概率论 (Probability) Level M [STATS 5024] - 讲座 1 笔记

主讲人: Xin Li
所属院系: 数学与统计学院，格拉斯哥大学

---

1. 课程管理 (Admin)

1.1 讲座安排 (Lectures)

- **周一 15:00 - 16:00** (9月22日至10月13日)
  - 9月22日: James McCune Smith Building - Room 745
  - 9月29日, 10月6日, 10月13日: Gilbert Scott Building - Room 255
- **周二 12:00 - 13:00** (9月23日至11月11日)
  - 地点: Gilbert Scott Building - Room 220
- **周三 9:00 - 10:00** (9月24日至11月12日)
  - 地点: Gilbert Scott Building - Room 220

1.2 考核方式 (Assessment)

- **期末考试:** 占总成绩的 80%
- **4次在线 (Moodle) 测验:** 占总成绩的 20%
- Moodle上已有一个**练习测验**（不计分）可供使用

1.3 辅导课 (Tutorials)

- 每周两小时，时间为周五 10:00 – 12:00
- 与“回归模型”和“统计推断”课程共享辅导资源
- 学生需要**提前完成问题单 (Problem Sheets)** 来为辅导课做准备
- 在辅导课上，助教 (Teaching Assistants) 将帮助解决学生在解题时遇到的困难

1.4 联系与支持 (Contact and Support)

- **邮箱:** Xin.Li@glasgow.ac.uk 。欢迎通过邮件提问 。
- **答疑时间 (Office hours):**
  - 11月13日之前：每周四 15:00 至 16:00
  - 地点：数学与统计大楼 335 办公室
  - 11月13日之后：需通过预约安排
- **Moodle 问答论坛:**
  - 提供一个可以提问（可匿名）和互相回答问题的平台
  - 讲师会每周检查一次论坛，确保所有问题都得到解答

---

2. 基础理论 (Foundations)

2.1 概率的意义与用途 (What Probability Means)

概率论的用途 (The Usefulness of Probability Theory)

- 概率论是一个数学体系，允许在存在不确定性（关于系统的信息不完全）的情况下进行一致的推理 。因此其用途非常广泛 。
- **统计推断:** 概率论是统计推断科学的基础 。它为分析和建模在许多不同研究领域中获得的结果提供了方法 。随机样本的属性被用来计算估计总体特征的精确度 。
- **遗传学:** 后代通过随机过程从父母那里继承特征，这些过程可以用概率模型来描述 。有高风险遗传疾病的父母可能会被提供基于概率计算的咨询，以帮助他们做出家庭计划决策 。
- **流行病学:** 病毒（如HIV，埃博拉）在人与人之间传播，但其行为具有随机性 。与感染者的接触有时会导致未感染者被感染，但并非总是如此 。概率模型被用来描述流行病的传播方式并预测其未来走向 。
- **金融:** 生命表显示了每个年龄的人在下一个生日前死亡的概率 。生命表被用来计算活到任何特定年龄的概率，或不同年龄人群的剩余预期寿命 。精算师使用这些概率来计算人寿保险的合理保费和评估养老基金的充足性 。

2.2 实验、结果和事件 (Experiments, Outcomes and Events)

基本定义 (Definitions)

- **实验 (Experiment):** 获取信息的任何过程（比通常的科学定义更广泛）
- **重复 (Replicate) 或 试验 (Trial):** 一次实验的单次执行
- **结果 (Outcome):** 一次实验重复所记录的信息

- **随机实验 (Random or Stochastic Experiment):** 即使在相同条件下进行，也存在多种可能结果，并且事先不知道下一次实验会产生哪种结果的实验

概率论的核心是为（随机）实验的结果建立数学模型 。

随机实验示例

- **示例 1(a) - 分诊:** 记录本地急诊部门下一位分诊患者的治疗需求类别 。类别分为：**非常紧急 (Very urgent)**、**紧急 (Urgent)** 或 **标准 (Standard)** 。
- **示例 1(b) - 网页点击量:** 记录从今晚午夜开始的24小时内，数学与统计学院主页的点击次数 。
- **示例 1(c) - 二氧化硫浓度:** 记录明天早上08:00至09:00在格拉斯哥中央火车站桥下的测量点，大气中 SO_2 的最高浓度（单位：ppm） 。
- **示例 1(d) - 双胞胎:** 记录在格拉斯哥皇家公主妇产医院下一对活产双胞胎是**同卵 (MZ)** 还是 **异卵 (DZ)**，以及每位婴儿的性别 。

更多定义 (Definitions)

- **样本空间 (Sample Space, S):** 包含随机实验所有可能结果的集合 。S可以是**有限的**、**可数的**或**不可数的** 。
  - **可数集 (Countable Set):** 元素可以被列出的无限集，如自然数 $\{s_1, s_2, ...\}$ 。
  - **不可数集 (Uncountable Set):** 不能被列出的无限集，如实数或实数线上的一个区间 。
- **事件 (Event):** 样本空间S的一个子集，即一组结果的集合 。
  - **简单事件 (Simple Event):** 只包含一个结果的事件 。
  - **复合事件 (Compound Event):** 不是简单事件的事件 。

示例中的样本空间与事件

- **示例 1(a) - 分诊:**
  - **有限样本空间:** $S =$ {'非常紧急', '紧急', '标准'}
  - **简单事件 E:** '治疗需求为非常紧急' = {'非常紧急'}
  - **复合事件 F:** '治疗需求不是非常紧急' = {'紧急', '标准'}
- **示例 1(b) - 网页点击量:**
  - **可数样本空间:** $S = \{0, 1, 2, ...\}$
  - **事件 E:** '至少有2000次点击' = $\{2000, 2001, 2002, ...\}$
  - **事件 F:** '至少有1000次点击' = $\{1000, 1001, 1002, ...\}$
  - **事件 G:** '点击数能被100整除' = $\{0, 100, 200, ...\}$
- **示例 1(c) - 二氧化硫浓度:**
  - **不可数样本空间:** $S = [0, 1000000]$
  - **事件 E:** 'SO_2 浓度最多为 1 ppm' = $[0, 1]$
  - **事件 F:** 'SO_2 浓度至少为 1 ppm' = $[1, 1000000]$
  - **事件 G:** 'SO_2 浓度高于 3 ppm' = $(3, 1000000]$
- **示例 1(d) - 双胞胎:**
  - **有限样本空间:** $S = \{(MZ, M, M), (MZ, F, F), (DZ, M, M), (DZ, F, F), (DZ, M, F), (DZ, F, M)\}$，其中记录的第一个性别是先出生的那个 。

