概率论与数理统计笔记

作者：刘承杰
组织：南京大学软件学院
时间：2021 年 7 月 10 日

目录

第 1 章 概率论的基本概念

1.1 随机试验

定义 1.1 (随机试验)

在概率论中，我们将具有以下三个特点的试验称为随机试验：

1. 可以在相同的条件下重复的进行；
2. 每次试验的可能结果不止一个，并且能事先明确试验的所有可能结果；
3. 进行一次试验之前不能确定哪个结果会出现

1.2 样本空间、随机事件

定义 1.2 (样本空间)

随机事件 E 的所有可能结果组成的集合，记为 S

定义 1.3 (样本点)

样本空间的元素，即 E 的每个结果

定义 1.4 (随机事件)

试验 E 的样本空间 S 的子集为 E 的随机事件，简称**事件**. 在每次试验中，当且仅当这一子集中的一个样本点出现时，称这一**事件发生**

定义 1.5 (基本事件)

由一个样本点组成的单点集，称为基本事件

定义 1.6 (必然事件和不可能事件)

样本空间 S 包含所有的样本点，它是 S 自身的子集，在每次试验中它总是发生的，S 称为**必然事件**. 空集 $\emptyset$ 不包含任何样本点，它也作为样本空间的子集，它在每次试验中都不可能发生，$\emptyset$ 称为**不可能事件**.

定义 1.7 (事件间的关系与事件的运算)

设试验 E 的样本空间为 S, 而 $A, B, A_k(k = 1, 2, \cdots)$ 是 S 的子集.

1. 若 $A \subset B$, 则称事件 B **包含**事件 A, 这指的是事件 A 发生，则事件 B 必然发生。若 $A \subset B$ 且 $B \subset A$, 即 $A = B$, 则称事件 A 与事件 B **相等**。
2. 事件 $A \cup B = \{x|x \in A或x \in B\}$ 称为事件 A 与事件 B 的**和事件**。当且仅当 A, B 中至少有一个发生时，事件 $A \cup B$ 发生。

   类似地，称 $\mathop{}\limits_{k=1}^{n} A_k$ 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的和事件；称 $\mathop{}\limits_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \cdots$ 的和事件
3. 事件 $A \cap B = \{x|x \in A且x \in B\}$ 称为事件 A 与事件 B 的**积事件**。当且仅当 A, B 同时发生时，事件 $A \cap B$ 发生。

   类似地，称 $\mathop{}\limits_{k=1}^{n} A_k$ 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的积事件；称 $\mathop{}\limits_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \cdots$ 的积事

件

4. 事件 $A - B = \{x|x \in A且x \notin B\}$ 称为事件 A 与事件 B 的**差事件**。当且仅当 A 发生、B 不发生时事件 $A - B$ 发生。
5. 若 $A \cap B = \emptyset$, 则称事件 A 与 B 是互不相容的，或**互斥的**。这指的是事件 A 与 B 不能同时发生。基本事件是两两互不相容的。
6. 若 $A \cup B = S$ 且 $A \cap B = \emptyset$, 则称事件 A 与事件 B 互为**逆事件**，又称事件 A 与事件 B 互为**对立事件**。这指的是对每次试验而言，事件 A、B 中必有一个发生，且仅有一个发生。A 的对立事件记为 $\overline{A}$

定理 1.1 (集合运算定律)

设 A、B、C 为事件，则有：
交换律：

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

结合律：

$$A \cup (B \cup C) = (A \cup B) \cup C$$

$$A \cap (B \cap C) = (A \cap B) \cap C$$

分配律：

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cap (A \cap C)$$

德摩根律：

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

1.3 频率与概率

定义 1.8 (频率)

在相同的条件下，进行了 n 次试验，在这 n 次试验中，事件 A 发生的次数 n_A 称为事件 A 发生的频数。比值 $\frac{n_A}{n}$ 称为事件 A 发生的频率，并记为 $f_n(A)$.

定义 1.9 (概率)

设 E 是随机试验，S 是它的样本空间，对于 E 的每一个事件 A 赋予一个实数，记为 $P(A)$，称为事件 A 的概率，如果集合函数 $P(\cdot)$ 满足下列条件：

1. **非负性**：对于每一个事件 A，有 $P(A) \geq 0$;
2. **规范性**：对于必然事件 S, 有 $P(S) = 1$;
3. **可列可加性**：设 $A_1, A_2, \cdots$ 是两两不相容的事件，即对于 $A_iA_j = \emptyset, i \neq j, i, j = 1, 2, \cdots$，有

$$P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$$

当 $n \to \infty$ 时频率 $f_n(A)$ 在一定意义下接近于概率 $P(A)$

1.4 等可能概型（古典概型）

定义 1.10 (等可能概型)

具有以下两个特点的试验称为等可能概型：

1. 试验的样本空间只包含有限个元素；
2. 试验中每个基本事件发生的可能性相同。

♣

定理 1.2 (等可能概型中事件 A 的概率计算公式)

