概率论笔记

作者：肖程哲

时间：July 5, 2022

苟日新，日日新，又日新

目录

第 1 章 概率基础

内容提要

- 事件
- 古典概型与几何概率
- 条件概率与独立
- 乘法法则
- 全概率公式
- Bayes 法则

1.1 概率空间

定义 1.1 (样本空间)

随机试验可能出现的结果称为**样本点** (sample point)，用 ω 表示。样本的全体构成**样本空间** (sample space)，用 Ω 表示。

定义 1.2 (事件的古典定义)

样本点 ω 的集合称为**事件** (event)。

我们关心的随机现象被抽象为集合, 逻辑运算 (且, 或, 非, etc.) 对应成集合论运算 (交, 并, 补, etc.)。

性质 集合的运算性质:

交换律 $A \cup B = B \cup A, \quad AB = BA$

结合律 $(A \cup B) \cup C = A \cup (B \cup C), (AB)C = A(BC)$

分配律

$$(A \cup B) \cup C = A \cup (B \cup C), \tag{1.1}$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C). \tag{1.2}$$

对偶律 (De Morgan's laws)

$$\text{事件并的对立等于对立的交:} \quad \overline{A \cup B} = \overline{A} \cap \overline{B}, \tag{1.3}$$

$$\text{事件交的对立等于对立的并:} \quad \overline{A \cap B} = \overline{A} \cup \overline{B}. \tag{1.4}$$

为方便概率的定义，并不把 Ω 的一切子集作为事件，应避免不可测集的出现。

定义 1.3 (事件域)

事件构成的全体称为**事件域** $\mathscr{F}$，是 Ω 的子集族 (collection of subsets)，应满足 σ 代数 的要求：

- $\varnothing \in \mathscr{F}$, 无事发生;
- $A \in \mathscr{F} \implies A^{\mathsf{C}} \in \mathscr{F}$, 即 $\mathscr{F}$ 对补集运算 (逻辑上的非) 封闭;
- $A_1, \ldots, A_n, \ldots \in \mathscr{F} \implies \bigcap_{n=1}^{\infty} A_n \in \mathscr{F}$, 即 $\mathscr{F}$ 对可数交运算 (逻辑上的可数多个且) 封闭.

笔记 可数性是为了在数学上能够恰当地处理无穷 的概念, 术语中的 σ 指的就是可数并。由对偶原理可得 σ 域同时对可数并运算封闭. 即 σ 域对逆, 并, 交, 差的可数次运算封闭.

事件域根据问题的不同要求适当选取, 在概率定义没有困难时, 应尽量选大, 通常以 Ω 的一切子集作为事件域. 当 Ω 给定后, 若某些子集必须作为事件处理, 能否找到包含他们的 σ 域?

命题 1.1

若给定 Ω 的一个非空集族 $\mathscr{G}$, 必存在 Ω 上唯一的 σ 域 $\mathrm{m}(\mathscr{G})$, 满足下列性质:

- 包含 $\mathscr{G}$
- 若其他 σ 域包含 $\mathscr{G}$, 则必包含 $\mathrm{m}(\mathscr{G})$

这个 $\mathrm{m}(\mathscr{G})$ 称为包含 $\mathscr{G}$ 的最小 σ 域, 或由 $\mathscr{G}$ 扩张而成的 σ 域.

扩张, 或者称为延拓, 是数学中很重要的一个概念, 大抵是将某映射的定义域适当扩大, 不改变在初始定义域上的映射取值 (注意值域可能是比较抽象的集合, 配备了某些操作之后被称为空间), 同时在扩大后的定义域上仍然保持某些优良的性质. 与此相对的概念是限制, 即关心局部上可能更加漂亮的性质, 把初始的定义域适当缩小.

证明 由于 Σ 的一切子集构成的集类包含 $\mathscr{G}$, 所以 m 存在. 再取 Σ 上满足此条件的 σ 域之交作为 $\mathrm{m}(\mathscr{G})$ 即可.

特别地, 实数集 $\mathbb{R}$ 的子集族 $\{(-\infty, x] : x \in \mathbb{R}\}$ 生成的 σ 代数 $\mathscr{B}_{\mathbb{R}}$ 称为 $\mathbb{R}$ 上的 *Borel* 代数.