关于实验、结果和事件的评论

- 实验结果不一定是数字 。
- 一个结果可能包含不止一条信息 。
- S 本身是一个事件，并且是**必然事件**（一定会发生） 。
- 空集 $\emptyset$ 被视为S的一个子集，代表一个**不可能事件** 。
- **注意:** 当S是有限或可数时，S的所有子集都被视为事件 。当S是不可数时，存在一些我们不希望归类为事件的子集（因为无法一致地为它们分配概率） 。本课程不会深入探讨此问题 。

---

3. 事件的运算与集合论 (Events and Set Theory)

3.1 事件的关系与运算

假设 E 和 F 是样本空间 S 的两个事件。

- S (全集): 必然事件
- $\emptyset$ (空集): 不可能事件
- E' (补集): 事件 E **不**发生
- $E \cup F$ (并集): 事件 E 发生 **或** 事件 F 发生 (或两者都发生)
- $E \cap F$ (交集): 事件 E **和** 事件 F **同时**发生
- $E \cap F = \emptyset$: E 和 F 是**互斥**或**不相交**的事件 。推论：E 和 E' 总是互斥的 。

- $E \subseteq F$ (子集): 如果事件 E 发生，那么事件 F **必然**发生 。
- $F \setminus E$ (差集): 如果 $E \subseteq F$，那么 $F \setminus E$ (等于 $F \cap E'$) 表示所有在 F 中但不在 E 中的结果所构成的事件 。

不相交并集 (Disjoint Union)

- 通常对于任何两个事件 E 和 F，F 是 $F \cap E$ 和 $F \cap E'$ 的不相交并集：$F = (F \cap E) \cup (F \cap E')$ 。

3.2 集合论定律 (Laws of Set Theory)

令 A, B, C 为三个事件：

- **交换律 (Commutative):** $A \cup B = B \cup A$ 并且 $A \cap B = B \cap A$
- **结合律 (Associative):** $(A \cup B) \cup C = A \cup (B \cup C)$ 并且 $(A \cap B) \cap C = A \cap (B \cap C)$
- **分配律 (Distributive):** $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ 并且 $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
- **德摩根定律 (De Morgan):** $(A \cup B)' = A' \cap B'$ 并且 $(A \cap B)' = A' \cup B'$

3.3 事件运算示例

- **示例 1(a) - 分诊:**
  - 事件 F ('不是非常紧急') 是事件 E ('非常紧急') 的补集, 所以 $F = E'$ 。因此，$E \cup F = S$ 且 $E \cap F = \emptyset$ 。
- **示例 1(b) - 网页点击量:**
  - 事件 E ('至少2000次') 是事件 F ('至少1000次') 的子集, 所以 $E \subseteq F$，因此 $E \cap F = E$ 。
  - $F \setminus E = \{1000, 1001, \ldots, 1999\}$
  - $E \cap G = \{2000, 2100, 2200, \ldots\}$
  - $F \cap G = \{1000, 1100, 1200, \ldots\}$

---

4. 概率的解释 (The Interpretation of a Probability)

一个事件E的概率，记作 $P(E)$，量化了当进行相关随机实验时E是否会发生的不确定性 。越有可能发生的事件被赋予越高的概率 。

- 概率的取值范围是 $[0, 1]$ 。
- $P(S) = 1$ (必然事件概率为1) 。
- $P(\emptyset) = 0$ (不可能事件概率为0) 。
- **注意:** 概率为1的事件不一定等同于必然事件（它们是**几乎必然**发生）；概率为0的事件也不一定是不可能事件（它们是**几乎永不**发生） 。

示例 2: 掷骰子

- **实验:** 掷一个标准骰子 。
- **样本空间:** $S = \{1, 2, 3, 4, 5, 6\}$
- **事件 E:** '点数至少为5' 。

解释 I: 古典概率 (等可能性)

- 假设样本空间S中的每个结果都是等可能出现的 。
- 对于掷骰子，每个点数出现的概率为 $\frac{1}{6}$，因此事件 E 的概率是 $\frac{2}{6} = \frac{1}{3}$ 。
- 这种解释基于实验的**对称性** 。
- **问题:** 大多数随机实验不具备这种对称性 。

解释 II: 相对频率

- 如果在n次重复实验中，事件E发生了 n_E 次，那么E的相对频率是 n_E/n 。
- E的概率可以被定义为当实验重复非常多次时，这个比例的长期或极限值 。
- 因此，$P(E) = \frac{1}{3}$ 意味着平均而言，每掷三次骰子会出现一次5或6 。
- **问题:** 这个定义在概念上与那些可以在相同条件下无限重复的实验绑定（但现实中可能不存在这样的实验） 。

解释 III: 主观概率

- 近年来流行一种观点，认为概率最好被看作是个人信念的陈述 。
- 这种观点使概率不再局限于可无限重复的实验 。
- $P(E) = \frac{1}{3}$ 这个陈述被看作是某人当前对骰子或实验的信念的陈述，或者更有说服力地，是他们对实验无知的陈述 。
- **影响:** 主观学派将概率的范围扩展到了一次性事件，但也打破了频率主义概率与客观现实之间的联系 。

概率论 (Probability) Level M [STATS 5024] - 讲座 2 笔记

---

1. 概率的公理和规则 (Axioms and Rules of Probability)

1.1 柯尔莫哥洛夫公理 (Kolmogorov's Axioms)

为了确保我们为不同事件分配的概率能够相互一致，我们需要遵守一套规则。 一组概率如果遵守由柯尔莫哥洛夫（Kolmogorov, 1933）estabelecidos 的概率公理，那么它将给出一致的结果。

