概率论笔记

作者：肖程哲

时间：July 3, 2022

苟日新，日日新，又日新

目录

第 1 章 概率基础 **1**

1.1 概率空间 1

1.2 古典概型与几何概率 3

1.3 条件概率 3

第 2 章 随机变量 **5**

2.1 随机变量的分布 5

2.2 随机变量的条件分布 5

2.3 随机变量的函数 5

第 3 章 随机变量的数值特征 **6**

3.1 期望与方差 6

3.2 矩母函数与特征函数 6

3.3 熵与信息 6

第 4 章 常见分布 **7**

4.1 离散分布 7

4.2 连续分布 7

4.3 正态分布及其导出分布 7

第 5 章 概率极限 **8**

5.1 收敛 8

5.2 大数定理 8

5.3 中央极限定理 8

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.

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 $A_1,\dots,A_n,\dots\in\mathscr{F}$ 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:

- $P(\Omega)=1$;
- $P(A^{\mathsf{C}})=1-P(A)$;
- 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:

$$P\left(A_1\cup A_2\cup\cdots\cup A_n\right)=\sum_{i=1,\cdots,n}P\left(A_i\right)-\sum_{\substack{i<j\\ i,j=1,\cdots,n}}P\left(A_iA_j\right)$$
$$+\sum_{\substack{i<j<k\\ i,j,k=1,\cdots,n}}P\left(A_iA_jA_k\right)-\cdots+(-1)^{n-1}P\left(A_1A_2\cdots A_n\right)$$

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

1.3 条件概率

定义 1.7 (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 $\bigcup_{i=1}^{n} B_i = \Omega, B_iB_j = \varnothing for i \neq j$ 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 $\bigcup_{i=1}^{n} B_i = \Omega, B_iB_j = \varnothing for i \neq j$ 如果 $P(A) >$

0, $P(B_i) > 0, i = 1, 2, \dots, n$, 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 1.8 (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 $A_1, A_2, \dots, A_n$, 若对于其中任意子集 $A_{i_1}, A_{i_2}, \dots, A_{i_n}$ 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$\underline{B\text{ 独立于 }A}$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.

$$\begin{tikzcd}
A \arrow[r, leftrightarrow] \arrow[dr, leftrightarrow] & B \arrow[dl, leftrightarrow] \\
A^{\complement} \arrow[r, leftrightarrow] & B^{\complement}
\end{tikzcd}$$

定义 1.9 (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

2.1 随机变量的分布

2.2 随机变量的条件分布

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