2个世纪末还活着



A: → 求: 已知A_1发生, 求A_2发生的概率



$$P(A_2|A_1) = \frac{P(A_1\cap A_2)}{P(A_1)} = \frac{P(A_2)}{P(A_1)}$$

$$= \frac{e^{-0.043\times2}}{e^{-0.043\times1}} = e^{-0.043} = P(A_1).$$

$\rightarrow P(A_2|A_1) = P(A_1)$

Reflection: 无记忆性: 未来的行为与过去无关, 过去不影响未来.

$P(T>s+t \mid T>s) = P(T>t)$ 　T是衰变时间.

条件概率

1. 重要性质

$\boxed{\text{归一化}}$

① $A \mapsto p(A|B)$ 是一个概率 → 合法的概率函数，满足概率定义

proof: $p(\Omega|B)=1$

$$\Rightarrow p(\Omega|B) = \frac{p(\Omega\cap B)}{p(B)} = \frac{p(B)}{p(B)} = 1$$

可以使用概率的所有性质

② 对于互斥事件：$\boxed{\text{可加性}}$

~~$p(A_1\cup A_2\cup\cdots\cup A_i|B)=p(A_1|B)+p(A_2|B)+\cdots+p(A_i|B)$~~

$$\Rightarrow p\left(\bigcup_{i=1}^{k} A_i \middle| B\right)=\sum_{i=1}^{k} p(A_i|B)$$

~~③ $p(A^c|B)=1-p(A|B)$~~

附属性质：<1> $p(A^c|B)=1-p(A|B)$.

推论 <2> $p(A|B)=p(A\cap C|B)+p(A\cap C^c|B)$.

⇒推论 $p(A|B)=\sum_{i=1}^{k} p\left(A \bigcap_{i=1}^{k} C_i \middle| B\right)$

<3> $p(A\cup C|B)=p(A|B)+p(C|B)-p(A\cap C|B)$.

<条件概率下的容斥原理>

③ $B\mapsto p(A|B)$. Divide and conquer. (Division by cases) ⇒ 分类法

$p(A)=p(A\cap B)+p(A\cap B^c)$

A B

把条件概率看作一个依赖于B的函数

$p(A\cap B)=p(A|B)\times p(B)$

$p(A\cap B^c)=p(A|B^c)\times p(B^c)$

全概率公式

Date.　　　Page.

③ $\boxed{\Rightarrow \text{概括：} p(A) = p(A|B) \times p(B) + p(A|B^c) \times p(B^c)}$

$\boxed{\Rightarrow \text{更一般的形式：} p(A) = \sum_{i=1}^{n} p(A|B_i)\, p(B_i)}$

eg：医疗测试：检测某个疾病：

P假阴：实际有病但检测阴性 = 5%

P假阳：健康但检测出阳性 = 10%

翻译数学语言

T：Test positive.

B：cow sick.

$p(T|B^c) = 0.3$ ①

$p(T^c|B) = 0.1$ ②

全概率公式适用 ⇒ 已知条件 ⇒ ①②③

已知 ~~p(T|B)~~ = 不服感病

$\Rightarrow p(T) = p(T|B) \times p(B) + p(T|B^c) \times p(B^c) = 0.9 \times 0.02 + 0.3 \times 0.98 = \underline{\underline{0.312}}$

$p(T|B) = 1 - p(T^c|B) = 1 - 0.1 = 0.9$

$p(T|B^c) = 0.3$

已给 0.02 ③

$1 - 0.02 = 0.98$

低

假阳概率太高

④贝叶斯公式：Beyes FORMULA！ 实现从 $p(T|B)$ 与 $p(B|T)$ 之间的转换

贝叶斯公式

$$p(B|T) = \frac{p(T \cap B)}{p(B)} = \frac{p(T|B) \times p(B)}{p(T|B) \times p(B) + p(T|B^c) \times p(B^c)}$$

⇒ 条件概率

全概率公式

~~⇒ 进一步算~~

Example：next page.

Date.　　Page.

Example：贝叶斯定理应用：检测阳性、假阳性

医学检测：　B：患病　　T：阳性

$\boxed{1}$　　¬B：未患病　　¬T：阴性

给出：$p(B)=0.02$　$p(\neg B)=0.98$　$p(T|B)=0.7$　$p(T|\neg B)=0.1$

⇒ 运用贝叶斯公式：$p(B|T)=\dfrac{p(T|B)\times p(B)}{p(T)}$

$\boxed{2}$

$=\dfrac{p(T|B)\times p(B)}{p(T|B)\times p(B)\times p(T|B^c)\times p(B^c)}=\dfrac{0.7\times0.02}{0.7\times0.02+0.1\times0.98}=0.125$

$\boxed{3}$　有70%准确率但假阳性很多 ⇒ 假阳性都可能

检测出阳性 但只有12.5%是真病了

$\boxed{4}$　改进：① 改进检测方法
② 重复测试

$p(B|TT)=\dfrac{p(TT|B)\times p(B)}{p(TT|B)\times p(B)+p(TT|B^c)\times p(B^c)}$

TT：两次检测都为阳性　$p(TT)=0.7^2=0.49$
$=p(T|B)\times p(T|B)=0.7^2$

$=0.5$ ⇒ 检测出阳性，有50%是真病了 ↑↑↑

显著提高!!!

Date. Page.

Example：检测官谬误 Prosecuter's Fallacy

[1] 犯罪发生，血液中存在某种疾病，与可能的嫌疑人

⇒ 被定罪：

[2] D：disorder 患病

I：innocent 无辜

$$p(D|I) = \frac{p(D \cap I)}{p(I)} = \frac{1}{10^6} \quad \times$$

[3] $$p(I|D) = \frac{p(D|I) \times p(I)}{p(D)} = \frac{10^{-6} \times 1}{10^{-5}} = 0.1$$ （多数是无辜的，接近1）

[4] 结论：将有罪与证据强度

混淆有罪 ⇒ 检察官谬误.

⑤ 独立事件：

$$p(A|B) = \frac{p(A \cap B)}{p(B)} = \frac{p(B|A) \times p(A)}{\sum_i p(A|B_i)\, p(B_i)}$$

$$p(B_i|A) = \frac{p(A \cap B_i)}{p(A)} = \frac{p(A|B_i) \times p(B_i)}{\sum_i p(A|B_i)\, p(B_i)}$$

当 $p(A|B) = p(A)$ 时 ⇒ B不影响A的概率 ⇒ AB独立

$(P(B)>0)$

⇒ 独立的定义：$p(A \cap B) = p(A) \cdot p(B)$

Date.　　　　Page.

~~独立的定义 ⇒ 两两独立(pairwise independent) & 整体独立~~

⊗ 假设 $A_1, A_2, A_3, \cdots A_n$ 相互独立，则充要条件是 $p(A_i \cap A_j) = p(A_i) \times p(A_j)$

⇒ 这样的条件还不够

假设 A_1, A_2 独立于 A_3

$$\Rightarrow p(A_3 \mid A_1 \cap A_2) = p(A_3) = \frac{p(A_1 \cap A_2 \cap A_3)}{p(A_1 \cap A_2)} = \frac{p(A_1 \cap A_2 \cap A_3)}{p(A_1) \cdot p(A_2)}$$

$\boxed{1}$ $\Rightarrow p(A_1 \cap A_2 \cap A_3) = p(A_1) \cdot p(A_2) \cdot p(A_3)$

⇒ We NEED: $p(A_1 \cap A_2 \cap A_3) = p(A_1) p(A_2) p(A_3)$

$\boxed{\text{整体独立}}$

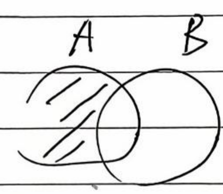
⇒ DEF: $A_1, A_2, \cdots A_n$ are independent 独立条件为

当且仅当 $\forall\, i_1, i_2 \cdots i_k$，$i_j \neq i_q$，$1 \leq i_q \leq N$，$\forall\, k \leq N$

$$\boxed{p(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = p(A_{i_1}) \cdot p(A_{i_2}) \cdots p(A_{i_k})}$$

$p(A_i \cap A_j) = p(A_i) p(A_j)$　$i \neq j$ ⇒ 两两独立 pairwise independent

$\boxed{2}$ $\boxed{\text{性质(讨论)：A、B独立，则 A与}B^c\text{也独立}}$



⇒ 证明: $p(A \cap B) = p(A) p(B)$　(1)

$p(A \cap B^c) = p(A) - p(A \cap B)$

$= p(A) - p(A) p(B) = p(A)(1 - p(B)) = p(A) p(B^c)$

∴ 得证.

Date.　　　Page.

[3] "$\perp\!\!\!\perp$" 为独立符号:

$A \perp\!\!\!\perp B \Rightarrow A \perp\!\!\!\perp B^C, A^C \perp\!\!\!\perp B, A^C \perp\!\!\!\perp B^C$.

$\Updownarrow$

A独立于B、

条件概率就此结束.

随机变量 Random Variable

引入: Gerolamo Cardano 意大利数学家. ⇒ 引入随机变量

引入: 掷两枚硬币, # Heads ⇒ 正面朝上的数量.

$\Omega = \{(H,H), (H,T), (T,H), (T,T)\}$　　#: $\Omega = \{0,1,2\}$ 样本变量.

$\frac{1}{4}$　$\frac{1}{4}$　$\frac{1}{4}$　$\frac{1}{4}$.　　$p(0) = \frac{1}{4}$

2　1　1　0　　$p(1) = \frac{1}{2}$　四种情况用

$p(2) = \frac{1}{4}$　几种情况整合.

卡丹诺方法: 权举. (定义随机变量)

引入: continuous. Roulette 俄罗斯轮盘赌

定义: $X: \Omega \to \mathbb{R}$: 随机变量 X 是从样本空间 Ω 映射到实数集 R 的函数

随机变量　样本空间　实数

$p(\omega_1 + \omega_2 = a) = 0$!! ⇒ 当 ω_1, ω_2 为连续型随机变量时, 在连续分布中单点概率值为0.

全 事件 prob

Date.　　Page.

定义：给定：(Ω, F, p) 是一个随机变量的函数

$X: \Omega \to R$.

∴所以 $[a,b]$ 的区间是对所有 $a \le b$ 的事件集合.

离散型变量 Discrete Random Variables 离散随机变量.

[1] 定义：给定 (Ω, F, p) 是一组离散随机变量的函数 $X: \Omega \to R$

对于随机变量 X 需满足 { ① 有限个值　② 可数无限值 不连续无限　$(X \to$ 无限个值$)$

⇒ 能列出来所有取值—~~连续型~~ 离散型　/ 列出来可以是任意小数—连续型.

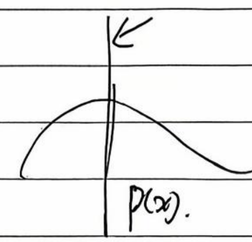
※ 随机变量用大写字母表示 eg: X, Y, Z, U, V

—— 而具体取值用小写字母 x, y, z 表示　eg: X has results x, y, z.

☆ $\{X=x\} = \{\omega \in \Omega : X(\omega)=x\}$

[2] PMF: Probability Mass Function 概率质量函数

定义：给定一个离散型随机变量 X，PMF 函数定义如下：



① 函数 $p: R \to [0,1]$　$(x \to x \in [0,1])$

输入　输出

② $p(x) = p(\{X=x\})$　$p(x)$ 的值等于事件 $X=x$ 的概率

eg: $p(0) = p(X=0) = p(TT) = \frac{1}{4}$

$p(1) = p(X=1) = p(HT \& TH) = \frac{1}{2}$

$p(2) = p(X=2) = p(HH) = \frac{1}{4}$

$\boxed{3}$ pmf 通用公式： $p(A) = \sum_{\{x\in A\}} p(x)$.

⇒ 事件A的概率等于所有属于A的x值的对应p(x)之和.

eg： H = H≥1 A={1,2}

$p(A) = p(1)+p(2) = p(X=1)+p(X=2) = p(HT或TH)+p(HH)$

$= \frac{1}{4}+\frac{1}{2} = \frac{3}{4}$.

$\boxed{4}$ CDF： Cumulative Distribution Function 累积分布函数

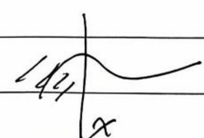
定义： 随机变量X的分布函数是一个函数（累积分布函数）

连续型也如此

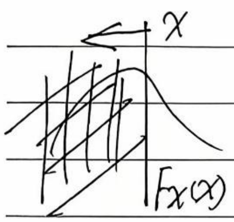
$F_X: \mathbb{R} \to [0,1]$ 不管X离散/连续，CDF都有定义

D→ 输入 输出

公式： $x \mapsto F_X(x) = p(X \le x)$



eg： ~~Fx(2)~~ $F_X(1) = P(X\le 1) = p(X=0)+p(X=1) = \frac{1}{4}+\frac{1}{2} = \frac{3}{4}$.



$p(a\le X\le b) = F_X(b) - F_X(a) = p(X\le b) - p(X\le a)$

$\boxed{5}$ CDF ⇔ pmf ⇒ $P(x) \Leftrightarrow F_X(x)$ 可以互相转换.

转换公式： $F_X(x) = \sum_{x_i \le x} p(x_i)$ CDF ⇐ PmF 求和

$P_X(x_i) = F_X(x_i) - F_X(x_{i-1})$ CDF ⇒ PmF.

pmf cdf $= p(X\le x_i) - p(X\le x_{i-1}) = P_X(x_i)$ 差分

类比数列 a_n S_n

⇒ $S_n = \sum_{i=1}^{n} a_n$ CDF ⇐ PmF

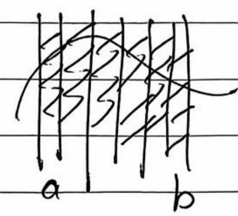
$a_n = S_n - S_{n-1}$ CDF ⇒ PmF.

Date.　　　Page.

Discrete Random Variable (Last Topic for the OCT 17 Test)

DEF 定义：X is Discrete Random Variable，概率质量函数 PMF.

$$\begin{cases} ① \ P_X(x_i) = p(X = x_i) = p(\{w \in \Omega : X(w) = x_i\}) \ (存在\ w). \\ ② \ p(a \le X \le b) = \sum_{x_i \in [a,b]} P_X(x_i) \end{cases}$$



DEF 定义：The Cumulative Distribution Function 累积分布函数 CDF.

① $F_X : R \to [0,1]$　$F_X(x) = p(X \le x)$　→　$X \le x$

某个特定的值

F_X 大写表示一个函数

x 小写表示一个值

表示函数的 X 值大写

* 如果 X 是离散的，则 PMF 与 CDF 等效

② $F_X(x) = \sum_{x_i \le x} p(x_i)$　~~F(x) ⇔~~　CDF ⇔ PMF 转换公式.