概率公理

1. **非负性与规范性:** 对于任何事件 $E \subseteq S$，其概率 $P(E)$ 满足 $0 \leq P(E) \leq 1$。
2. **全集概率:** $P(S) = 1$。
3. **可加性 (两个互斥事件):** 如果 E 和 F 是互斥事件 (disjoint events)，那么 $P(E \cup F) = P(E) + P(F)$。
4. **可数可加性:** 如果 $E_1, E_2, \dots$ 是一系列互斥事件，那么 $P(\bigcup_i E_i) = \sum_i P(E_i)$。

关于公理3和公理4的说明:

- 乍一看，公理4似乎使公理3变得多余。
- 实际上，公理3只能通过归纳法推广到有限个互斥事件的集合 $(E_1, \dots, E_k)$，而公理4可以应用于无限个事件的集合。
- 另一方面，可以通过设置 $E_1 = E, E_2 = F$ 并且 $E_3 = E_4 = \dots = \emptyset$ 从公理4推导出公理3。

示例 3: 验证等可能性结果
假设一个实验有 k 个等可能的结果 $s_1, s_2, \dots, s_k$。 对于任何事件 E, 其概率定义为 $P(E) = n(E)/k$，其中 $n(E)$ 是 E 中包含的结果数量。 我们可以证明这种定义满足概率公理：

- **公理 1:** 因为 $0 \leq n(E) \leq k$，所以 $0 \leq P(E) = n(E)/k \leq 1$。
- **公理 2:** 因为 $n(S) = k$，所以 $P(S) = n(S)/k = 1$。
- **公理 3:** 如果 E 和 F 互斥，它们没有共同的结果，所以 $n(E \cup F) = n(E) + n(F)$。因此，$P(E \cup F) = P(E) + P(F)$。
- **公理 4:** 在这个有限样本空间中，只有有限个互斥事件的集合，公理3可以通过归纳法推广到任意有限数量的互斥事件。

1.2 概率的规则 (Rules of Probability)

如果每次计算特定事件的概率都必须直接从公理出发，那将非常繁琐。 我们可以从公理中轻松推导出其他更有用的通用结果，这些通常被称为“概率规则”。

示例 4: 补集规则 (Complement Rule)
证明: $P(E') = 1 - P(E)$

- 我们知道 $E \cup E' = S$，根据公理2, $P(E \cup E') = P(S) = 1$。
- 因为 E 和 E' 是互斥的，根据公理3, $P(E \cup E') = P(E) + P(E')$。
- 因此，$P(E) + P(E') = 1$，移项可得 $P(E') = 1 - P(E)$。

示例 5: 空集概率
证明: $P(\emptyset) = 0$

- 因为空集 $\emptyset$ 是全集 S 的补集，即 $\emptyset = S'$。
- 使用补集规则：$P(\emptyset) = P(S') = 1 - P(S) = 1 - 1 = 0$。

示例 6: 单调性 (Monotonicity)
证明: 如果 $E \subseteq F$，那么 $P(E) \leq P(F)$。

- 我们可以将 F 表示为两个互斥事件的并集：$F = E \cup (F \setminus E)$。
- 根据公理3，$P(F) = P(E) + P(F \setminus E)$。
- 根据公理1，$P(F \setminus E) \geq 0$。
- 因此，$P(F) \geq P(E)$。

示例 7: 加法规则 (Addition Rule for Two Events)
证明: 对于任意两个事件 E 和 F， $P(E \cup F) = P(E) + P(F) - P(E \cap F)$。

- $E \cup F$ 可以分解为三个互斥事件的并集: $E \setminus (E \cap F)$, $F \setminus (E \cap F)$ 和 $E \cap F$。
- 根据公理4, $P(E \cup F) = P(E \setminus (E \cap F)) + P(F \setminus (E \cap F)) + P(E \cap F)$。
- 同时，E 可以分解为 $(E \cap F)$ 和 $(E \setminus (E \cap F))$ 的互斥并集，所以 $P(E) = P(E \cap F) + P(E \setminus (E \cap F))$，可得 $P(E \setminus (E \cap F)) = P(E) - P(E \cap F)$。
- 同理，$P(F \setminus (E \cap F)) = P(F) - P(E \cap F)$。
- 将后两式代入第一式，得到 $P(E \cup F) = [P(E) - P(E \cap F)] + [P(F) - P(E \cap F)] + P(E \cap F)$。
- 化简后即为 $P(E \cup F) = P(E) + P(F) - P(E \cap F)$。

示例 8: 三个事件的容斥原理 (Inclusion-Exclusion for Three Events)
证明: 对于任意三个事件 E_1, E_2, E_3
$P(E_1 \cup E_2 \cup E_3) = P(E_1) + P(E_2) + P(E_3) - P(E_1 \cap E_2) - P(E_1 \cap E_3) - P(E_2 \cap E_3) + P(E_1 \cap E_2 \cap E_3)$。

- **证明过程:**

1. 令 $A = E_1 \cup E_2$，则 $P(E_1 \cup E_2 \cup E_3) = P(A \cup E_3)$。
2. 应用两个事件的加法规则: $P(A \cup E_3) = P(A) + P(E_3) - P(A \cap E_3)$。
3. 将 A 替换回 $E_1 \cup E_2$:
   $P(E_1 \cup E_2) + P(E_3) - P((E_1 \cup E_2) \cap E_3)$。
4. 再次应用加法规则和分配律:
   $[P(E_1) + P(E_2) - P(E_1 \cap E_2)] + P(E_3) - P((E_1 \cap E_3) \cup (E_2 \cap E_3))$。
5. 对最后一项再次应用加法规则:
   $P((E_1 \cap E_3) \cup (E_2 \cap E_3)) = P(E_1 \cap E_3) + P(E_2 \cap E_3) - P(E_1 \cap E_2 \cap E_3)$。
6. 将所有项合并并化简，即可得到最终结果。

概率论 (Probability) Level M [STATS 5_0_2_4] - 讲座 3 笔记

---

1. 条件概率 (Conditional Probability)

