



# 概率论与数理统计讲义

## Probability and Statistics

作者：薛冰

时间：November 20, 2024

# 目录

<b>第 1 章 随机事件与概率</b>	<b>1</b>
1.1 随机事件及其运算	1
1.2 概率的定义及其确定方法	3
1.2.1 概率的公理化定义	3
1.2.2 排列组合公式	3
1.2.3 确定概率的频率方法	4
1.2.4 确定概率的古典方法	4
1.2.5 确定概率的几何方法	4
1.2.6 确定概率的主观方法	4
1.3 概率的性质	5
1.3.1 概率的可加性	5
1.3.2 概率的单调性	5
1.3.3 概率的加法公式	6
1.3.4 概率的连续性	7
1.4 条件概率	9
1.4.1 条件概率的定义	9
1.4.2 乘法公式	9
1.4.3 全概率公式	10
1.4.4 Bayes 公式	10
1.5 独立性	11
1.5.1 两个事件的独立性	11
1.5.2 多个事件的相互独立性	12
1.6 试验的独立性	12
<b>第 2 章 随机变量及其分布</b>	<b>14</b>
2.1 基本概念	14
2.1.1 随机变量的定义	14
2.1.2 分布函数	14
2.1.3 离散随机变量的概率分布列	15
2.1.4 连续随机变量的概率密度函数	16
2.2 数学期望	16
2.2.1 数学期望的定义	16
2.2.2 数学期望的性质	17
2.2.3 Markov 不等式	18
2.3 方差与标准差	18
2.3.1 方差与标准差的定义	18
2.3.2 方差的性质	19
2.3.3 Chebyshev 不等式	19
2.3.4 随机变量的中心化与标准化	20
2.4 常用离散分布	21
2.4.1 二项分布	21

# 第 1 章 随机事件与概率

## 1.1 随机事件及其运算

在一定的条件下,并不总是出现相同结果的现象称为**随机现象**.反之,只有一个结果的现象称为**确定性现象**.对在相同条件下可以重复的随机现象的观察、记录、实验称为**随机试验**.

### 定义 1.1 (样本空间)

随机现象的一切可能基本结果组成的集合称为**样本空间**,记为  $\Omega = \{\omega_i\}$ ,其中  $\omega$  表示基本结果,称为**样本点**. $i$  是指标值,不同的样本空间中样本点的个数不一定相同,指标  $i$  的取值范围就不同.

注

- (1) 样本点可以是数,也可以不是数;
- (2) 随机现象的样本空间至少含有两个样本点;
- (3) 样本空间中样本点的个数为至多可数个时,称为**离散样本空间**,样本点个数不可数时,称为**连续样本空间**.随机现象的某些样本点组成的集合称为**随机事件**,简称**事件**,常用  $A, B, C, \dots$  表示.

由样本空间  $\Omega$  中的单个元素组成的子集称为**基本事件**.样本空间  $\Omega$  的最大子集(即  $\Omega$  本身)称为**必然事件**.样本空间  $\Omega$  的最小子集(即空集  $\emptyset$ )称为**不可能事件**.

用来表示随机现象结果的变量称为**随机变量**,常用  $X, Y, Z$  表示.

事件是一类集合,因此对事件间的关系和事件间的运算,可由集合间的关系和集合间的运算推知.

### 定义 1.2 (包含关系)

如果属于  $A$  的样本点必属于  $B$ ,则称  $A$  被包含在  $B$  中,或称  $B$  包含  $A$ ,记作  $A \subset B$  或  $B \supset A$ .

注 用概率论的语言说:事件  $A$  的发生必然导致事件  $B$  发生.

注 对任一事件  $A$ ,必有

$$\emptyset \subset A \subset \Omega.$$

### 定义 1.3 (相等关系)

如果事件  $A$  与事件  $B$  满足:

$$A \subset B \text{ 且 } B \subset A,$$

则称事件  $A$  与  $B$  相等,记作  $A = B$ .

### 定义 1.4 (互不相容)

如果事件  $A$  和  $B$  没有相同的样本点,则称  $A$  与  $B$  互不相容.

注 用概率论的语言说:  $A$  与  $B$  不可能同时发生.

事件  $A$  与  $B$  的并,记作  $A \cup B$ ,含义为“由事件  $A$  与  $B$  中所有的样本点组成的新事件”.或用概率论的语言说:  $A$  与  $B$  至少有一个发生.

事件  $A$  与  $B$  的交,记作  $A \cap B$  或  $AB$ ,含义为“由事件  $A$  与  $B$  中公共的样本点组成的新事件”.或用概率论的语言说:  $A$  与  $B$  同时发生.

事件  $A$  对  $B$  的差,记作  $A - B$ ,含义为“由在事件  $A$  中而不在事件  $B$  中的样本点组成的新事件”.或用概率论的语言说:事件  $A$  发生而  $B$  不发生.

事件  $A$  的对立事件,记作  $\bar{A}$ ,含义为“由在  $\Omega$  中而不在  $A$  中的样本点组成的新事件”.或用概率论的语言说:  $A$  不发生.

## 注

(1) 对立事件一定互不相容, 但互不相容的事件不一定对立.

$$(2) A - B = A - AB = A\bar{B}$$

集合的运算性质一样可以迁移到事件中来.

## 定理 1.1

(1) (交换律)  $A \cup B = B \cup A, A \cap B = B \cap A$ .

(2) (结合律)  $A \cup (B \cap C) = (A \cup B) \cap C, A \cap (B \cup C) = (A \cap B) \cup C$ .

(3) (分配律)  $A \cap (B \cup C) = (A \cap B) \cup (A \cap C), A \cap (\bigcup_{\alpha \in \Lambda} B_{\alpha}) = \bigcup_{\alpha \in \Lambda} (A \cap B_{\alpha})$ .

(4)  $A \cup A = A, A \cap A = A, \bar{\bar{A}} = A$ .



## 定理 1.2 (De Morgan 公式)

若  $\{A_{\alpha} | \alpha \in \Lambda\}$  是至多可数个事件, 则

$$(1) \overline{\left(\bigcup_{\alpha \in \Lambda} A_{\alpha}\right)} = \bigcap_{\alpha \in \Lambda} \bar{A}_{\alpha};$$

$$(2) \overline{\left(\bigcap_{\alpha \in \Lambda} A_{\alpha}\right)} = \bigcup_{\alpha \in \Lambda} \bar{A}_{\alpha}.$$



## 定义 1.5 (事件域)

设  $\Omega$  为一样本空间,  $\mathcal{F}$  为  $\Omega$  的某些子集所组成的集族. 如果  $\mathcal{F}$  满足:

1.  $\Omega \in \mathcal{F}$ ;

2. 若  $A \in \mathcal{F}$ , 则  $\bar{A} \in \mathcal{F}$ ;

3. 若  $A_n \in \mathcal{F}, n = 1, 2, \dots$ , 则  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ ,

则称  $\mathcal{F}$  为一个事件域. 显然, 它是一个  $\sigma$  代数.