$P_X(x_i) = p(X = x_i) = \cancel{\Sigma_x}\ p(X \le x_i) - p(X \le x_{i-1}) = F_X(x_i) - F_X(x_{i-1})$

PMF ⇔ CDF 转换公式.

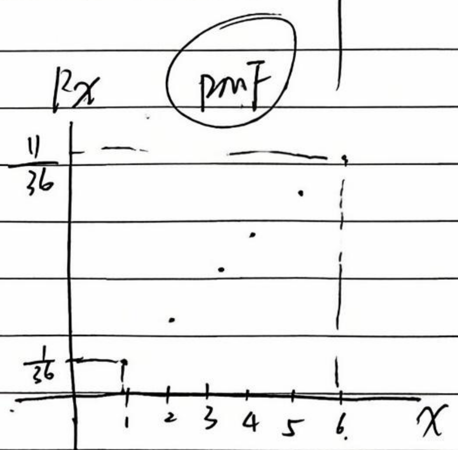
=36.

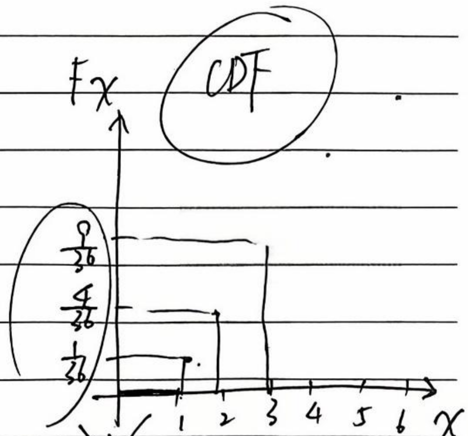
Example: 抛两枚骰子，样本空间 $\Omega=\{1,2,\dots 6\}^2$，每种 (ω_1,ω_2) 的概率均等为 $p(\omega_1,\omega_2)=\frac{1}{36}$

1) 列表、

X	P_X	
1	$\frac{1}{36}$	(1,1)
2	$\frac{3}{36}$	(1,2) (2,1) (2,2).
3	$\frac{5}{36}$	(1,3) (2,3) (3,3) (3,2) (3,1).
4	$\frac{7}{36}$	(1,4) (2,4) (3,4) (4,4) (4,3) (4,2) (4,1)
5	$\frac{9}{36}$	以此类推.
6	$\frac{11}{36}$	

2) 画图 CDF PMF





缓慢，线性

两者图像不同

⇓

解释区别

陡快 线性

eg: ~~P(X)~~ $F_X(3)=P(X\le 3)$
$=P(X=1)+P(X=2)+P(X=3)$
$=\frac{1+3+5}{36}=\frac{9}{36}$.

Cumulative distribution function

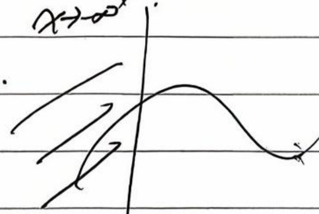
累积分布函数的性质

性质: General Properties of $\cancel{F_X}$ CDF

F1) $F_X(x)=0$ if $x <$ lowest Result x_i.

$$\lim_{x\to-\infty} F_X(x) = 0.$$

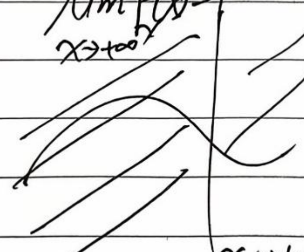
$\lim_{x\to-\infty} F_X(x) \to 0.$



F2) $F_X(x)=1$ if $x \geq$ largest Result x_i

$$\lim_{x\to+\infty} F_X(x) = 1$$

$\lim_{x\to+\infty} F_X(x) = 1$



F3) F_X is increasing [$a \le b \Rightarrow F_X(a) \le F_X(b)$].

$\Rightarrow F_X(a) = P(X \le a) \le P(X \le b) = F_X(b) \Rightarrow \quad a \le b \Rightarrow F_X(a) \le F_X(b)$

$F_X(x)$单调增

F4) F_X is right continuous

$\Rightarrow F_X(x)$函数是右连续的

$$\lim_{h\to 0} F_X(x+h) = F_X(x)$$

F5) F_X Has Jump at a $\Rightarrow$ size of jump $= p_X(a)$

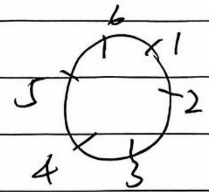
如果CDF在$x=a$处有一个跳跃(不连续点), 那么跳跃的大小就等于$x=a$的概率, 即$p(X=a)$

[只要$F(x)$满足F_1-F_5, 则其一定是 Cumulative Distribution Function
CDF 累积分布函数]

[illegible]

Date. Page.

Example 2: 轮盘赌 Roulette. ⇒ 连续随机变量. CDF.



$\Omega = [0,1]$ $\xi : \Omega \to \mathbb{R}$. $\xi(\omega) = \omega \in [0,1)$

$RV : \Omega \to R$.

?

~~DEF: 2 Discrete R.V. x And y 有相同的~~ ~~离散概率~~.

LAW — R.V. 的所有概率行为

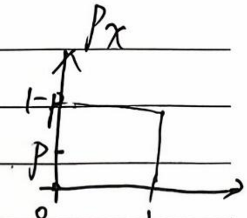
⇒ Catalogue of popular Discrete R.V. Laws 引入.

应用的 ~~随机~~ 离散随机变量概率行为的目录.

(1) Bernouli Law: ~~"0,1" 分布定律~~ 伯努力分布

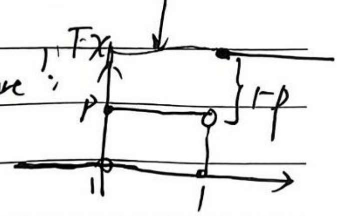
对 $\operatorname{Im} X = \{0,1\}$ $P_X(1) = p$ $P_X(0) = 1-p$.

↳ X的取值集合



eg: Coin F = prob(h) 1 = success 0 = Failure.

p: prob of success 1-p: prob of failure



notation ⟶ $X \sim BER(p)$

服从

(2) Uniform law ⇒ $\operatorname{Im} X = \{x_1, \ldots x_N\}$ ⇒ $\boxed{P_X(x_i) = \frac{1}{N}}$

?

notation: $X \sim U(N)$

$O \sim X \sim U(x_i \ldots x_N)$

离散均匀分布 → 所有可能出现的结果的概率都一样

$X \sim U(N)$ X服从离散均匀分布

Binary 二进制

$E[X] = np$.

$D[X] = np(1-p)$.

Date. Page.

<3) Binomial Law. 二项分布 $X \sim B(N,p)$ / $Bin(N,p)$.

$Im(X) = \{0, 1, \dots N\} \Rightarrow p_X(k) = \binom{N}{k} p^k (1-p)^{N-k}$

$= C_N^k p^k (1-p)^{N-k}$.

eg: Coin 投五次, 恰好三次正面概率 N—次数

$\Rightarrow p_X(3) = C_5^3 \cdot (\frac{1}{2})^3 \cdot (\frac{1}{2})^2 = 0.3125$. p—成功概率

eg: $X \sim B(10, \frac{1}{2})$ 10次, $P(success) = \frac{1}{2}$.

x	$p_X(x)$
0	$(\frac{1}{2})^{10} = \binom{10}{0}(\frac{1}{2})^0(\frac{1}{2})^{10}$
1	$\binom{10}{1}(\frac{1}{2})^1(\frac{1}{2})^9$
2	$\binom{10}{2}(\frac{1}{2})^2(\frac{1}{2})^8$
3	以此类推
⋮	
10	$\binom{10}{10}(\frac{1}{2})^{10}(\frac{1}{2})^0$

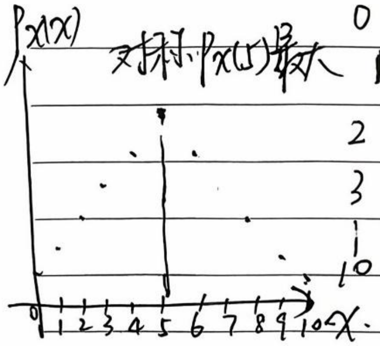
p概率相等对称

因为symmetric对称

$\Rightarrow p_X(3) = p_X(7)$

$p_X(i) = p_X(10-i)$.

对称: $p_X(5)$ 最大



<4) GEOMETRIC Law. $X \sim Geom(p)$. X服从几何分布.

$p(success) = p$ $p(failure) = 1-p$.

X → first flip it is head / try until 1st success

$Im X = \{1, 2, 3 \dots\} = N^*$

$\boxed{p(k) = (1-p)^{k-1} \cdot p}$ → 第k次成功

直到第k次才成功的概率   前k-1次都失败.

Date.　　Page.

(5) poisson Law：泊松分布

当 ① P小 ② 试验次数n很大 ③事件之间相互独立.

~~[illegible]~~ ④ λ 稳定.

① 附：$x \sim Pois(\lambda)$ / $x \sim P(\lambda)$

λ表示单位时间内/空间内的平均发生率(期望值).

$I_m(x) = \{0, 1, 2, 3, \ldots\}$ 非负整数.

② PMF：概率质量函数 泊松分布

$$p(x=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k = 0, 1, 2 \ldots$$

泊松分布作为二项分布极限所得的结果

$\lambda = n \cdot p$

∗. λ：平均发生次数. $e \approx 2.71828$ k → 事件发生的次数.

含义：在给定λ下，事件恰好发生k次的概率.

③ $E[x] = \lambda$ 期望 均值

$Var(x) = \lambda$ 方差 σ^2 ∗ 均值 = 方差

$\sigma_x = \sqrt{\lambda}$ 标准差

次数 概率

np

$\lambda = n \cdot p = 52 \cdot 10^{-6} = 52 \times 10^{-6}$ 类似二项分布

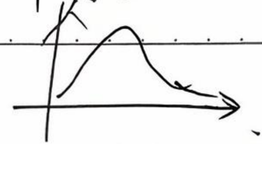
④ eg：每周买一次彩票，中奖概率 $p = 10^{-6}$，求第一年(52周)从未中奖的概率

$$p(x=0) = \frac{\lambda^0 e^{-\lambda}}{0!} = \frac{1 \times e^{-\lambda}}{1} = e^{-52 \times 10^{-6}} \approx 0.999948$$

$\lambda = 52 \times 10^{-6}$

⇒ 结论：第一年没中奖的概率高达 99.9948%

⑤ 泊松分布图像.



λ 越小，峰值越靠左.

λ 越大，越对称(越接近于正态分布)

Date.　　　Page.

Geometric Distribution Memoryless Property

从0开始，第s次成功的概率

DEF: A R.V. X is memoryless if $\underbrace{p(X>t+s \mid X>t)}$ $=$ $p(X>s)$ //

我已经等了t次未成功，我到还需第s次才成功的概率

⇒ 无记忆性. 系统不会记得你等了多久，概率不会变，和一开始一样

Theorem: 一个R.V. X → memoryless ⇒ X的分布为几何分布，当且仅当

几何分布 Geometric distribution 是唯一的离散的、无记忆的分布统计分布

证明: R.V. $X \sim Geo(p)$

Then $p(X>t+s \mid X>t) = \dfrac{p(X>t+s)\cap p(X>t)}{p(X>t)}$

$= \dfrac{p(X>t+s)}{p(X>t)} = \dfrac{(1-p)^{t+s}}{(1-p)^t} = (1-p)^s = p(X>s)$

相等. 证毕.

eg: 我扔一枚硬币. 我扔了100次未出现正面. 我 $p\{(X>100)=H\} = p\{(X>1)=H\}$

⇒ 我未来扔出正面的概率与我第一次一样 ⇒ 只适用于几何分布的无记忆性

Poisson Law 泊松分布 $X \sim pois(\lambda)$, $Im(X)=\mathbb{N}$, $P_X(k)=\frac{\lambda^k e^{-\lambda}}{k!}$

Situation: Arrival of individuals/Customers
arrivals rare (never more than 1 at a time)
independent
X: number of arrivals in a unit of time.

Individuals	Unit of Time
People	hour
folders	Hour
muders	year
Horse kicks	10 years.
Bombs	Day
Meteorites	year (Not Really).

wtf is talking bro?

Proof:

Model: $X_n \sim Bin(n, \frac{\lambda}{n})$

$$p(X_n=k) = \binom{n}{k}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k} = \frac{\overbrace{n(n-1)\cdots(n-k+1)}^{n项}}{k!}\frac{\lambda^k}{n^k}\frac{(1-\frac{\lambda}{n})^n}{(1-\frac{\lambda}{n})^k}$$

$$= \frac{\lambda^k}{k!}\cdot\underbrace{\frac{n}{n}\cdot\frac{n-1}{n}\cdots\frac{n-k+1}{n}}_{\to 1\ 极限}\cdot\frac{(1-\frac{\lambda}{n})^n}{\underbrace{(1-\frac{\lambda}{n})^k}_{\to 1\ 极限}}$$

$\underbrace{n\to\infty\to}$

$$= \frac{\lambda^k}{k!}\cdot\underbrace{\lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n}_{极限\ =\ e^{-\lambda}}$$

$$= \frac{\lambda^k e^{-\lambda}}{k!} \Rightarrow 从二项分布推出泊松分布$$

* 泊松分布是二项分布的一种极限情况, $n\to+\infty$ 极限.

Applications: All death of soldiers by horse kicks.

German CAVALRY: 10 units

During 20 years = 200 unit time.

Why is this example?

Applications 2: London WII 2. V1-BOMBS 德军伦敦空袭轰炸.

Divided in 576 squares of 0.5km of length.

~~$x(i)$ = number of Bombs~~ $\downarrow$ ~~square = 576.~~

$x(i)$ = number of squares.
$\downarrow$
number of Bombs

#Bombs	#squares	$576 \times pois(0.93) = 576 \times \frac{0.93^k e^{-0.93}}{k!}$
0	229	217.5
1	211	211.3.
2	93.	98.1
3	35	30.4
4	7	7.1
5	1	1.3.

此泊松分布算出来的非常接近.

Proposal: $X \sim Pois(\lambda)$ If True, Bombing at Random.

$$\lambda = \frac{0\times229 + 1\times211 + 2\times93 + 3\times35 + 4\times7 + 5\times1}{576} = 0.93.$$

(单位体积的国土上炸弹的数量 ⇒ λ的期望值)

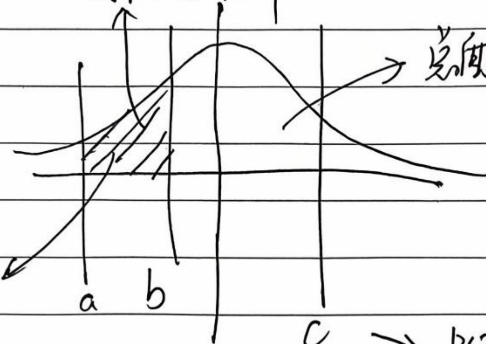
单位长度内的概率质量的集中程度

PDF Probability Density Function (PDF) 概率密度函数

⇒ 微积分基本定理 $p(a\le X\le b) = F_X(b) - F_X(a) = \int_a^b f_X(x)\,dx$

[1]

面积代表概率



总面积为1.

