概率论与数理统计

Guotao He

2025-05-05

目录

1 概率论的基本概念

1.1 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 **样本空间**.
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 $\varnothing$ 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，$\varnothing$ 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 $A \cup B = \{x \mid x \in A \text{ 或} x \in B\}$ 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 $A \cap B = \{x \mid x \in A \text{ 且} x \in B\}$ 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. $A \cap B$ 也记作 AB. 类似地，称 $\bigcap\limits_{k=1}^{n} A_k$ 为 n 个事件 $A_1, A_2, \cdots, A_n$ 的 **积事件**；称 $\bigcap\limits_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \cdots$ 的积事件.
4. 事件 $A - B = \{x \mid x \in A \text{ 且} x \notin B\}$ 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 $A \cap B = \varnothing$，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 $A \cap B = \varnothing$，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：$\overline{A \cup B} = \bar{A} \cap \bar{B}$；$\overline{A \cap B} = \bar{A} \cup \bar{B}$.

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 $\dfrac{n_A}{n}$ 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

Definition 1.1 (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 $A_iA_j = \varnothing$，$i \neq j$，有

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

Remark 1.1. 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 $A_1, \ldots, A_n$，有：

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

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.

- 性质：
  1. $P(\varnothing)=0$.
  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，所以也称为 **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

Definition 1.2 (条件概率)**.** 设 A,B 是两个事件，且 $P(A)>0$，称

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

为在事件 A 发生下事件 B 发生的 **条件概率**.

Definition 1.3 (乘法定理)**.** 设 $P(A)>0$，则有

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

Theorem 1.1 (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

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

Theorem 1.2 (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

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

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 $P(B\mid A)=P(B)$. 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.

Remark 1.2. 一般而言，两两独立和相互独立并不相同. 具体而言，两两独立指从 $A_1,\ldots,A_n$ 中任取两个，均有 $P(A_iA_j)=P(A_i)P(A_j)$. 而相互独立指的是 $P(A_1\ldots A_n)=P(A_1)\ldots P(A_n)$. 相互独立一定两两独立，但两两独立不一定相互独立.

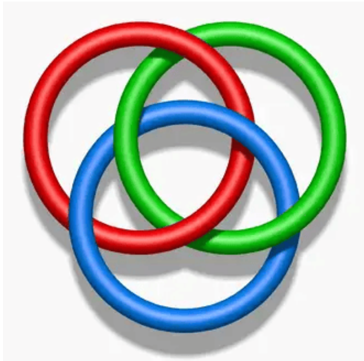

图 1: Borronmean Ring，任意两个环两两不相交，但整体无法解开

2 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

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

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

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

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

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

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

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

一般，当 $n \geq 20$，$p \leq 0.05$ 时，用 $\dfrac{\lambda^k e^{-\lambda}}{k!}$（$\lambda = np$）作为 $\dbinom{n}{k} p^k (1-p)^{n-k}$ 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

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

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

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

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

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

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

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

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

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

其中 μ, σ（$\sigma > 0$）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 $\varphi(x)$ 和 $\Phi(x)$ 表示，即有

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

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

易知

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

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

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

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

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

3 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

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

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
     (a) 对于任意固定的 y，$F(-\infty, y) = 0$；
     (b) 对于任意固定的 x，$F(x, -\infty) = 0$；
     (c) $F(-\infty, -\infty) = 0$，$F(\infty, \infty) = 1$.
  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)，若 $x_1 < x_2$，$y_1 < y_2$，则

$$F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2) \geq 0.$$

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**.

称 $P\{X = x_i, Y = y_j\} = p_{ij}$（$i, j = 1, 2, \cdots$）为二维离散型随机变量 (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

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

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**.

- 性质：

1. $f(x, y) \geq 0$；
2. $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \, dx \, dy = F(\infty, \infty) = 1$；
3. 设 G 是 xOy 平面上的区域，点 (X, Y) 落在 G 内的概率为

$$P\{(X, Y) \in G\} = \iint_G f(x, y) \, dx \, dy;$$

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 **维随机向量**或 n 维随机变量.