在概率论中, 又称  $(\Omega, \mathcal{F})$  为可测空间, 在可测空间上才定义概率.

对于含有全体实数的样本空间  $\Omega = (-\infty, \infty) = \mathbb{R}$ , 我们可以取 Borel 代数的一维情形来定义事件域. 由测度论相关知识, Borel 集都是 Lebesgue 可测的.

## 定义 1.6 (样本空间的分割)

对样本空间  $\Omega$ , 如果有  $n$  个互不相容的事件  $D_1, D_2, \dots, D_n$  满足:

$$\bigcup_{i=1}^n D_i = \Omega,$$

则称  $\{D_i | i = 1, 2, \dots, n\}$  为样本空间  $\Omega$  的一组分割. 这里的事件个数也可以是可数个.



一般场合, 若分割  $\mathcal{D} = \{D_1, D_2, \dots, D_n\}$  由  $n$  个事件组成, 则其产生的事件域  $\sigma(\mathcal{D})$  共含有  $2^n$  个不同的事件.

## 1.2 概率的定义及其确定方法

### 1.2.1 概率的公理化定义

#### 定义 1.7 (概率的公理化定义)

设  $\Omega$  为一个样本空间,  $\mathcal{F}$  为  $\Omega$  的某些子集组成的一个事件域. 如果对任一事件  $A \in \mathcal{F}$ , 定义在  $\mathcal{F}$  上的一个实值函数  $P(A)$  满足:

- (1) 非负性: 若  $A \in \mathcal{F}$ , 则  $P(A) \geq 0$ ;
- (2) 正则性:  $P(\Omega) = 1$ ;
- (3) 可列可加性: 若  $A_1, A_2, \dots, A_n$  互不相容, 则

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

则称  $P(A)$  为事件  $A$  的概率, 称  $(\Omega, \mathcal{F}, P)$  为概率空间.



**注** 显然, 概率是一类测度, 概率空间是一类测度空间.

**注** 不可能事件概率一定为零, 但概率为零的事件不一定是不可能事件.

### 1.2.2 排列组合公式

下面我们介绍排列与组合公式. 首先, 介绍下面两条计数原理.

**乘法原理:** 如果某件事需经  $k$  个步骤才能完成, 做第  $i$  ( $i = 1, 2, \dots, k$ ) 步有  $m_i$  种方法, 那么完成这件事共有  $m_1 \times m_2 \times \dots \times m_k$  种方法.

**加法原理:** 如果某件事可由  $k$  类不同途径之一去完成, 在第  $i$  ( $i = 1, 2, \dots, k$ ) 类途径中有  $m_i$  种方法, 那么完成这件事共有  $m_1 + m_2 + \dots + m_k$  种方法.

#### 定义 1.8 (排列)

从  $n$  个不同元素中任取  $r$  ( $r \leq n$ ) 个元素排成一列 (考虑元素先后出现次序), 称此为一个排列. 此种排列的总数记为  $P_n^r$ . 按乘法原理, 取出的第一个元素有  $n$  种取法, 取出的第二个元素有  $n-1$  种取法, 以此类推, 取出的第  $r$  个元素有  $n-r+1$  种取法, 所以有

$$P_n^r = n \times (n-1) \times \dots \times (n-r+1) = \frac{n!}{(n-r)!}.$$

若  $r = n$ , 则称为全排列, 记为  $P_n$ . 显然,  $P_n = n!$ .



#### 定义 1.9 (重复排列)

从  $n$  个不同元素中每次取出一个, 放回后再取下一个, 如此连续取  $r$  次所得的排列称为重复排列, 此种重复排列数共有  $n^r$  个. 注意: 这里的  $r$  允许大于  $n$ .



#### 定义 1.10 (组合)

从  $n$  个不同元素中任取  $r$  ( $r \leq n$ ) 个元素并成一组 (不考虑元素间的先后次序), 称此为一个组合, 此种组合的总数记为  $\binom{n}{r}$  或  $C_n^r$ . 按乘法原理此种组合的总数为

$$\binom{n}{r} = \frac{P_n^r}{r!} = \frac{n!}{r!(n-r)!}.$$

在此规定  $0! = 1$  与  $\binom{n}{0} = 1$ . 组合具有性质:

$$\binom{n}{r} = \binom{n}{n-r}.$$



**定义 1.11 (重复组合)**

从  $n$  个元素中每次取出一个, 放回后再取下一个, 如此连续取  $r$  次所得的组合称为**重复组合**, 此种组合总数为  $\binom{n+r-1}{r}$ . 注意: 这里的  $r$  也允许大于  $n$ .

**1.2.3 确定概率的频率方法**

确定概率的频率方法是在大量重复试验中, 用频率的稳定性去获得概率的一种方法, 其基本思想是:

- (1) 与考察事件  $A$  有关的随机现象可大量重复进行.
- (2) 在  $n$  次重复试验中, 记  $n(A)$  为事件  $A$  出现的次数, 又称  $n(A)$  为事件  $A$  的**频数**. 称

$$f_n(A) = \frac{n(A)}{n}$$

为事件  $A$  出现的**频率**.

- (3) 长期实践表明: 随着试验重复次数  $n$  的增加, 频率  $f_n(A)$  会稳定在某一常数  $a$  附近, 我们称这个常数为**频率的稳定值**. 这个频率的稳定值就是我们所求的概率.

容易验证: 用频率方法确定的概率满足公理化定义, 它的非负性与正则性是显然的, 而可加性只需注意到: 当  $A$  与  $B$  互不相容时, 计算  $A \cup B$  的频数可以分别计算  $A$  的频数和  $B$  的频数, 然后再相加, 这意味着  $n(A \cup B) = n(A) + n(B)$ , 从而有

$$\begin{aligned} f_n(A \cup B) &= \frac{n(A \cup B)}{n} = \frac{n(A) + n(B)}{n} \\ &= \frac{n(A)}{n} + \frac{n(B)}{n} = f_n(A) + f_n(B). \end{aligned}$$

**1.2.4 确定概率的古典方法**

确定概率的古典方法是概率论历史上最先开始研究的情形. 其基本思想是:

- (1) 所涉及的随机现象只有有限个样本点, 譬如有  $n$  个.
- (2) 每个样本点发生的可能性相等 (称为等可能性).
- (3) 若事件  $A$  含有  $k$  个样本点, 则事件  $A$  的概率为

$$P(A) = \frac{\text{事件 } A \text{ 所含样本点的个数}}{\Omega \text{ 中所有样本点的个数}} = \frac{k}{n}.$$

容易验证: 用古典方法确定的概率满足公理化定义, 它的非负性与正则性是显然的, 而满足可加性的理由与频率方法类似. 当  $A$  和  $B$  互不相容时, 计算  $A \cup B$  的样本点个数可以分别计算  $A$  的样本点个数和  $B$  的样本点个数, 然后再相加, 从而有可加性  $P(A \cup B) = P(A) + P(B)$ .

