

# 人工智能中的数学讲义

方聪

北京大学

## 摘要

本讲义收录了人工智能中的数学课程中的主要概念与课程习题。概率与统计讲义内容摘录于陈家鼎、郑忠国《概率与统计》教材与复熹和张原概率与统计课程课件。图论内容摘录于耿素云、屈婉玲、王捍贫《离散数学教程》。本讲义版权归上述作者，不会出版。讲义仅供于上该课程的同学学习参考，讲义的错误会不断修正。感谢张乙沐、张海涵对讲义整理的帮助。

## 1.1 随机事件及其运算

### 1.1.1 随机事件

**样本空间和样本点：**随机实验  $E$  中所有可能结果组成的集合称为  $E$  的**样本空间**，记为  $\Omega$ 。样本空间中的元素称为**样本点**，记为  $\omega$

- $E_1$ ：抛掷硬币，观察正面  $H$ ，反面  $T$  出现的情况。

$$\Omega_1 = \{H, T\}.$$

- $E_2$ ：抛掷一枚硬币 3 次，观察正面  $H$ ，反面  $T$  出现的情况。

$$\Omega_2 = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}.$$

- $E_3$ ：抛掷一枚硬币 3 次，观察正面出现的次数。

$$\Omega_3 = \{0, 1, 2, 3\}.$$

随机现象的某些样本点组成的集合称为**随机事件**，简称为事件，常用  $A, B, C, \dots$  表示

例如， $E$  为抛掷一枚骰子，事件  $A = \text{“出现奇数点”}$ ，即  $A = \{1, 3, 5\}$ ，是样本空间  $\Omega = \{1, 2, 3, 4, 5, 6\}$  的一个子集

**事件的频率：**设  $\mu$  是  $n$  次实验中事件  $A$  发生的次数，则事件  $A$  发生的频率  $\frac{\mu}{n}$ ，随着实验次数  $n$  增大，频率会在某一数值  $p$  附近摆动，称为该事件的**概率**，记为  $P(A) = p$

由于频率  $\frac{\mu}{n}$  总在  $0, 1$  之间，我们有：

$$0 \leq P(A) \leq 1$$

例如投一枚硬币  $n$  次，出现  $\mu$  次正面，则  $\frac{\mu}{n} \xrightarrow{n \rightarrow \infty} p$ 。其中，主观概率  $p$  为事件的置信度，概率是可能性大小的度量。大概率事情易发生，小概率事情不易发生。

#### 1.1.1.1 事件的交和并

**定义 2.1** 设有事件  $A$  和事件  $B$ ，如果  $A$  发生，则  $B$  必发生，那么称事件  $B$  包含事件  $A$ （或称事件  $A$  在  $B$  中），并记为

$$A \subset B \text{ (或 } B \supset A)$$

**定义 2.2** 如果事件  $A$  包含事件  $B$ ，同时事件  $B$  包含事件  $A$ ，则事件  $A$  和事件  $B$  相等，并记为

$$A = B$$

**定义 2.3** 设  $A$  和  $B$  都是事件，则“ $A$  或  $B$ ”表示这样的事件  $C$ ： $C$  发生当且仅当  $A$  或  $B$  中至少有一个发生，该事件  $C$  叫做  $A$  与  $B$  的并，记为  $A \cup B$ 。

**例 2.1**（对应郑书例 2.1）在桌面上，投掷两枚匀称的硬币， $A$  表示“恰好一枚国旗朝上”， $B$  表示“两枚国旗朝上”， $C$  表示“至少一枚国旗朝上”，则  $C = A \cup B$ 。

对于并运算，有以下性质，我们恒记必然事件为  $U$ ，不可能事件为  $V$ ：

$$\begin{aligned} A \cup B &= B \cup A \\ A \cup U &= U, \quad A \cup V = V \end{aligned}$$

**定义 2.4** 设  $A$  和  $B$  都是事件，则“ $A$  且  $B$ ”表示这样的事件  $C$ ： $C$  发生当且仅当  $A$  和  $B$  都发生，该事件  $C$  叫做  $A$  与  $B$  的交，记为  $A \cap B$ ，也简记为  $AB$ 。

在例 2.1 中， $A \cap C = A$ ， $B \cap C = C$ ， $A \cap B = A$

对于交运算，有以下性质：

$$\begin{aligned} A \cap B &= B \cap A \\ A \cap U &= A, \quad A \cap V = V \end{aligned}$$