1.1 条件概率的定义

- 知道一个事件已经发生，如何影响另一个事件的概率?
- **定义:** 假设 E, F 是样本空间 S 中的事件，且 $P(E) > 0$ 。那么，在事件 E 发生的条件下，事件 F 发生的条件概率定义为：
  $P(F|E) = \frac{P(E \cap F)}{P(E)}$

示例 9: 抽球问题

- 一个袋子里有100个除了颜色外完全相同的球：67个黑球，33个红球 。 实验为从袋中不放回地随机抽球 。
  - (a) 第一个抽出的球是黑色的概率是多少?
  - (b) 第一个球是黑色且第二个球是红色的概率是多少?
  - (c) 已知第一个球是黑色的，求第二个球是红色的条件概率 。
- **解答:**
  - 令 E = '第一个球是黑色' ，F = '第二个球是红色' 。
  - (a) $P(E) = \frac{67}{100} = 0.67$
  - (b) 抽取两个球的方式有 100×99 种 。 先抽黑球再抽红球的方式有 67×33 种 。 所以 $P(E \cap F) = \frac{67 \times 33}{100 \times 99} \approx 0.2233$ 。
  - (c) $P(F|E) = \frac{P(E \cap F)}{P(E)} = \frac{(67 \times 33)/(100 \times 99)}{67/100} = \frac{33}{99} = \frac{1}{3}$ 。(直观解释：已知E发生，袋中还剩99个球，其中33个是红球 。)

1.2 概率的乘法定理 (Multiplication Theorem)

- 也称为“乘法法则” (Product rule) 。
- 假设 E, F 是事件且 $P(E) > 0$ ，则:
  $P(E \cap F) = P(F|E)P(E)$

示例 10: 投资问题

- 一个投资者考虑投资股票和黄金 。 金融专家告知，来年股票上涨的概率为 0.6，黄金上涨的概率为 0.8，两者都上涨的概率为 0.5 。
  - **解答:**
    - 令 E = '股票上涨'，F = '黄金上涨' 。
    - 已知 $P(E) = 0.6, P(F) = 0.8, P(E \cap F) = 0.5$ 。
    - (a) 至少一项资产上涨的概率:
      $P(E \cup F) = P(E) + P(F) - P(E \cap F) = 0.6 + 0.8 - 0.5 = 0.9$ 。
    - (b) 黄金上涨但股票不上涨的概率:
      $P(F) = P(E \cap F) + P(E' \cap F)$，所以 $P(E' \cap F) = P(F) - P(E \cap F) = 0.8 - 0.5 = 0.3$ 。
    - (c) 在黄金上涨的条件下股票上涨的条件概率:
      $P(E|F) = P(E \cap F)/P(F) = 0.5/0.8 = 0.625$ 。
- **问题续:**
  - 投资者也考虑投资房产 。 专家称，如果股票上涨，那么房产上涨的概率是 0.9 。
  - **解答:**
    - 令 G = '房产上涨' 。 已知 $P(G|E) = 0.9$ 。
    - (d) 股票和房产都上涨的概率:
      $P(E \cap G) = P(G|E) \times P(E) = 0.9 \times 0.6 = 0.54$ 。
    - (e) 关于房产上涨概率的已知信息:
      因为 $G \supseteq E \cap G$，所以 $P(G) \geq P(E \cap G) = 0.54$ 。
      同时，$P(E \cup G) = P(E) + P(G) - P(E \cap G) = P(G) + 0.06 \leq 1$ ，所以 $P(G) \leq 0.94$ 。

---

2. 事件的独立性 (Independence of Events)

- **非正式定义:** 如果事件 F 的发生或不发生不改变事件 E 的概率（反之亦然），那么 E 和 F 是独立的 。
- **正式定义:** 事件 E 和 F 被定义为独立的，如果：
$P(E \cap F) = P(E) \times P(F)$
- **与条件概率的关系:**
    - 如果 E 和 F 独立且 $P(E) > 0$，那么 $P(F|E) = P(F)$ 。这意味着，当 $P(E) > 0$ 时，E 和 F 独立的充要条件是 F 在 E 发生下的条件概率与 F 的无条件概率相同 。
    - 同理，如果 $P(F) > 0$，E 和 F 独立的充要条件是 $P(E|F) = P(E)$ 。
- **实践中的应用:** 独立性通常是基于对实验条件的了解而做出的一个**模型假设**，而不是通过计算来发现的 。

示例 11: 掷两个骰子

- 掷两个"公平的"标准骰子，求总点数为7的概率 。
- **解答:**
    - 令 E = '总点数为7' = $\{(1,6),(2,5),(3,4),(4,3),(5,2),(6,1)\}$ 。
    - 两个骰子的点数是独立的 。 因此，任何特定组合（如(1,6)）的概率是
    $P(\{(1,6)\}) = P('第一个骰子为1') \times P('第二个骰子为6') = \frac{1}{6} \times \frac{1}{6} = \frac{1}{36}$ 。
    - 所有6种组合的概率都相同，所以 $P(E) = 6 \times \frac{1}{36} = \frac{1}{6}$ 。

2.1 多个事件的独立性

- 三个事件 E, F, G 被称为独立的，如果:
    1. 它们**两两独立** (pairwise independent) 。
    2. $P(E \cap F \cap G) = P(E)P(F)P(G)$ 。
- 这两个条件都是必需的，因为一个并不必然能推导出另一个 。

示例 12: 掷三个骰子

- 掷三个公平的标准骰子 。 共有 $6^3 = 216$ 种等可能的结果 。
    - (a) 三个骰子点数相同的概率:
    E_1 = {(1,1,1), (2,2,2), ..., (6,6,6)} 。 $P(E_1) = \frac{6}{216}$ 。
    - (b) 三个骰子点数都不同的概率:
    可能的结果数 = $6 \times 5 \times 4 = 120$ 。 $P(E_2) = \frac{120}{216}$ 。
    - (c) 两个点数相同，第三个不同的概率:
    该事件 E_3 是 E_1 和 E_2 的并集的补集 。 由于 E_1 和 E_2 互斥， $P(E_3) = 1 - [P(E_1) + P(E_2)] = 1 - [\frac{6}{216} + \frac{120}{216}] = \frac{90}{216}$ 。