定义 1.4 (Borel 集)

设 $\mathbb{R}^1$ 为全集, 形为 $[a, b)$ 构成的集类产生的 σ 域称为**一维 Borelσ 域**, 记为 $\mathscr{B}_1$, 其中的元素称为**一维 Borel 集**

若 x, y 为任意实数, 由于:

$$\{x\} = \bigcap_{n=1}^{\infty}\left[x, x+\frac{1}{n}\right)$$

$$(x, y) = [x, y) - \{x\}$$

$$[x, y] = [x, y) + \{y\}$$

$$(x, y] = [x, y) + \{y\} - \{x\}$$

因此 $\mathscr{B}_1$ 包含一切开区间, 闭区间, 单个实数, 可列个实数, 以及他们经可列次逆, 并, 交, 差运算的集合.

定义 1.5 (概率空间)

定义在事件域(非样本空间) 上的集合函数 $P : \mathscr{F} \to \mathbb{R}$ 称为**概率**的条件是:

非负性 $P(A) \geq 0, \forall A \in \mathscr{F}$

规范性 $P(\Omega) = 1$; (如果没有这条就是一般的有限测度)

可列可加性 若 $A_1, \ldots, A_n, \ldots \in \mathscr{F}$ 两两不交, 即 $A_i \cap A_j = \varnothing$, $\forall i \neq j$, 则 $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n)$.

我们称 $(\Omega, \mathscr{F}, P)$ 为一个**概率空间** (probability space)

性质 概率的性质:

- $P(\Omega) = 1$;
- $P(A^{\complement}) = 1 - P(A)$;
- 若 $A \subseteq B$ 则 $P(A) \leq P(B)$;

推论 1.1 (加法公式)

基础形式:

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

一般形式:

$$P\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) = \sum_{i=1,\cdots,n} P\left(A_i\right) - \sum_{\substack{i<j \\ i,j=1,\cdots,n}} P\left(A_i A_j\right) + \sum_{\substack{i<j<k \\ i,j,k=1,\cdots,n}} P\left(A_i A_j A_k\right) - \cdots + (-1)^{n-1} P\left(A_1 A_2 \cdots A_n\right)$$

特别地, 若事件出现个数相同时概率相等, 则可简化为:

$$P\left(A_1 \cup A_2 \cup \cdots \cup A_n\right) = nP_1 - \binom{n}{2}P_2 + \binom{n}{3}P_3 - \cdots + (-1)^{n-1}P_n$$

显然, 可列可加性可以推出有限可加性. 但是一般来讲, 由有限可加性并不能推出可列可加性. 设 $A_i \in \mathscr{F}, i = 1, 2, \cdots$ 且两两互不相容, 若希望由有限可加性推出可列可加性, 则需要下式成立:

$$\lim_{n\to\infty} P(\sum_{i=1}^{n} A_i) = P(\lim_{n\to\infty} \sum_{i=1}^{n} A_i)$$

定义 1.6

对于 $\mathscr{F}$ 上的集合函数 P, 若它对 $\mathscr{F}$ 中任何一个单调不减的集序列 $\{S_n\}$(即 $S_n \in \mathscr{F}, S_n \subseteq S_{n+1}$) 均满足:

$$\lim_{n\to\infty} P(S_n) = P(\lim_{n\to\infty} S_n)$$

则称它是**下连续的**.

故若令 $S_n = \sum_{i=1}^{n} A_n$, 则可列可加性条件等价于有限可加性加下连续.

1.2 古典概型与几何概率

1.2.1 古典概型

古典概型的基本思想:

有限个样本点 所涉及的随机现象只有有限个样本点, 譬如为 n 个, 且这些事件是两两互不相容的;

等可能性 每个样本点发生的可能性相等

定义 1.7

若事件 A 含有 k 个样本点, 则事件 A 的概率为

$$P(A) = \frac{\text{事件}A\text{ 所含样本点的个数}}{\Omega\text{ 中所有样本点的个数}} = \frac{k}{n}$$

笔记 事实上, 古典概型的大部分问题都能形象化地用摸球模型来描述以后我们经常研究摸球模型, 意义即在于此.

1.3 条件概率

定义 1.8 (条件概率)

令 $A, B \in \mathscr{F}$, 且 $P(B) > 0$ 称

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

为**基于于 B 的条件概率** (probability conditional on B), 这仍然是一个概率测度.

定理 1.1 (乘法法则)

令 $A, B \in \mathscr{F}$, 且 $P(B) > 0$, 则

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

泛化后有:

$$P(A_1 \cap A_2 \cap \cdots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)\cdots$$

定理 1.2 (全概率公式)

设 $B_1, B_2, \ldots, B_n$ 为样本空间 Ω 的一个分割, 且互不相容, 即 $\bigcup_{i=1}^n B_i = \Omega, B_iB_j = \varnothing fori \neq j$ 如果 $P(B_i) > 0, i = 1, 2, \ldots, n$, 则对任一事件 A 有

$$P(A) = \sum_{i=1}^{n} P(B_i)P(A|B_i). \tag{1.5}$$

笔记 $P(A|B_i)$ 可视为事件 A 在 B_i 上的平均, $P(B_i)$ 则为其权重.