对于任意 n 个实数 $x_1, x_2, \cdots, x_n$，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

二维随机变量 (X, Y) 作为一个整体，具有分布函数 $F(x, y)$. 而 X 和 Y 都是随机变量，各自也有分布函数，分别记为 $F_X(x), F_Y(y)$，依次称为二维随机变量 (X, Y) 关于 X 和关于 Y 的 **边缘分布函数**.

对于离散型随机变量可得

$$F_X(x) = F(x, \infty) = \sum_{x_i \leq x} \sum_{j=1}^{\infty} p_{ij},$$

X 的分布律为

$$P\{X=x_i\}=\sum_{j=1}^{\infty}p_{ij},\quad i=1,2,\cdots.$$

同样，Y 的分布律为

$$P\{Y=y_j\}=\sum_{i=1}^{\infty}p_{ij},\quad j=1,2,\cdots.$$

记

$$p_{i\cdot}=\sum_{j=1}^{\infty}p_{ij}=P\{X=x_i\},\quad i=1,2,\cdots,$$

$$p_{\cdot j}=\sum_{i=1}^{\infty}p_{ij}=P\{Y=y_j\},\quad j=1,2,\cdots,$$

分别称 $p_{i\cdot}$（$i=1,2,\cdots$）和 $p_{\cdot j}$（$j=1,2,\cdots$）为 (X,Y) 关于 X 和关于 Y 的 **边缘分布律**.

对于连续型随机变量 (X,Y)，设它的概率密度为 $f(x,y)$，由于

$$F_X(x)=F(x,\infty)=\int_{-\infty}^{x}\left[\int_{-\infty}^{\infty}f(x,y)\,dy\right]dx,$$

X 是一个连续型随机变量，其概率密度为

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)\,dy.$$

同样，Y 也是一个连续型随机变量，其概率密度为

$$f_Y(y)=\int_{-\infty}^{\infty}f(x,y)\,dx.$$

分别称 $f_X(x)$、$f_Y(y)$ 为 (X,Y) 关于 X 和关于 Y 的 **边缘概率密度**.

3.3 条件分布

设 (X,Y) 是二维离散型随机变量，对于固定的 j，若 $P\{Y=y_j\}>0$，则称

$$P\{X=x_i\mid Y=y_j\}=\frac{P\{X=x_i,Y=y_j\}}{P\{Y=y_j\}}=\frac{p_{ij}}{p_{\cdot j}},\quad i=1,2,\cdots$$

为在 $Y=y_j$ 条件下随机变量 X 的 **条件分布律**. 同样，对于固定的 i，若 $P\{X=x_i\}>0$，则称

$$P\{Y=y_j\mid X=x_i\}=\frac{P\{X=x_i,Y=y_j\}}{P\{X=x_i\}}=\frac{p_{ij}}{p_{i\cdot}},\quad j=1,2,\cdots$$

为在 $X=x_i$ 条件下随机变量 Y 的 **条件分布律**.

设二维随机变量 (X,Y) 的概率密度为 $f(x,y)$，(X,Y) 关于 Y 的边缘概率密度为 $f_Y(y)$. 若对于固定的 y，$f_Y(y)>0$，则称 $\dfrac{f(x,y)}{f_Y(y)}$ 为在 $Y=y$ 的条件下 X 的 **条件概率密度**，记为

$$f_{X|Y}(x\mid y)=\frac{f(x,y)}{f_Y(y)},$$

类似地，可以定义

$$f_{Y|X}(y\mid x)=\frac{f(x,y)}{f_X(x)}.$$

3.4 相互独立的随机变量

设 $F(x,y)$ 及 $F_X(x), F_Y(y)$ 分别是二维随机变量 (X,Y) 的分布函数及边缘分布函数. 若对于所有 x,y 有

$$P\{X \le x, Y \le y\} = P\{X \le x\}P\{Y \le y\},$$

即

$$F(x,y) = F_X(x)F_Y(y),$$

则称随机变量 X 和 Y 是相互独立的.