**1.2.5 确定概率的几何方法**

确定概率的几何方法, 其基本思想是:

- (1) 如果一个随机现象的样本空间  $\Omega$  充满某个区域, 其度量可用  $S_\Omega$  表示.
- (2) 任意一点落在度量相同的子区域内是等可能的.
- (3) 若事件  $A$  为  $\Omega$  中某个子区域, 且其度量大小可用  $S_A$  表示, 则事件  $A$  的概率为

$$P(A) = \frac{S_A}{S_\Omega}.$$

这个概率称为**几何概率**, 它满足概率的公理化定义.

**1.2.6 确定概率的主观方法**

在现实世界中有一些随机现象是不能重复的或不能大量重复的, 这时有关事件的概率该如何确定呢?



统计界的贝叶斯学派认为：一个事件的概率是人们根据经验对该事件发生的可能性所给出的个人信念. 这样给出的概率称为主观概率.

人们常用自己的经验或者他人的经验来确定主观概率. 主观给定的概率也要符合公理化的定义.

## 1.3 概率的性质

概率是一种测度，所以概率的很多性质都可以由测度的性质迁移过来.

### 定理 1.3

$$P(\emptyset) = 0.$$



**证明** 由于任何事件与不可能事件之并仍是此事件本身，于是

$$\Omega = \Omega \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots.$$

因为不可能事件与任何事件是互不相容的，故由可列可加性公理得

$$P(\Omega) = P(\Omega \cup \emptyset \cup \cdots \cup \emptyset \cup \cdots),$$

从而由  $P(\Omega) = 1$  得

$$P(\emptyset) + P(\emptyset) + \cdots = 0,$$

再由非负性公理，必有

$$P(\emptyset) = 0.$$



### 1.3.1 概率的可加性

#### 定理 1.4 (有限可加性)

若有限个事件  $A_1, A_2, \dots, A_n$  互不相容，则有

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



**证明** 对  $A_1, A_2, \dots, A_n, \emptyset, \emptyset, \dots$  应用可列可加性即得.



由有限可加性，我们就可以得到下面求对立事件概率的公式.

### 定理 1.5

对任一事件  $A$ ，有

$$P(\bar{A}) = 1 - P(A).$$



**证明** 由概率的正则性和有限可加性即得.



### 1.3.2 概率的单调性

#### 定理 1.6

若  $A \supset B$ ，则

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



**证明** 因为  $A \supset B$ , 所以

$$A = B \cup (A - B).$$

又  $B$  与  $A - B$  互不相容, 因此

$$P(A) = P(B) + P(A - B),$$

即

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

■

#### 推论 1.1 (单调性)

若  $A \supset B$ , 则  $P(A) \geq P(B)$ .

♡

#### 定理 1.7

对任意两个事件  $A, B$ , 有

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

♡

**证明**  $P(A - B) = P(A - AB)$ , 且  $AB \subset A$ , 由上条性质即得结论.

■

### 1.3.3 概率的加法公式

#### 定理 1.8

对任意两个事件  $A, B$ , 有

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

对任意  $n$  个事件  $A_1, A_2, \dots, A_n$ , 有

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \dots + (-1)^{n-1} P(A_1 A_2 \dots A_n).$$

♡

**证明** 因为

$$A \cup B = A \cup (B - AB),$$

且  $A$  和  $B - AB$  互不相容, 由有限可加性得

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

由归纳法可以证明任意  $n$  个事件的情况.

■

#### 推论 1.2 (半可加性)

对任意两个事件  $A, B$ , 有

$$P(A \cup B) \leq P(A) + P(B).$$

对任意  $n$  个事件  $A_1, A_2, \dots, A_n$ , 有

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

♡



### 1.3.4 概率的连续性

中学时期我们已经学习了离散的概率, 对于连续概率我们前所未闻. 为了讨论概率的连续性, 我们先对事件序列的极限给出定义 (类比于集合列的极限).

#### 定义 1.12 (极限事件)

对  $\mathcal{F}$  中任一增的事件序列  $F_1 \subset F_2 \subset \cdots \subset F_n \subset \cdots$ , 称可列并  $\bigcup_{n=1}^{\infty} F_n$  为  $\{F_n\}$  的极限事件, 记为

$$\lim_{n \rightarrow \infty} F_n = \bigcup_{n=1}^{\infty} F_n.$$

对  $\mathcal{F}$  中任一减的事件序列  $E_1 \supset E_2 \supset \cdots \supset E_n \supset \cdots$ , 称可列交  $\bigcap_{n=1}^{\infty} E_n$  为  $\{E_n\}$  的极限事件, 记为

$$\lim_{n \rightarrow \infty} E_n = \bigcap_{n=1}^{\infty} E_n.$$



有了以上极限事件的定义, 我们就可给出概率函数的连续性定义.

#### 定义 1.13

对  $\mathcal{F}$  上的一个概率  $P$ , 若它对  $\mathcal{F}$  中任一增的事件序列  $\{F_n\}$  均成立

$$\lim_{n \rightarrow \infty} P(F_n) = P(\lim_{n \rightarrow \infty} F_n),$$

则称概率  $P$  是下连续的.

若它对  $\mathcal{F}$  中任一减的事件序列  $\{E_n\}$  均成立

$$\lim_{n \rightarrow \infty} P(E_n) = P(\lim_{n \rightarrow \infty} E_n),$$

则称概率  $P$  是上连续的.



有了以上的定义, 我们就可以证明概率的连续性了.

#### 定理 1.9 (概率的连续性)

若  $P$  为事件域  $\mathcal{F}$  上的概率, 则  $P$  既是下连续的, 又是上连续的.



**证明** 先证  $P$  的下连续性. 设  $\{F_n\}$  是  $\mathcal{F}$  中一个单调增的事件序列, 即

$$\bigcup_{i=1}^{\infty} F_i = \lim_{n \rightarrow \infty} F_n.$$

若定义  $F_0 = \emptyset$ , 则

$$\bigcup_{i=1}^{\infty} F_i = \bigcup_{i=1}^{\infty} (F_i - F_{i-1}).$$

由于  $F_{i-1} \subset F_i$ , 显然诸  $(F_i - F_{i-1})$  两两不相容, 再由可列可加性得

$$P\left(\bigcup_{i=1}^{\infty} (F_i - F_{i-1})\right) = \lim_{n \rightarrow \infty} \sum_{i=1}^n P(F_i - F_{i-1}).$$

又由有限可加性得

$$\sum_{i=1}^n P(F_i - F_{i-1}) = P\left(\bigcup_{i=1}^n (F_i - F_{i-1})\right) = P(F_n).$$

所以

$$P(\lim_{n \rightarrow \infty} F_n) = \lim_{n \rightarrow \infty} P(F_n).$$

这就证得了  $P$  的下连续性.

