概率论与数理统计笔记

Guotao He

2025-01-29

"All knowledge degenerates into probability."

— David Hume

目录

Chapter 1 概率论的基本概念

1.1 基本概念

样本空间与样本点 对于一个随机的试验 E，其实验的所有可能的结构构成一个集合，此集合称为随机实验 E 的**样本空间**，记为 S，样本空间的每一个元素被称为一个**样本点**。

随机事件 随机试验 E 的样本空间 S 的子集 A 被称为 E 的一个**随机事件**。特别的，由一个样本点组成的单点集，称为 **基本事件**，样本空间 S 是自身的子集，且每次实验中必然发生，称 S 为 E 的**必然事件**，空集 $\varnothing$ 不包含任何样本点，也为样本空间 S 的子集，其在每次实验中必然不发生，$\varnothing$ 称为 E 的**不可能事件**。

事件空间 在随机试验 E 中的所有随机事件构成一个集合 $\mathscr{F}$（集合的每一个元素也是集合），此集合被称为**事件空间**

事件本质上是样本空间的一个子集，因此事件之间的运算自然按集合论中集合之间的运算处理。具体而言，有：

1. $A \subseteq B$ 表示事件 B 包含事件 A，若 A 发生则 B 一定也发生
2. $A = B$ 表示 $A \subseteq B$ 且 $B \subseteq A$
3. $A + B = A \cup B = \{x | x \in A \text{ or } x \in B\}$
4. $AB = A \cap B = \{x | x \in A \text{ and } x \in B\}$
5. $A - B = \{x | x \in A \text{ and } x \notin B\}$
6. $A \cap B = \varnothing$，称 A 与 B **互斥**
7. $A \cup B = S, A \cap B = \varnothing$，称 A 与 B 互为**逆事件**，互为**对立事件**
8. 记 A 的对立事件为 $\bar{A} = S - A$

摩根率（对偶率） $\overline{A+B} = \bar{A}\bar{B}, \overline{AB} = \bar{A} + \bar{B}$

Definition 1.1.1 (概率)：设随机试验 E 的样本空间为 S，$\mathscr{F}$ 为 S 的某些子集组成的一个事件空间，如果对任一事件 $A \in \mathscr{F}$，定义在 F 上的一个实值函数 $P(A)$ 满足如下性质：

1. **非负性**：$\forall A \in \mathscr{F}, P(A) \geq 0$
2. **规范性（归一性）**：$P(S) = 1$
3. **可列可加性**：$\forall A_1, A_2, ... \in \mathscr{F}$, 且满足 $A_i A_j = \varnothing, i \neq j, i, j \in \mathbb{R}$，有：

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

则我们称 P 是 $\mathscr{F}$ 上的**概率测度**，称 $P(A)$ 是 A 的**概率**。

Remark: 这里的“可列可加”实际上是由测度论带来的，这里的“可列”要求事件序列 A_i 是无穷长的，且是可数无穷。

当只有有限个事件序列时（比如只有 n 个）我们只需要简单地将 $A_i(i > n)$ 全部定义为空集即可,换句话说，对于有限个互不相容的事件序列 $A_1, ..., A_n$，有：

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

实际上，上述性质被称为“有限可加性”，而“可列可加性”是其到无穷情况的推广，“可列可加性”可以推出“有限可加性”，但“有限可加性”不能推导得到“可列可加性”。因此，我们不可用“有限可加性”去替代“可列可加性”，即使我们大多数情况下只处理有限的情况。

换句话说，所谓概率，就是一个从事件空间 $\mathscr{F}$ 到 $\mathbb{R}$ 上的一个映射，且满足上面三个条件。实际上，概率的公理化条件并不直接告诉我们在实际问题中如何计算 $P(A)$，其只告诉了我们什么是 $P(A)$（实际上很多数学上的公理化定义都如此），$P(A)$ 的具体计算要根据问题的条件和背景得到。

概率空间 所谓概率空间是一个三元组，包含样本空间 S，事件集合 $\mathscr{F}$ 和概率测度 P, 记为 $(S, \mathscr{F}, P)$

Definition 1.1.2 (条件概率)：设 A, B 是两个事件，且 $P(A) > 0$，称：

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

为事件 A 发生的条件下事件 B 发生的**条件概率**。

独立性 对于试验 E 中的两个事件 A, B，若事件 A 发生的概率对事件 B 发生的概率无影响，即 $P(AB) = P(A)P(B)$，则我们称事件 A 与 B 互相**独立**。更一般地，若 $P(A_1A_2...A_n) = P(A_1)P(A_2)...P(A_n)$ 则我们称 $A_1, A_2, ..., A_n$ 相互独立。