#### 1.1.1.2 事件的余和差

**定义 2.5** 设  $A$  是事件，称“非  $A$ ”是  $A$  的对立事件（或称余是事件），其含义为，“非  $A$ ”发生当且仅当  $A$  不发生，常常用  $\bar{A}$  表示“非  $A$ ”，也用  $A^c$  表示“非  $A$ ”。

由定义知  $\overline{\bar{A}} = A$ ， $\bar{U} = V$ ， $\bar{V} = U$

**定义 2.6** 设  $A$  和  $B$  都是事件，则两个事件的差“ $A$  减去  $B$ ”表示这样的事件  $C$ ： $C$  发生当且仅当  $A$  发生而  $B$  不发生，该事件  $C$  记为  $A - B$ （或  $A \setminus B$ ）

由定义知， $A - B = A \cap \bar{B}$

画图法确定关系。

### 1.1.1.3 事件运算的性质

事件的基本运算还有以下性质：

- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$  “并”的结合律
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  “交”的结合律
- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$  分配律
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$  分配律
- $A \cup A = A, A \cap A = A$
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$  对偶律
- $\overline{A \cap B} = \overline{A} \cup \overline{B}$  对偶律

多个事件的交和并：

设  $A_1, A_2, \dots, A_n$  是  $n$  个事件，则“ $A_1, A_2, \dots, A_n$ ”的并是指这样的事件：它发生当且仅当  $A_1, A_2, \dots, A_n$  中至少一个发生，常常用  $\bigcup_{i=1}^n A_i$  表示  $A_1, A_2, \dots, A_n$  的并

设  $A_1, A_2, \dots, A_n$  是  $n$  个事件，则“ $A_1, A_2, \dots, A_n$ ”的交是指这样的事件：它发生当且仅当  $A_1, A_2, \dots, A_n$  这  $n$  个事件都发生，常常用  $\bigcap_{i=1}^n A_i$  表示  $A_1, A_2, \dots, A_n$  的交，也用  $A_1 A_2 \dots A_n$  表示这个“交”

实际应用中，还需定义无穷多事件的并与交

设  $A_1, A_2, \dots, A_i, \dots$  是一列事件，则  $B$  是指这样的事件： $B$  发生当且仅当这些  $A_i (i = 1, 2, \dots)$  中至少一个发生，这个  $B$  叫做诸  $A_i$  的并，记为  $\bigcup_{i=1}^{\infty} A_i$ ，有时也写为  $A_1 \cup A_2 \cup \dots$

设  $A_1, A_2, \dots, A_i, \dots$  是一列事件，则  $C$  是指这样的事件： $C$  发生当且仅当这些  $A_i (i = 1, 2, \dots)$  都发生，这个  $C$  叫做诸  $A_i$  的交，记为  $\bigcap_{i=1}^{\infty} A_i$ ，有时也写为  $A_1 A_2 \dots$

**例：**取  $X \in \mathbb{R}$ ，事件  $A_i$  为  $X \in [\frac{1}{i+1}, \frac{1}{i}]$ ，事件  $B_i$  为  $X \in [0, \frac{1}{i}]$ 。则事件  $\bigcup_{i=1}^n A_i$  发生等价于  $X \in [\frac{1}{n+1}, 1]$ ，事件  $\bigcap_{i=1}^n B_i$  发生等价于  $X \in [0, \frac{1}{n}]$ 。进而当  $n \rightarrow \infty$  时事件  $\bigcup_{i=1}^{\infty} A_i$  发生等价于  $X \in (0, 1]$ ，事件  $\bigcap_{i=1}^{\infty} B_i$  发生等价于  $X = 0$ 。

并的更一般定义是，设  $\{A_a, a \in \Gamma\}$  是一族事件（其中  $\Gamma$  是任何非空集，每个  $a \in \Gamma$  对应一个事件  $A_a$ ），这些事件  $A_a$  的“并”是指这样的事件  $B$ ： $B$  发生当且仅当至少一个  $A_a$  发生，这个  $B$  常常记为  $\bigcup_{a \in \Gamma} A_a$ ，类似可以定义一族事件的交  $\bigcap_{a \in \Gamma} A_a$

