概率论与数理统计总结

刘阳

2019 年 6 月 6 日

目录

1 第一章 概率论的基本概念

1.1 随机试验

随机试验:
1. 可以在相同条件下重复地进行;
2. 每次试验的可能结果不止一个，并且能事先明确实验的所有可能结果;
3. 进行一次实验之前不能确定哪一个结果会出现.

1.2 样本空间、随机事件

1.2.1 样本空间

随机试验 E 的所有可能结果组成的集合称为 **随机试验**.
样本空间的元素，即 E 的每个结果称为 **样本点**.

1.2.2 随机事件

试验 E 的样本空间 S 的子集称为 **随机事件**，简称 **事件**.
每次试验中，当且仅当这一子集的一个样本点出现称为 **事件发生**.
有一个样本点组成的单点集称为 **基本事件**.
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 成为 **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，他在每次试验中都不发生，$\emptyset$ 称为 **不可能事件**.

1.2.3 事件间的关系与事件的运算

1. 若 $A \subset B$ ，则称事件 B 包含事件 A ，这指的是事件 A 发生必导致事件 B 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 $A \cup B = \{x | x \in A\ x \in B\}$ 称为事件 A 与事件 B 的 **和事件**. 当且仅当 A, B 中至少一个发生时，事件 $A \cup B$ 发生.
   类似地，称 $\bigcup\limits_{k=1}^{n} A_k$ 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的 **和事件**；称 $\bigcup\limits_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \cdots$ 的和事件.
3. 事件 $A \cap B = \{x | x \in A\ x \in B\}$ 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. $A \cap B$ 也记作 AB .

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

4. 事件 $A - B = \{x | x \in A\ x \notin B\}$ 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 与事件 B 是 **互不相容的**，或 **互斥的**. 这指的是事件 A 与事件 B 不能同时发生. 基本事件也是两两互不相容的.

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 $\bar{A}$. $\bar{A} = S - A$

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: $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$;
$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.

德摩根律: $A \bar{\cup} B = \bar{A} \cap \bar{B}$; $A \bar{\cap} B = \bar{A} \cup \bar{B}$.

1.3 频率与概率

1.3.1 频率

在相同条件下，进行了 n 次试验，在这 n 次试验中，事件 A 发生的次数为 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**，并记成 $f_n(A)$

性质:

1. $0 \leq f_n(A) \leq 1$;
2. $f_n(S) = 1$;
3. 若 $A_1, A_2, \cdots, A_k$ 是两两互不相容的事件，则

$$f_n(A_1 \cup A_2 \cup \cdots \cup A_k) = f_n(A_1) + f_n(A_2) + \cdots + f_n(A_k).$$

1.3.2 概率

设 E 是随机试验，S 是它的样本空间. 对于 E 的每一事件 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 $P(\cdot)$ 满足以下条件:

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 $A_1, A_2, \cdots$ 是两两互不相容的事件，即对于 $A_iA_j = \emptyset$ ，$i \neq, i, j = 1, 2, \cdots$, 有

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

性质: 1. $P(\emptyset) = 0$.

2. 若 $A_1, A_2, \cdots, A_n$ 是两两互不相容的事件，则有

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

3. 设 A, B 是两个事件，若 $A \subset B$ ，则有

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

$$P(B) \geq P(A).$$

4. 对于任一事件 A ，有

$$P(A) \leq 1.$$

5. 对于任意两事件 A, B 有

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

1.4 等可能概型（古典概型）

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.

具有以上两个特点的试验称为 **等可能概型**. 它在概率论发展初期曾是主要的研究对象，所以也成为 **古典概型**.

设试验的样本空间为 $S=\{e_1,e_2,\cdots,e_n\}$ 若 A 包含 k 个基本事件则有:

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

超几何分布的概率公式:

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

1.5 条件概率

1.5.1 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 发生下事件 B 发生的 **条件概率**.