- 定理：设 $(X_1, X_2, \cdots, X_m)$ 和 $(Y_1, Y_2, \cdots, Y_n)$ 相互独立，则 X_i（$i = 1, 2, \cdots, m$）和 Y_j（$j = 1, 2, \cdots, n$）相互独立. 若 h, g 是连续函数，则 $h(X_1, X_2, \cdots, X_m)$ 和 $g(Y_1, Y_2, \cdots, Y_n)$ 相互独立.

4 随机变量的数字特征

4.1 数学期望

Definition 4.1 (期望)**.** 设离散型随机变量 X 的分布律为

$$P\{X = x_k\} = p_k, \quad k = 1, 2, \cdots.$$

若级数

$$\sum_{k=1}^{\infty} x_k p_k$$

绝对收敛，则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 X 的 **数学期望**，记为 $E(X)$，即

$$E(X) = \sum_{k=1}^{\infty} x_k p_k.$$

设连续型随机变量 X 的概率密度为 $f(x)$，若积分

$$\int_{-\infty}^{\infty} x f(x)\, dx$$

绝对收敛，则称积分 $\int_{-\infty}^{\infty} x f(x)\, dx$ 的值为随机变量 X 的 **数学期望**，记为 $E(X)$，即

$$E(X) = \int_{-\infty}^{\infty} x f(x)\, dx.$$

数学期望简称 **期望**，又称 **均值**.

- 性质：

  1. 设 C 是常数，则有 $E(C) = C$；
  2. 设 X 是随机变量，C 是常数，则有

$$E(CX) = CE(X);$$

3. 设 X, Y 是两个随机变量，则有

$$E(X+Y) = E(X) + E(Y);$$

这一性质可以推广到任意有限多个随机变量之和的情况；

4. 设 X, Y 是相互独立的随机变量，则有

$$E(XY) = E(X)E(Y);$$

这一性质可以推广到任意有限个相互独立的随机变量之积的情况.

4.2 方差

Definition 4.2 (方差). 设 X 是一个随机变量，若 $E\{[X-E(X)]^2\}$ 存在，则称 $E\{[X-E(X)]^2\}$ 为 X 的 **方差**，记为 $D(X)$ 或 $Var(X)$，即

$$D(X) = Var(X) = E\{[X-E(X)]^2\}.$$

在应用上还引入量 $\sqrt{D(X)}$，记为 $\sigma(X)$，称为 **标准差** 或 **均方差**.

- 性质：

1. 设 C 是常数，则有 $D(C) = 0$；
2. 设 X 是随机变量，C 是常数，则有

$$D(CX) = C^2 D(X),$$
$$D(X+C) = D(X);$$

3. 设 X, Y 是两个随机变量，则有

$$D(X+Y) = D(X) + D(Y) + 2E\{[X-E(X)][Y-E(Y)]\};$$

特别地，若 X, Y 相互独立，则有

$$D(X+Y) = D(X) + D(Y);$$

这一性质可以推广到任意有限多个相互独立的随机变量之和的情况；

4. $D(X) = 0$ 的充要条件是 X 以概率为 1 取常数 $E(X)$，即

$$P\{X = E(X)\} = 1.$$

4.3 常见分布的期望和方差

常见分布的期望和方差如下表：

- (0-1) 分布 $B(1, p)$：$EX = p$，$DX = p(1-p)$
- 二项分布 $B(n, p)$：$EX = np$，$DX = np(1-p)$
- Poisson 分布 $\pi(\lambda)$：$EX = \lambda$，$DX = \lambda$
- 均匀分布 $U(a, b)$：$EX = \dfrac{a+b}{2}$，$DX = \dfrac{(b-a)^2}{12}$
- 正态分布 $N(\mu, \sigma^2)$：$EX = \mu$，$DX = \sigma^2$
- 指数分布 $Exp(\theta)$：$EX = \theta$，$DX = \theta^2$

4.4 协方差及相关系数

Definition 4.3 (协方差、相关系数). $E\{[X-E(X)][Y-E(Y)]\}$ 称为随机变量 X 与 Y 的 **协方差**，记为 $Cov(X,Y)$，即

$$Cov(X,Y)=E\{[X-E(X)][Y-E(Y)]\}.$$

而

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量 X 与 Y 的 **相关系数**.