精确定义:

$$f_X(x)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} f_X(y)\,dy$$

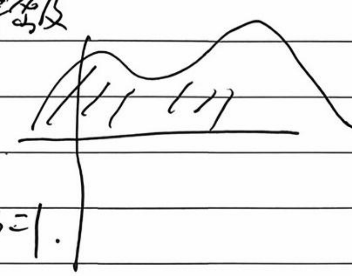
$$p(a\le X\le b)=\int_a^b f_X(x)\,dx$$

$$= f_X(x)\Big|_a^b = F_X(b)-F_X(a).$$

c → $p(X=c)=0$

单点积分面积即概率为0.

$$f_X(x)=\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\int_{x-\frac{\varepsilon}{2}}^{x+\frac{\varepsilon}{2}} f_X(y)\,dy \Rightarrow \text{平均概率密度}$$



* 2) $p(X\in R)=\int_{-\infty}^{+\infty} f_X(x)\,dx=1$

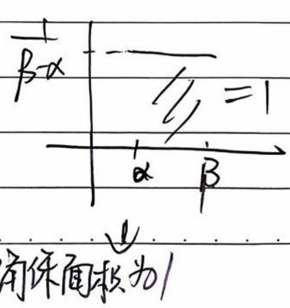
概率总和/密度总和/积分=1.

[2] catalogue of continuous laws 连续定律

1) Uniform law = (Uniform distribution)

$Im\,X=[\alpha,\beta]$

离散连续分布



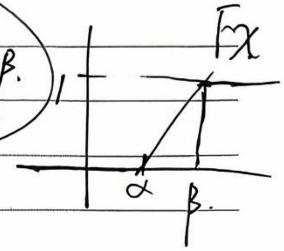
$$f_X(x)=\begin{cases}\frac{1}{\beta-\alpha}, & \alpha\le x\le\beta\\ 0, & \text{otherwise}\end{cases}$$

矩形面积为1

$\boxed{X\sim U[\alpha,\beta]}$ 标记.

What the fuck! 这什么分布规律??

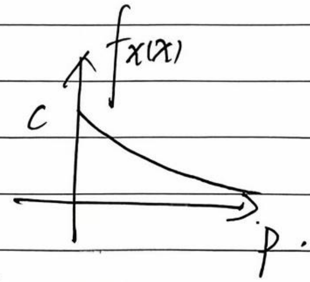
$$F_X(x) = \int_{-\infty}^{+\infty} f_X(x) = \begin{cases} 0 & x<\alpha. \\ \int_\alpha^x \frac{1}{\beta-\alpha}\,dy = \frac{x-\alpha}{\beta-\alpha}, & \alpha\le x\le\beta. \\ \int_\alpha^\beta \frac{1}{\beta-\alpha}\,dy = 1, & x\ge\beta. \end{cases}$$



Exponential distribution 指数分布

2) exponential Law (Exp. Distribution).

$$f_X(x) = \begin{cases} ce^{-\lambda x}, & x\ge 0 \\ 0 & x<0. \end{cases}$$



$$\cancel{1 = \int_{-\infty}^{+\infty} f_X(x)dx = \int_0^{+\infty} ce^{-\lambda x}dx =}$$

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x\ge 0 \\ 0, & x<0. \end{cases} \quad X \sim Exp(\lambda). \quad \boxed{\lambda e^{-\lambda x}}$$

1x

cal c place out 考试.

应用: X = survival time of $\begin{cases}\text{Random atoms} \\ \text{species} \\ \text{- -}\end{cases}$

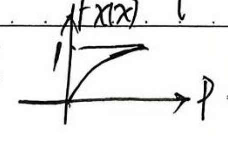
$$\boxed{P(X\ge x) = e^{-\lambda x}}$$

key: $P(X\ge x)$ = prob of survival up to x

$$P(X\ge x) = \int_x^{+\infty} f_X(y)\,dy = \int_x^{+\infty} \lambda e^{-\lambda y}dy = -e^{-\lambda y}\Big|_x^{\infty}$$

$$= 0-(-e^{-\lambda x}) = e^{-\lambda x}.$$

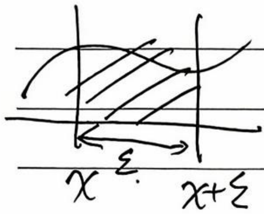
$$F_X(x) = P(X\le x) = 1-P(X\ge x) = 1-e^{-\lambda x}.$$



Application: $\lambda = ?$ = Instantaneous Death Rate.

$$= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} p(x \le X \le x+\varepsilon \mid X \ge x) \quad \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \int_x^{x+\varepsilon} f_X(x)$$

条件概率



$$= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[p(X \ge x) - p(X \ge x+\varepsilon) \right] \Big/ p(X \ge x).$$

$$= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[e^{-\lambda x} - e^{-\lambda(x+\varepsilon)} \right] \Big/ e^{-\lambda x}$$

$$= \lim_{\varepsilon \to 0} \frac{\frac{1}{\varepsilon} \cancel{e^{-\lambda x}} \left[1 - e^{-\lambda \varepsilon} \right]}{\cancel{e^{-\lambda x}}}$$

$$= \lim_{\varepsilon \to 0} \frac{1 - e^{-\lambda \varepsilon}}{\varepsilon} \Rightarrow \left. \frac{d\,e^{-\lambda x}}{dx} \right|_{x=0} = \lambda.$$

用洛必达.

$$\hookrightarrow \frac{\lambda e^{-\lambda \varepsilon}}{1} = \frac{\lambda e^{0}}{1} = \lambda.$$

Memoryless Property $X \sim Exp(\lambda)$ X服从指数分布

$$p(X > t+s \mid X > t) = \frac{p(X > t+s \text{ AND } X > t)}{p(X > t)} = \frac{p(X > t+s)}{p(X > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}}$$

$$= e^{-\lambda s} = p(X > s).$$

※ ⇒ 指数分布也可以具有无记忆性，之前的几何分布也具有无记忆性.

Date. Page.

Theorem: X is continuous, $\operatorname{Im} X = [0, +\infty)$

Then: X memoryless $\Leftrightarrow$ $X \sim \operatorname{Exp}(\lambda)$, some $\lambda > 0$.

无记忆性 $\Leftrightarrow$ 指数分布 是充要条件

proof: ($\Leftarrow$) ABOVE.

By Hypothesis

STEP 0 $\quad p(X > t+s) = p(X > t)\, p(X > s)$.

证明 但是没听懂

Iterating 迭代.

$\cancel{p(X > s_1 + s_2 + \cdots + s_n) = p(X > s_1)\, p(X > s_2) \cdots p(X > s_n)}$.

FUCK YOU

STEP 1 Apply (*) For $s_1 = s_2 = \cdots = s_n = 1$

$p(X > n) = p(X > 1)^n$.

To simplify Denote $g(t) = p(X > t)$

$g(n) = g(1)^n$.

~~STEP 2: (*) $s_1 = \cdots = s_n = \frac{1}{n}$~~

$g(1) = g(\frac{1}{n})^n$

$g(\frac{1}{m}) = g(1)^{\frac{1}{m}}$

step 4:

$x \in R(>0)$,

$\exists\, x_1, x_2 \cdots \to x$.

$g(x) = p(X > x)$ is continuous

$g(x) = g($

what the fuck ???

STEP 3: $g(\frac{p}{q}) = g(p \cdot \frac{1}{q}) \underset{(1)}{=} g(\frac{1}{q})^p \underset{(2)}{=} (g(1)^{\frac{1}{q}})^p = g(1)^{p/q}$.

~~From (1)–(3) $g(x) = g(1)^x$~~ $\quad x \in Q$.

conclusion:

$p(X > x) = g(1)^x$ ↗ $g(1) = 0 \Rightarrow X = 0$ ↘ $\lambda = -\ln g(1)$

Date. Page.

3) Paretto Law、帕累托分布 / 幂律分布

Fraction of people with income $> x \sim \frac{c}{x^{\alpha}}$

in probability teams $c = ?$

~~P(X≤x~~ $p(X > x) = \frac{c}{x^{\alpha}}$ $\quad x \geq 1$

$F_X(x) = p(X \leq x) = 1 - \frac{c}{x^{\alpha}} \quad x \geq 1$

$$f_X(x) = \begin{cases} \frac{\alpha c}{x^{\alpha+1}}, & x \geq 1 \\ 0, & \text{Else.} \end{cases}$$

$$c: \ 1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_1^{\infty} c\frac{\alpha}{x^{\alpha+1}}\,dx = c\alpha\int_1^{\infty} \frac{1}{x^{\alpha+1}}\,dx$$

$$= \alpha\left(-\frac{1}{x^{\alpha}}\right)\Big|_1^{\infty} = c[0-(-1)] = c.$$

DEF: X has par(α) if

$$f_X(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}}, & x \geq 1 \\ 0, & \text{else} \end{cases} \Rightarrow \ \therefore \text{Paretto law}$$

$p(X > \alpha) = \frac{1}{x^{\alpha+1}}$?

Date. Page.

Properties:

P) memoriless ? $x > x_0$

$$p(X>x \mid X>x_0) = \frac{p(X>x)}{p(X>x_0)} = \frac{\frac{1}{x^{\alpha}}}{\frac{1}{x_0^{\alpha}}} = \left(\frac{x_0}{x}\right)^{\alpha}.$$

DEF: $X \sim Par(x_0, 2)$ $\begin{cases} X \to \frac{X}{x_0} \\ =1 \to x_0 \text{ initial.} \end{cases}$

$$f_X(x) = \begin{cases} \frac{\alpha \, x_0^{\alpha+1}}{x^{\alpha+1}}, & x \geq x_0 \\ 0 & \text{Else.} \end{cases}$$

OR. $p(X>x) = \left(\frac{x_0}{x}\right)^{2}$

Date. Page.

* $X \sim pois(\lambda) \Rightarrow Y = \ln(X) \sim Exp(\lambda)$

4) Normal Law 正态分布

Le Moivre 1783: limit of Binomial → normal

PDF = X normal Distribution

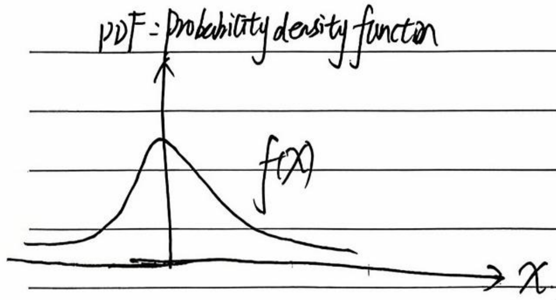
对称轴由中心对称轴定、

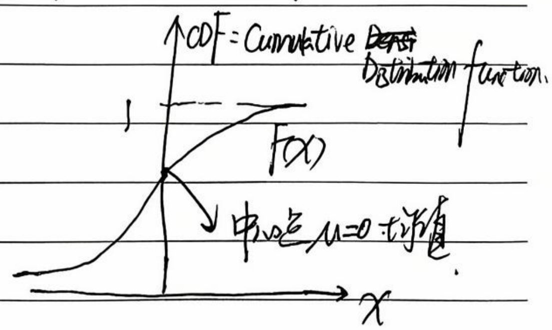
① $X \sim N(\mu, \sigma^2)$

标记 mean 均值 variations 方差

② $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ for $-\infty < x < \infty$) $= \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

③ σ^2 variation (VAR) 越大越胖矮，越小越瘦高.





④ $F_X(x) = \int_{-\infty}^{x} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma}$ ↗

$F_X(+\infty) = \int_{-\infty}^{+\infty} \frac{e^{-\frac{(x-\mu)^2}{2\sigma^2}}}{\sqrt{2\pi}\sigma} = 1$

⑤ 标准正态分布 $X \sim N(0,1)$

↑ 转换公式

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

$Z = \frac{X-\mu}{\sigma}$.

proof: $P(a \le Z \le b) = P(a \le \frac{X-\mu}{\sigma} \le b) = P(\mu + a\sigma \le X \le \mu + b\sigma)$

$= \frac{1}{\sqrt{2\pi}\sigma} \int_{\mu+\sigma a}^{\mu+\sigma b} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$. (Given $Z = \frac{X-\mu}{\sigma}$)

$\Downarrow$

$x = z\sigma + \mu$

$dx = d(z\sigma + \mu) = dz$

$= \frac{1}{\sqrt{2\pi}} \int_a^b e^{-\frac{z^2}{2}} dx$

$\Rightarrow \boxed{Z \sim N(0,1) \quad f_Z = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}} \Rightarrow$ 标准正态分布 PDF 公式.

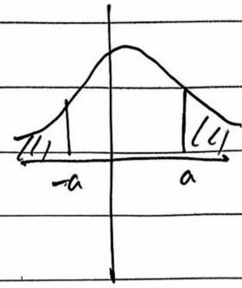
⑥ Tables of Calculators.

Either: $F_X(x) = P(X \le x)$ CDF: Cumulative Distribution Function.

$\Phi_X(x) = P(X \ge x)$. Tail Distribution Function.

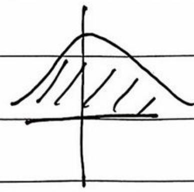
① key properties.

1) 对称性



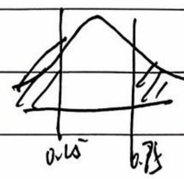
$p(z \le -a) = p(z \ge a)$

2) total area = 1



$p(x \le 0) = 0.5$

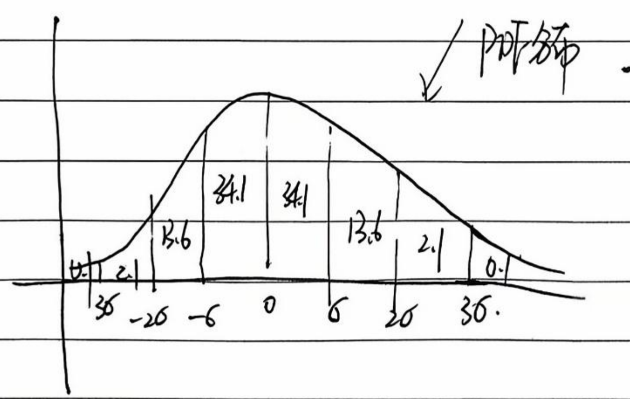
$p(x \le 0.25) = 1 - \psi(0.75)$.