定理 1.3 (Bayes 定理)

设 $B_1, B_2, \ldots, B_n$ 为样本空间 Ω 的一个分割, 且互不相容, 即 $\bigcup_{i=1}^n B_i = \Omega, B_iB_j = \varnothing fori \neq j$ 如果 $P(A) > 0, P(B_i) > 0, i = 1, 2, \ldots, n$, 则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)} \tag{1.6}$$

定义 1.9 (事件的独立性)

如果 $A, B \in \mathscr{F}$ 满足

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

则称 A 与 B **独立** (independent), 记为 $A \perp\!\!\!\perp B$.

对于事件集 $A_1, A_2, \ldots, A_n$, 若对于其中任意子集 $A_{i_1}, A_{i_2}, \ldots, A_{i_n}$ 有:

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

则称此事件集**相互独立** (mutually independent)

笔记 当 $P(A) > 0$ 时, 我们有 $P(B|A) = P(B) \iff B \perp\!\!\!\perp A$, 由此可得到B 独立于 A 的直观理解

性质 独立性是对称的, 即 $A \perp\!\!\!\perp B \iff B \perp\!\!\!\perp A$. 若两事件独立, 则其补集也独立.

$$\begin{array}{ccc} A & \longleftrightarrow & B \\ & \times & \\ A^{\mathsf{C}} & \leftrightarrow & B^{\mathsf{C}} \end{array}$$

定义 1.10 (事件域的独立性)

若 $\mathscr{G} \subset \mathscr{F}$ 与 $\mathscr{H} \subset \mathscr{F}$ 满足

$$A \perp\!\!\!\perp B, \quad \forall A \in \mathscr{G},\ B \in \mathscr{H},$$

则称 $\mathscr{G}$ 与 $\mathscr{H}$ 独立, 记为 $\mathscr{G} \perp\!\!\!\perp \mathscr{H}$

测度论告诉我们一个重要结果: 如果 $\mathscr{G}$ 对交集运算封闭, 那么成立 $\mathscr{G} \perp\!\!\!\perp \mathscr{H} \implies \sigma(\mathscr{G}) \perp\!\!\!\perp \mathscr{H}$

第 2 章 随机变量

内容提要

- ❑ 离散与连续随机变量
- ❑ 一元与多元
- ❑ cdf, pmf, pdf
- ❑ 条件分布
- ❑ 独立随机变量
- ❑ 随机变量函数的分布
- ❑ 次序随机变量

在概率论中, 主要关心 X 取值于数值集合 $\mathcal{X}$ 中某个子集 B 的可能性, 即希望得到 $\mathbb{P}(\{\omega \in \Omega : X(\omega) \in B\})$. 概率论不关心具体的样本点 $\omega \in \Omega$, 将其记为 $\{X \in B\} = X^{-1}(B)$. 由于 $\mathbb{P}$ 定义在 $\mathscr{F}$ 上, 故需 $X^{-1}(B) \in \mathscr{F}$.

定义 2.1 (可测性)

设所有值得关心的 $B \subset \mathcal{X}$ 组成 $\mathscr{F}_{\mathcal{X}}$, 且 $\forall B \in \mathscr{F}_{\mathcal{X}}$ 都满足 $\{X \in B\} \in \mathscr{F}$, 则称 X 为 $\mathscr{F}/\mathscr{F}_{\mathcal{X}}$ **可测的**. 当 $\mathscr{F}_{\mathcal{X}}$ 不引起混淆时, 简记为关于 $\mathscr{F}$ **可测**, 写作 $X \in \mathscr{F}$. ♣

由于原像保持交、并、补等集合运算, 且 $\mathscr{F}$ 是 σ 代数, 可将 $\mathscr{F}_{\mathcal{X}}$ 扩张为合适的最小的 σ 代数, 即 $\sigma(\mathscr{F}_{\mathcal{X}})$, 因此可测映射的定义不妨只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数 的情况.

定义 2.2 (随机变量)

为了表示因随机性而变动的量, 称可测映射(measurable mapping)

$$X : (\Omega, \mathscr{F}, \mathbb{P}) \to (\mathcal{X}, \mathscr{F}_{\mathcal{X}}), \quad \omega \in \Omega \mapsto X(\omega) \in \mathcal{X}$$

为**随机元** (random element), 也称**随机变量** (random variable). 其中 $\mathscr{F}_{\mathcal{X}}$ ♣

由于只考虑 $\mathscr{F}_{\mathcal{X}}$ 是 σ 代数的情况, 可将随机变量看作将原概率空间映射到新概率空间的方式. 新样本空间只由数值构成, 对应的概率测度等于原象的.