再证  $P$  的上连续性. 设  $\{E_n\}$  是减的事件序列, 则  $\{\overline{E_n}\}$  为单调增的事件序列, 由概率的下连续性得

$$\begin{aligned} 1 - \lim_{n \rightarrow \infty} P(E_n) &= \lim_{n \rightarrow \infty} [1 - P(E_n)] = \lim_{n \rightarrow \infty} P(\overline{E_n}) \\ &= P\left(\bigcup_{i=1}^{\infty} \overline{E_n}\right) = P\left(\overline{\bigcap_{i=1}^{\infty} E_n}\right) \\ &= 1 - P\left(\bigcap_{i=1}^{\infty} E_n\right). \end{aligned}$$

故

$$\lim_{n \rightarrow \infty} P(E_n) = P\left(\bigcap_{i=1}^{\infty} E_n\right).$$

从上面的讨论可知, 由可列可加性可以推出有限可加性和下连续性, 但仅由有限可加性并不能推出可列可加性, 于是我们会想, 缺少的条件是否就是下连续性? 答案是肯定的, 下面定理说明了这一性质.

#### 定理 1.10

若  $P$  是  $\mathcal{F}$  上满足  $P(\Omega) = 1$  的非负集合函数, 则它具有可列可加性的充要条件是它是有限可加的且是下连续的.

**证明** 必要性前面已证, 下证充分性.

设  $A_i \in \mathcal{F}$ ,  $i = 1, 2, \dots$  是两两不相容的事件序列, 由有限可加性, 有

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

这个等式的左边不超过 1, 因此正项级数  $\sum_{i=1}^{\infty} P(A_i)$  收敛, 即

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

记

$$F_n = \bigcup_{i=1}^n A_i,$$

则  $\{F_n\}$  为增的事件序列, 所以由下连续性得

$$\lim_{n \rightarrow \infty} P\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} P(F_n) = P\left(\bigcup_{n=1}^{\infty} F_n\right) = P\left(\bigcup_{n=1}^{\infty} A_n\right).$$

综上, 即得可列可加性.

可见, 概率的公理化定义可以将可列可加性改为有限可加性和下连续性.

## 1.4 条件概率

### 1.4.1 条件概率的定义

#### 定义 1.14 (条件概率)

设  $A$  与  $B$  是样本空间  $\Omega$  中的两事件, 若  $P(B) > 0$ , 则称

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

为在  $B$  发生下  $A$  的条件概率, 简称条件概率.



**注** 由定义易见, 所谓条件概率, 是指在  $B$  发生的条件下  $A$  发生的概率.

我们定义了条件概率, 需要说明这是一类概率. 即我们通过验证概率的公理化定义来证明条件概率是概率.

#### 定理 1.11

条件概率是概率, 即若设  $P(B) > 0$ , 则满足

- (1) 非负性:  $P(A|B) \geq 0, A \in \mathcal{F}$ .
- (2) 正则性:  $P(\Omega|B) = 1$ .
- (3) 可列可加性: 若  $\mathcal{F}$  中的  $A_1, A_2, \dots, A_n, \dots$  互不相容, 则

$$P\left(\bigcup_{n=1}^{\infty} A_n|B\right) = \sum_{n=1}^{\infty} P(A_n|B).$$



**证明** 用条件概率的定义易证 (1)(2), 下面来证明 (3).

因为  $A_1, A_2, \dots, A_n, \dots$  互不相容, 所以  $A_1B, A_2B, \dots, A_nB, \dots$  也互不相容, 故

$$\begin{aligned} P\left(\bigcup_{n=1}^{\infty} A_n|B\right) &= \frac{P\left(\left(\bigcup_{n=1}^{\infty} A_n\right)B\right)}{P(B)} = \frac{P\left(\bigcup_{n=1}^{\infty} (A_nB)\right)}{P(B)} \\ &= \sum_{n=1}^{\infty} \frac{P(A_nB)}{P(B)} = \sum_{n=1}^{\infty} P(A_n|B) \end{aligned}$$

故满足可列可加性, 于是条件概率是概率. ■

下面我们给出条件概率特有的三个实用公式: 乘法公式、全概率公式和 Bayes 公式.

### 1.4.2 乘法公式

#### 定理 1.12 (乘法公式)

1. 若  $P(B) > 0$ , 则

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

2. 若  $P(A_1A_2 \cdots A_{n-1}) > 0$ , 则

$$P(A_1A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1A_2) \cdots P(A_n|A_1A_2 \cdots A_{n-1}).$$



**证明** (1) 可由条件概率的定义直接推出, 下证 (2).

因为

$$P(A_1) \geq P(A_1A_2) \geq \cdots \geq P(A_1A_2 \cdots A_{n-1}) > 0,$$

所以式中的条件概率均有意义，且按条件概率的定义，等号右边等于

$$P(A_1) \cdot \frac{P(A_1 A_2)}{P(A_1)} \cdot \frac{P(A_1 A_2 A_3)}{P(A_1 A_2)} \cdots \frac{P(A_1 A_2 \cdots A_n)}{P(A_1 A_2 \cdots A_{n-1})} = P(A_1 A_2 \cdots A_n).$$

从而 (2) 中等式成立. ■

### 1.4.3 全概率公式

#### 定理 1.13 (全概率公式)

设  $B_1, B_2, \dots, B_n$  为样本空间  $\Omega$  的一个分割，即  $B_1, B_2, \dots, B_n$  互不相容，且  $\bigcup_{i=1}^n B_i = \Omega$ ，如果  $P(B_i) > 0, i = 1, 2, \dots, n$ ，则对任一事件  $A$  有

$$P(A) = \sum_{i=1}^n P(B_i)P(A|B_i).$$



**证明** 因为

$$A = A\Omega = A\left(\bigcup_{i=1}^n B_i\right) = \bigcup_{i=1}^n (AB_i),$$

且  $AB_1, AB_2, \dots, AB_n$  互不相容，所以由可加性得

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

再将  $P(AB_i) = P(B_i)P(A|B_i), i = 1, 2, \dots, n$  代入上式即得. ■

**注** 全概率公式的最简单形式：如果  $0 < P(B) < 1$ ，则

$$P(A) = P(B)P(A|B) + P(\bar{B})P(A|\bar{B}).$$

**注** 条件  $B_1, B_2, \dots, B_n$  为样本空间  $\Omega$  的一个分割可改成  $B_1, B_2, \dots, B_n$  互不相容，且  $A \subset \bigcup_{i=1}^n B_i$ ，全概率公式仍然成立.

**注** 对可列个事件  $B_1, B_2, \dots, B_n, \dots$  互不相容，且  $A \subset \bigcup_{i=1}^{\infty} B_i$ ，全概率公式仍然成立，只需将等式右边写成可列项之和.

### 1.4.4 Bayes 公式

在乘法公式和全概率公式的基础上立即可推得下面的公式.