2.2 互斥事件与独立事件的比较

- 如果两个概率非零的事件 E 和 F 是**互斥**的，那么它们**不可能是独立**的 。 因为知道 E 发生就意味着 F 不可能发生 。 形式上，$P(E \cap F) = 0$，而 $P(E) \times P(F) > 0$ 。
- **总结 (对于 $P(E) > 0, P(F) > 0$)**
    - **互斥事件:**
        - $P(E \cap F) = 0$
        - $P(E \cup F) = P(E) + P(F)$
        - $P(F|E) = 0$
    - **独立事件:**
        - $P(E \cap F) = P(E)P(F)$
        - $P(E \cup F) = P(E) + P(F) - P(E)P(F)$
        - $P(F|E) = P(F)$

---

3. 划分与全概率定律 (Partitions and Total Probability)

- **划分 (Partition):** 一系列事件 $E_1, E_2, ...$ 构成样本空间 S 的一个划分，如果：
    1. $\bigcup_i E_i = S$
    2. 当 $i \neq j$ 时，$E_i \cap E_j = \emptyset$ (互斥)
    3. 对每个 i, $P(E_i) > 0$
- 特别地，如果 E 是一个事件且 $0 < P(E) < 1$，那么 E 和 E' 就构成 S 的一个划分 。
- **全概率定律 (Law of Total Probability):**
    - 假设事件 $E_1, E_2, ...$ 划分了样本空间 S 。令 F 为 S 中的任意事件 。则:
    $P(F) = \sum_i P(F|E_i)P(E_i)$
    - **证明思路:** F 可以表示为一系列互斥事件 $(F \cap E_i)$ 的并集 。 因此 $P(F) = \sum_i P(F \cap E_i)$ ，再根据乘法定理

$P(F \cap E_i) = P(F|E_i)P(E_i)$ 即可得证 。

示例 13: 幽门螺杆菌治疗

- 一种药物疗法对耐药性患者的根除成功率为 75%，对非耐药性患者的成功率为 95% 。 在某个感染人群中，25% 的人是耐药的 。 问该人群中治疗失败的比例是多少?
- **解答:**
    - 令 E_1 = '患者耐药', E_2 = '患者不耐药' 。 F = '治疗失败' 。
    - $P(E_1) = 0.25, P(E_2) = 0.75$ 。
    - 治疗成功的条件概率: $P(F'|E_1) = 0.75, P(F'|E_2) = 0.95$ 。
    - 治疗失败的条件概率: $P(F|E_1) = 1 - 0.75 = 0.25, P(F|E_2) = 1 - 0.95 = 0.05$ 。
    - 根据全概率定律:
      $P(F) = P(F|E_1)P(E_1) + P(F|E_2)P(E_2) = (0.25 \times 0.25) + (0.05 \times 0.75) = 0.1$ 。
    - 因此，该人群中有十分之一的患者治疗失败 。

---

4. 贝叶斯定理 (Bayes' Theorem)

- **定理:** 假设事件 $E_1, E_2, \ldots$ 划分了样本空间 S 。 令 F 为 S 中的任意事件且 $P(F) > 0$ 。则对于任意 j:
  $P(E_j|F) = \frac{P(F|E_j)P(E_j)}{\sum_i P(F|E_i)P(E_i)}$
- **证明思路:** 分子 $P(F|E_j)P(E_j)$ 等于 $P(F \cap E_j)$ 。 分母 $\sum_i P(F|E_i)P(E_i)$ 等于 $P(F)$ （根据全概率定律） 。 定理可由条件概率的定义 $P(E_j|F) = P(F \cap E_j)/P(F)$ 得出 。

示例 13 (续):

- 如果一个患者治疗失败，那么该患者是耐药性的概率是多少?
- **解答:**
    - 我们需要计算 $P(E_1|F)$。
    - 根据贝叶斯定理:
      $P(E_1|F) = \frac{P(F|E_1)P(E_1)}{P(F)} = \frac{0.25 \times 0.25}{0.1} = 0.625$
    - **结论:** 尽管人群中只有 25% 的人是耐药的（先验概率），但在治疗失败的人群中，有 62.5% 的人是耐药的（后验概率） 。

概率论 (Probability) Level M [STATS 5024] - 讲座 4 笔记

---

1. 随机变量与概率分布 (Random Variables and Probability Distributions)

1.1 随机变量的定义 (The Definition of a Random Variable)

在很多实验中，我们感兴趣的不是结果本身，而是与结果相关联的某个数值 。

- **示例 1 (内在数值型):** 记录明天富时100指数中价值下跌的股票数量 。 结果本身就是一个数字 。 样本空间为 $S = \{0, 1, 2, \ldots, 100\}$ 。
- **示例 2 (编码为数值型):** 请一名格拉斯哥大学学生对“MyCampus在记录学生数据方面运行良好”这一说法做出回应，选项为“非常不同意”、“不同意”、“中立”、“同意”、“非常同意” 。 样本空间是分类的 ，但我们可以用数字代码 $\{-2, -1, 0, 1, 2\}$ 来代表这些结果 。
- **示例 3 (从函数型结果中派生数值):** 记录24小时的连续温度图 。 实验结果是一个函数 。 我们可以从中提取一个我们感兴趣的数值，比如一天中记录的最高温度 。
- **示例 4 (连续数值型):** 在一种特殊的淘汰赛制足球比赛中，比赛时长（分钟）是 [90, 120] 区间内的一个数字 。

正式定义

- **随机变量 (Random Variable):** 是一个函数， $X : S \to \mathbb{R}$，它为样本空间 S 中的每一个结果 s 关联一个唯一的数值 $X(s)$ 。
- **实现 (Realisation):** 每进行一次实验，只观察到一个结果 $s \in S$，因此也只观察到随机变量的一个值 $X(s)$ 。 这个值被称为随机变量的一次实现 。
- **符号约定:** 通常用大写字母（如X）表示随机变量，用相应的小写字母（如x）表示它的一般实现 。

值域空间 (Range Space)