由定义知

$$\mathrm{Cov}(X,Y)=\mathrm{Cov}(Y,X),$$

$$\mathrm{Cov}(X,X)=D(X).$$

对于任意两个随机变量 X 和 Y，下列等式成立：

$$D(X+Y)=D(X)+D(Y)+2\mathrm{Cov}(X,Y).$$

将 $\mathrm{Cov}(X,Y)$ 的定义式展开，易得

$$\mathrm{Cov}(X,Y)=E(XY)-E(X)E(Y).$$

常用此式计算协方差.

- 性质：
  1. $\mathrm{Cov}(aX,bY)=ab\mathrm{Cov}(X,Y)$，其中 a,b 是常数；
  2. $\mathrm{Cov}(X_1+X_2,Y)=\mathrm{Cov}(X_1,Y)+\mathrm{Cov}(X_2,Y)$.
- 定理：
  1. $|\rho_{XY}|\le 1$；
  2. $|\rho_{XY}|=1$ 的充要条件是存在常数 a,b，使得

$$P\{Y=a+bX\}=1.$$

当 $|\rho_{XY}|=1$ 时，称 X 和 Y **不相关**.

4.5 矩、协方差矩阵

设 X 和 Y 是随机变量：

- 若 $E(X^k)$（$k=1,2,\cdots$）存在，称它为 X 的 k **阶原点矩**；
- 若 $E\{[X-E(X)]^k\}$（$k=2,3,\cdots$）存在，称它为 X 的 k **阶中心矩**；
- 若 $E(X^kY^l)$（$k,l=1,2,\cdots$）存在，称它为 X 和 Y 的 $k+l$ **阶混合矩**；
- 若 $E\{[X-E(X)]^k[Y-E(Y)]^l\}$（$k,l=1,2,\cdots$）存在，称它为 X 和 Y 的 $k+l$ **阶混合中心矩**.

X 的数学期望 $E(X)$ 是其一阶原点矩，方差 $D(X)$ 是二阶中心矩，协方差 $\mathrm{Cov}(X,Y)$ 是二阶混合中心矩.

设 n 维随机变量 $(X_1,X_2,\cdots,X_n)$ 的二阶混合中心矩

$$c_{ij}=\mathrm{Cov}(X_i,X_j)=E\{[X_i-E(X_i)][X_j-E(X_j)]\},\quad i,j=1,2,\cdots,n$$

都存在，则称矩阵

$$C=\begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 n 维随机变量的 **协方差矩阵**. 由于 $c_{ij}=c_{ji}$，该矩阵为对称矩阵.

5 大数定律及中心极限定理

5.1 大数定律

Theorem 5.1 (伯努利大数定理). 设 f_A 为 n 次独立试验中事件 A 发生的次数，p 为每次试验中 A 发生的概率，则对任意 $\epsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{f_A}{n}-p\right|<\epsilon\right\}=1.$$

Theorem 5.2 (弱大数定理（辛钦大数定理）). 设 $X_1,X_2,\cdots$ 是独立同分布的随机变量序列，且 $E(X_k)=\mu$，则对任意 $\epsilon>0$，有

$$\lim_{n\to\infty} P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\epsilon\right\}=1.$$

Definition 5.1. 称随机变量序列 $Y_1,Y_2,\cdots$ **依概率收敛于** a，若对任意 $\epsilon>0$，有

$$\lim_{n\to\infty} P\{|Y_n-a|<\epsilon\}=1,$$

记为 $Y_n\xrightarrow{P}a$.

