

MATH1207 Probability

Xiyuan Yang

2025.10.21

Lecture and Summary Notes for MATH 1207: Probability

Contents

1. Introduction	1
2. Chapter1: Random Events and Probability	1
2.1. 基本概念	1
2.1.1. Probabilities as Sets	2
2.1.1.1. 事件运算的化简	2
2.1.2. Axiomatic Probability Theory	2
2.2. Classical Probability	2
2.3. Geometric Probability	3
2.4. 概率的性质	3
2.4.1. 加法原理	4
2.5. 条件概率	4
2.5.1. 条件概率的性质保持	4
2.5.2. 完备事件组	5
2.5.3. 全概率公式	5
3. 随机变量及其分布	5
3.1. 随机变量及其分布函数	5
3.1.1. 分布函数	5
3.1.1.1. 分布函数的定义	6
3.1.1.2. 分布函数的性质	6
3.2. 离散型随机变量及其概率分布	6
3.2.1. 常见离散型随机变量的分布	6
3.2.1.1. Poisson Distribution	7
3.3. 连续性随机变量及其概率分布	7
3.3.1. 概率密度函数	7
3.3.2. 均匀分布	8
3.3.3. 指数分布	8
3.3.4. 高斯分布	9
3.4. 随机变量的函数及其分布	9
3.4.1. 离散型随机变量函数的概率分布	9
3.4.2. 连续型随机变量函数的概率分布	9
4. Conclusion	10

§1. Introduction

§2. Chapter1: Random Events and Probability

§2.1. 基本概念

Axiomatic Probability Theory

§2.1.1. Probabilities as Sets

Definition 2.1.1.1 (Basic Definition for probability).

- 样本点
 - 单个样本点之间存在互斥性，类似于集合中的 partition。
- 样本空间：由样本点组成的集合 Ω
- 基本事件：单个样本点组成的集合 $\{x\} \in \mathcal{P}(\Omega), x \in \Omega$
- 随机事件 $A \in \mathcal{P}(\Omega)$

Recordings (Probabilities as sets).

用集合论的眼光来看概率的定义

§2.1.1.1. 事件运算的化简

- 巧用各种公式进行集合运算
- 注意括号和集合之间的分配率

$$A - B = A\overline{B} = A - AB$$

- **Attention!**，注意 \overline{AB} and $\overline{A} \overline{B}$
 - $\overline{AB} = \overline{A \cap B}$
 - $\overline{A} \overline{B} = \overline{A \cup B}$

§2.1.2. Axiomatic Probability Theory

- 概率本身来源于测度论，或者说概率就是一种概率测度，并且是一种满足 $P(\Omega) = 1$ 的特殊测度。
- 测度论的三个公理条件
 - 非负性
 - 规范性
 - 可列可加性

§2.2. Classical Probability

在分析古典概型使用不要忘了下面几点：

Recordings (classical probability).

- 古典概型必须保证两个前提条件
 - Ω 为有限集
 - 测度就是元素的个数
 - 保证每个样本点的概率都是相同的
- 解题思路：
 - 判断是否等可能，考虑使用几何概型
 - 这一步最关键的就是建模出有限的样本空间和等可能的样本点
 - 计算不同时间的测度
 - 计算 Ω 的测度
 - 计算 A 的测度

- 相关模型
 - 抽球问题：
 - 放回：二项分布
 - 不放回：超几何分布
 - 抽签模型
 - 等可能性
 - 使用排列组合求解
 - 球入盒子模型
 - 可以考虑计算对立事件的概率
 - 看清楚题目是否指定盒子 和 指定球，这将决定是否需要再额外乘一个组合数（从不指定要指定涉及一个先选球的排列问题）
 - 生日悖论

§2.3. Geometric Probability

几何概型作为古典概型的延伸，具体的延伸在于将样本空间的数量提升到了无限。但是此时基本事件的概率变为 0，因此无法比较等可能性，因此定义如下：

Definition 2.3.1 (Geometric Probability).

- 存在有限区域 Ω
 - 区域可以是 1D 2D 3D
- 对于区域测度的度量就是对应的面积、体积和长度等等
- 关键：落入子区域 G 的概率和区域 G 的测度成正比。

Recordings (Recordings for Geometric Probability).

- 有关几何概型的具体结论和思考
- 同样的，识别出几何概型之后，往往需要将题目中的事件约束条件刻画为一个不等式或者其他约束条件
 - 再将这个约束条件的测度转化为一个区域
 - 有点类似于线性规划问题
- 因此，针对几何概型的建模问题多半是多变量的约束的规划问题
 - 多变量导致了几何概型建模是在一维还是更高维的空间中进行。

Example (投针问题).

- 投针问题是典型的通过变量约束的规划条件转化为几何概型进行求解的
- 不过要注意明确自由变量的整体样本空间的范围，不要多数或者少数。
 - 例如这里的角度 θ 只是在 $[0, \pi]$ 中的范围内。
- 最终转化为一个集合区域的测度

§2.4. 概率的性质

- 概率的若干不等式：

$$P(A) \leq P(\Omega) = 1$$

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

- 差事件的概率

$$B - A = B\bar{A} = B - AB$$

$$P(B - A) = P(B) - P(AB)$$

$$A \subset B, P(B - A) = P(B) - P(A)$$

Recordings (概率的性质).