$p(|z| \ge a) = p(z \ge a) + p(z \le a) = 2p(z \ge a)$

$p(z \ge 1)$

$p(z \ge 2)$



分布复习 (4种分布) Summary

Date. Page.

① 离散随机分布

$U[\alpha,\beta]$ $f_X(x) = \begin{cases} \frac{1}{\beta-\alpha}, & \alpha\le x\le\beta \\ 0, & \text{ELSE} \end{cases}$

② 指数分布

$Exp(\lambda)$ $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x\ge 0 \\ 0 & \text{else} \end{cases}$ $[p(X\ge x) = e^{-\lambda x}]$

③ 帕雷托/幂律分布

$par(\alpha) \to par(1,\alpha)$ $f_X(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}}, & x\ge 1 \\ 0 & \text{otherwise} \end{cases}$ $[p(X\ge x) = \frac{1}{x^\alpha}]$

$par(x_0,\alpha)$ $f_X(x) = \begin{cases} \alpha\left(\frac{x_0}{x}\right)^{\alpha+1}, & x\ge x_0 \\ 0, & \text{otherwise} \end{cases}$

④ 正态分布

$N(\mu,\sigma^2)$ $f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ $-\infty<x<+\infty$

$N(0,1)$ (标准正态分布) $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $-\infty<x<+\infty$

σ越小，越瘦高 σ越大，越矮胖；μ决定对称轴位置

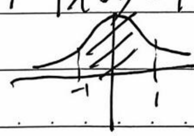
computations with normal law: 有关正态分布的计算

$x\sim N(\mu,\sigma^2) \Leftrightarrow z = \frac{x-\mu}{\sigma}$ 标准正态分布的转化.

μ=mean (平均) σ^2=variance 方差 σ = standard deviation 标准差

$T(\alpha) = p(X\ge\alpha) = 1 - F_X(\alpha) = 1 - p(X\le\alpha)$ "="取不取一样因为单点概率

$x\sim N(0,1)$



$\Rightarrow p(|Z|\le 1) = p(-1\le Z\le 1)$

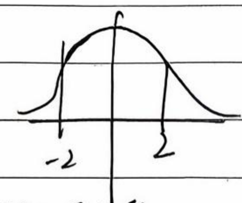
$= 1-p(Z\ge 1 \text{ or } Z\le -1) = 1-2p(Z\ge 1)$

$= 1-2T(1) = 1-2\cdot 0.1587 = 0.6826$

⊕ 正态分布的对称性

$$-1 \le z \le 1 \xrightarrow{z=\frac{x-\mu}{\sigma}} -1 \le \frac{x-\mu}{\sigma} \le 1 \rightarrow -\sigma+\mu \le x \le \sigma+\mu.$$

$$p(|z| \le 1) = p\left(\left|\frac{x-\mu}{\sigma}\right| \le 1\right) = p(|x-\mu| \le \sigma).$$



$$p(-2 \le z \le 2) = p(|z| \le 2) = 1-2p(z \ge 2)$$

$$= \underbrace{1-2T(2) = 1-2\times 0.02}_{\text{读正态分布表.}} = 0.9544 = p(-2\sigma \le x-\mu \le 2\sigma).$$

※ $T(\alpha) = p(x \ge \alpha)$，$\Phi(\alpha) = p(x \le \alpha)$ 两种表达范围相反

Example: 分布是否指数？

S1) Quantities 分位数

Radium $p(x > t) = e^{-\lambda t}$

x = survival probability of 1 ATOM.

$x \sim Exp(\lambda)$ $\lambda = 0.043$ centuries^{-1} t in centuries.

In General, if initial Quantity M_0.

$M(t)$ = mass surviving at time t

$M(t) = M_0 e^{-\lambda t}$

$$\boxed{M(t) = \underset{\text{原始}}{M_0}\, e^{-\underset{\text{参数}}{\lambda} \underset{\text{时间}}{t}}} \Rightarrow \text{指数衰减模型.}$$

需要多少世纪才能只剩10%　　　　指数衰减模型.

How long for 90% of the mass to disappear?

$\Rightarrow A:$ $m(t) = 0.9m_0$ $\qquad \Rightarrow 0.1m_0 = m_0 e^{-\lambda t_{0.9}}$ ln运算

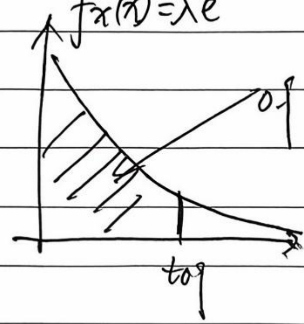
$p(x>t) = 0.1$

$\Rightarrow t_{0.9} = -\frac{1}{\underset{0.043}{\lambda}} \ln 0.1 \approx 53.5$ centuries

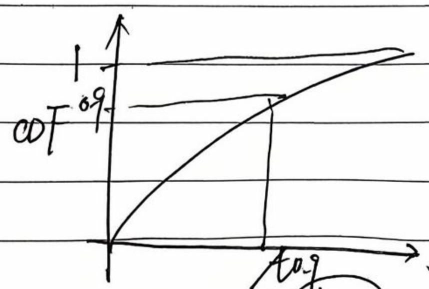
$t_{0.9}$?

$f_X(x) = \lambda e^{-\lambda t}$.

PDF



CDF



$t_{0.9}$ is the solution of $p(x \le t_{0.9}) = 0.9$

$F_X(t_{0.9}) = 0.9$

⑤ $x > 3$ $F_X(x) = \cdots = 1$.

$\Rightarrow$ 推广到: $p(x \le t_\alpha) = \alpha \Leftrightarrow F_X(t_\alpha) = \alpha$.

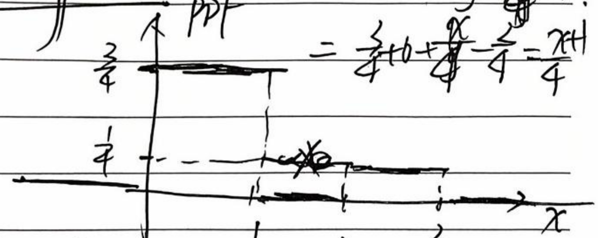
④ $2 \le x \le 3$ $F_X(x) = \cancel{\int_2^x \frac{1}{4} dx}$

$= \int_0^1 \frac{3}{4} dx + \int_1^2 0\, dx + \int_2^x \frac{1}{4} dx$

$= \frac{3}{4} + 0 + \frac{x}{4} - \frac{2}{4} = \frac{x+1}{4}$

Warning example: X with Density

$$f_X(x) = \begin{cases} 3/4 & 0 \le x \le 1 \\ 1/4 & 2 \le x \le 3 \\ 0 & \text{else} \end{cases}$$



$F_X(x) = \int_{-\infty}^{x} f_X(y)\, dy = \int_{-\infty}^{y} f_X(y)\, dy$

① $x < 0$ $F_X(x) = 0$.

② $0 \le x \le 1$ $F_X(x) = \int_0^x \frac{3}{4} dx = \frac{3}{4} x$

③ $0 \le x \le 2$ $F_X(x) = \int_0^1 \frac{3}{4} dx + \int_1^x 0\, dx$

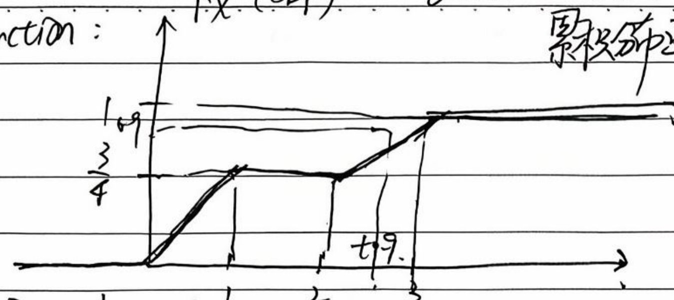
$= \frac{3}{4}$.

不是 atom 混合的.

Date. Page.

Cumulative Distribution Function 累积分布函数

The CDF Function:



$t(0.9) = ?$ $p(X \le t_{0.9}) = 0.9 \Leftrightarrow F_X(t_{0.9}) = 0.9$.

$\frac{3}{4} + \frac{1}{4} t_{0.9} = 0.9 \Rightarrow t_{0.9} = 2.6$.

$t(0.75)$? $p(X \le t_{0.75}) = 0.75 \Leftrightarrow F_X(t_{0.9}) = 0.9$.

commentation: take smallest solution.

DEF: 定义. If X is a ~~solution~~ continuous RV and $p \in [0,1]$, The p-th Quantile or 100 p-th percentile is the ~~result of~~ value x_p, which the smallest solution of $F_X(x_p) = p$

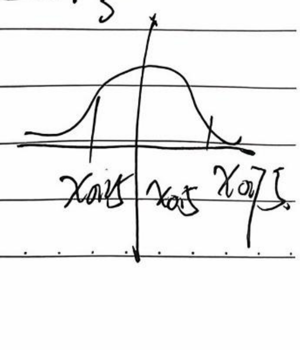
连续随机变量的分位数 (Quantile) $\Rightarrow$ $p(X \le t_p) = F_X(t_p) = p$.

重要分位数 Quantile

$x_{\frac{1}{2}}$ = Medium

25th percentile = 1st Quartile 第25百分位数 = 第1四分位数

75th percentile = 3rd Quartile .75 —— 3



Date. Page.

5.2) Expectations 期望 — mean 平均值 / variance 方差 / ⋮

[1] X is Discrete R.V $\Rightarrow$ $\bar{X} = \sum_i x_i \, p_X(x_i) = \sum_i x_i \times p(X = x_i)$ — 数值 权重.

~~$\Rightarrow$ 加权平均数 Weighted Average.~~

eg: $\bar{X} = \frac{1}{7}\cdot 8 + \frac{2}{7}\times 9 + \frac{2}{7}\times 10 + \frac{1}{7}\times 7 + \frac{1}{7}\times 6$ — 数值

权重.

$\boxed{\text{DEF}}$: If X is a R.V, The mean value of X / expected value of X / expectation of X

is $E(X) = \begin{cases} \sum_i x_i p_X(x_i) & , X \text{ Discrete} \\ \int_{-\infty}^{+\infty} x f_X(x)\, dx & , X \text{ continuous} \end{cases}$

$\Rightarrow$ key: find $E[X]$ for different laws 不同分布期望的计算方式不同

$\begin{pmatrix} 0 & 1 \\ p(X=0) & p(X=1) \\ 1-p & p \end{pmatrix}$

E1) $X \sim Ber(p)$ X服从伯努力分布

$E(X) = 1\cdot p_X(1) + 0\cdot p_X(0) = 1\cdot p(X=1) + 0\cdot p(X=0)$

$= 0\cdot(1-p) + 1\cdot p = p.$

E2) $X \sim Geom(p)$ $x_i = 1, 2, 3 \cdots \Rightarrow p(k) = (1-p)^{k-1}p.$

$E(X) = \sum_{k=1}^{\infty} k\, p_X(k) = \sum_{k=1}^{\infty} k(1-p)^{k-1}p = p\sum_{k=1}^{+\infty} k(1-p)^{k-1}.$

$= p\sum_{k=0}^{+\infty}\left[-\frac{d}{dp}(1-p)^k\right] = p\left[\frac{d}{dp}\sum_{k=0}^{\infty}(1-p)^k\right] =$

(需求导形式 才好化简公式)

Date. Page.

级数/极限: $\sum_{k=0}^{n} \alpha^k = \frac{1-\alpha^{n+1}}{1-\alpha}$.

PF: $S_n = 1+a+a^2+ \cdots +a^n$

$aS_n = a+a^2+a^3+ \cdots +a^n+a^{n+1}$

$S_n - a(S_n) = 1-a^{n+1}$

$\Rightarrow S_n = \frac{1-a^{n+1}}{1-a} = \frac{1-\alpha^{n+1}}{1-\alpha}$

计 $|a|<1$ $a^N \xrightarrow[N\to+\infty]{} 0$ → 指数函数 → 0.

Hence $\sum_{k=0}^{+\infty} a^k = \lim_{n\to+\infty} \sum_{k=0}^{n} a^k = \lim_{n\to+\infty} \frac{1-a^{n+1}}{1-a} = \frac{1}{1-a}$. 常见结论

书接上回: $\sum_{k=0}^{+\infty} (1-p)^k = \frac{1}{1-(1-p)} = \frac{1}{p}$

$E[X] = p\left[-\frac{d}{dp} \sum_{0}^{k\to+\infty} (1-p)^k\right] = p\left[-\frac{d}{dp} \frac{1}{p}\right]$

$= p \cdot \frac{1}{p^2} = \frac{1}{p}$

$\Rightarrow$ 几何分布 $X \sim Geom(p)$ $E[X] = \frac{1}{p}$.

✓

Date.　　Page.

E3) $X \sim Bin(n,p)$ 二项分布的期望计算

$$E[X] = \sum_{k=0}^{m} k \cdot \binom{m}{k} p^k (1-p)^{n-k} = \sum_{x=0}^{m} k \cdot \underbrace{p(x=k)}_{\text{二项分布概率}}$$

$$s(x,y) = \sum_{k=0}^{n} k \cdot \binom{n}{k} x^k y^{n-k} \qquad \underbrace{\sum_{k=0}^{m} \binom{n}{k} x^k y^{n-k} = (x+y)^n}_{\text{二项式定理}}$$

$$= x \sum_{k=1}^{m} k \binom{n}{k} x^{k-1} \cdot y^{n-k}$$

$$= x \frac{d}{dx} \sum_{k=0}^{n} \binom{n}{k} x^k \cdot y^{n-k} \times \frac{d}{dx}[(x+y)^n]$$

$$= x\, n (x+y)^{n-1}$$

$$E[X] = s(p, 1-p) = pn[\cancel{p+1-p}]^{n-1} = np.$$

⇒ 二项分布期望 $X \sim B(n,p)$, $E[X] = np$.

E4) $X \sim pois(\lambda)$ 泊松分布的期望计算

$$E[X] = \sum_{n=0}^{\infty} n \cdot \frac{e^{-\lambda} \lambda^n}{n!} = e^{-\lambda} \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} = e^{-\lambda} \sum_{n=1}^{\infty} n \frac{\lambda^n}{n!}$$

期望的定义 ⇒ $\sum_{n=0}^{\infty} n \cdot p(x=n)$

$$= e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^n}{(n-1)!} = e^{-\lambda} \sum_{n=1}^{+\infty} \frac{\lambda^{n-1}}{(n-1)!}$$

⇒ 泊松分布期望 $X \sim pois(\lambda)$, $E[X] = \lambda$

$$= \lambda e^{-\lambda} \underbrace{\sum_{m=0}^{\infty} \frac{\lambda^n}{n!}}_{\text{级数}} = \lambda \cdot e^{-\lambda} \cdot e^{\lambda}$$

$$= \lambda$$

Date. Page.