- **定义:** 随机变量X的值域空间 R_X 是X所有可能实现值的集合 。
- **示例 1 (续):** 随机变量X是价值下跌的股票数量 。 $R_X = S = \{0, 1, 2, \ldots, 100\}$ 。
- **示例 2 (续):** 随机变量Y代表学生的回应编码 。 $R_Y = \{-2, -1, 0, 1, 2\}$ 。
- **示例 3 (续):** 随机变量Z是24小时内的最高温度 。 $R_Z = [-273.15, \infty)$ 。

与随机变量相关的概率

- 随机变量的实现是由底层实验的结果完全决定的 。
- 这使得我们可以讨论与X相关的概率，例如 $P(X = x)$ 或 $P(X \leq x)$ 。

- 这些概率是从原始样本空间的概率派生出来的 。例如，$P(X \le x)$ 是事件 $\{s \in S : X(s) \le x\}$ 的概率 。

示例 5: 老鼠走迷宫

- **实验:** 一只老鼠被放入一个有两条路径的迷宫（一条通向食物，一条不通） 。每只老鼠尝试三次 。
- **模型1 (无学习):**
  - 样本空间 S = {SSS, SSF, SFS, SFF, FSS, FSF, FFS, FFF} 。
  - 假设老鼠不会从经验中学习，那么8个结果都是等可能的，概率为 $\frac{1}{8}$ 。
  - 定义随机变量X为老鼠成功找到食物的次数 。$R_X = \{0, 1, 2, 3\}$ 。
  - $P(X = 0) = P(\{FFF\}) = \frac{1}{8}$ 。
  - $P(X = 1) = P(\{SFF, FSF, FFS\}) = \frac{3}{8}$ 。
  - $P(X = 2) = P(\{SSF, SFS, FSS\}) = \frac{3}{8}$ 。
  - $P(X = 3) = P(\{SSS\}) = \frac{1}{8}$ 。
- **模型2 (有学习):**
  - 假设老鼠第一次成功的概率是0.5 。一旦成功，后续尝试将永远成功 。在首次成功前，每次尝试成功的概率保持0.5不变 。
  - $P(X = 0) = P(\{FFF\}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$ 。
  - $P(X = 1) = P(\{FFS\}) = \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{8}$ 。
  - $P(X = 2) = P(\{FSS\}) = \frac{1}{2} \times \frac{1}{2} \times 1 = \frac{1}{4}$ 。
  - $P(X = 3) = P(\{SSS\}) = \frac{1}{2} \times 1 \times 1 = \frac{1}{2}$ 。
- 注意，在两种模型中，与X相关的四个概率之和都为1 。

1.2 分布函数与百分位数 (The Distribution Function and Percentiles)

分布函数 (Distribution Function, d.f.)

- **定义:** 随机变量X的分布函数（d.f.），有时也称为累积分布函数（c.d.f.），定义为：
  $F_X(x) = P(X \le x), \quad x \in \mathbb{R}$
- 分布函数包含了我们需要了解的关于X的所有概率信息 。例如，$P(a < X \le b) = F(b) - F(a)$ 。

分布函数的性质

- 任何有效的分布函数 F_X 都具有以下性质：
  1. $F_X(-\infty) = 0$ 且 $F_X(\infty) = 1$ 。
  2. 如果 $a \le b$，则 $F_X(a) \le F_X(b)$（即 F_X 在 $\mathbb{R}$ 上是**非递减**的） 。
  3. 对于所有 $x \in \mathbb{R}, \lim_{h \to 0^+} F_X(x + h) = F_X(x)$（即 F_X 是**右连续**的） 。

随机变量的类型

- **离散随机变量 (Discrete Random Variable):**
  - 定义：值域空间是有限或可数的随机变量 。示例5中的X就是离散的 。
  - 其分布函数由一系列在 $x \in R_X$ 处的阶梯跳跃组成 。
- **连续随机变量 (Continuous Random Variable):**
  - 定义：值域空间是不可数无限的，并且其分布函数在各处都是连续的（不仅是右连续） 。
  - **示例 6 (随机数):** 从 [0, 1] 中抽取一个"随机数"X 。其分布函数为：
    $F_X(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x \ge 1 \end{cases}$
- **混合随机变量 (Mixed Random Variable):**
  - 定义：其分布函数可以写成一个离散分布函数和一个连续分布函数的加权和形式：
    $F_X(x) = \lambda F_d(x) + (1 - \lambda) F_c(x)$, 其中 $0 < \lambda < 1$ 。
  - **示例 4 (足球比赛) (续):** 比赛时长X就是一个混合随机变量 。它的分布函数在 x=90 和 x=120 处有跳跃，但在 [90, 120) 区间内是连续增加的 。

百分位数 (Percentiles)

- **定义:** 对于 $0 < \alpha < 1$，随机变量X（或其概率分布）的第 100α 百分位数是使 $F_X(x) \ge \alpha$ 成立的最小的x值 。
- **中位数 (Median):** 是第50百分位数 。
- **计算示例:**
  - **示例 5 (离散):** 分布函数图显示 $F_X(2) = 0.5$ 而当 $x < 2$ 时 $F_X(x) < 0.5$ 。因此，中位数是2 。
  - **示例 6 (连续):** 解方程 $F_X(x) = 0.5$，即 $x = 0.5$ 。因此，中位数是0.5 。
  - **示例 4 (混合):** 不存在x使得 $F_X(x) = 0.5$ 。在 $x = 90$ 处，$F_X(x)$ 从0跳跃到0.7 。因此，使 $F_X(x) \ge 0.5$ 成立的最小x值是90 。所以中位数是90 。

概率论 (Probability) Level M [STATS 5024] - 讲座 5 笔记

---

1. 随机变量的一些性质 (Some Properties of Random Variables)

虽然分布函数 (d.f.) 包含了随机变量的所有概率信息，但它通常不是最直观的总结方式 。 对于离散和连续随机变量，我们分别有更具信息性的函数：概率质量函数 (p.m.f.) 和概率密度函数 (p.d.f.)。

---

2. 概率质量函数 (The Probability Mass Function, p.m.f.)