**例 2.3:** (对应郑书例 2.3) 一射手向一个目标连续射击, 设  $A_1 =$  “第一次射击, 命中”,  $A_i =$  “前  $i-1$  次射击都未命中, 第  $i$  次射击命中” ( $i = 2, 3, \dots$ ),  $B =$  “终于命中”, 则  $B = \bigcup_{i=1}^{\infty} A_i$

**例 2.4:** (对应郑书例 2.4) 一射手向一个目标连续射击, 设  $A_i =$  “第  $i$  次射击, 未命中目标” ( $i = 2, 3, \dots$ ) 则  $\bigcap_{i=1}^{\infty} A_i =$  “每次均未命中目标”

不难验证, 对可列个事件的并和交有以下规律:

- $A \cup \left( \bigcap_{i=1}^{\infty} B_i \right) = \bigcap_{i=1}^{\infty} (A \cup B_i)$  分配律
- $A \cap \left( \bigcup_{i=1}^{\infty} B_i \right) = \bigcup_{i=1}^{\infty} (A \cap B_i)$  分配律
- $\overline{\left( \bigcup_{i=1}^{\infty} A_i \right)} = \bigcap_{i=1}^{\infty} \overline{A_i}$  对偶律
- $\overline{\left( \bigcap_{i=1}^{\infty} A_i \right)} = \bigcup_{i=1}^{\infty} \overline{A_i}$  对偶律

#### 1.1.1.4 互斥事件

##### 互不相容的事件

如果事件  $A$  和事件  $B$  不能都发生, 即  $A \cap B = \emptyset$ , 则称  $A$  和  $B$  是互不相容的事件 (也称互斥的事件)

称事件  $A_1, \dots, A_n$  互不相容, 若对任何  $i \neq j (i, j = 1, \dots, n)$ ,  $A_i$  与  $A_j$  互不相容

例如, 抛掷两枚硬币, 事件 “恰好一枚国徽朝上” 和事件 “两枚都是国徽朝上” 是互不相容的。不难看出, 对任何事件  $A$ ,  $A$  和  $\overline{A}$  是互不相容的

- 加法公式:  $A_1, A_2, \dots$  互不相容, 则:

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

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

## 1.2 概率的公理化定义

**概率空间子类:** 设  $\Omega$  为样本空间,  $\mathcal{F}$  为  $\Omega$  的一些子集构成的集类。若  $\mathcal{F}$  满足以下三个条件: (1)  $\Omega \in \mathcal{F}$ , (2)  $A \in \mathcal{F} \Rightarrow A^c \in \mathcal{F}$ , (3)  $\{A_n\}_{n \in \mathbb{N}} \subseteq \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$ , 则称  $\mathcal{F}$  为概率空间子类

例:

- $\mathcal{F} = \{\emptyset, \Omega\}$  平凡概率空间子类
- $\mathcal{F} = \{\emptyset, \Omega, A, \bar{A}\}$  包含  $A$  的最小概率空间子类
- $\mathcal{F} = \{A | A \subset \Omega\}$   $\Omega$  上的最大概率空间子类
- $\Omega = \{\omega_1, \dots, \omega_n\}$ , 则  $\Omega$  所有子集构成的概率空间子类共有  $2^n$  个元素

**定义:** 设  $\mathcal{F}$  是满足上述条件的概率空间子集类。概率  $P = P(\cdot)$  是  $\mathcal{F}$  上面定义的实值函数, 满足:

- 非负性:  $P(A) \geq 0$  对于一切  $A \in \mathcal{F}$
- 规范性:  $P(\Omega) = 1$
- 可列可加性: 若  $A_n \in \mathcal{F} (n = 1, 2, \dots)$  两两不相交, 则

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

$(\Omega, \mathcal{F}, P)$  为概率空间

**例 1:** 假定  $\Omega = \{\omega_1, \dots, \omega_n\}$ ,  $\mathcal{F}$  为全体子集构成的概率空间子类。

设  $p_1, \dots, p_n$  为  $n$  个非负实数, 且满足  $\sum_{i=1}^n p_i = 1$ 。令

$$\mathbb{P}(\emptyset) = 0, \quad \mathbb{P}(A) = \sum_{j=1}^k p_{i_j}, \quad A = \{\omega_{i_1}, \dots, \omega_{i_k}\}, k = 1, \dots, n$$

则  $\mathbb{P}$  为  $(\Omega, \mathcal{F})$  上概率。

概率  $P$  有以下性质:

- (1)  $P(\emptyset) = 0$ ;
- (2) 若  $A \in \mathcal{F}$ , 则  $P(A^c) = 1 - P(A)$ ;
- (3) 若  $A_1, \dots, A_n$  都属于  $\mathcal{F}$  且两两不相交, 则

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) \quad (1.2.1)$$

(4) 若  $A \subset B$ ,  $A \in \mathcal{F}$ ,  $B \in \mathcal{F}$ , 则  $P(A) \leq P(B)$ , 且

$$P(B - A) = P(B) - P(A) \quad (1.2.2)$$

(5) 若  $A_n \subset A_{n+1}$ ,  $A_n \in \mathcal{F} (n = 1, 2, \dots)$ , 则

$$P(\bigcup_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \quad (1.2.3)$$

(6) 若  $A_n \supset A_{n+1}$ ,  $A_n \in \mathcal{F} (n = 1, 2, \dots)$ , 则

$$P(\bigcap_{n=1}^{\infty} A_n) = \lim_{n \rightarrow \infty} P(A_n) \quad (1.2.4)$$

(7) 若  $A_n \in \mathcal{F} (n = 1, 2, \dots)$ , 则

$$P(\bigcup_{n=1}^{\infty} A_n) \leq \sum_{n=1}^{\infty} P(A_n) \quad (1.2.5)$$

## 2.1 古典概型

**模型定义：**若随机现象有如下两个特征：

(1) 在实验中它的全部可能性只有有限个；

(2) 基本事件发生或出现是等可能的；

则称其对应的数学模型为古典概型

取

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}, \quad \mathcal{F} = \{A | A \subset \Omega\}$$

令  $P$  为  $(\Omega, \mathcal{F})$  上的概率测度，满足

$$P(\{\omega_1\}) = \dots = P(\{\omega_n\})$$

则  $(\Omega, \mathcal{F}, P)$  为古典概型对应的概率空间。

**计算公式：**对  $A = \{\omega_{i_1}, \dots, \omega_{i_k}\} \in \mathcal{F}$ ，利用概率的有限可加性可知：

$$P(A) = \sum_{j=1}^k P(\{\omega_{i_j}\}) = \frac{k}{n} = \frac{|A|}{|\Omega|}$$

**排列：**从含有  $n$  个不同元素的总体中抽取  $r$  个进行排列

(1) 放回情形：共有  $n^r$  种排列方式

(2) 不放回情形：共有  $A_n^r := n(n-1)\cdots(n-r+1)$  种排列方式

当  $r = n$  时，为全排列，此时  $A_n^n = n!$ 。

**组合：**(1) 从  $n$  个不同元素中取出  $r$  个而不考虑其顺序，称为组合，其总数为  $C_n^r = \frac{n!}{r!(n-r)!} = \frac{A_n^r}{r!}$

(2) 把  $n$  个不同元素分成  $k$  个部分，且第  $i$  个部分有  $r_i$  个元素， $1 \leq i \leq k$ ，且  $r_1 + r_2 + \dots + r_k = n$ ，则有  $\frac{n!}{r_1! r_2! \cdots r_k!}$  种方法

(3) 把  $n$  个元素全部带有标注，其中  $n_1$  个带标注 1， $n_2$  个带标注 2， $\dots$ ， $n_k$  个带标注  $k$ 。现在从此  $n$  个元素中取出  $r$  个，使得带有标注  $i$  的元素有  $r_i$  个，其中  $1 \leq i \leq k$  且  $r_1 + r_2 + \dots + r_k = r$ 。则不同取法的总数为  $C_{n_1}^{r_1} C_{n_2}^{r_2} \cdots C_{n_k}^{r_k}$ 。

(4) 从  $n$  个不同元素中有重复的取出  $r$  个，不计顺序，则不同的取法有  $C_{n+r-1}^r$ （有重复组合数）

**组合公式：**对一切正整数  $a, b$ ,

$$\sum_{i=0}^n C_a^i C_b^{n-i} = C_{a+b}^n$$



约定当  $k > n$  时,  $C_n^k = 0$ 。特别地,

$$\sum_{i=0}^n (C_n^i)^2 = C_{2n}^n$$

**例 1:** (对应郑书例 3.1) 某人同时抛掷两枚骰子, 问: 得到 7 点 (两颗骰子的点数之和的概率是多少?)

**解:** 我们用甲乙分别表示这两颗骰子, 每颗骰子共有 6 种可能的点数: 1, 2, 3, 4, 5, 6, 两颗骰子共有  $6 \times 6 = 36$  种可能结果:  $(i, j) (i = 1, \dots, 6) (j = 1, \dots, 6)$ , 这里  $i$  表示骰子甲的点数,  $j$  表示骰子乙的点数, 显然这些结果出现的机会是相等的, 它们构成了等概完备事件组, 事件“得到 7 点”由 6 种结果 (基本事件) 组成:  $(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)$ , 故事件“得到 7 点”的概率为  $\frac{6}{36} = \frac{1}{6}$   
□