若事件 A 包含 K 个基本事件，即 $A = e_{i_1} \cup e_{i_2} \cup \cdots \cup e_{i_k}$，这里 $i_1, i_2, \cdots, i_k$ 是 $1, 2 \cdots, n$ 中某 k 个不同的数，则有：

$$P(A) = \sum_{j=1}^{k} P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A\text{包含的基本事件数}}{S\text{中基本事件的总数}}$$

♡

定理 1.3 (超几何分布的概率公式)

设共有 N 件产品，其中有 D 件次品，从中取 n 件，其中恰好有 $k(k \leq D)$ 件次品的概率为：

$$p = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

♡

1.5 条件概率

定义 1.11 (条件概率)

设 A, B 是两个事件，且 $P(A) > 0$，称

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

为在事件 A 发生的条件下事件 B 发生的条件概率。

♣

定理 1.4 (乘法定理)

设 $P(A) > 0$, 则有

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

上式可以推广到多个事件的积事件的情况。例如，设 A, B, C 为事件，且 $P(AB) > 0$, 则有

$$P(ABC) = P(C|AB)P(B|A)P(A)$$

一般地，设 $A_1, A_2, \cdots, A_n$ 为 n 个事件，$n \geq 2$, 且 $P(A_1A_2\cdots A_{n-1}) > 0$, 则有

$$P(A_1A_2\cdots A_n) = P(A_n|A_1A_2\cdots A_{n-1})P(A_{n-1}|A_1A_2\cdots A_{n-2})\cdots P(A_2|A_1)P(A_1)$$

♡

定义 1.12 (样本空间的划分)

设 S 为试验 E 的样本空间，$B_1, B_2, \cdots, B_n$ 为 E 的一组事件，若

1. $B_iB_j = \emptyset, i \neq j, i, j = 1, 2, \cdots, n$
2. $B_1 \cup B_2 \cup \cdots \cup B_n = S$

则称 $B_1, B_2, \cdots, B_n$ 为样本空间的一个划分。

若 $B_1, B_2, \cdots, B_n$ 为样本空间的一个划分，那么对于每次试验，事件 $B_1, B_2, \cdots, B_n$ 中必有且只有一个发

生。

定理 1.5 (全概率公式)

设试验 E 的样本空间为 S，A 为 E 的事件，$B_1, B_2, \cdots, B_n$ 为 S 的一个划分，且 $P(B_i) > 0, i = 1, 2, \cdots, n$，则

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \cdots + P(A|B_n)P(B_n)$$

定理 1.6 (贝叶斯 (Bayes) 公式)

设试验 E 的样本空间为 S,A 为 E 的事件，$B_1, B_2, \cdots, B_n$ 为 S 的一个划分，且 $P(A) > 0$, $P(B_i) > 0, i = 1, 2, \cdots, n$, 则

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum\limits_{j=1}^{n} P(A|B_j)P(B_j)}, i = 1, 2, \cdots, n$$

1.6 独立性

定义 1.13 (独立性)

设 A, B 是两事件，如果满足等式

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

则称事件 A, B 相互独立

同理，对于 A, B, C 三个事件，如果满足等式

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

$$P(BC) = P(B)P(C)$$

$$P(AC) = P(A)P(C)$$

$$P(ABC) = P(A)P(B)P(C)$$

则称事件 A, B, C 相互独立。

一般地，设 $A_1, A_2, \cdots, A_n$ 是 $n(n \geq 2)$ 个事件，如果对于其中任意 2 个，任意 3 个，$\cdots$，任意 n 个事件的积事件的概率都等于各事件概率的积，则称事件 $A_1, A_2, \cdots, A_n$ 相互独立。

第 2 章 随机变量及其分布

2.1 随机变量

定义 2.1 (随机变量)

设随机试验的样本空间为 $S = \{e\}, X = X(e)$ 是定义在样本空间 S 上的实值单值函数，称 $X = X(e)$ 为随机变量

2.2 离散型随机变量及其分布律

定义 2.2 (离散型随机变量)

全部可能取到的值是有限个或可列无限多个的随机变量，称为离散型随机变量。

定义 2.3 ((0 − 1) 分布)

设随机变量 X 只能取 0, 1 两个值，它的分布律是

$$P\{X = k\} = p^k(1-p)^{1-k}, k = 0, 1 \quad (0 < p < 1)$$

定义 2.4 (伯努利试验)

设试验 E 只有两个可能结果：A 和 $\overline{A}$, 则称 E 为伯努利 (Bernoulli) 试验. 设 $P(A) = p(0 < p < 1)$, 此时 $P(\overline{A}) = 1 - p$, 将 E 独立重复进行 n 次，则称这一串重复的独立试验为 n 重伯努利试验。

定义 2.5 (二项分布)

以 X 表示 n 重伯努利试验中事件 A 发生的次数，事件 A 在指定的 $k(0 \le k \le n)$ 次试验中发生的概率为

$$P\{X = k\} = \binom{n}{k} p^k(1-p)^{n-k}, k = 0, 1, 2, \cdots, n$$

随机变量 X 服从参数为 n, p 的二项分布，并记为 $X \sim b(n, p)$.