#### 定理 1.14 (Bayes 公式)

设  $B_1, B_2, \dots, B_n$  是样本空间  $\Omega$  的一个分割，如果  $P(A) > 0, P(B_i) > 0, i = 1, 2, \dots, n$ ，则

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_{j=1}^n P(B_j)P(A|B_j)}.$$



**证明** 由条件概率的定义

$$P(B_i|A) = \frac{P(AB_i)}{P(A)}.$$

对上式的分子用乘法公式，分母用全概率公式即得. ■

**注** 在 Bayes 公式中，如果称  $P(B_i)$  为  $B_i$  的先验概率，称  $P(B_i|A)$  为  $B_i$  的后验概率，则 Bayes 公式是专门用于计算后验概率的，也就是通过  $A$  的发生这个新信息，来对  $B_i$  的概率作出的修正.

条件概率的三个公式中，乘法公式是求事件交的概率，全概率公式是求一个复杂事件的概率，Bayes 公式是求一个条件概率.

**例题 1.1 Monty Hall Problem 三门问题** 假设有三个门，标号为 1、2、3. 门后分别藏有一辆车，两只羊，选手选定其中一扇门，主持人打开另外两扇门中其中一扇藏有羊的门，请问选手是否要换另一扇门？

解

## 1.5 独立性

独立性是概率论中又一个重要概念，利用独立性可以简化概率的计算.


### 1.5.1 两个事件的独立性

两个事件之间的独立性是指：一个事件的发生不影响另一个事件的发生. 另外，从概率的角度看，事件  $A$  的条件概率  $P(A|B)$  与无条件概率  $P(A)$  的差别在于：事件  $B$  的发生改变了事件  $A$  发生的概率，如果事件  $A$  与  $B$  是相互独立的，则有  $P(A|B) = P(A)$ ， $P(B|A) = P(B)$ ，它们都等价于

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

另外对  $P(B) = 0$  或  $P(A) = 0$  上式仍然成立，为此我们用上式来定义两个事件相互独立.


#### 定义 1.15 (独立)

如果  $P(AB) = P(A)P(B)$ ，则称事件  $A$  与  $B$  相互独立，简称  $A$  与  $B$  独立. 否则称  $A$  与  $B$  不独立或相依. 

在许多实际问题中，两个事件相互独立大多是根据经验（相互有无影响）来判断的. 但在某些问题中有时也用定义来判断.

注意：独立和互不相容没有必然的联系，请勿混淆.

#### 定理 1.15

概率为零的事件与任意事件相互独立，概率为 1 的事件与任意事件相互独立. 

**证明** 设任意事件  $A$ ，事件  $B, C$ ，且  $P(B) = 0$ ， $P(C) = 1$ ，则

$$0 \leq P(AB) \leq P(B) = 0,$$

故由定义可知  $A$  与  $B$  是独立的.

由加法公式，有


$$P(A \cup C) = P(A) + P(C) - P(AC),$$

而  $P(A \cup C) = P(C)$ ，故有

$$P(AC) = P(A) \times 1 = P(A)P(C),$$

故由定义可知  $A$  与  $C$  是独立的. 

#### 定理 1.16

以下四个命题等价：事件  $A$  与  $B$  独立， $A$  与  $\bar{B}$  独立， $\bar{A}$  与  $B$  独立， $\bar{A}$  与  $\bar{B}$  独立. 

**证明** 由概率的性质知

$$P(A\bar{B}) = P(A) - P(AB).$$

又由  $A$  与  $B$  的独立性知

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

所以

$$P(A\bar{B}) = P(A) - P(A)P(B) = P(A)[1 - P(B)] = P(A)P(\bar{B}).$$

这表明  $A$  与  $\bar{B}$  独立. 类似可证其余结论. ■

**注** 对上述性质的理解是直观的:  $A$  与  $B$  独立, 则  $A$  的发生不影响  $B$  的发生与不发生,  $A$  的不发生也不会影响  $B$  的发生与不发生.

### 1.5.2 多个事件的相互独立性

首先研究三个事件的相互独立性, 我们先给出以下的定义.


#### 定义 1.16

设  $A, B, C$  是三个事件, 如果有

$$\begin{cases} P(AB) = P(A)P(B), \\ P(AC) = P(A)P(C), \\ P(BC) = P(B)P(C), \end{cases} \quad (1.1)$$

则称  $A, B, C$  两两独立. 若还有

$$P(ABC) = P(A)P(B)P(C), \quad (1.2)$$

则称  $A, B, C$  相互独立. 


**注** 由 1.1 式不一定能推出 1.2 式, 同样由 1.2 式不一定能推出 1.1 式.

由此我们可以定义三个以上事件的相互独立性.

#### 定义 1.17 (相互独立)

设有  $n$  个事件  $A_1, A_2, \dots, A_n$ , 对任意的  $1 \leq i < j < k < \dots \leq n$ , 如果以下等式

$$\begin{cases} P(A_i A_j) = P(A_i)P(A_j), \\ P(A_i A_j A_k) = P(A_i)P(A_j)P(A_k), \\ \dots\dots\dots \\ P(A_1 A_2 \dots A_n) = P(A_1)P(A_2) \dots P(A_n), \end{cases} \quad (1.3)$$

均成立, 则称此  $n$  个事件  $A_1, A_2, \dots, A_n$  相互独立. 

**注** 由上述定义可以看出,  $n$  个相互独立的事件中的任意一部分内仍是相互独立的, 而且任意一部分与另一部分也是独立的. 可以证明: 将相互独立事件中的任一部分换为对立事件, 所得的各个事件仍为相互独立的.

#### 命题 1.1

设  $A, B, C$  三事件相互独立, 则  $A \cup B$  与  $C$  独立,  $AB$  与  $C$  独立,  $A - B$  与  $C$  独立. 

**证明** 只证  $A \cup B$  与  $C$  独立, 其余情况类似可证明.

因为

$$\begin{aligned} P((A \cup B)C) &= P(AC \cup BC) = P(AC) + P(BC) - P(ABC) \\ &= P(A)P(C) + P(B)P(C) - P(A)P(B)P(C) \\ &= (P(A) + P(B) - P(A)P(B))P(C) = P(A \cup B)P(C), \end{aligned}$$

所以  $A \cup B$  与  $C$  独立. ■

## 1.6 试验的独立性

利用事件的独立性可以定义两个或更多个试验的独立性.

**定义 1.18 (两个试验的独立性)**

设有两个试验  $E_1$  和  $E_2$ , 假如试验  $E_1$  的任一结果 (事件) 与试验  $E_2$  的任一结果 (事件) 都是相互独立的事件, 则称这两个试验相互独立.



类似地可以定义  $n$  个试验的独立性.

**定义 1.19 ( $n$  个试验的独立性)**

设有  $n$  个试验  $E_1, E_2, \dots, E_n$ , 如果  $E_i$   $i = 1, 2, \dots, n$  的任一结果都是相互独立的事件, 则称试验  $E_1, E_2, \dots, E_n$  相互独立.



**注** 如果这  $n$  个独立试验还是相同的, 则称其为  $n$  重独立重复试验. 如果在  $n$  重独立重复试验中, 每次试验的可能结果为两个:  $A$  或  $\bar{A}$ , 则称这种试验为  $n$  重 Bernouli 试验.