Ex) $X \sim \exp(\lambda)$ 指数分布 (连续分布)

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_0^{+\infty} \underset{\underset{x}{\downarrow}}{x}\,\underset{\underset{\text{PDF } f_X(x)}{\downarrow}}{\lambda e^{-\lambda x}} = \lambda \int_0^{\infty} \underbrace{x}_{u}\,\underbrace{e^{-\lambda x}dx}_{dv}$$

$u=x \quad du=dx \;\Big|\; dv = e^{-\lambda x}dx \quad v = \dfrac{e^{-\lambda x}}{-\lambda}$

$$\Rightarrow \lambda\left[\frac{xe^{-\lambda x}}{-\lambda}\Big|_0^{+\infty} - \int_0^{+\infty}\frac{e^{-\lambda x}}{-\lambda}dx\right]$$

$$\Rightarrow E[X] = \lambda\left[\frac{xe^{-\lambda x}}{-\lambda}\right] \ldots\ldots = \frac{1}{\lambda}$$

$\Rightarrow$ 指数分布期望. $X \sim Exp(\lambda) \Rightarrow E(X) = \frac{1}{\lambda}$

Ex) $X \sim N(\mu, \sigma^2)$ 正态分布 (连续分布)

$$E[X] = \int_{-\infty}^{+\infty} x e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi}\sigma} \qquad z = \frac{x-\mu}{\sigma} \quad dz = \frac{dx}{\sigma}$$

$$= \int_{-\infty}^{+\infty} \cancel{(\mu+\sigma z)}\, \mu e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{+\infty} \sigma z e^{-\frac{z^2}{2}} \frac{dz}{\sqrt{2\pi}}$$

$$= \underbrace{\mu\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}}_{=1} + \underbrace{\sigma\int_{-\infty}^{+\infty} z e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}}_{0}$$

$E[X] = \mu$ 就是均值 μ.

$\Rightarrow$ 正态分布期望 $X \sim N(\mu, \sigma^2) \Rightarrow E[X] = \mu$.

Date. Page.

Ex) $X \sim Par(\alpha)$ 帕累托分布（连续分布） $Par(1,\alpha)$ 默认 $x_m = 1$ $\alpha > 0$

$$E[X] = \int_{-\infty}^{+\infty} x f_X(x)\,dx = \int_1^{+\infty} x \frac{\alpha}{x^{\alpha+1}}\,dx = \alpha \int_1^{+\infty} \frac{dx}{x^{\alpha}} = \alpha \left(-\frac{1}{\alpha-1} \frac{1}{x^{\alpha-1}} \right) \Big|_1^{+\infty}$$

$\frac{1}{x^{\alpha-1}}$ → $x \to \infty$ 0 对 $x>1$

→ $x \to +\infty$ $+\infty$ $0 < x \le 1$

帕累托分布期望 $X \sim par(1,\alpha)$

$$\Rightarrow E[X] = \begin{cases} \frac{\alpha}{\alpha-1} & x > 0 \\ D.N.E & 0 < \alpha \le 1 \end{cases}$$ (Do careful $\alpha = 1$)

D.N.E: Do not Exist

More Generally: 期望通式.

$$E(g(X)) = \begin{cases} \sum_i g(x_i)\, p_X(x_i), & X \text{ discrete} \\ \int_{-\infty}^{+\infty} g(x_i) f_X(x_i)\,dx, & X \text{ continuous} \end{cases}$$

~~Expectation of g:~~ 没看懂 😊.

~~useful expectations:~~

~~$E(X^m)$ $m \ge 2$~~

$D(X) = E(X^2) - E^2(X)$?

Date. Page.

$VAR(X) = E\left[(X - E(X))^2\right]$ = variance of X 方差 $= \sigma^2$

DEF: The variance of A R.V X is ~~$Var(X) = E((X-\mu$~~

$$Var(X) = E\left((X - E(X))^2\right)$$

Example: $X \sim N(\mu, \sigma^2)$ X服从正态分布 $E(X) = \mu$ $\Rightarrow Var(X) = E[(X-\mu)^2]$

$$Var(X) = \int_{-\infty}^{+\infty} (x-\mu)^2 e^{-\frac{(x-\mu)^2}{2\sigma^2}} \frac{dx}{\sqrt{2\pi}\sigma}$$

$=$ —— $=$ 一顿不太复杂的计算之后脑补推理 $= \sigma^2$.

$\Rightarrow$ 正态分布 $X \sim N(\mu, \sigma^2)$ $Var(X) = \sigma^2$ 方差.

DEF: $\sigma = \sqrt{\sigma^2}$ standard Deviation of X X的标准差.

Properties of Mean and Variance.

(1) $\boxed{Y = aX + b}$ ~~$\Rightarrow E(Y) = aE(X) + b$~~

$\boxed{E(Y)} = E(aX+b) = \int_{-\infty}^{+\infty} (ax+b) f_X(x)\,dx = a\int_{-\infty}^{+\infty} x f(x)\,dx + b\int_{-\infty}^{+\infty} f_X(x)\,dx$

? ?

$= \dots = \boxed{aE(X) + b}$

$Var(aX+b) = a^2 Var(X)$.

$Z = X - \mu$.

$E(Z) = E(X) - \mu = 0$.

$Z = \frac{X-\mu}{\sigma}$

$E(Z) = E\left(\frac{X}{\sigma} - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma}E(X) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0$. 期望为0. $\Rightarrow$ 标准正态分布

以后说~..

Summary

$$E(g(x)) = \begin{cases} \sum_{x_i} g(x_i) \cdot f_X(x_i) & , X \text{ Discrete} \\ \int_{-\infty}^{+\infty} g(x) f_X(x) dx & , X \text{ continious} \end{cases}$$

Date. Page.

Mean $E(X) = \mu$.

$E(aX+b) = aE(X)+b$.

Variance $= Var(X) = E[(X-E(X))^2]$ → 标准公式

1) $X \sim N(\mu, \sigma^2)$

eg: $X \sim N(\mu, \sigma^2)$, $\mu = E(X)$, $\sigma^2 = Var(X)$

standard variation $= \sqrt{\sigma^2} = \sigma$.

2) $X \sim Par(\alpha)$ $\alpha > 0$ $f_X(x) = \begin{cases} \frac{\alpha}{x^{\alpha+1}} , & x \geq 1 \\ 0 . & \text{else} \end{cases}$

? 没看懂

$0 < \alpha \leq 1$ $E(X)$ Not Declined (?)

$1 < \alpha \leq 2$ $E(X)$ DEF But $E(X^2)$ 23 NOT ?

$\vdots$

$n \leq \alpha \leq \alpha+1$ $E[(X)] \geq E(X^2)$ ---- $E(X^2)$ DEF

But $E(X^{n+1})$ NOT

Date. Page.

Properties of Var(X): 方差的属性、 P1、P2、P3 求.

P1) $y = ax+b$ Above $E[Y] = aE[X]+b$, $Var(y)$?

proof: $\underbrace{E(ax+b) = aE(X)+b \; ; \; Var(ax+b) = a^2 Var(X)}$

$\Rightarrow E(X) = \mu \quad E(y) = a\mu+b$

$$Var(y) = E[(y-E(y))^2] = E\left[\left[(ax+b)-(a\mu+b)\right]^2\right] = E\left[\left[a(x-\mu)\right]^2\right]$$

$$= E[a^2(x-\mu)^2] = a^2 E[(x-\mu)^2] = a^2 Var(X)$$

P2). let X with mean μ and variance σ^2

consider: $z = \frac{x-\mu}{\sigma} = \frac{1}{\sigma}x - \frac{\mu}{\sigma}$

$$E(z) = \frac{1}{\sigma}E(x-\mu) = \frac{1}{\sigma}E(x) - \frac{\mu}{\sigma} = \frac{\mu}{\sigma} - \frac{\mu}{\sigma} = 0.$$

$$Var(z) = Var(\underbrace{\frac{1}{\sigma}}_{a}x - \underbrace{\frac{\mu}{\sigma}}_{b}) = \frac{1}{\sigma^2}Var(X) = \frac{1}{\sigma^2}\sigma^2 = 1$$

☆ 标准分布的期望/均值为0. 方差/标准差为1

P3) Alternative Formular for ~~E(X)~~ Var(X)

$$Var(X) = E[(x-\mu)^2] = E(x^2 - 2\mu x + \mu^2) = E[x^2] - 2\mu \underbrace{E[x]}_{=\mu} + \mu^2$$

$$= E(x^2) - 2\mu^2 + \mu^2 = E(x^2) - \mu^2 = E(x^2) - [E(x)]^2.$$

☆ 方差替代公式: $Var(X) = E(X^2) - [E(X)]^2$.

OK 别停啊，推导一点没看懂。 → 这个很高级 方差关联上了

Date.　　Page.

Moment-generating Function　MGF 矩生成函数

— The MGF of a RV X is

$$\phi_X(t) = E(e^{tX}) \quad \text{矩生成函数}$$

comments:

<1> not every law has $\phi_X(t)$

<2> can be used to compute all moments through Derivatives

Indeed:

$$\phi'_X\Big|_{t=0} = \frac{d}{dt}E(e^{tX}) = E\left(\frac{d}{dt}e^{tX}\right)\Big|_{t=0}$$

$$= E(Xe^{tX})\Big|_{t=0} = E(X)$$

$$\frac{d^2}{dt^2}\phi_X\Big|_{t=0} = E\left(\frac{d^2}{dt^2}e^{tX}\right)\Big|_{t=0} = E(X^2e^{tX})\Big|_{t=0}.$$

$$= E(X^2)$$

⇒ 类比

$$\frac{d^n}{dt^n}\phi_X(t)\Big|_{t=0} = E(X^n e^{tX})\Big|_{t=0} = E(X^n)$$

E1) $X \sim \text{Exp}(\lambda)$　$[E(X) = \frac{1}{\lambda}\quad Var(X) = ?]$

$$E(X) = \phi'(t)\Big|_{t=0} = \frac{\lambda}{(\lambda-t)^2}\Big|_{t=0} = \frac{\lambda}{\lambda^2} = \frac{1}{\lambda}$$

$$E(X^2) = \phi''(t)\Big|_{t=0} = \frac{d}{dt}[\ '(t)]\Big|_{t=0} = \frac{d}{dt}\left(\frac{\lambda}{(\lambda-t)^2}\right)\Big|_{t=0}.$$

$$= \frac{2\lambda}{(\lambda-t)^3}\Big|_{t=0} = \frac{2\lambda}{\lambda^3} = \frac{2}{\lambda^2}$$

Date. Page.

$Var(X)=E(X^2)-E(X)^2=\frac{2}{\lambda^2}-(\frac{1}{\lambda})^2=\frac{1}{\lambda^2}$

指数分布 $E(X)=\frac{1}{\lambda}$, $Var(X)=\frac{1}{\lambda^2}$ $X\sim Exp(\lambda)$

※ 补充: $E(X^n)=\frac{n!}{\lambda^n}$. 指数函数 $E(X^n)$ 通式.

Ex) $X\sim Bin(n,p)$ $p_X(k)=\binom{n}{k}p^k(1-p)^{n-k}$

$E(X)=np$

$\phi_X(t)=E(e^{tX})=\sum_{k=0}^{n}e^{tx}p_X(k)=\sum_{k=0}^{n}e^{tk}\binom{n}{k}p^k(1-p)^{n-k}$

$=\sum_{k=0}^{n}\binom{n}{k}(pe^t)^k(1-p)^{n-k}=[pe^t+(1-p)]^n=[1+p(e^t-1)]^n$.

[用二项式定理: $(x+y)^n=\sum_{k=0}^{n}\binom{n}{k}x^ky^{n-k}$]

$E(X)=\frac{d}{dt}\phi_X(t)\Big|_{t=0}=n[1+\underbrace{p(e^t-1)}_{0}]^{n-1}\cdot \underbrace{pe^t}_{1}\Big|_{t=0}=np$.

$E(X^2)=\frac{d^2}{dt^2}\phi_X(t)\Big|_{t=0}=\frac{d}{dt}[n[1+p(e^t-1)]^{n-1}pe^t]\Big|_{t=0}$

$=n(n-1)[1+p(e^t-1)]^{n-2}pe^t+n[1+p(e^t-1)]^{n-1}pe^t\Big|_{t=0}$.

$=n(n-1)p^2+np=n^2p^2+np(1-p)$.

$Var(X)=E(X^2)-E(X)^2=n^2p^2+np(1-p)-(np)^2=np(1-p)$.

二项分布 $E(X)=np$. $Var(X)=np(1-p)$ $X\sim Bin(n,p)$.

我始终没搞明白$\phi_X(t)$到底是什么?和谁有什么关系.?指数也看麻了

Date.　　　Page.

E3) $X \sim pois(\lambda)$　~~$p_X(x)$~~ $p_X(k) = e^{-\lambda}\frac{\lambda^k}{k!}$, $k \geq 0$.

$E(X) = \lambda$

$$\phi_X(t) = E(e^{tX}) = \sum_{k=0}^{+\infty} e^{tk}\frac{e^{-\lambda}\lambda^k}{k!} = e^{-\lambda}\sum_{k=0}^{\infty}\frac{(e^t\lambda)^k}{k!}$$

由 $\sum_{k\geq 0}\frac{x^k}{k!} = e^x$ $= e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$

$$E(X) = \phi'(t)\big|_{t=0} = \lambda e^t e^{\lambda(e^t-1)}\big|_{t=0} = \lambda$$

$$E(X^2) = \frac{d}{dt}\left[\lambda e^t e^{\lambda(e^t-1)}\right]\Big|_{t=0} = \lambda e^t e^{\lambda(e^t-1)} + \lambda e^t\lambda e^t e^{\lambda(e^t-1)}\Big|_{t=0}$$

$= \lambda + \lambda^2$

$Var(X) = E(X^2) - E(X)^2 = \lambda + \lambda^2 - \lambda^2 = \lambda$.

$\Rightarrow$ 泊松分布 $E(X) = \lambda$　$Var(X) = \lambda$　$X \sim pois(\lambda)$

E4) $X \sim N(0,1)$ 标准正态分布

$$\phi_X(t) = E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx}e^{-\frac{x^2}{2}}\frac{dx}{\sqrt{2\pi}} = \int_{-\infty}^{+\infty} e^{-\frac{x^2}{2}+tx}\frac{dx}{\sqrt{2\pi}}$$

$= -\frac{x^2}{2} + tx = -\frac{1}{2}[x^2 - 2tx] = \frac{1}{2}[x^2 - 2tx + t^2 - t^2]$

$= -\frac{1}{2}[(x-t)^2 - t^2] = \frac{t^2}{2} - \frac{1}{2}(x-t)^2$.

$$\phi_X(t) = \int_{-\infty}^{+\infty} e^{\frac{t^2}{2} - \frac{(x-t)^2}{2}}dx = e^{\frac{t^2}{2}}\int_{-\infty}^{+\infty} e^{-(x-t)^2/2}\frac{dx}{\sqrt{2\pi}}$$

$= e^{t^2/2}\int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}}\frac{dz}{\sqrt{2\pi}}$ (分部积分法) $= e^{\frac{t^2}{2}}$.