2.1 定义与性质

- **定义:**
  - 设 X 是一个**离散随机变量** 。 其概率质量函数 (p.m.f.) 定义为：
    $p_X(x) = P(X = x), \quad x \in \mathbb{R}$
  - 由于对于所有不在值域空间 R_X 内的 x， $P(X = x) = 0$，因此在讨论 p.m.f. 时，通常只关注值域空间内的 x 值 。
- **p.m.f. 的性质:**
  - (a) 对于所有 $x \in \mathbb{R}$，有 $0 \leq p_X(x) \leq 1$，因为 $p_X(x)$ 本身就是一个概率 。
  - (b) $\sum_{x \in R_X} p_X(x) = 1$ 。 这是因为对于 $x \in R_X$ 的所有事件 $\{s : X(s) = x\}$ 构成了样本空间 S 的一个划分 。

2.2 示例与应用

示例 7: 塑料棒硬度测试

- **实验:** 用已知力反复敲击一根塑料棒直到其断裂 。
- **随机变量:** X 是使棒断裂所需的敲击次数 。
- **假设:** 每次敲击时棒断裂的概率为 0.5 ，且每次敲击的结果是独立的 。
- **推导 p.m.f.:**
  - X 的值域空间是 $R_X = \{1, 2, 3, \ldots\}$ 。
  - $p_X(1) = P(X = 1) = P(\text{第1次敲击断裂}) = 0.5$ 。
  - $p_X(2) = P(X = 2) = P(\text{第1次未断裂且第2次断裂}) = 0.5 \times 0.5 = 0.25$ 。
  - $p_X(3) = P(X = 3) = P(\text{前2次未断裂且第3次断裂}) = 0.5 \times 0.5 \times 0.5 = 0.125$ 。
  - 推广可得：$p_X(x) = (\frac{1}{2})^x$, 对于 $x = 1, 2, 3, \ldots$ 。

知识点: 几何级数 (Geometric Series)

- **无限和:** 当公比 r 满足 $-1 < r < 1$ 时，无限几何级数 $T = a + ar + ar^2 + \ldots$ 的和为 $T = \frac{a}{1-r}$ 。
- **有限和:** 前 k 项的和为 $T_k = \frac{a(1-r^k)}{1-r}$ 。

示例 7 (续):

- **验证 p.m.f.:**
  - $\sum_{x \in R_X} p_X(x) = \sum_{x=1}^{\infty} (\frac{1}{2})^x = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \ldots$ 。
  - 这是一个首项 $a = \frac{1}{2}$、公比 $r = \frac{1}{2}$ 的几何级数 。
  - 其和为 $T = \frac{0.5}{1-0.5} = 1$，符合 p.m.f. 的性质 。
- **从 p.m.f. 计算分布函数 (d.f.):**
  - $P(X \leq 3) = p_X(1) + p_X(2) + p_X(3) = 0.5 + 0.25 + 0.125 = 0.875$ 。
  - 对于任意 $x \in R_X$， $P(X \leq x) = \sum_{i=1}^{x} (\frac{1}{2})^i = 1 - (\frac{1}{2})^x$ 。
  - 知道 p.m.f. 和 d.f. 中的任何一个，总能推导出另一个 。

---

3. 概率密度函数 (The Probability Density Function, p.d.f.)

3.1 定义与性质

- **一个重要事实:**
  - 如果 X 是一个**连续随机变量**，那么对于所有的 x，都有 $P(X = x) = 0$ 。
  - **证明概要:** 我们可以通过一个围绕 x 的不断缩小的区间序列 $E_n = [x - \frac{1}{n}, x + \frac{1}{n}]$ 来证明 。 由于连续随机变量的分布函数 F 是连续的，当 $n \to \infty$ 时，这个区间的概率 $P(E_n) = F(x + \frac{1}{n}) - F(x - \frac{1}{n})$ 会趋近于 $F(x) - F(x) = 0$ 。
  - 因此，对于连续随机变量，p.m.f. 的概念是无意义的 。
- **定义:**
  - 设 X 是一个连续随机变量 。 其概率密度函数 (p.d.f.) 定义为分布函数的导数：

$f_X(x) = \frac{d}{dx}F_X(x), \quad x \in \mathbb{R}$

  - 反之，分布函数可以通过对 p.d.f. 积分得到：
    $F_X(x) = \int_{-\infty}^{x} f_X(t)dt$

- **p.d.f. 的含义:**
  - $f_X(x)$ 本身不是一个概率 。
  - 它的值与 X 落在以 x 为中心的一个微小区间内的概率成正比 。即 $P(x - \frac{\delta}{2} < X \le x + \frac{\delta}{2}) \approx \delta \times f_X(x)$ 。
- **p.d.f. 的性质:**
  - (a) 对于所有 $x \in \mathbb{R}$，有 $f_X(x) \ge 0$ 。这是因为它是非递减函数 F_X 的导数 。
  - (b) $\int_{-\infty}^{\infty} f_X(x)dx = 1$ 。这是因为 $F_X(\infty) = 1$ 。

3.2 计算与应用

- **用 p.d.f. 计算概率:**
  - 连续随机变量 X 落在区间 (a, b) 内的概率可以通过对 p.d.f. 在该区间上积分得到：
    $P(a < X \le b) = \int_a^b f_X(x)dx$
  - 对于连续变量，由于单点概率为0，因此开区间和闭区间的概率是相同的： $P(a < X < b) = P(a \le X \le b)$ 。

示例 6 (续): 均匀分布 (Uniform Distribution)

- 随机变量 X 在 (0, 1] 上均匀分布，其分布函数为：
  $F_X(x) = \begin{cases} 0, & x \le 0 \\ x, & 0 < x \le 1 \\ 1, & x > 1 \end{cases}$
- 对 F(x) 求导，得到其 p.d.f.：
  $f_X(x) = \begin{cases} 1, & 0 < x \le 1 \\ 0, & \text{其他} \end{cases}$
- 这个分布被称为在0到1上的均匀分布，记为 $X \sim U(0, 1)$ 。
- **一般均匀分布:**
  - 如果一个连续随机变量 X 在区间 [a, b] 上服从均匀分布，记为 $X \sim U(a, b)$，其 p.d.f. 和 d.f. 分别为：
    $f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x \le b \\ 0, & \text{其他} \end{cases} \quad \text{和} \quad F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$