Mark: 这里需要区分 A, B 互相独立和 A, B 互相对立的区别。实际上，事件 A, B 的独立性和对立性不可能同时成立，若已知 A, B 对立，且 A 不成立，则我们可以马上得到 B 成立，显然这不符合 A, B 互相独立的定义。

另外，A, B 对立很容易在 Venn 图中表示，但 A, B 独立不然，究其原因是因为 Venn 图并无体现数量关系，故无法表示 A, B 独立这一数量关系。

1.2 概率公式的基本计算

Theorem 1.2.1 (加法公式（Jordan 公式）): 对于概率空间 $(S,\mathscr{F},P)$ 若 $A_1,...,A_n\in\mathscr{F}$，则：

$$P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k-1}\sum_{1\le i_1<...<i_n\le n}P\left(A_{i_1}...A_{i_k}\right)$$

Theorem 1.2.2 (减法公式): 对于概率空间 $(S,\mathscr{F},P)$，若 $A,B\in\mathscr{F}$，则：

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

Theorem 1.2.3 (乘法公式): 对于概率空间 $(S,\mathscr{F},P)$，若 $A_1,...,A_n\in\mathscr{F}$，则：

$$P(A_1...A_n)=\prod_{i=1}^{n}P(A_i|A_1...A_{n-1})=\sum_{k=1}^{n}\sum_{1\le i_1<i_n\le n}(-1)^{k-1}P\left(A_{i_1}\cup...\cup A_{i_k}\right)$$

Theorem 1.2.4 (全概率公式): 对于概率空间 $(S,\mathscr{F},P)$，若 $B,A_1,...,A_n\in\mathscr{F}$，其中 A_i 是 S 的分割，则：

$$P(B)=\sum_{j}P\left(A_j\right)P\left(B|A_j\right)$$

Theorem 1.2.5 (Bayes 公式): 对于概率空间 $(S,\mathscr{F},P)$，若 $B,A_1,...,A_n\in\mathscr{F}$，其中 A_i 是 S 的分割，则：

$$P(A_i|B)=\frac{P(A_i)P(B|A_i)}{\sum_j P\left(A_j\right)P\left(B|A_j\right)}$$

1.3 古典概率模型

古典概型 若试验 E 满足两个条件：1. 样本空间 S 是有限的集合。2. S 中的每个样本点发生的可能性相同。则我们称这种试验为**等可能概型**，又称**古典概型**。

在古典概型中，取样本空间 $S=\{e_1,...,e_n\}$，其中 $e_1,...,e_n$ 代表该随机试验的 N 个结果，事件域 $\mathscr{F}$ 取为 2^S（即 S 的所有子集都是事件），且每个基本事件发生的可能性相同，即

$$P(\{e_1\}) = P(\{e_2\}) = ... = P(\{e_n\})$$

若事件 A 包含 K 个基本事件，即 $A = \left\{e_{i_1}\right\} \cup ... \cup \left\{e_{i_n}\right\}$，则给出概率测度 P 如下：

$$P(A) = \frac{K}{N} = \sum_{e \in A} \frac{1}{N}$$

由于古典概型的概率与计数直接相关，因此下文中简要介绍一些基本的组合计数方法：

乘法原理 设某个试验共包含 r 个依次执行的阶段，其中：1. 第一个阶段总共有 n_1 个可能的结果。2. 在完全前面的 $i-1$ 个阶段并得到一个相应的结果后，第 i 个阶段将总共有 n_i 个可能的结果。则该试验一共有 $n_1 \cdot n_2 \cdot ... \cdot n_r$ 个可能的结果。

加法原理 设完成某个试验有 r 种不同的途径，且只能在这 r 种途径中选一种来完成该试验。若采取第 i 种途径时总共有 n_i 种可能的结果，则该试验一共有 $n_1 + n_2 + ... + n_r$ 种可能的结果。