Date.　　　Page.

$X \sim N(0,1) \Rightarrow \phi_X(t) = e^{t^2/2}$

$E(X) = \phi'(t)\Big|_{t=0} = te^{t^2/2}\Big|_{t=0} = 0.$

$E(X^2) = \frac{d}{dt}\phi'(t)\Big|_{t=0} = e^{t^2/2} + t^2e^{t^2/2}\Big|_{t=0} = (1+t^2)e^{t^2/2}\Big|_{t=0}$

$= 1$

$Var(X) = E(X^2) - E(X)^2 = 1^2 - 0^2 = 1$

※ 标准正态分布 $E(X) = \mu = 0$; ~~E(X)~~ $Var(X) = \sigma^2 = 1$
$X \sim N(0,1)$

※补充 $E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) = E(X^3) = \frac{d}{dt}\phi''(t)\Big|_{t=0}$

$= \frac{d}{dt}\left[(1+t^2)e^{t^2/2}\right]\Big|_{t=0} = 0$

$E\left(\left(\frac{t-\mu}{\sigma}\right)^n\right) = E(X^n) = \frac{d^n}{dt^n}\phi_X(t)\Big|_{t=0} = ?$

看不懂!!!! 解答一下!!!

$\cancel{E(e^{tX}) = \int_{-\infty}^{+\infty} e^{tx} f_X(x)}$ Variance & expectation.

Var(X) E(X)

Moment Generating function mGF.

law	$\phi_X(t)$
Ber(p)	$1-p(e^t-1)$
Geom(p)	$\frac{pe^t}{1-(1-p)e^t}$
Bin(n,p)	$[1-p(e^t-1)]^n$
Pois(λ)	$e^{\lambda(e^t-1)}$
exp(λ)	$\frac{\lambda}{\lambda-t}$
N(0,1)	$e^{x^2/2}$
N(μ,σ²)	$e^{\mu t+\sigma^2 t^2/2}$
par(α)	D.N.E

Date. Page.

Joint laws 联合分布: 卡方

引入: Two R.V x and y, seperately → $p(x)$ $f_X(x)$ $p(y)$, $f_Y(y)$

Jointly 是啥意思?

example: consider 10000 people, 调查收入与教育水平的关系.
Y X

$y = \begin{cases} 1 & \text{income} \le 1 \text{ min salary} \\ 2 & \text{income between 1 and 5 min salary} \\ 3 & > 5 \text{ min salary} \end{cases}$

$x = \begin{cases} 1 & \text{primary} \\ 2. & \text{High school} \\ 3. & \text{Terciary 职校/大专.} \\ 4 & \text{university} \end{cases}$

How to organize these information?

marginals 大数据.

$x \backslash y$	1	2	3	N_X
1	2820	1275	5	4100
2	1620	3005	125	4800
3	50	200	270.	600
4	10	160	250.	480
N_y	4500	5200	800	10000

300 people
middle income
Terciary ed.

概率表、联合随机

$X \backslash Y$	1	2	3	P_X
1	0.282	0.1275	0.005	0.41
2	0.162	0.3005	0.0275	0.48
3	0.005	0.03	0.027	0.062
4	0.001	0.012	0.035	0.028
P_Y	0.4	0.52	0.08	1

DEF: i) X, Y Discrete R.V. X, Y为离散随机变量.

The joint probability mass is the function

$$p_{X,Y} = \quad Im(X) \times Im(Y) \to [0,1]$$

$$\underbrace{p_{X,Y}}_{\text{大写}}\underbrace{(x,y)}_{\text{小写}} = P(\underset{大}{X} = \underset{小}{x}, \underset{大}{Y} = \underset{小}{y})$$

且 and.

marginal masses:

$$p_X(x) = \sum_y p_{X,Y}(x,y)$$

$$p_Y(y) = \sum_x p_{X,Y}(x,y)$$

comments:

<1> $p_{X,Y}$ or $f_{X,Y}$ characterize Joint laws

<2> normalization Conditions

ii) X, Y continuous RV $\Rightarrow$ Their joint probability

is a function $f_{X,Y}^{(x,y)} = R \times R \Longrightarrow \mathbb{R}_{\geq 0} = [0, +\infty)$

such that

$$P(X \in [a_1, b_1], Y \in [a_2, b_2]) = \int_{a_1}^{b_1} dx \int_{a_2}^{b_2} dy\, f(x,y)$$

Date. Page.

comments:

<2> normalization condition

$$\sum_{x,y} p_{x,y}(x,y) = 1$$

$$\int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy \, f_{x,y}(x,y) = 1$$

这讲什么?

<3> in General

$$p\big((x,y) \in \underset{\hat{R}^2}{C}\big) = \begin{cases} \sum_{x,y:(x,y)\in C} p_{x,y}(x,y) \\ \int dx \int dy \, f_{x,y}(x,y) \end{cases}$$

Example:

$$p(x \le y) = \begin{cases} \sum_{a} \sum_{b \ge a} p_{x,y}(x,y) \\ \int_{-\infty}^{+\infty} dx \int_{x}^{\infty} dy \, f_{x,y}(x,y) \end{cases}$$

DEF: Given $p_{x,y}$ or $f_{x,y}$, the marginal laws are defined by

$$p_x(x) = \sum_y p_{x,y}(x,y)$$

$$p(x \in [a,b]) = \int_a^b dx \underbrace{\int_{-\infty}^{+\infty} f_{x,y}(x,y) \, dy}$$

$$f_x(x) = \int_{-\infty}^{+\infty} f_{x,y}(x,y) dy$$

DEF: The joint Cumulative Distribution Function is CDF.

$F_{X,Y} : \mathbb{R} \times \mathbb{R} \to [0,1]$

$$F_{X,Y}(x,y) = P(X \le x, Y \le y)$$

* Knowing $F_{X,Y}$ is Enough to Knowing $P_{X,Y}$ or f_X

$$F_{X,Y}(a,b) = \begin{cases} \sum_{x \le a} \sum_{y \le b} P_{X,Y}(x,y) & x,y \text{ discrete} \\ \int_a^{+\infty} dx \int_b^{+\infty} dy\, f_{X,Y}(x,y), & x,y \text{ continuous} \end{cases}$$

$F \to f$ < Differences / Derivatives

marginal Dist Function.

$$F_X(x) = P(X \le x) = P(X \le x, Y < \infty) = ?$$

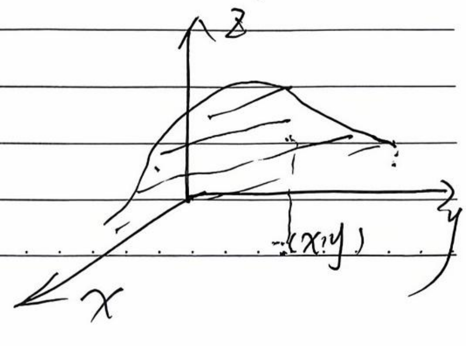
$$\boxed{P(X_1 \in [a_1,b_1] \cdots X_n \in [a_n,b_n]) = \int_{a_1}^{b_1} dx_1 \int_{a_n}^{b_2} dx_2 \cdots \int_{a_n}^{b_n} dx_n\, f_{X_1 \cdots X_n}(x_1 \cdots x_n)}$$

Joint distribution function:

$$\boxed{F_{X_1 \cdots X_n}(x_1 \cdots x_n) = P(X_1 \le x_1, X_2 \le x_2 \cdots X_n \le x_n)}$$?

2R.V $f_{X,Y}(x,y)$ is a surface.

是个平面.



Date. Page.

Crash course on $\boxed{\text{MVC.}}$

Derivatives Integrals

Example:

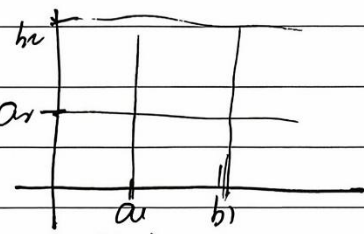
$$f(x,y) = \cancel{x^2y} \; x^2y^2 + xy + y$$

notation $\dfrac{\partial f}{\partial x} = 2xy^2 + y$ 偏导

$$\frac{\partial f}{\partial y} = 2x^2y + x + 1$$

Example:

$$\iint_{\square} f(x,y)\,dxdy$$



$a_1 \le x \le b_1$

$a_2 \le x \le b_2.$

$a_2 \le y \le b_2$

$a_2 \le x \le b_1.$

$$\int_{a_1}^{b_1} dx \int_{a_2}^{b_2} dy \; f(x,y)$$

Date.　　Page.

Summary: Joint expectations $x_1 \cdots x_n$ R.V

① $E[g(x_1 \cdots x_n)] = \begin{cases} \sum_{x_1} \cdots \sum_{x_n} g(x_1 \cdots x_n) p_{x_1 \cdots x_n}(x_1 \cdots x_n) & 离散 \\ \int_{-\infty}^{+\infty} dx_1 \cdots \int_{-\infty}^{+\infty} dx_n \, g(x_1 \cdots x_n) f_{x_1 \cdots x_n}(x_1 \cdots x_n) & 连续 \end{cases}$

② R.V的独立性: $x_1 \cdots x_n$ 是独立的当以下的条件为真

General 通式.

i) $\{a_1 \le x_1 \le b_1\} \cdots \{a_n \le x_n \le b_n\}$ 独立 for all 对于所有的 a_i, b_i

ii) 特别的.

$$F_{x_1 \cdots x_n}(x_1 \cdots x_n) = F_{x_1}(x_1) \cdots F_{x_n}(x_n)$$

iii) $E(F_1(x_1) \cdots F_n(x_n)) = E(F_1(x_1)) \cdots E(F_n(x_n))$ 对于所有的 $F_1 \cdots F_n$ 函数

iv) If jointly discrete / continuous 联合分布下的离散/独立.

$$p_{x_1 \cdots x_n}(x_1 \cdots x_n) = p_{x_1}(x_1) \cdots p_{x_n}(x_n)$$

$$f_{x_1 \cdots x_n}(x_1 \cdots x_n) = f_{x_1}(x_1) \cdots f_{x_n}(x_n)$$

③ $cov(x,y)$ 协方差 ~~$cov(x,y) = E((x-\mu_x$~~

$\Rightarrow \mu_x = E(x)$ / 线性性质.

协方差 $cov(x,y) = E((x-\mu_x)(y-\mu_y)) = E(xy) - E(x)E(y)$

解释:

$cov(x,y) > 0$　x,y 同向前进. 正相关

$cov(x,y) < 0$.　x,y 反向　负相关.

x,y 不相关 当且仅当 协方差 $cov(x,y) = 0$.　$cov(x,y) = 0 \Leftrightarrow$ 不相关 $\not\Rightarrow$ 独立.

* 不相关未必独立, 但 独立一定不相关, 不相关就是协方差 $=0$ ($cov(x,y) = 0$)

$cov(x,y) = E(xy) - E(x)E(y) = E[(x-\mu x)(y-\mu y)]$

Date. Page.

Properties 性质: ~~Var(x,y)~~ = $= E(xy - x\mu y - y\mu x + \mu x \mu y)$

$= E(xy) - E(x)E(y) - E(y)E(x) + E(x)E(y)$

? 拍没看懂

P1) $Var(x_1 + \cdots + x_n) = \sum_{i=1}^{n} Var(x_i) + \sum_{\substack{i\neq j \\ i,j=1}}^{n} cov(x_i, x_j)$

$x_1 \cdots x_n$ 不相关(比如说两两独立) $\Rightarrow Var(\sum x_i) = \sum Var(x_i)$

P2) $cov(x,x) = E[(x-\mu x)(x-\mu x)] = E[(x-\mu x)^2] = Var(x)$

(特殊情况)

P3) 因为: $E(ax+b) = aE(x)+b$ $\Rightarrow$ E(x) 具有线性性质

$Var(ax+b) = a^2 Var(x)$ $\Rightarrow$ Var(x) 系数变平方

$cov[(a_1x+b_1),(a_2y+b_2)] = E[[a_1x+b_1 - E(a_1x+b_1)][a_2y+b_2 - E(a_2y+b_2)]]$

$= E[(a_1x+b_1 - (a_1E(x)+b_1)][a_2x+b_2 - (a_2E(y)+b)]$

$= E[a_1(x-\mu x)a_2(y-\mu y)] = a_1a_2 E[(x-\mu x)(y-\mu y)]$

$= a_1a_2 cov(x,y)$

对比: $cov(x,y) = E(xy) - E(x)E(y) = E[(x-\mu x)(y-\mu y)]$

$cov[(a_1x+b_1),(a_2y+b_2)] = a_1a_2 cov(x,y)$

eg: $x \sim Bin(n,p)$ $x = x_1 + \cdots + x_n$ $x_i \sim Ber(p)$ independent

$E(x_i) = p$ $Var(x_i) = p(1-p)$

$E(x) = E(x_1) + \cdots + E(x_n) = np$.

$Var(x) = \sum_{i=1} Var(x_i) = np(1-p)$.

多次伯努利分布 $\xrightarrow{n次}$ 二项分布

$E(x) = p \longrightarrow np$

$Var(x) = p(1-p) \longrightarrow np(1-p)$.