概率论 (Probability) Level M [STATS 5024] - 讲座 6 笔记

---

1. 矩 (Moments)

矩是用于概括和描述概率分布特性的数值度量，其中最重要的是期望值和方差。

1.1 期望值 (Expected Value)

定义
随机变量 X 的期望值（如果存在）定义为：

- **对于离散随机变量:**
  $E(X) = \sum_{x \in R_X} x p_X(x)$
- **对于连续随机变量:**
  $E(X) = \int_{x \in R_X} x f_X(x)dx$
- **存在条件:** 期望值只有在相应的求和 $\sum |x| p_X(x)$ 或积分 $\int |x| f_X(x)dx$ 收敛到一个有限值时，才被认为是明确定义的 。

期望值的解释

- $E(X)$ 可以被解释为，如果底层随机实验被大量重复，那么 X 的观测值的长期平均值将是 $E(X)$ 。(这将在课程稍后通过强大数定律和弱大数定律进行精确说明) 。
- 通常，如果 X 的分布是对称的，并且期望值存在，那么 $E(X)$ 就位于对称中心 。

计算示例

- **示例 5 (离散型):**
  对于老鼠走迷宫的例子，$R_X = \{0, 1, 2, 3\}$ ，其概率质量函数为 $p_X(0) = 0.125, p_X(1) = 0.125, p_X(2) = 0.25, p_X(3) = 0.5$ 。
  $E(X) = (0 \times 0.125) + (1 \times 0.125) + (2 \times 0.25) + (3 \times 0.5) = 2.125$ 。
  这个例子表明，对于离散随机变量，其期望值不一定需要是其可能实现的值之一 。
- **示例 8 (连续型 - 均匀分布):**

如果 $X \sim U(a,b)$，那么 $f_X(x) = \frac{1}{b-a}$ 对于 $a < x < b$ 。
$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{1}{b-a}[\frac{1}{2}x^2]_a^b = \frac{a+b}{2}$ 。
由于均匀分布是对称的，其期望值和中位数都位于区间的中心 $\frac{a+b}{2}$ 。

随机变量函数的期望

- **定义:** 如果 $g(X)$ 是随机变量 X 的一个函数，那么它的期望值定义为：
  - **离散情况:** $E[g(X)] = \sum_{x \in R_X} g(x)p_X(x)$ 。
  - **连续情况:** $E[g(X)] = \int_{x \in R_X} g(x)f_X(x)dx$ 。
  - 前提是相应的求和或积分是绝对收敛的 。
- **示例 8.5 (函数期望):** 如果 $X \sim U(a,b)$，那么 $E[e^X] = \int_a^b e^x \frac{1}{b-a} dx = \frac{e^b - e^a}{b-a}$ 。
- **期望的线性性质 (示例 9):**
  对于任意常数 a 和 b， $E[aX+b] = aE(X)+b$ 。
  证明 (离散情况):
  $E[aX+b] = \sum(ax+b)p_X(x) = a\sum xp_X(x) + b\sum p_X(x) = aE(X)+b$ 。
  这个结果对于连续随机变量同样成立 。

1.2 中心矩与方差 (Central Moments & Variance)

矩的定义

- **k阶矩 (关于c):** $E[(X-c)^k]$ 。
- **k阶原点矩 (或称k阶矩):** $\mu_k \equiv E(X^k)$ 。 因此，一阶矩就是期望值 。
- **k阶中心矩:** $E[(X-\mu)^k]$，其中 $\mu = E(X)$ 。

方差 (Variance)

- **定义:** 方差是**二阶中心矩**，记作 $Var(X)$。$Var(X) = E[(X-\mu)^2]$ 。
- **解释:**
  - 方差是衡量随机变量 X 分布**离散程度**或**扩展性**的度量 。
  - 它是 X 的值与其期望值之间距离平方的平均值 。
  - 由于它是一个非负量的期望值，所以方差本身必须是非负的，$Var(X) \geq 0$ 。
- **方差的性质 (示例 9 续):**
  - 对于任意常数 a 和 b， $Var(aX+b) = a^2Var(X)$ 。
  - **证明:**
    $Var(aX+b) = E([(aX+b) - E(aX+b)]^2) = E([(aX+b) - (aE(X)+b)]^2) = E([a(X-E(X))]^2) = a^2E[(X-E(X))^2] =$
    。
- **方差的计算公式:**
  - 一个更便捷的计算公式是：$Var(X) = E(X^2) - [E(X)]^2$ 。
  - **推导:**
    $Var(X) = E[(X-E[X])^2] = E[X^2 - 2XE[X] + (E[X])^2] = E[X^2] - 2E[X]E[X] + (E[X])^2 = E[X^2] - [E(X)]^2$ 。
- **示例 8 (均匀分布的方差):**
  对于 $X \sim U(a,b)$，我们已经知道 $E(X) = \frac{a+b}{2}$ 。
  首先计算 $E(X^2) = \int_a^b x^2 \frac{1}{b-a} dx = \frac{1}{3}(a^2+ab+b^2)$ 。
  因此，$Var(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}$ 。

标准差 (Standard Deviation)

- 方差的单位是随机变量 X 单位的平方 。
- 为了得到一个与 X 具有相同单位的离散程度度量，我们使用标准差 。
- **定义:** $\sigma(X) = \sqrt{Var(X)}$ 。

学习路径回顾

1. **实验结果 (Outcomes)**
2. **事件 (Events)**: 结果的集合
3. **概率 (Probabilities)**: 分配给事件的数值 $P(E)$
4. **随机变量 (Random Variable)**: 将结果映射到数值的函数 $X(s)$
5. **概率分布 (Probability Distributions)**: 描述随机变量概率行为的函数 ($p_X(x)$ 或 $f_X(x)$)
6. **矩 (Moments)**: 概率分布的数值摘要 (如期望值、方差)