定义 2.6 (泊松分布)

设随机变量 X 所有可能取的值为 $0, 1, 2, \cdots$, 而取各个值的概率为

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, 2, \cdots,$$

其中 $\lambda > 0$ 是常数。则称 X 服从参数为 λ 的泊松分布，记为 $X \sim \pi(\lambda)$

定理 2.1 (泊松定理)

设 $\lambda > 0$ 是一个常数，n 是任意正整数，设 $np_n = \lambda$, 则对于任一固定的非负整数 k, 有

$$\lim_{n\to\infty} \binom{n}{k} {p_n}^k(1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

2.3 随机变量的分布函数

定义 2.7 (随机变量的分布函数)

设 X 是一个随机变量，x 是任意实数，函数

$$F(X) = P\{X \le x\}, -\infty < x < \infty$$

称为 X 的分布函数

定理 2.2 (分布函数性质)

1. $F(x)$ 是一个不减函数
2. $0 \le x \le 1$, 且 $F(-\infty) = \lim_{x\to-\infty} F(x) = 0, F(\infty) = \lim_{x\to\infty} F(x) = 1$
3. $F(x+0) = F(x)$, 即 $F(x)$ 是右连续的

反之，具备上述三条性质的函数必是某个随机变量的分布函数。

2.4 连续型随机变量及其概率密度

定义 2.8 (连续型随机变量、概率密度)

如果对于随机变量 X 的分布函数 $F(x)$，存在非负可积函数 $f(x)$，使对于任意实数 x 有

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

则称 X 为**连续型随机变量**，$f(x)$ 称为 X 的概率密度函数，简称**概率密度**。

定义 2.9 (均匀分布)

若连续型随机变量 X 具有概率密度

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

则称 X 在区间 (a, b) 上服从均匀分布，记为 $X \sim U(a, b)$

X 的分布函数为

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & x \ge b \end{cases}$$

定义 2.10 (指数分布)

若连续型随机变量 X 的概率密度为

$$f(x) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

其中 $\theta > 0$ 为常数，则称 X 服从参数为 θ 的指数分布

X 的分布函数为

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\theta}}, & x > 0 \\ 0, & \text{其他} \end{cases}$$

定义 2.11 (正态分布)

若连续型随机变量 X 的概率密度为

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

其中 $\mu,\sigma(\sigma>0)$ 为常数，则称 X 服从参数为 μ,σ 的正态分布或高斯分布，记为 $X\sim N(\mu,\sigma^2)$

X 的分布函数为

$$F(x)=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}\,dt$$

特别地，当 $\mu=0,\sigma=1$ 时，称随机变量 X 服从**标准正态分布**，其概率密度和分布函数分别用 $\varphi(x),\Phi(x)$ 表示，有

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

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

对于一般的正态分布，只需通过一个线性变换就能化为标准正态分布：

$$\text{若}X\sim N(\mu,\sigma^2),\text{则}Z=\frac{X-\mu}{\sigma}\sim N(0,1)$$

♣

2.5 随机变量的函数的分布

定理 2.3

设随机变量 X 具有概率密度 $f_X(x),-\infty<x<\infty$, 又设函数 $g(x)$ 处处可导且恒有 $g'(x)>0$（或恒有 $g'(x)<0$，则 $Y=g(X)$ 是连续型随机变量，其概率密度为

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

其中 $\alpha=\min\{g(-\infty),g(\infty)\},\beta=\max\{g(-\infty),g(\infty)\}$,$h(y)$ 是 $g(x)$ 的反函数

♡

第 3 章 多维随机变量及其分布

3.1 二维随机变量

定义 3.1 (联合分布函数)

设 (X,Y) 是二维随机变量，对于任意实数 x,y，二元函数

$$F(x,y) = P\{(X \le x) \cap (Y \le y)\} \equiv P\{X \le x, Y \le y\}$$

称为二维随机变量 (X,Y) 的**分布函数**，或称为随机变量 X 和 Y 的**联合分布函数**。

如果将二维随机变量 (X,Y) 看成是平面上随机点的坐标，那么分布函数 $F(x,y)$ 在 (x,y) 处的函数值就是随机点 (X,Y) 落在以点 (x,y) 为顶点而位于该点左下方的无穷矩形区域内的概率。

所以，随机点 (X,Y) 落在矩形区域 $\{(x,y)|x_1 < x \le x_2, y_1 < y \le y_2\}$ 的概率为

$$P\{x_1 < x \le x_2, y_1 < y \le y_2\} = F(x_2,y_2) - F(x_2,y_1) + F(x_1,y_1) - F(x_1,y_2)$$

♣

定义 3.2 (联合概率密度)

对于二维随机变量 (X,Y) 的分布函数 $F(x,y)$，如果存在非负可积函数 $f(x,y)$ 使对于任意 x,y 有

$$F(x,y) = \int_{-\infty}^{y}\int_{-\infty}^{x} f(u,v)\,du dv$$

则称 (X,Y) 是**连续型的二维随机变量**，函数 $f(x,y)$ 称为二维随机变量 (X,Y) 的**概率密度**，或称为随机变量 X 和 Y 的**联合概率密度**。

♣

3.2 边缘分布

定义 3.3

♣