## 第2章 随机变量及其分布

### 2.1 基本概念

#### 2.1.1 随机变量的定义

##### 定义 2.1 (随机变量)

定义在样本空间  $\Omega$  上的实值函数  $X = X(\omega)$  称为随机变量, 常用大写字母  $X, Y, Z$  等表示随机变量, 其取值用小写字母  $x, y, z$  等表示.

假如一个随机变量仅可能取有限个值或可列个值, 则称其为离散随机变量. 假如一个随机变量的可能取值充满数轴上的一个区间  $(a, b)$ , 则称其为连续随机变量, 其中  $a$  可以是  $-\infty$ ,  $b$  可以是  $+\infty$ .



由定义, 随机变量  $X$  是样本点  $\omega$  的一个映射. 这个映射的原像 (定义域) 可以是数, 也可以不是数, 但像一定是实数.

与微积分中的变量不同, 概率论中的随机变量  $X$  是一种“随机取值的变量且伴随一个分布”. 以离散随机变量为例, 我们不仅要知道  $X$  可能的取值, 而且还要知道去这些值的概率各是多少, 这就需要分布的概念. 有没有分布也是区分一般变量和随机变量的主要标志.

#### 2.1.2 分布函数

##### 定义 2.2 (分布函数)

设  $X$  是一个随机变量, 对任意实数  $x$ , 称

$$F(x) = P(X \leq x)$$

为随机变量  $X$  的分布函数. 且称  $X$  服从  $F(x)$ , 记为  $X \sim F(x)$ . 有时也用  $F_X(x)$  以表明是  $X$  的分布函数.



由定义可知, 任一随机变量  $X$  (离散的或连续的) 都有一个分布函数. 有了分布函数, 就可据此计算与随机变量  $X$  有关的事件的概率.

下面给出了分布函数的三个性质.

##### 定理 2.1 (分布函数的基本性质)

设分布函数  $F(x)$ , 则它有基本性质:

(1) 单调性  $F(x)$  是定义在整个实轴  $(-\infty, +\infty)$  上的单调增函数, 即对任意  $x_1 < x_2$ , 有  $F(x_1) \leq f(x_2)$ .

(2) 有界性 对任意  $x$ , 有  $0 \leq F(x) \leq 1$ , 且

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0.$$

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1.$$

(3) 右连续性  $F(x)$  是  $x$  的右连续函数, 即对任意的  $x_0$ , 有

$$\lim_{x \rightarrow x_0 + 0} F(x) = F(x_0).$$

即

$$F(x_0 + 0) = F(x_0)$$



**证明** (1) 是显然的. 下证 (2).

由于  $F(x)$  是事件  $\{X \leq x\}$  的概率, 所以  $0 \leq F(x) \leq 1$ . 由  $F(x)$  的单调性知, 对任意整数  $m$  和  $n$ , 有

$$\lim_{x \rightarrow -\infty} F(x) = \lim_{m \rightarrow -\infty} F(m), \quad \lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow +\infty} F(n)$$

都存在. 又由概率的可列可加性得

$$\begin{aligned} 1 &= P(-\infty < X < +\infty) = P\left(\bigcup_{i=-\infty}^{+\infty} \{i-1 < X \leq i\}\right) \\ &= \sum_{i=-\infty}^{+\infty} P(i-1 < X \leq i) = \lim_{\substack{n \rightarrow +\infty \\ m \rightarrow -\infty}} \sum_{i=m}^n P(i-1 < X \leq i) \\ &= \lim_{n \rightarrow +\infty} F(n) - \lim_{m \rightarrow -\infty} F(m), \end{aligned}$$

由此可得

$$\lim_{x \rightarrow -\infty} F(x) = 0, \quad \lim_{x \rightarrow +\infty} F(x) = 1.$$

再证 (3). 因为  $F(x)$  是单调有界增函数, 所以其任一点  $x_0$  的右极限  $F(x_0+0)$  必存在. 为证右连续性, 只要对任意单调减的数列  $x_1 > x_2 > \cdots > x_n > \cdots > x_0$ , 当  $x_n \rightarrow x_0$  ( $n \rightarrow \infty$ ) 时, 证明  $\lim_{n \rightarrow \infty} F(x_n) = F(x_0)$  成立即可. 因为

$$\begin{aligned} F(x_1) - F(x_0) &= P(x_0 < X \leq x_1) = P\left(\bigcup_{i=1}^{\infty} \{x_{i+1} < X \leq x_i\}\right) \\ &= \sum_{i=1}^{\infty} P(x_{i+1} < X \leq x_i) = \sum_{i=1}^{\infty} [F(x_i) - F(x_{i+1})] \\ &= \lim_{n \rightarrow \infty} [F(x_1) - F(x_n)] = F(x_1) - \lim_{n \rightarrow \infty} F(x_n), \end{aligned}$$

由此得

$$F(x_0) = \lim_{n \rightarrow \infty} F(x_n) = F(x_0+0).$$

至此三条基本性质证毕. ■

我们还可以证明满足上述三条性质的函数是某个随机变量的分布函数. 从而这三条基本性质是某个函数是分布函数的充要条件.

### 例题 2.1 函数

$$F(x) = \frac{1}{\pi} \left( \arctan x + \frac{\pi}{2} \right), \quad -\infty < x < +\infty$$

满足分布函数的三条基本性质, 所以它是一个分布函数, 特别地, 它称为 Cauchy 分布函数.

## 2.1.3 离散随机变量的概率分布列

### 定义 2.3 (概率分布列)

设  $X$  是一个离散随机变量, 如果  $X$  的所有可能取值是  $x_1, x_2, \cdots, x_n, \cdots$ , 则称  $X$  取  $x_i$  的概率

$$p_i = p(x_i) = P(X = x_i), \quad i = 1, 2, \cdots, n, \cdots$$

为  $X$  的概率分布列或分布列, 记为  $X \sim \{P_i\}$ . ♣

分布列也可用如下列表的方式来表示.

$X$	$x_1$	$x_2$	$\cdots$	$x_n$	$\cdots$
$P$	$p(x_1)$	$p(x_2)$	$\cdots$	$p(x_n)$	$\cdots$

**定理 2.2**

分布列有以下基本性质:

- (1) 非负性  $p(x_i \geq 0), i = 1, 2, \dots$   
 (2) 正则性  $\sum_{i=1}^{\infty} p(x_i) = 1$ .



由离散随机变量  $X$  的分布列很容易写出  $X$  的分布函数

$$F(x) = \sum_{x_j \leq x} p(x_j).$$

它的图形是至多可数级的阶梯函数. 不过在离散场合, 常用来描述分布的是分布列, 很少用到分布函数.

特别地, 常量  $c$  可看作仅取一个值的随机变量  $X$ , 即

$$P(X = c) = 1.$$

这个分布常称为**单点分布**或**退化分布**, 它的分布函数是

$$F(x) = \begin{cases} 0, & x < c \\ 1, & x \geq c. \end{cases}$$

在具体求随机变量  $X$  的分布列时, 关键是求出  $X$  的所有可能取值及取这些值的概率.