Correlation Coefficient ~~相关系~~ 相关系数

~~协方差依赖于单位??~~

这里在讲例子
但我没听懂

~~$x \Rightarrow$ Heights in country A in meters~~

~~$y \Rightarrow$~~

Def: The correlation coefficient of x, y is [x,y的相关系数]

$$\rho_{x,y} = \begin{cases} 0, & \text{if } Var(x)=0 \text{ or } Var(y)=0 \text{ 两者之一方差为0} \\ \dfrac{cov(x,y)}{\sqrt{Var(x)} \cdot \sqrt{Var(y)}}, & \text{else } [cov(x,y)=E(xy)-E(x)E(y)] \end{cases}$$

properties 性质:

1) $x=y=1$ $\rho_{x,y}=1$ $\Rightarrow$ 完全正相关，相关系数为1

$x=-y=0$ $\rho_{x,y}=-1$ $\Rightarrow$ 完全负相关，相关系数为-1

全体化成标准分布 (正态)

2) $-1 \leq \rho_{x,y} \leq 1$

3) Let $z_1 = \dfrac{x-\mu_x}{\sigma_x}$ $z_2 = \dfrac{y-\mu_y}{\sigma_y}$ [$E(z_1)=0=E(z_2)$, $Var(z_1)=Var(z_2)=1$]

Then $\Rightarrow$ $\rho=1 \Leftrightarrow E[(z_1-z_2)^2]=0$

$\rho=-1 \Leftrightarrow E[(z_1+z_2)^2]=0$

i) 证明: Let $x=\pm y$

$Var(x)$ 得到

$$\rho_{x,y} = \frac{cov(x,\pm x)}{\sqrt{Var(x)}\sqrt{Var(\pm x)}} = \frac{\pm cov(x,x)}{\sqrt{Var(x)}\sqrt{Var(x)}} = \pm 1$$

$\Rightarrow$ $x=\pm y$ 时 分别为相关系数+1/-1，完全正/负相关

$$Var(x+y) = Var(x) + Var(y) + 2cov(x,y)$$

Date. Page.

$$\Downarrow$$

iv)

$$0 \le Var\left(\frac{x}{\sqrt{Var(x)}} \pm \frac{y}{\sqrt{Var(y)}}\right) =$$

$$= Var\left(\frac{x}{\sqrt{Var(x)}}\right) + Var\left(\frac{y}{\sqrt{Var(y)}}\right) + 2cov\left(\frac{x}{\sqrt{Var(x)}}, \pm\frac{y}{\sqrt{Var(y)}}\right)$$

系数提到前
变平方

$$= \frac{Var(x)}{Var(x)} + \frac{Var(y)}{Var(y)} \pm \frac{2cov(x,y)}{\sqrt{Var(x)}\sqrt{Var(y)}}$$

$$= 1 + 1 \pm 2\rho_{x,y}$$

$$0 \le 2(1 \pm \rho_{x,y}) \quad \begin{cases} 1+\rho_{x,y} \ge 0 \Rightarrow \rho_{x,y} \ge -1 \\ 1-\rho_{x,y} \ge 0 \Rightarrow \rho_{x,y} \le 1 \end{cases}$$

vii) $E[(Z_1 \pm Z_2)^2] = Var(Z_1 \pm Z_2) = Var(Z_1) + Var(\pm Z_2) + 2cov(Z_1, \pm Z_2)$

$$= 1 + 1 \pm 2\rho_{x,y}$$

$[cov(Z_1, Z_2) = \rho_{x,y}$ Exercise$]$

这一段完全没听懂

$$E[(Z_1 \pm Z_2)^2] = 2[1 \pm \rho_{x,y}]$$

$$\underset{0}{\|} \Leftrightarrow \underset{0}{\|}$$

Date. Page.

大数定律 Law of large numbers 经典投硬币正面的频率→概率=0.5.

① Observation: Repeated sampling leads to stable result of averages $= \mu$.

② 有 $x_1 \cdots x_n$ 独立 $x_{(i)} \sim X^{\sigma_X}_{\mu_X}$ independent identically Distributed

$$\bar{x}_n = \frac{x_1 + \cdots + x_n}{n}.$$

$$E(\bar{x}_n) = E\left(\frac{x_1 + \cdots + x_n}{n}\right) = \frac{1}{n}\left[E(x_1) + \cdots E(x_n)\right] = \frac{n\mu_X}{n} = \mu_X.$$

实际算下来的 就等于 μ_X

大数定律就是这么来的

$$Var(\bar{x}_n) = Var\left(\frac{x_1 + \cdots + x_n}{n}\right) = \frac{1}{n^2}\left(Var(x_1) + \cdots + Var(x_n)\right) = \frac{n\sigma_X^2}{n^2} = \frac{\sigma_X^2}{n}$$

由于 $x_1 \cdots x_n$ 独立, 方差可直接加.

$$\Rightarrow \boxed{\begin{array}{l} \mu_X \Rightarrow E(\bar{x}_n) = \mu_X \\ \sigma_X \Rightarrow Var(\bar{x}_n) = \dfrac{\sigma_X^2}{n} \end{array}}$$

③ Theorem (Chebyshev) X with mean μ and var σ^2

$$P(|X-\mu| \geq a) \leq \frac{\sigma^2}{a^2}.$$ 切比雪夫定理/大数定律.

证明: $\sigma^2 = Var(X)$ 没听懂.

④ Theorem: $x_1 \cdots x_n$ IID with

新 $E(x_i)=\mu$ $Var(x_i)=\sigma^2$ Then

For and $\varepsilon>0$

定理. $P(|\bar{x}_n-\mu| \geq \varepsilon) \xrightarrow[n\to+\infty]{} 0$ 契比雪夫不等式.

证明: 契比雪夫 Chebyshev. ?

example: coin fair?

$x \sim Ber(p)$ $p=?$

好了, 从契比雪夫到这个例子都听不懂.

$x_1 \cdots x_n$ IID $x_i \sim x$

$E(x_i)=p$ $Var(x_i)=p(1-p)$

By (*)

$$P(|\bar{x}_n-p| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2}$$

Determine n so that with 99.9% probability $\bar{x}_n$ is within 10^{-3} of p.

(1) with $\varepsilon=10^{-3}$ And $\frac{p(1-p)}{n\varepsilon^2} \leq 0.999$. $\frac{p(1-p)}{n \cdot 10^{-6}} \leq 0.999$

claim $p(1-p) \leq \frac{1}{4}$

Pf: $\frac{1}{4}-p(1-p) = p^2-p+\frac{1}{4} = (p-\frac{1}{2})^2 \geq 0$.

In (1)

$$\frac{1}{4n \cdot 10^{-6}} \leq 0.999 \qquad n \geq \frac{10^6}{4+0.999} = 250000.$$

strong LLN $x_1 \cdots x_n$ IID $E(x_i)=\mu$ $P(\{\bar{x}_n-\mu \to 0\})=1$

Summary:

Date. Page.

I. X & Y are independent if one of them following:

$P_{X,Y}$, $f_{X,Y}$

i) $P(a_1 \le X \le b_1, a_2 \le Y \le b_2) = P(a_1 \le X \le b_1) \times P(a_2 \le Y \le b_2)$

$\forall a_1, b_1, a_2, b_2$

ii) $E(F(X) G(Y)) = E(F(X)) \cdot E(G(Y))$

For all $F(X)$ and $G(Y)$ — P_X, f_X — P_Y, f_Y

iii) $P_{X,Y}(x,y) = P_X(x) \cdot P_Y(y) \rightarrow$ discrete 离散 — independent

$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y) \rightarrow$ continuous 连续 — 独立.

$\Rightarrow$ same for $X_1 \cdots X_n$ independent if:

more General

- $\{a_i \le X_i \le b_i\}$ independent any a_i, b_i
- $E(F_1(X_1) \cdots F_n(X_n)) = E(F_1(X_1)) \cdots E(F_n(X_n))$
- $P_{X_1 \cdots X_n}(x_1 \cdots x_n) = P_{X_1}(x_1) \cdots P_{X_n}(x_n)$

$f_{X_1 \cdots X_n}(x_1 \cdots x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$

II. $\mathrm{cov}(X,Y) = E((X-\mu_X)(Y-\mu_Y)) = E(XY) - E(X)E(Y)$

$[\mu_X = E(X), \mu_Y = E(Y)]$

$\Rightarrow$ X, Y are uncorrelated $\longleftarrow$ independent — 相关系、不相关、独立 三者关系

$\Updownarrow$ 充要条件 — 只能单向, 必要条件

$\mathrm{cov}(X,Y) = 0$ — 必要条件.

Recall: $\mathrm{cov}(X,X) = E(X^2) - E^2(X) = \mathrm{Var}(X)$

$\mathrm{cov}[a_1X + b_1, a_2Y + b_2] = a_1a_2 \,\mathrm{cov}(X,Y)$

Correlation coefficient 相关系数

$$\rho_{X,Y} = \begin{cases} 0, & \text{若 } Var(X)=0 \text{ or } Var(Y)=0 \\ \dfrac{cov(X,Y)}{\sqrt{Var(X)} \cdot \sqrt{Var(Y)}}, & \text{else} \end{cases}$$

· ∴ 相关系数：$-1 \le \rho_{X,Y} \le$

· $\rho_{X,X} = 1 \quad \rho_{X,-X} = -1$

· $Z_1 = \dfrac{X-\mu_X}{\sigma_X} \quad Z_2 = \dfrac{Y-\mu_Y}{\sigma_Y}$

$\rho_{X,Y} = 1 \iff E((Z_1 - Z_2)^2) = 0$.

Ⅱ. 大数定律：law of Large numbers 大数定律

①

sample: $X_1 \cdots\cdots X_n$ $\begin{cases} E(X_i) = \mu \text{ (mean)} \\ Var(X_i) = \sigma^2 \text{ (variance)} \end{cases}$

$\bar{X}_n = \dfrac{X_1 + \cdots + X_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$ Average (≠mean) 均值≠方差.

$E(\bar{X}_n) = \mu$ mean

$Var(\bar{X}_n) = \dfrac{\sigma^2}{n}$ ← !!! variance.

独立性 随机变量 同分布.

② Chebyshev 切比雪夫大数定律 + 不等式. DO UNDERSTAND !!!

X: R.V with $E(X) = \mu$, $Var(X) = \sigma^2$

$p(|X-\mu| \ge \varepsilon) \le \dfrac{\sigma^2}{\varepsilon^2}$. 切比雪夫不等式.

→ 弱大数定律 证明：$p(|\bar{X}_n - \mu| \ge \varepsilon) \le \dfrac{\sigma^2}{n\varepsilon^2}$ WLLN

Date. Page.

拓展：LLN（大数定律）can be used to determine laws of R.V.

Issue: R.V. X want.

$$p(X \in C) = p(a \le X \le b)$$

Key: $Y = \begin{cases} 1, & X \in C \ [a \le X \le b] \\ 0, & \text{else} \end{cases}$ [indicator function]

Ber(p) p: $1 \cdot p(Y=1) + 0 \cdot p(Y=0) = p(X \in C)$

$E(Y) = p = \mu_y \quad Var(y) = p(1-p) = \sigma_y^2$

LLN: $y_1 - y_n \Rightarrow \overline{y_n} = \dfrac{y_1 + \cdots + y_n}{n} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$

$=$ Fraction of times $X_i \in C$

$p(|\overline{y_n} - p| \ge \varepsilon) \le \dfrac{p(1-p)}{n\varepsilon^2} \le \dfrac{1}{4n\varepsilon^2} \le S$ 切比雪夫大数定律

（S是允许出现误差的最大值）

其实我也不是很清楚

对切比雪夫不等式的理解需加强，这S是什么？？

$n \ge \dfrac{1}{4S\varepsilon^2}$

Example: DIE: Throw it n times

$p_i = p(X_n = i) \quad i = 1 - 6$

对 $n \ge \dfrac{1}{4S\varepsilon^2} \quad \overline{y_n}^{(i)} = \dfrac{\# X_k = i}{n} = \dfrac{n_i}{n}$

$p(|\overline{y_n}^{(i)} - p_i| \ge \varepsilon) \le S$

？没啥意

中心极限定理

Date. Page.

central limit Therom：中心极限定理 $\frac{1}{n}\sum_{i=1}^{n} X_i$.

—motivation：$\bar{X}_n \approx \mu$ $\mu \to \mu$

$$\frac{\bar{X}_n-\mu}{\sigma_{\bar{X}_n}} = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$$ $\sigma^2 \to \frac{\sigma^2}{n}$

$\sigma \to \sqrt{\frac{\sigma}{n}}$

中心极限定理：iid. 标准化样本均值的抽样分布也趋向于标准正态分布

Therom (central limit Therom)

let $X_1 \cdots X_n$ IID (独立同分布的随机变量) $[X_i \sim X]$

$E[X_i]=\mu$ $Var(X_i)=\sigma^2$

Then, if：

$$Z_n = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}} \quad P(Z_n\in[a,b]) \xrightarrow[n\to+\infty]{} \int_a^b \frac{e^{-z^2/2}dz}{\sqrt{2\pi}}$$ 标准正态分布

Equivalently.

$$E(f(Z_n)) \xrightarrow[n\to+\infty]{} \int_{-\infty}^{+\infty} f(x)\frac{e^{-x^2/2}dx}{\sqrt{2\pi}} \quad (*)$$

Comments：注释.

(1) (*) or Equivalently is stated as：

$$Z_n \xrightarrow[n\to+\infty]{} N(0,1) \text{ in law}$$ N(0,1) 标准正态分布

(2) no control on errors

· Rule of Thumbs ($n\geq 30$, ok)

· Tests of normality

Date. Page.

$Y = \frac{X-\mu}{\sigma}$

(3) alternative statements

CLT (中心极限定理) $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow{law} N(0,1)$

Recall: $X \sim N(\mu, \sigma^2)$; $Y = aX + b$

$\Rightarrow Y \sim N(a\mu + b, a^2\sigma^2)$

线性性质　　平方.

$$Z_n = \frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \Rightarrow \bar{X}_n = \mu + \frac{\sigma}{\sqrt{n}} Z_n.$$

(3.1) $Z_n \to N(0,1) \longrightarrow \bar{X}_n \sim (\mu + \frac{\sigma}{\sqrt{n}} \times 0, \frac{\sigma^2}{n})$

$\sim (\mu, \frac{\sigma^2}{n})$ 和前面所切比雪夫大数定律非常相似.

$$\boxed{\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})}$$

(3.2) $Z_n = \frac{\frac{X_1 + \cdots + X_n}{n} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}}{n}\left[\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma}\right]$

同乘n

$$Z_n = \frac{X_1 + X_2 + \cdots + X_n - \mu n}{\sqrt{n}\sigma}$$

$\Rightarrow X_1 + X_2 + \cdots + X_n = n\mu + \sqrt{n}\sigma Z_n$

$X_1 + X_2 + \cdots + X_n \sim N(n\mu, n\sigma^2)$.

$$\boxed{X_1 + \cdots + X_n \sim N(n\mu, n\sigma^2)}$$

中心极限定理的应用 Applications of CLT (central limit Theorem)

Date. Page.

A1) "Improvement" over Chebyshev:

Cheb: $P(|\bar{X}_n-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$ # 切比雪夫不等式+大数定律

CLT: $P(|\bar{X}_n-\mu| \geq \varepsilon) = P\left(\left|\frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}\right| \geq \frac{\varepsilon}{\sigma/\sqrt{n}}\right)$

($Z_n = \frac{\bar{X}_n-\mu}{\sigma/\sqrt{n}}$) $= P\left(|Z_n| \geq \frac{\varepsilon\sqrt{n}}{\sigma}\right)$ ~ Probability for $N(0,1)$

(Tables) $= 2P\left(Z_n \geq \frac{\varepsilon\sqrt{n}}{\sigma}\right) = 2\underbrace{\int_{\frac{\varepsilon\sqrt{n}}{\sigma}}^{+\infty} e^{-z^2/2}\frac{dz}{\sqrt{2\pi}}}_{\text{From tables 查表可知}}$

$\Rightarrow P(|\bar{X}_n-\mu| \geq \varepsilon) = 2\int_{\frac{\varepsilon\sqrt{n}}{\sigma}}^{+\infty} e^{z^2/2}\frac{dz}{\sqrt{2\pi}}$ (查表可知)

A2) Case of $X \sim Ber(p)$.