注 使用随机变量 X 时, 有两个可能的含义:

- X 的 (随机) 取值
- X 的分布

2.1 随机变量的分布

定义 2.3

随机元 X 的**分布** (distribution/law) 是 X 诱导的概率测度 $\mathbb{P}\{X \in \bullet\}$, $\bullet \in \mathscr{F}_{\mathcal{X}}$. ♣

	离散	连续
一元随机变量	pmf	pdf
	cdf	cdf
	mgf/chf	mgf/chf
多元随机变量	joint pmf	joint pmf
	joint cdf	joint cdf
	joint mgf/chf	joint mgf/chf

定义 2.4 (离散与连续随机变量)

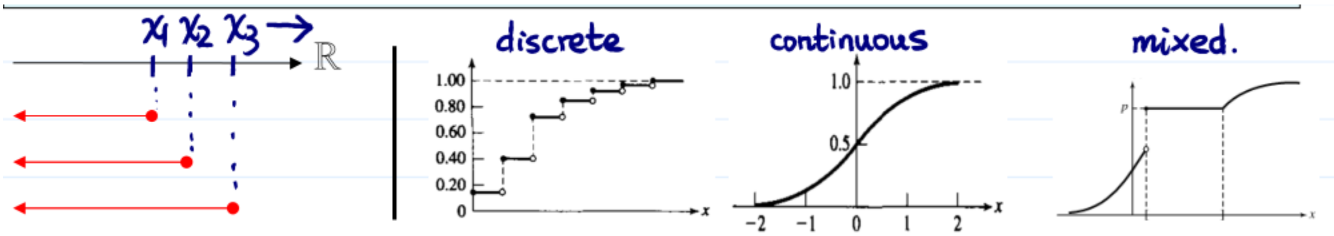
假如一个随机变量仅取有限个或可列个值, 则称其为**离散随机变量** (discrete random varible). 假如一个随机变量的可能取值充满数轴上的一个区间 (a, b), 则称其为**连续随机变量** (continuous random varible), 其中 a 可以是 $-\infty$, b 可以是 $+\infty$.

注 可测性定义中的 B 常取作 Borel 点集

定义 2.5 (累积分布函数)

对于随机变量 X, 定义其**累积分布函数** (cumulative random varible) 为:

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



定义 2.6 (概率质量函数)

当 $\mathcal{X}$ 是 (至多可数的) 离散点集, 设 $\mathscr{F}_{\mathcal{X}}$ 由 $\mathcal{X}$ 的所有子集组成, 此时 X 的分布由**概率质量函数** (probability mass function, p.m.f.)

$$p_X(x) = \mathbb{P}\{X = x\} = \mathbb{P}(\{\omega \in \Omega : X(\omega) = x\}), \quad x \in \mathcal{X}$$

唯一刻画.

注 这是概率的哪个性质保证的?

定义 2.7

当 $\mathcal{X} = \mathbb{R}^n$, 考虑 $\mathscr{F}_{\mathcal{X}}$ 为 $\{\prod_{i=1}^n (-\infty, x_i] : x_1, \dots, x_n \in \mathbb{R}\}$ 生成的 Borel 代数 (最小的 σ 代数), 此时 $X = (X_1, \dots, X_n)^\top$ 的分布由 (累积) **分布函数** (cumulative distribution function, c.d.f.)

$$F_X(x_1, \dots, x_n) = \mathbb{P}\{X_1 \le x_1, \dots, X_n \le x_n\}, \quad x_1, \dots, x_n \in \mathbb{R}.$$

唯一刻画. 若 $F_X : \mathbb{R}^n \to [0, 1]$ 可微 (或者更一般地, 绝对连续), 称

$$f_X := \frac{\partial^n F_X}{\partial x_1 \cdots \partial x_n}$$

为 X 的**概率密度函数** (probability density function, p.d.f.), 此时称 X 为**连续型随机向量**.