- 性质：若 $X_n\xrightarrow{P}a$，$Y_n\xrightarrow{P}b$，且函数 $g(x,y)$ 在 (a,b) 连续，则

$$g(X_n,Y_n)\xrightarrow{P}g(a,b).$$

5.2 中心极限定理

Theorem 5.3 (独立同分布的中心极限定理). 设 $X_1,X_2,\cdots$ 是独立同分布序列，$E(X_k)=\mu$，$D(X_k)=\sigma^2>0$，则标准化变量

$$Y_n=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 满足：

$$\lim_{n\to\infty} F_n(x) = \Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt.$$

Theorem 5.4 (棣莫弗-拉普拉斯定理). 设 $\eta_n \sim b(n,p)$，则对任意 x，有

$$\lim_{n\to\infty} P\left\{\frac{\eta_n - np}{\sqrt{np(1-p)}} \le x\right\} = \Phi(x).$$

6 样本及抽样分布

6.1 随机样本

总体中抽取的独立同分布样本 $X_1, X_2, \cdots, X_n$ 称为 **简单随机样本**，其观察值 $x_1, x_2, \cdots, x_n$ 称为 **样本值**.

6.2 抽样分布

统计量是不含未知参数的样本函数. 常用统计量包括：

- 样本均值：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$
- 样本方差：$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$
- 样本标准差：$S = \sqrt{S^2}$
- 样本 k 阶矩：$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

经验分布函数：

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I_{\{X_i \le x\}}$$

其中 I 为示性函数.

6.2.1 三大抽样分布

1. χ^2 **分布：**
   (a) 定义：$X_1, \cdots, X_n \sim N(0,1)$，则 $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2(n)$
   (b) 性质：可加性；$E(\chi^2) = n$，$D(\chi^2) = 2n$
2. t **分布：** 定义：$X \sim N(0,1)$，$Y \sim \chi^2(n)$ 独立，则 $t = \frac{X}{\sqrt{Y/n}} \sim t(n)$
3. F **分布：** 定义：$U \sim \chi^2(n_1)$，$V \sim \chi^2(n_2)$ 独立，则 $F = \frac{U/n_1}{V/n_2} \sim F(n_1, n_2)$

6.2.2 正态总体抽样定理

- 样本均值分布：$\bar{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$
- 样本方差分布：$\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
- t 分布构造：$\dfrac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$
- 方差比分布：$\dfrac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} \sim F(n_1-1, n_2-1)$

7 参数估计

7.1 点估计

7.1.1 矩估计法

通过匹配样本矩与总体矩来估计参数. 设总体前 k 阶矩为 $\mu_l(\theta_1, \cdots, \theta_k)$，解方程组：

$$\begin{cases} \mu_1 = A_1 \\ \vdots \\ \mu_k = A_k \end{cases}$$

得矩估计量 $\hat{\theta}_i$.

7.1.2 最大似然估计

似然函数 $L(\theta) = \displaystyle\prod_{i=1}^{n} f(x_i;\theta)$ 或 $L(\theta) = \displaystyle\prod_{i=1}^{n} p(x_i;\theta)$，通过求解 $\dfrac{d}{d\theta}\ln L(\theta) = 0$ 得极大似然估计量.

7.2 估计量评价标准

- 无偏性：$E(\hat{\theta}) = \theta$
- 有效性：方差最小的无偏估计量
- 相合性：$\lim\limits_{n\to\infty} P(|\hat{\theta}-\theta| < \epsilon) = 1$

7.3 区间估计

构造置信区间 $(\underline{\theta}, \overline{\theta})$ 使得 $P(\underline{\theta} < \theta < \overline{\theta}) \geq 1-\alpha$，常用枢轴量法：

1. 寻找与参数相关的枢轴量 $W(X,\theta)$，其分布已知
2. 确定区间 (a,b) 使得 $P(a < W < b) = 1-\alpha$
3. 转化为参数不等式

7.4 正态总体区间估计

- 均值估计（σ^2 已知）：$\bar{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$
- 均值估计（σ^2 未知）：$\bar{X} \pm t_{\alpha/2}(n-1)\dfrac{S}{\sqrt{n}}$
- 方差估计：$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$