**例 2:** 甲口袋有 5 个白球, 3 个黑球, 乙口袋中有 4 个白球, 6 个黑球, 从两个口袋中各任取一球, 求取到的两个球颜色相同的概率。

**解:** 从两个口袋中各取一球, 共有  $C_8^1 C_{10}^1$  种可能取法。两球颜色相同可能情况为: 从甲乙口袋均取出白球, 从甲乙口袋均取出黑球, 共有  $C_5^1 C_4^1 + C_3^1 C_6^1$  种取法, 于是

$$P(\text{取到的两个球颜色相同}) = \frac{C_5^1 C_4^1 + C_3^1 C_6^1}{C_8^1 C_{10}^1} = \frac{19}{40}$$

□

**例 3:** (巴拿赫问题) 某数学家有两盒火柴, 每盒有  $n$  根, 每次使用时, 他任取一盒并从中抽出一根, 问他发现一盒空而同时另一盒还有  $r (0 \leq r \leq n)$  的概率为多少 (发现为空表示最后一次抽到空盒)?

**解:** 设两盒火柴分别为  $A, B$ , 由对称性, 所求概率为事件  $E = \text{“发现 } A \text{ 盒空而 } B \text{ 盒还有 } r \text{ 根”}$  的概率的 2 倍。

先计算样本空间中的样本点个数, 由于共取了  $2n - r + 1$  次, 故有  $2^{2n-r+1}$  个样本点。

考察事件  $E$ , 等效为前  $2n - r$  次  $A$  盒恰好取  $n$  次, 次序不论, 最后一次必定取到  $A$  盒, 此种样本点共有  $C_{2n-r}^n$  个, 因此

$$P(E) = \frac{C_{2n-r}^n}{2^{2n-r+1}}.$$

所求概率为  $\frac{C_{2n-r}^n}{2^{2n-r}}.$

## 2.2 条件概率与独立性

### 2.2.1 条件概率

**条件概率：**设  $(\Omega, \mathcal{F}, P)$  为概率空间， $B \in \mathcal{F}$  满足  $P(B) > 0$ 。称

$$P(A|B) = \frac{P(AB)}{P(B)}, A \in \mathcal{F}$$

为  $B$  发生条件下  $A$  发生的条件概率。

条件概率  $P(\cdot|B)$  为  $\mathcal{F}$  上的概率，即满足：

- $P(A|B) \geq 0, \forall A \in \mathcal{F}$
- $P(\Omega|B) = 1$
- $\forall \{A_n\} \subset \mathcal{F}, A_n \cap A_m = \emptyset, \forall n \neq m,$

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

容易得到， $P(B|\Omega) = P(B)$ 。

**乘法公式：** $P(AB) = P(B|A)P(A)$

**乘法公式的推广：** $P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1})$ ，其中  $P(A_1 A_2 \cdots A_{n-1}) > 0$ 。

**例 1：**将 52 张扑克牌（不含大王、小王）随机地分为 4 堆，每堆 13 张，问：各堆都含有 A 牌（即 1 点）的概率是多少？

**解：**将 4 堆扑克牌编号：第 1 堆，第 2 堆，第 3 堆，第 4 堆，用  $A_1, A_2, A_3, A_4$  依次表示 4 个 A 牌，设  $i_1, i_2, i_3, i_4$  是 1, 2, 3, 4 的一个排列，令  $E_{i_1 i_2 i_3 i_4} =$  “第  $i_1$  堆有  $A_1$  但没有  $A_2, A_3, A_4$ ，第  $i_2$  堆有  $A_2$  但没有  $A_1, A_3, A_4$ ，第  $i_3$  堆有  $A_3$  但没有  $A_1, A_2, A_4$ ，第  $i_4$  堆有  $A_4$  但没有  $A_1, A_2, A_3$ ”， $E =$  “各堆都含有 A”，则