2.2 随机变量的条件分布

记 $\sigma(X) := X^{-1}(\mathscr{F}_{\mathcal{X}}) := \{X^{-1}(B) : B \in \mathscr{F}_{\mathcal{X}}\}$. 根据前面的定义, X 可测当且仅当 $\sigma(X) \subset \mathscr{F}$. 若随机元 $X_1 : (\Omega, \mathscr{F}) \to (\mathcal{X}_1, \mathscr{F}_{\mathcal{X}_1})$ 与随机元 $X_2 : (\Omega, \mathscr{F}) \to (\mathcal{X}_2, \mathscr{F}_{\mathcal{X}_2})$ 满足 $\sigma(X_1) \perp\!\!\!\perp \sigma(X_2)$, 则称 X_1 与 X_2 **独立**, 记为 $X_1 \perp\!\!\!\perp X_2$.

注: 对任意可测映射 g_1, g_2 成立 $X_1 \perp\!\!\!\perp X_2 \implies g_1(X_1) \perp\!\!\!\perp g_2(X_2)$. 事实上, $\sigma(g(X)) \subset \sigma(X)$.

当 X_1 和 X_2 都是实值随机向量时, $X_1 \perp\!\!\!\perp X_2$ 当且仅当 (思考: $\sigma(X^{-1}(\bullet)) = X^{-1}(\sigma(\bullet))$, $\bullet \subset \mathscr{F}_{\mathcal{X}}$?)

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1}(x_1) F_{X_2}(x_2), \ \forall x_1, x_2.$$

对于连续型随机向量, 刻画独立性只需要

$$f_{X_1,X_2}(x_1,x_2) = f_{X_1}(x_1)f_{X_2}(x_2),\ \forall x_1,x_2.$$

设随机向量 X 和 Y 有联合 (joint) 概率密度函数 $f_{X,Y}$, 可以证明 X 有概率密度函数

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

称为 (X,Y) 中 X 的边际 (margin) 概率密度函数. Y **条件于** $X=x$ **的概率密度函数** $f_{Y|X}(\cdot|x)$ 满足

$$f_{X,Y}(x,\cdot) = f_{Y|X}(\cdot|x)f_X(x).$$

2.3 随机变量的函数

第 3 章 随机变量的数值特征

3.1 期望与方差

3.2 矩母函数与特征函数

3.3 熵与信息

第 4 章 常见分布

4.1 离散分布

4.2 连续分布

4.3 正态分布及其导出分布

第 5 章 概率极限

5.1 收敛

5.2 大数定理

5.3 中央极限定理

附录 A 基本数学工具

A.1 排列与组合

全部组合分析公式的推导基于下列两条原理：

乘法原理 若进行 A_1 过程有 n_l 种方法，进行 A_2 过程有 n_2 种方法，则进行 A_1 过程后再接着进行 A_2 程共有 $n_1 \cdot n_2$ 种方法

加法原理 若进行 A_1 过程有 n_l 种方法，进行 A_2 过程有 n_2 种方法，假定 A_1 过程与 A_2 过程是并行的，则进行过程 A_1 或过程 A_2 的方法共有 $n_1 + n_2$ 种

排列与组合的定义及其计算公式如下.

1. 排列: 从 n 个不同元素中任取 $r(r \le n)$ 个元素排成一列 (考虑元素先后出现次序), 称此为一个排列, 此种排列的总数记为 P_n^r, 按乘法原理, 取出的第一个元素有 n 种取法, 取出的第二个元素有 $n-1$ 种取法…取出的第 r 个元素有 $n-r+1$ 种取法, 所以有
$$P_n^r = n \times (n-1) \times \cdots \times (n-r+1) = \frac{n!}{(n-r!)}. \tag{A.1}$$
若 $r = n$, 则称为全排列, 记为 $_n$. 显然, 全排列 $P_n = n!$.
2. 重复排列: 从 n 个不同元素中每次取出一个, 放回后再取下一个, 如此连续取 r 次所得的排列称为重复排列, 此种重复排列数共有 n^r 个. 注意这里的 r 允许大于 n.
3. 组合: 从 n 个不同元素中任取 $r(r \le n)$ 个元素并成一组 (不考虑元素间的先后次序), 称此为一个组合, 此种组合的总数记为 $\binom{n}{r}$ 或 C_n^r. 按乘法原理此种组合的总数为
$$\binom{n}{r} = \frac{P_n^r}{r!} = \frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{r!(n-r)!}. \tag{A.2}$$
在此规定 $0! = 1$ 与 $\binom{n}{0} = 1$.
4. 重复组合: 从 n 个不同元素中每次取出一个, 放回后再取下一个, 如此连续取 r 次所得的组合称为重复组合, 此种重复组合总数为 $\binom{n+r-1}{r}$. 注意这里的 r 也允许大于 n.

上述四种排列组合及其总数计算公式，在确定概率的古典方法中经常使用，但在使用中要注意识别有序与无序、重复与不重复。