1.5.2 乘法定理

设 $P(A)>0$ 则有

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

1.5.3 全概率公式和贝叶斯公式

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

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

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

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

1.6 独立性

设 A, B 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**，简称 A, B **独立**.

定理:
1. 设 A, B 是两事件，且 $P(A) > 0$. 若 A, B 相互独立，则 $P(B|A) = P(B)$. 反之亦然.
2. 若事件 A 与 B 相互独立，则下列各对事件也相互独立: A 与 $\bar{B}$ ，$\bar{A}$ 与 B ，$\bar{A}$ 与 $\bar{B}$.

一般，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，都等于各事件概率之积，则称事件 $A_1, A_2, \cdots, A_n$ 相互独立.

推论:
1. 若事件 $A_1, A_2, \cdots, A_n (n \geq 2)$ 相互独立，则其中任意 $k(2 \leq k \leq n)$ 个事件也是相互独立的.
2. 若 n 个事件 $A_1, A_2, \cdots, A_n (n \geq 2)$ 相互独立，则将 $A_1, A_2, \cdots, A_n$ 中任意多个事件换成它们各自的对立事件，所得的 n 个事件仍相互独立.

2 第二章 随机变量及其分布

2.1 随机变量

设随机试验的样本空间为 $S = \{e\}$. $X = X(e)$ 是定义在样本空间上的单值函数. 称 $X = X(e)$ 为 **随机变量**

2.2 离散型随机变量及其分布率

2.2.1 (0-1) 分布

设随机变量 X 只可能取 0 与 1 两个值，它的分布律是

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

则称 X 服从以 p 为参数的 $(0-1)$ **分布**或 **两点分布**

2.2.2 伯努利试验、二项分布

设试验 E 只有两个可能结果：A 及 $\bar{A}$ ，则称 E 为 **伯努利** (Bernoulli) **试验**.

将 E 独立重复地进行 n 次，则称这一串重复的独立试验为 n **重伯努利试验**.

以 X 表示 n 重伯努利试验中事件 A 发生的次数,每次伯努利试验中 A 事件发生的概率为 p 称随机变量 X 服从参数为 n, p 的 **二项分布**, 并记为 $X \sim b(n, p)$. 它的分布律是

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

2.2.3 泊松分布

设随机变量 X 的所有可能取的值为 $0, 1, 2, \cdots$, 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 $X \sim \pi(\lambda)$

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

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

一般，当 $n \geq 20, p \leq 0.05$ 时用 $\frac{\lambda^k e^{-\lambda}}{k!}(\lambda = np)$ 作为 $\binom{n}{k} p^k (1-p)^{n-k}$ 的近似值效果颇佳.

2.3 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

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

称为 X 的 **分布函数**.

2.4 连续型随机变量及其概率密度

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

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

则称 X 为 **连续型随机变量**, 其中函数 $f(x)$ 称为 X 的 **概率密度函数**, 简称 **概率密度**.

2.4.1 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 $X\sim U(a,b)$

2.4.2 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**.

2.4.3 正态分布

若连续型随机变量 X 的概率密度为

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

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** (Gauss) 分布, 记为 $X\sim N(\mu,\sigma^2)$.

性质: 1. 曲线关于 $x=\mu$ 对称. 这表明对于任意 $h>0$ 有

$$P\{\mu-h<X\leq\mu\}=P\{\mu<X\leq\mu+h\}.$$

2. 当 $x=\mu$ 时取到最大值

$$F(\mu)=\frac{1}{\sqrt{2\pi}\sigma}.$$

特别, 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, 即有

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

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

易知

$$\Phi(-x) = 1 - \Phi(x).$$

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

2.5 随机变量的函数的分布

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.

3 第三章 多维随机变量及其分布

3.1 二维随机变量

一般，设 E 是一个随机试验，它的样本空间是 $S = \{e\}$ ，设 $X = X(e)$ 和 $Y = Y(e)$ 是定义在 S 上的随机变量，由它们构成的一个向量 (X, Y) ，叫做 **二维随机向量**或 **二维随机变量**.