### 2.1.4 连续随机变量的概率密度函数

**定义 2.4**

设随机变量  $X$  的分布函数为  $F(x)$ , 如果存在实数轴上的一个非负可积函数  $p(x)$ , 使得对任意实数  $x$  有

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

则称  $p(x)$  为  $X$  的**概率密度函数** (probability density function), 简称**密度函数**或**密度**. 同时称  $X$  为**连续随机变量**, 称  $F(x)$  为**连续分布函数**.



在  $F(x)$  的可导点上有

$$F'(x) = p(x).$$

其中  $F(x)$  是(累积) 概率函数, 其导函数  $F'(x)$  是概率密度函数, 这也是  $p(x)$  被称为概率密度函数的理由.

**定理 2.3 (密度函数基本性质)**

- (1) 非负性  $p(x) \geq 0$ .  
 (2) 正则性  $\int_{-\infty}^{+\infty} p(x) dx = 1$ .



## 2.2 数学期望

### 2.2.1 数学期望的定义

下面我们先定义离散随机变量的数学期望.

**定义 2.5 (离散随机变量的数学期望)**

设离散随机变量  $X$  的分布列为

$$p(x_i) = P(X = x_i), i = 1, 2, \dots, n, \dots$$

如果

$$\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty,$$

则称

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

为随机变量  $X$  的数学期望 (expectation), 或称为该分布的数学期望, 简称期望或均值. 若级数  $\sum_{i=1}^{\infty} |x_i| p(x_i)$  不收敛, 则称  $X$  的数学期望不存在.



**注** 以上定义中, 要求级数绝对收敛的目的在于使数学期望唯一. 因为随机变量的取值可正可负, 取值次序可先可后, 由无穷级数理论可知, 如果无穷级数绝对收敛, 则可保证其和不受次序变动的影响.

**注** 由于有限项的和不受次序变动的影响, 故取有限个可能值的随机变量的数学期望总是存在的.

类似地, 我们可以定义连续随机变量的数学期望.

### 定义 2.6 (连续随机变量的数学期望)

设连续随机变量  $X$  的密度函数为  $p(x)$ . 如果

$$\int_{-\infty}^{+\infty} |x| p(x) dx < \infty,$$

则称

$$E(x) = \int_{-\infty}^{+\infty} x p(x) dx$$

为  $X$  的数学期望 (expectation), 或称为该分布  $p(x)$  的数学期望, 简称期望或均值. 若  $\int_{-\infty}^{+\infty} |x| p(x) dx$  不收敛, 则称  $X$  的数学期望不存在.



数学期望的理论意义是深刻的, 它是消除随机性的主要手段.

## 2.2.2 数学期望的性质

下面均假设所涉及的数学期望存在, 给出数学期望的一些性质.

### 定理 2.4

若随机变量  $X$  的分布用分布列  $p(x_i)$  或用密度函数  $p(x)$  表示, 则  $X$  的某一函数  $g(X)$  的数学期望为

$$E[g(X)] = \begin{cases} \sum_i g(x_i) p(x_i), & \text{在离散场合,} \\ \int_{-\infty}^{+\infty} g(x) p(x) dx, & \text{在连续场合.} \end{cases} \quad (2.1)$$



**证明**

### 定理 2.5

若  $c$  是常数, 则  $E(c) = c$ .



**证明** 如果将  $c$  看作仅取一个值的随机变量  $X$ , 则有  $P(X = c) = 1$ , 从而其数学期望  $E(c) = E(X) = c \times 1 = c$ . ■

**定理 2.6**

对任意常数  $a$ , 有

$$E(aX) = aE(X).$$



**证明** 在2.1式中令  $g(x) = ax$ , 然后把  $a$  从求和号或积分号中提出来即得. ■

**定理 2.7**

对任意的两个函数  $g_1(x)$  和  $g_2(x)$ , 有

$$E[g_1(X) \pm g_2(X)] = E[g_1(X)] \pm E[g_2(X)].$$



**证明** 在2.1式中令  $g(x) = g_1(x) \pm g_2(x)$ , 然后把和式分解成两个和式, 或把积分分解成两个积分即得. ■

**2.2.3 Markov 不等式****定理 2.8 (Markov 不等式)**

设非负随机变量  $X$  的数学期望  $E(X)$  存在, 则对任意常数  $a > 0$ , 有

$$P(X \geq a) \leq \frac{E(X)}{a}.$$



**证明** 设连续的非负随机变量  $X$  的密度函数为  $p(x)$ , 数学期望为  $E(X)$ , 则

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xp(x) dx = \int_0^{\infty} xp(x) dx = \int_0^a xp(x) dx + \int_a^{\infty} xp(x) dx \\ &\geq \int_a^{\infty} xp(x) dx \geq \int_a^{\infty} ap(x) dx = a \int_a^{\infty} p(x) dx = aP(X \geq a). \end{aligned}$$

即

$$P(X \geq a) \leq \frac{E(X)}{a}.$$



Markov 不等式的一个实际应用是, 超过  $n$  倍平均工资的人数不会超过总人数的  $\frac{1}{n}$ .

**2.3 方差与标准差****2.3.1 方差与标准差的定义**

方差和标准差是度量随机变量取值波动大小最重要的两个特征数.

**定义 2.7 (方差与标准差的定义)**

若随机变量  $X^2$  的数学期望  $E(X^2)$  存在, 则称偏差平方  $(X - E(X))^2$  的数学期望  $E(X - E(X))^2$  为随机变量  $X$  或相应分布的方差 (variance), 记为

$$\text{Var}(X) = E(X - E(X))^2 = \begin{cases} \sum_i (x_i - E(X))^2 p(x_i), & \text{在离散场合,} \\ \int_{-\infty}^{+\infty} (x - E(X))^2 p(x) dx, & \text{在连续场合.} \end{cases}$$

称方差的正平方根  $\sqrt{\text{Var}(X)}$  为随机变量的标准差 (standard deviation), 记为  $\sigma(X)$  或  $\sigma_X$ .



方差与标准差功能相似, 它们都可以用来描述随机变量取值的集中和分散程度的两个特征数. 方差和标准差越小, 随机变量的取值越集中; 反之, 方差和标准差越大, 随机变量的取值就越分散.

方差和标准差之间的差别主要在量纲上. 由于标准差与数学期望有相同的量纲, 因此  $E(x) \pm k\sigma$  是有意义的 ( $k$  为正实数), 所以在实际中, 人们比较乐意选用标准差, 但标准差的计算必须通过方差才能算得.

### 2.3.2 方差的性质

以下均假定随机变量的方差是存在的.

#### 定理 2.9

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

**证明** 因为

$$\text{Var}(X) = E[X - E(X)]^2 = E(X^2 - 2X \cdot E(X) + [E(X)]^2),$$

由数学期望的性质, 得

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

**注** 在实际计算方差时, 这个性质往往比定义  $\text{Var}(X) = E[X - E(X)]^2$  更常用.

#### 定理 2.10

常数的方差为 0.