$X_i \sim Ber(p)$ $E(X_i)=p$ $Var(X_i)=p(1-p)$

Question: Determine $P(|\bar{X}_n-\mu| \geq \varepsilon) \leq \delta$

A2.1) Chebyshev:

$P(|\bar{X}_n-\mu| \geq \varepsilon) \leq \frac{p(1-p)}{n\varepsilon^2} \leq \frac{1}{4n\varepsilon^2} \leq \cancel{\delta}\ \delta$

$n \geq \frac{1}{4\varepsilon^2\delta}$

$\mu = p \quad \sigma^2 = p(1-p) \quad \sigma = \sqrt{p(1-p)}$

A2.2) CLT 中心极限定理. $\bar{x}_n \to z_n = \frac{\bar{x}_n - p}{\sqrt{p(1-p)}/\sqrt{n}} = \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}}$

$P(|\bar{x}_n - \mu| \ge \varepsilon) = P\left(|z_n| \ge \frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}}\right) = 2P\left(z_n \ge \frac{\varepsilon\sqrt{n}}{p(1-p)}\right)$

[一样的假设, 见上]

From Tables: GET z_0 s.t.

$P(z \ge z_0) = \delta$ $\Rightarrow$ 查表.

Ask:

$\frac{\varepsilon\sqrt{n}}{\sqrt{p(1-p)}} \ge z_0 \Rightarrow n \ge \frac{z_0^2}{4\varepsilon^2}$.

$\sqrt{n} \ge \frac{z_0\sqrt{p(1-p)}}{\varepsilon}$ 恒成立.

$\sqrt{n} \ge \frac{z_0}{\sqrt{4}\,\varepsilon} \Rightarrow n \ge \frac{z_0^2}{4\varepsilon^2}$.

$\Rightarrow n \ge \frac{z_0^2}{4\varepsilon^2}$

z_0 是什么?

Example: $\varepsilon = 10^{-2}$, $\delta = 10^{-3}$. $n \ge$

$\boxed{\text{Chebyshev:}}$ $n \ge \frac{1}{4\times10^{-4}\times10^{-3}} = \frac{10^7}{4} = 2500000$.

$\delta/2)$

$\boxed{\text{CLT:}}$ z_0 s.t. $P(z \ge \frac{\delta}{2}) = 0.0005$. $z_0 = 3.32$

$\delta/2)$. $\Rightarrow n \ge \frac{3.32^2}{4\times10^{-4}} = 30000$

Date. Page.

A3): $x_1 + \cdots + x_n \sim N(np, n\sigma^2)$

$x_i \sim Ber(p)$

$\underbrace{x_1 + \cdots + x_n}_{} \sim N(np, np(1-p))$

$$\boxed{Bin(n,p) \sim N(np, np(1-p))}$$

$\Rightarrow p(X = 378) = \binom{n}{378} p^{378} (1-p)^{n-378}$

probability of 8th in 10 fair Flips:

$x \sim Bin(10, \frac{1}{2})$

By CLT:

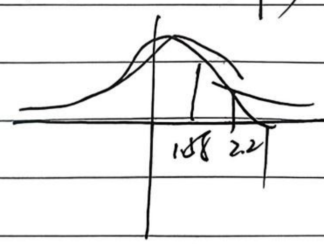
$x \sim N(10 \cdot \frac{1}{2}, 10 \cdot \frac{1}{2} \cdot \frac{1}{2}) = N(5, \overset{\sigma^2}{\overset{\shortparallel}{2.5}})$

$z = \frac{x-5}{\sqrt{2.5}} \sim N(0,1) \Rightarrow \underline{x = \sqrt{2.5}\, z + 5}$

$p(7.5 \le \sqrt{2.5}\, z + 5 \le 8.5)$

$\Rightarrow p\left(\frac{2.5}{\sqrt{2.5}} \le z \le \frac{3.5}{\sqrt{2.5}}\right) = p(1.58 \le z \le 2.21)$

$= p(z \ge 1.58) - p(z \ge 2.21)$



$= 0.0435$.

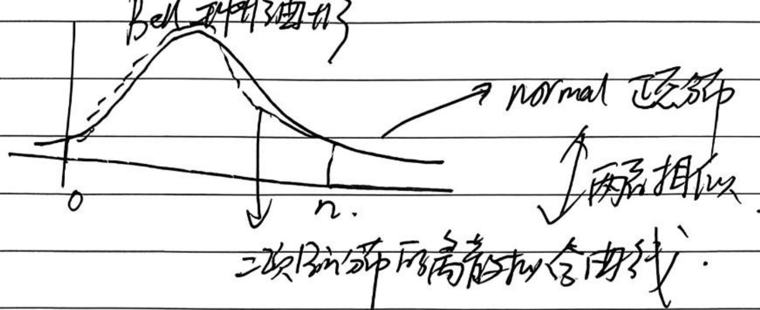
Exact $X \sim Bin(10, \frac{1}{2})$ $p(X=8) = \binom{10}{8} \frac{1}{2}^8 (\frac{1}{2})^2 = 0.04 34 \approx 0.0435$

* 二项分布可以近似成正态分布

Bin与normal的近似条件

为什么可以这样? 可以的,当且仅当正态分布钟形曲线都落在 $(0,n)$ 范围内..?

Bin(n,p) → normal is Good if most of the normal Bell is in $(0,n)$



$[\mu-3\sigma, \mu+3\sigma] \subset [0,n]$

in our example

(a) $\mu-3\sigma = np-3\sqrt{np(1-p)} > 0$

(b) $\mu+3\sigma = np+3\sqrt{np(1-p)} < n.$

Can compute for $n=10$, $p=\frac{1}{2}$ and see that it is true.

(a) $\Rightarrow n > 9\frac{(1-p)}{p}$

(b) $\Rightarrow n > 9\frac{p}{1-p}.$

normal approx of Bin(n,p) is OK!

Here $p=\frac{1}{2}$ $\frac{p}{1-p} = 1 = \frac{1-p}{p}.$

(a) $n>9$

(b) $n>9$ we have $n>10$.

Date、 Page.

A4) Round off Errors

Round off comes

$100.35 \to 100$

$100.75 \to 101$.

Error is $\sim$ Unif $(-0.5, 0.5)$

Sum of Errors

$x_1 - x_2 -$

$E(x_i) = 0$

$Var(x_i) = 1/12$.

Probability That $|x_1 + \cdots + x_{100}| \geq 10$

Date. Page.

In General to compute Med $(x_1 \cdots x_N)$

step 1: get $k = \lfloor \frac{N+1}{2} \rfloor$

$$\text{Med}: \frac{x(k) + x(k+1)}{2} = x(k+1) + \left[k - \left(\frac{N+1}{2}\right)\right](x_{k+1} - x_k)$$

$$= x(k) + \left[\frac{N+1}{2} - k\right](x(k+1) - x(k))$$

· MODE 众数

22) Dispersion: 分散:

· Sample (Empical) Variance: 样本、经验方差

$$S_n = \underbrace{\frac{1}{N-1}}_{\uparrow\ \text{有偏估计}} \sum_{i=1}^{n} (x_i - \bar{x}_N)^2 \quad 经验方差$$

MAD: $= MED\left[|x_1 - MED|, \cdots, |x_N - MED|\right]$

每个数与其中位数的绝对值数列的中位数

3) Quantiles: 百分位数

$Q(1/4)$ (lower).

$Q(1/2)$ = median.

$Q(3/4)$ = Upper.

~~Q(1/4) ideally~~

Formally:

step 1: $k = \left\lfloor \frac{N+1}{4} \right\rfloor$ 向下取整.

$$Q(1/4) = x_{(k)} + \left[\frac{N+1}{4} - k\right]\left(x_{(k+1)} - x_{(k)}\right)$$

$$Q(3/4) = \quad k = (N+1)\frac{3}{4}$$

In General: p-th Quartile

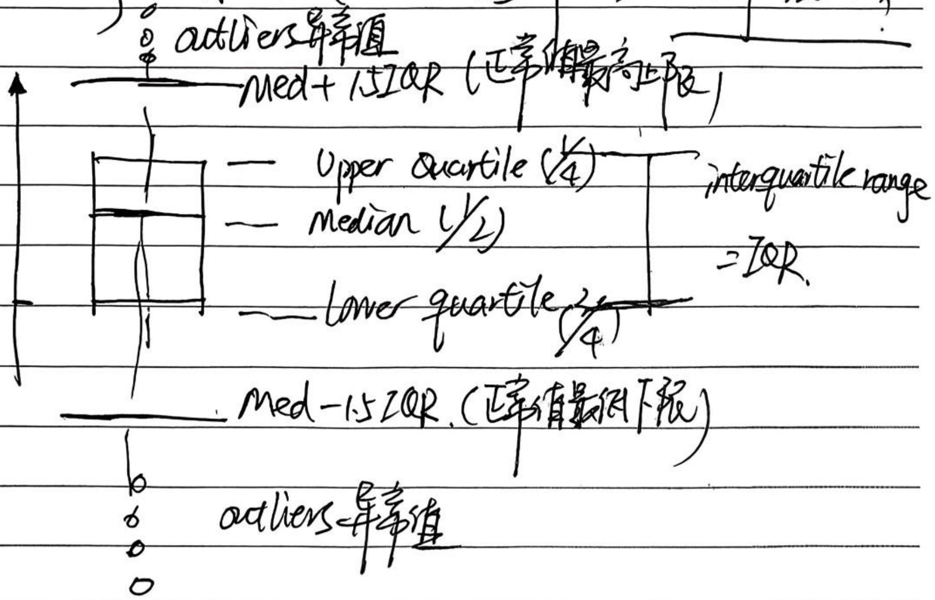
分位数求法 $\Longrightarrow$

$$\boxed{\begin{array}{l} k = \lfloor (N+1)p \rfloor \quad (\text{floor 向下取整}) \\ Q(p) = x_{(k)} + [(N+1)p - k][x_{(k+1)} - x_{(k)}] \end{array}}$$

Tuke's summary = Box-and-whistles plots cutf 求that ?

箱形图

Boxplot.



Date. Page.

STATISTICAL MODELS

$x_i \to X_i$ sample: $X_1 - X_N$ iid. 独立同分布.

$x_i \to X$

Q: law of X?

Usually

- law proposed from experience

Meas Errors $X \sim N(\mu, \sigma^2)$

population par(α)

population growth $\exp(\lambda)$

proposal = statistical model

- or settle for less

want mean variance.

Always: parameters to determine ($\mu, \sigma^2, \lambda, \alpha$ ——

How? STATISTICS = $H_n(X_1 - - X_n)$

H_n as estimator for a parameter

θ If

$H_n(X_1 - X_N)$ related carefacts to θ.

That is realizations

$H_N(x_1 - x_N) \to \hat{\theta}$ Estimator of θ.

examples:

1) estimator of $\mu = E(X)$

$\bar{X}_n$

know (LLN) $\bar{X}_N \to \mu$

so $\bar{X}_N = \hat{\theta}_N$ Estimator of μ

2) sample Frequencies:

$$\underbrace{\frac{\#\{i : x_i = a\}}{N}}_{\text{estimator of } P_X(a)} \xrightarrow[N\to\infty]{} P_X(\alpha) \quad [LLN]$$

3) ~~Histor~~ Histograms $(x_1 \cdots x_N)$ estim of f_X

4) sample variance

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^{2} \; ?$$

$$S_n^2 = \frac{1}{N-1}\sum(x_i - \overline{X_N})^2 \quad \text{Estimator of } \sigma^2$$

By LLN

$$S_n^2 = \frac{1}{N-1}\sum_{i=1}^{n}\left(x_i^2 - 2x_i\overline{x_n} + \bar{x}_n^2\right)$$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}_n \sum x_i + N\bar{x}_n^2\right]$$

$$= \frac{N}{N-1}\left[\frac{1}{N}\sum x_i^2 - \text{———}\right]$$

$$= \sigma^2.$$

5) empirical Quantiles: (Quantile Median)

$x_1 - x_n \to x_{(1)} - x_{(n)}$ order statistics

p-th Quantile.

$$k = \lfloor (n+1)\cdot p \rfloor$$

$$Q(p) = X_{(k)} + [(n+1)p - k] \cdot (X_{(k+1)} - X_{(k)})$$

6) $MAD(x_1 - - x_N)$

$= MED[|x_1 - MED| - - |x_N - MED|]$

Estimator of actual MAD.

Quality of Estimators?

Desired properties?

- Unbiasedeness
- Efficiency
- Low mean ~~sota~~ square Error

unbiased estimators

DEF = an estimator T_n of θ is unbiased if
$E(T_N) = \theta$ All N

$E(T_N) - \theta = $ Bias of T_N ≥ 0 < 0

Application: $x_i \sim X$ $E(X) = \mu$.
$\overline{x_N}$ unbiased estimator of μ.
Indeed $E(\overline{x_N}) = \mu$.
unbiased estimator for σ^2 = variance?

Naive choice

$$T_N^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

$$E(T_N^2) = \frac{1}{N}\sum_{i=1}^{N} E\left[(x_i-\bar{x})^2\right]$$

OBS 1: $E(x_i-\bar{x}_N)=0$.

Then $E\left[(x_i-\bar{x}_N)^2\right] = Var(x_i-\bar{x}_N)$.

$$= Var\left[x_i - \frac{x_1+\cdots+x_i+\cdots+x_N}{n}\right]$$

$$= Var\left[(1-\tfrac{1}{N})\underset{\uparrow}{x_i} - \sum_{j\neq i}\underset{\uparrow}{x_j}\right]$$

独立.

$$= Var\left[(1-\tfrac{1}{N})x_i\right] + \sum_{j\neq i} Var(-x_j)$$

$$= (1-\tfrac{1}{N})^2 Var(x_i) + \sum_{j\neq i} Var(x_j).$$

$$= (1-\tfrac{1}{N})^2\sigma^2 + (N-1)\sigma^2$$

$$= \sigma^2\left[N-1+\left(\tfrac{N-1}{N}\right)^2\right]$$

$$= \sigma^2\left(\frac{N-1}{N}\right)$$

$\Rightarrow E(T_N^2) = \sigma^2\cdot\frac{N-1}{N}$ negative Bias

$S_N^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x}_N)^2$ as unbiased