设 (X, Y) 是二维随机变量，对于任意实数 x, y ，二元函数

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

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**.

性质:

1. $F(x, y)$ 是变量 x 和 y 的不减函数，即对于固定的 y ，当 $x_2 > x_1$ 时，$F(x_2, y) \ge F(x_1, y)$ ；对于任意固定的 x ，当 $y_2 > y_1$ 时，$F(x, y_2) \ge F(x, y_1)$.
2. $0 \le F(x, y) \le 1$ ，且

$$\text{对于任意固定的} y, F(-\infty, y) = 0,$$

$$\text{对于任意固定的}y, F(x,-\infty)=0,$$
$$F(-\infty,-\infty)=0, F(\infty,\infty)=1$$

.

3. $F(x+0,y)=F(x,y), F(x,y+0)=F(x,y)$ ，即 $F(x,y)$ 关于 x 右连续，关于 y 也右连续.
4. 对于任意 (x_1,y_1) ，(x_2,y_2) ，$x_1<x_2$ ，$y_1<y_2$ ，下述不等式成立：
$$F(x_2,y_2)-F(x_2,y_1)+F(x_1,y_1)-F(x_1,y_2)\geq 0.$$

如果二维随机变量 (X,Y) 全部可能取到的值是有限对或可列无限多对，则称 (X,Y) 是 **离散型的随机变量**.

我们称 $P\{X=x_i,Y=y_i\}=p_{ij}, i,j=1,2,\cdots$ 为二位离散型随机变量 (X,Y) 的 **分布律**，或称随机变量 X 和 Y 的 **联合分布律**.

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
$$F(x,y)=\int_{-\infty}^{y}\int_{-\infty}^{x}f(u,v)\mathrm{d}u\mathrm{d}v,$$
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**.

概率密度 $f(x,y)$ 具有以下性质：

1. $f(x,y)\geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\mathrm{d}x\mathrm{y}=F(\infty,\infty)=1.$
3. 设 G 是 xOy 平面上的区域，点 (X,Y) 落在 G 内的概率为
$$P\{(X,Y)\in G\}=\iint\limits_{G}f(x,y)\mathrm{d}x\mathrm{d}y.$$
4. 若 $f(x,y)$ 在点 (x,y) 连续，则有
$$\frac{\partial^2 F(x,y)}{\partial x\partial y}=f(x,y).$$

设 E 是一个随机试验，它的样本空间是 $S=\{e\}$ ，设 $X_1=X_1(e), X_2=X_2(e),\cdots,X_n=X_n(e)$ 是定义在 S 上的随机变量，由它们构成的一个 n 维向量 $(X_1,X_2,\cdots,X_n)$ 叫做 n **维随即向量**或 n 维随机变量.

对于任意 n 个实验 $x_1,x_2,\cdots,x_n,n$ 元函数
$$F(x_1,x_2,\cdots,x_n)=P\{X_1\leq x_1,X_2\leq x_2,\cdots,X_n\leq x_n\}$$
称为 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的 **分布函数**，或随机变量 $X_1,X_2,\cdots,X_n$ 的 **联合分布函数**.

3.2 边缘分布

3.3 条件分布

3.4 相互独立的随机变量

3.5 两个随机变量的函数的分布

4 第四章 随机变量的数字特征

4.1 数学期望

4.2 方差

4.3 协方差及相关系数

4.4 矩、协方差矩阵

5 第五章 大数定律及中心极限定理

5.1 大数定律

5.2 中心极限定理

6 第六章 样本及抽样分布

6.1 随机样本

6.2 直方图和箱线图

6.3 抽样分布

7 第七章 参数估计

7.1 点估计

7.2 基于截尾样本的最大似然估计

7.3 估计量的评选标准

7.4 区间估计

7.5 正态总体均值与方差的区间估计

7.6 （0-1）分布参数的区间估计

7.7 单侧置信区间