**证明** 设  $c$  为常数, 则

$$\text{Var}(c) = E[c - E(c)]^2 = E(c - c)^2 = 0.$$

#### 定理 2.11

若  $a, b$  是常数, 则  $\text{Var}(aX + b) = a^2 \text{Var}(X)$ .

**证明** 因  $a, b$  是常数, 则

$$\text{Var}(aX + b) = E[aX + b - E(aX + b)]^2 = E[a(X - E(X))]^2 = a^2 \text{Var}(X).$$

### 2.3.3 Chebyshev 不等式

前面我们介绍了 Markov 不等式, 实际上它给出的界太过宽松, 为进一步精细化, 我们介绍下面的 Chebyshev 不等式. 事实上, 它是 Markov 不等式的一种特殊情形.

#### 定理 2.12 (Chebyshev 不等式)

设随机变量  $X$  的数学期望和方差都存在, 则对任意常数  $\varepsilon > 0$ , 有

$$P(|X - E(X)| \geq \varepsilon) \leq \frac{\text{Var}(X)}{\varepsilon^2},$$

或

$$P(|X - E(X)| < \varepsilon) \geq 1 - \frac{\text{Var}(X)}{\varepsilon^2}.$$

**证明** 只需证第一个不等式. 设  $X$  是一个连续型随机变量, 其密度函数为  $p(x)$ . 记  $E(X) = a, A = \{x : |x - a| \geq \varepsilon\}$ . 我们有

$$P(|X - a| \geq \varepsilon) = \int_A p(x) dx \leq \int_A \frac{(x - a)^2}{\varepsilon^2} p(x) dx \leq \frac{1}{\varepsilon^2} \int_{-\infty}^{\infty} (x - a)^2 p(x) dx = \frac{\text{Var}(X)}{\varepsilon^2}.$$

对于离散型随机变量亦可类似进行证明. ■

在概率论中, 事件  $\{|X - E(X)| \geq \varepsilon\}$  称为**大偏差**, 其概率  $P(|X - E(X)| \geq \varepsilon)$  称为**大偏差发生概率**. Chebyshev 不等式给出了大偏差发生概率的上界, 这个上界与方差成正比, 方差越大上界也越大.

下面定理进一步说明了方差为 0 就意味着随机变量的取值几乎集中在一点上.

### 定理 2.13

若随机变量  $X$  的方差存在, 则  $\text{Var}(X) = 0$  的充要条件是  $X$  几乎处处为某个常数  $a$ , 即  $P(X = a) = 1$ . ♥

**证明** 充分性显然成立. 下证必要性.

设  $\text{Var}(X) = 0$ , 这时  $E(X)$  存在. 因为

$$\{|X - E(X)| > 0\} = \bigcup_{n=1}^{\infty} \left\{ |X - E(X)| \geq \frac{1}{n} \right\},$$

所以有

$$\begin{aligned} P(|X - E(X)| > 0) &= P\left(\bigcup_{n=1}^{\infty} \left\{ |X - E(X)| \geq \frac{1}{n} \right\}\right) \\ &\leq \sum_{n=1}^{\infty} P\left(|X - E(X)| \geq \frac{1}{n}\right) \\ &\leq \sum_{n=1}^{\infty} \frac{\text{Var}(X)}{(1/n)^2} = 0, \end{aligned}$$

其中最后一个不等式用到了 Chebyshev 不等式. 由此可知

$$P(|X - E(X)| > 0) = 0,$$

因而有

$$P(|X - E(X)| = 0) = 1,$$

即

$$P(X = E(X)) = 1,$$

这就证明了结论, 且其中的常数  $a$  就是  $E(X)$ . ■

**注** 设零测集  $E_0 \subseteq E$ ,  $P$  是关于  $E$  中元素的命题, 若对  $\forall x \in E \setminus E_0$ , 命题  $P$  成立, 那么我们说命题  $P$  在  $E$  上**几乎处处** (almost everywhere) 成立.

## 2.3.4 随机变量的中心化与标准化

下面我们介绍随机变量的中心化和标准化.

### 定义 2.8 (随机变量的中心化)

已知  $X$  是任意的随机变量, 当  $E(x)$  和  $\text{Var}(x)$  存在时, 对  $X$  作变换

$$X_* = X - E(x),$$

这个变换称为**随机变量的中心化**. ♣

由期望和方差的性质推得:

$$E(X_*) = E[X - E(X)] = E(X) - E[E(X)] = 0,$$

$$\text{Var}(X_*) = \text{Var}[X - E(X)] = \text{Var}(X).$$

即中心化后的随机变量, 期望为 0, 方差不变.

中心化的性质解释:



- (1) 期望归零化：中心化随机变量将其中心点（期望点）平移至原点，使其分布不偏左也不偏右，其期望为零。  
 (2) 分布波动不变性：平移不影响波动的分布程度，方差不变。

**定义 2.9 (随机变量的标准化)**

已知  $X$  是任意的随机变量，当  $E(X)$  和  $\text{Var}(X)$  存在且  $\text{Var}(X) \neq 0$  时，对  $X$  作变换

$$X^* = \frac{X - E(x)}{\sqrt{\text{Var}(X)}},$$

这个变换称为随机变量的标准化。



由期望和方差的性质推得：

$$E(X^*) = E\left[\frac{X - E(X)}{\sqrt{\text{Var}(X)}}\right] = \frac{E[X - E(X)]}{E(\text{Var}(x))} = 0,$$

$$\text{Var}(X^*) = \text{Var}\left[\frac{X - E(X)}{\sqrt{\text{Var}(X)}}\right] = \left(\frac{1}{\sqrt{\text{Var}(X)}}\right)^2 \text{Var}(X) = 1.$$

即标准化后的随机变量，期望为 0，方差为 1。

标准化的性质解释：

- (1) 期望归零化：标准化随机变量将其中心点（期望点）平移至原点，使其分布不偏左也不偏右，其期望为 0。  
 (2) 分布波动归一化：标准化将随机变量的取值按照标准差等比压缩，使其分布不疏也不密，压缩改变了分布的波动程度，方差变为 1。

## 2.4 常用离散分布

### 2.4.1 Bernoulli 分布

**定义 2.10 (Bernoulli 分布)**

如果  $X$  只取值 0 或 1，概率分布是

$$P(X = 1) = p, \quad P(X = 0) = q, \quad p + q = 1,$$

就称  $X$  服从 Bernoulli 分布 (Bernoulli Distribution)，或称二点分布或 0-1 分布记作  $X \sim b(1, p)$ 。



任何试验，当只考虑成功与否时，就可以用 Bernoulli 分布的随机变量描述。

### 2.4.2 二项分布

**定义 2.11 (二项分布)**

如果随机变量有如下的概率分布：

$$P(X = k) = C$$