- 核心在于灵活的使用概率的公理化定义, 尤其是第三条。有时候我们需要通过事件集合之间的运算灵活地构造这样的互斥事件。
- 使用加法原理可以将和事件的求解转化为积事件。
- 例如排列问题或者“至少...”
 - 这类问题除了考虑对立事件之外, 也可以考虑把这个概率的时间拆成若干个子事件的事件

§2.4.1. 加法原理

对于两个事件的加法原理:

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

对于一般性的加法原理见书上。

Problem 2.4.1.1 (Matching Problem).

给定 n 个编号的盒子和 n 个编号的小球, 每个盒子和每个小球存在一一对应的关系, 现在固定盒子并让小球随机排列, 求至少有一对小球和盒子成功配对的概率。

使用一般性的加法原理, 最终计算得到:

$$P(A) = \sum_{i=1}^n \frac{(-1)^{i-1}}{i!} = 1 - \frac{1}{e}$$

§2.5. 条件概率

在古典概型中, 可以根据集合的计数运算从“条件”出发给出条件概率的表达式。使用缩减样本空间的方法。

因此, 对于一般的条件概率是使用算式定义的。

§2.5.1. 条件概率的性质保持

- 可列可加性仍然保持
- 加法原理和基本的概率性质仍然保持

$$P(A_1 \cup A_2 \mid B) = P(A_1 \mid B) + P(A_2 \mid B) - P(A_1 A_2 \mid B)$$

$$P(A_1 - A_2 \mid B) = P(A_1 \mid B) - P(A_1 A_2 \mid B)$$

- 联合积事件的概率可以使用基本的条件概率改写成连乘积

- ▶ 这样的模型可以实现多次实验之间存在相关性的场景（第一次实验的结果会影响第二次实验结果的概率分布，例如不放回摸球实验）
- 高中的基本不能忘：定义基本事件，把题目中的若干事件改写成若干基本事件的集合事件运算

§2.5.2. 完备事件组

Definition 2.5.2.1 (Partition).

- 这样的完备事件组和概率公理化定义中的可列可加性保持一致
- 对于事件 B_1, B_2, \dots, B_n

$$\Omega = \cup_{i=1}^n B_i$$

§2.5.3. 全概率公式

首先是对样本空间的划分，将原始事件转化为一系列积事件的和事件。

$$\Omega = \cup_{i=1}^n B_i$$

$$A = \cup_{i=1}^n AB_i$$

Proposition 2.5.3.1 (全概率公式).

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

Proposition 2.5.3.2 (Bayes).

- 先验概率由过去的经验（过去的统计结果）

§3. 随机变量及其分布

§3.1. 随机变量及其分布函数

- 随机变量是从样本空间到实数域上的一个映射
 - ▶ 因此理论上你可以定义一些随机变量，将样本空间的一些基本时间点映射到一个实数域
 - ▶ 对于一些事件，这些定义是显然的
- 随机变量具有概率性，可以表示更复杂的随机事件
- 对于离散型随机变量，可以看做是连续性随机变量的特殊情况（取点）

Recordings (随机变量).

- 这里的变量是广义上的
- 一般来说，只需要对研究问题的样本空间的所有基本事件构成了一个映射，就可以认为是一个随机变量。
- 随机变量使用映射将难以使用数学语言刻画的样本空间转化为了易于表示的实数 R

§3.1.1. 分布函数

§3.1.1.1. 分布函数的定义

Definition 3.1.1.1.1 (随机变量的分布函数).

X 为随机变量, x 是任意实数:

$$F(x) = P(X \leq x), -\infty < x < \infty$$

$F(x)$ 为随机变量的分布函数。

§3.1.1.2. 分布函数的性质

- $F(x)$ 单调递增
- 极限:

$$\lim_{x \rightarrow +\infty} F(x) = 1$$

$$\lim_{x \rightarrow -\infty} F(x) = 0$$

- $F(x)$ 右连续

$$P(a < X \leq b) = F(b) - F(a)$$

$$P(X = x_0) = \lim_{\Delta x \rightarrow 0} P(x_0 - \Delta x < X \leq x_0) = F(x_0) - \lim_{\Delta x \rightarrow 0} F(x_0 - \Delta x) = F(x_0) - F(x_0 - 0)$$

- 右连续, 但是左极限未知!
- 根据右连续可以推出下面的一些二级结论:

$$P(a < X \leq b) = F(b) - F(a)$$

$$P(X = a) = F(a) - F(a - 0)$$

$$P(a \leq X \leq b) = F(b) - F(a - 0)$$

$$P(a < X < b) = F(b - 0) - F(a)$$

$$P(a \leq X < b) = F(b - 0) - F(a - 0)$$

§3.2. 离散型随机变量及其概率分布

$$P(X = x_k) = p_k$$

- 非负性
- 规范性

离散型随机变量对应的分布函数图像是阶梯函数。

§3.2.1. 常见离散型随机变量的分布

- 两点分布
- n 重伯努利实验

Recordings (伯努利实验).

$$X \sim B(n, p)$$

$$P(X = k) = C_n^k p^k (1-p)^{n-k}$$

- 当 $(n+1)p$ 为整数的时候, 在 $(n+1)p, (n+1)p-1$ 两处取得相同的最大值。
- 如果不是整数, $\lfloor (n+1)p \rfloor$

- 帕斯卡分布

Example (帕斯卡分布).