排列 假设有 n 个不同的个体，从中选出 n 个并将其排成一个序列，则总共有：

$$n(n-1)(n-2)...(n-m+1) = \frac{n!}{(n-m)!}$$

个可能的组合方式。

组合 假设一个集合包含 n 个不同的元素，我们需要从中选出 m 个元素构成一个子集，则可能得到的子集总共有：

$$\frac{n(n-1)...(n-m+1)}{m!} = \frac{n!}{m!(n-m)!}$$

个，上式给出的整数常常又被称为**二项式系数**，记作 $\binom{n}{m}$

多项式系数 给定一组自然数 $n_1, n_2, ..., n_r$，以及 $n = n_1 + ... + n_r$，定义多项式系数为：

$$\binom{n}{n_1, n_2, ..., n_r} = \frac{n!}{n_1!n_2!...n_r!}$$

Chapter 2 随机变量及其分布

2.1 随机变量

"随机变量"可以看成是将"底层"的概率空间 $(S, \mathscr{F}, P)$ "包装"成分布函数，分布列，概率密度等方便使用微积分等分析工具的"接口"，从而为我们对概率进行建模提供了莫大的便利。

Definition 2.1.1 (随机变量): 设 $(S, \mathscr{F}, P)$ 为一概率空间，我们将定义在 S 上的实值函数 $X : S \to \mathbb{R}$ 称作一个**随机变量**，若对于任意 $\mathbb{R}$ 中的区间 I，均有：

$$\{e \in S | X(e) \in I\} \in \mathscr{F}$$

上述定义中，要求 $\{e \in S | X(e) \in I\} \in \mathscr{F}$，其保证了 $\{e \in S | X(e) \in I\}$ 的概率都是有定义的，也就是对任意区间 I，我们都可以谈论"随机变量的取值落在 I 中"这一事件的概率。

Remark: 上述条件实际上是从可测性而来，即对于实数轴上的任意的"合理"的子集（*Borel* 集），其原像都必须是一个可测事件（即属于事件域）同时，上述要求的一个常见的等价描述如下："对于任意实数 x，集合 $\{e | X(e) \leq x\}$ 有确定的概率"。

另外，不满足上述条件的情况实际应用中极少出现，相关例子基本属于高等概率论的内容，其基本都是数学上卡 *BUG* 而来，一般可忽略不理。

对于给定的随机变量 X 以及关于实数 x 的命题 $\varphi(x)$，我们常用记号 $\{\varphi(X)\}$ 表示集合 $\{e \in S | \varphi(X(x))$ 成立$\}$，例如：

$$\{X > 2\} = \{e \in S \mid X(x) > 2\}$$

随机变量之间可以进行运算，具体而言，给定 n 个随机变量 $X_1, X_2, ..., X_n$ 以及函数 $f : \mathbb{R}^n \to \mathbb{R}$，我们用 $f(X_1, ..., X_n)$ 表示将 $e \in S$ 映射为 $f(X_1(e), X_2(e), ..., X_n(e))$ 的函数。

2.2 离散型随机变量及其分布律

Definition 2.2.1 (离散型随机变量): 设 X 为一随机变量，若 X 只有可数多个可能的取值，则称 X 为**离散型随机变量**。

另外，对于离散型随机变量，我们定义分布列 $p_X : \mathbb{R} \to [0, 1]$ 为：

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

显然，分布列满足如下性质：

Theorem 2.2.1：设 X 为一离散型随机变量，p_X 为其分布列，则：
1. 对任意 $x \in \mathbb{R}$，均有 $p_X(x) \geq 0$
2. $\{x \in \mathbb{R} | p_X(x) > 0\}$ 为可数集
3. $\sum_x p_{X(x)} = 1$

此定理的逆定理同样成立。

2.3 常见离散型随机变量

2.3.1 Bernoulli 分布

2.3.2 二项分布

2.3.3 超几何分布

2.3.4 Poisson 分布

2.3.5 几何分布

2.3.6 帕斯卡分布

2.3.7 多项分布

2.4 连续型随机变量及其分布律

2.5 常见连续型随机变量

2.5.1 均匀分布

2.5.2 正态分布

2.5.3 指数分布

2.5.4 Erlang 分布

2.5.5 Γ 分布

2.5.6 x^2 分布

2.5.7 多元均匀分布

2.5.8 多元正态分布

Chapter 3 多维随机变量

Chapter 4 杂七杂八

4.1 概率的连续性

4.2 测度论简介

Chapter 5 Typst 测试章节

5.1 Test Section

§5.1

5.2 Test Test Test 02