一门大炮对目标进行射击, 假定此目标必须被击中 r 次才能被摧毁. 若每次击中目标的概率为 p ($0 < p < 1$), 且各次射击相互独立, 一次一次地射击直到摧毁目标为止. 求所需射击次数 X 的概率分布

- 可以分成两段来看, 保证最后一次肯定击中目标

$$P(X = k) = C_{k-1}^{r-1} p^{r-1} (1-p)^{k-r} p$$

§3.2.1.1. Poisson Distribution

Recordings (Poisson Distribution).

- 泊松分布本质上可以看成是一种针对二项分布的近似过程
 - $X \sim B(n, p)$, n is large and p is small, and $np = \lambda$
- 在使用泊松分布的使用, 不要忘了这个式子, 有时可以化简无穷项的求和:

$$\sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1$$

$$X \sim P(\lambda)$$

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

§3.3. 连续性随机变量及其概率分布

§3.3.1. 概率密度函数

可以认为是对分布函数的微分处理, 严格的数学定义如下:

Definition 3.3.1.1 (概率密度函数).

$$F(x) = \int_{-\infty}^x f(t) dt$$

Recordings (有关概率密度函数).

- 概率密度函数, 概率, 概率分布函数是完全不同的三个概念!

- 因为对于连续性随机变量而言很多情况下会出现单点概率 $P(x = x_0) = 0$ 的情况，在这种情况下需要一个概念来衡量：
 - 从定义上，概率密度函数是连续性随机变量的求导结果，或者说概率密度函数的变上限积分就是分布函数
- 描述的是 x 附近单位长度的区间内取值的概率

$$P(x < X < x + \Delta x) \approx f(x)\Delta x$$

- 存在既不连续也不离散的随机变量，对于间断点，此处不存在概率密度函数
 - 例如单点跳跃的情况

Recordings (不影响区间的概率).

- 因为概率值不会收到端点值的影响，因此讨论断点的概率密度函数是无意义的

$$P(a \leq X \leq b) = P(a < X < b) = P(a \leq X < b) = P(a < X \leq b)$$

- 连续性概率密度函数因为分布函数保证连续，因此不存在 $P(x = x_0) \neq 0$ 的情况。

§3.3.2. 均匀分布

$$x \sim U(a, b)$$

概率密度函数：

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } x \in (a, b) \\ 0 & \text{else} \end{cases}$$

分布函数

$$F(x) = \begin{cases} 0 & \text{if } x < a \\ \frac{x-a}{a-b} & \text{if } x \in [a, b) \\ 1 & \text{if } x \geq b \end{cases}$$

概率函数

$$P(c < X < d) = \frac{d-c}{a-b}, \forall (c, d) \subset (a, b)$$

§3.3.3. 指数分布

$$X \sim E(\lambda)$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{if } x < 0 \end{cases}$$

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

Recordings (有关指数分布).

- 指数分布具有无记忆性

$$P(X > s+t | X > s) = P(X > t)$$

- 指数分布的形式和泊松分布很像!

§3.3.4. 高斯分布

$$X \sim N(\mu, \sigma^2)$$

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

标准正态分布 $X \sim N(0, 1)$:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < \infty$$

$$\Phi(x) = \int_{-\infty}^x \varphi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

Recordings (标准正态分布和正态分布的性质).

- 对于正态分布:

$$P(|X| < x) = 2\Phi(x) - 1$$

- 转化为标准正态分布:

$$\Phi(x) = \Phi_0\left(\frac{x-\mu}{\sigma}\right)$$

§3.4. 随机变量的函数及其分布

用函数来描绘两个随机变量之间的映射关系

$$Y = g(X)$$

§3.4.1. 离散型随机变量函数的概率分布

Definition 3.4.1.1.

Given the definition of distribution X :

$$P(X = x_k) = p_k$$

We can calculate:

$$P(Y = y_i) = \sum_{k: g(x_k) = y_i} P(X = x_k)$$

§3.4.2. 连续型随机变量函数的概率分布

Theorem 3.4.2.1.

- 随机变量 X 的密度函数分布为 $f_X(x), x \in (-\infty, +\infty)$
- $y = g(x)$ 为 $(-\infty, \infty)$ 上的单调可导函数

$$f_Y(y) = \begin{cases} |h'(y)| \cdot f_X[h(y)] & \text{if } \alpha < y < \beta \\ 0 & \text{else} \end{cases}$$

推广到若干子区间的分段:

$$f_Y(y) = \sum_{i=1}^n |h'_{i(y)}| \cdot f_X[h_{i(y)}]$$

§4. Conclusion