$$E = \bigcup_{i_1 i_2 i_3 i_4} E_{i_1 i_2 i_3 i_4}$$

这些事件两两不相容，易知  $P(E) = 4!P(E_{1234})$ ，令  $E_k = \{ \text{第 } k \text{ 堆含有 } A_k \text{ 但不含有其他的 } A_j (j \neq k) \} \quad (k = 1, 2, 3, 4)$ ，则

$$P(E_{1234}) = P(E_1)P(E_2|E_1)P(E_3|E_1 E_2)P(E_4|E_1 E_2 E_3)$$

易知

$$P(E_1) = C_{48}^{12}/C_{52}^{13}, \quad P(E_2|E_1) = C_{36}^{12}/C_{39}^{13},$$

$$P(E_3|E_1E_2) = C_{24}^{12}/C_{26}^{13}, \quad P(E_4|E_1E_2E_3) = 1,$$

于是

$$P(E_{1234}) = \frac{C_{48}^{12}C_{36}^{12}C_{24}^{12}}{C_{52}^{13}C_{39}^{13}C_{26}^{13}} = \frac{13^4}{52 \times 51 \times 50 \times 49},$$

$$P(E) = 4!P(E_{1234}) \approx 0.105$$

□

**例 2:** (罐子模型) 设罐中有  $b$  个黑球,  $r$  个红球, 每次随机取出一个球, 取出后将原球放回, 还加进  $c$  个同色球和  $d$  个异色球, 记  $B_i$  为“第  $i$  次取出的是黑球”,  $R_j$  为“第  $j$  次取出的是红球”. 若连续从罐中取出三个球, 其中有两个红球, 一个黑球, 则由乘法公式我们可得

$$P(B_1R_2R_3) = P(B_1)P(R_2|B_1)P(R_3|B_1R_2) = \frac{b}{b+r} \cdot \frac{r+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d},$$

$$P(R_1B_2R_3) = P(R_1)P(B_2|R_1)P(R_3|R_1B_2) = \frac{r}{b+r} \cdot \frac{b+d}{b+r+c+d} \cdot \frac{r+d+c}{b+r+2c+2d},$$

$$P(R_1R_2B_3) = P(R_1)P(R_2|R_1)P(B_3|R_1R_2) = \frac{r}{b+r} \cdot \frac{r+c}{b+r+c+d} \cdot \frac{b+2d}{b+r+2c+2d},$$

以上概率与黑球在第几次被抽出有关. 罐子模型也称波利亚 (Polya) 模型, 这个模型的各种变化如下:

(1) 当  $c = -1, d = 0$  时, 为不返回抽样, 此时前次抽取结果会影响后次抽取结果, 但只要抽取的黑球和红球个数确定, 则概率不依赖其抽出球的次序, 有

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br(r-1)}{(b+r)(b+r-1)(b+r-2)}$$

(2) 当  $c = 0, d = 0$  时, 为返回抽样, 此时前次抽取结果不会影响后次抽取结果, 上述三种概率相等, 有

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br^2}{(b+r)^3}$$

(3) 当  $c > 0, d = 0$  时, 为传染病模型, 此时每次取出球后会增加下一次取到同色球的概率, 或者说, 每发现一个传染病患者, 以后都会增加再传染的概率. 同样的, 上述三种概率相等, 且都等于

$$P(B_1R_2R_3) = P(R_1B_2R_3) = P(R_1R_2B_3) = \frac{br(r+c)}{(b+r)(b+r+c)(b+r+2c)}$$

可以看出, 当  $d = 0$  时, 只要取出的黑球和红球个数确定, 则概率不依赖于其抽出球的顺序.

(4) 当  $c = 0, d > 0$  时, 为安全模型, 可以解释为, 每当事故发生, 会抓紧安全工作, 从而下一次发生事故的的概率会减少, 而当事故未发生时, 安全工作会松懈, 下一次发生事故的的概率会增大, 上述三种概率分别为:

$$P(B_1R_2R_3) = \frac{b}{(b+r)} \cdot \frac{r+d}{b+r+d} \cdot \frac{r+d}{b+r+2d},$$

$$P(R_1 B_2 R_3) = \frac{r}{(b+r)} \cdot \frac{b+d}{b+r+d} \cdot \frac{r+d}{b+r+2d},$$

$$P(R_1 R_2 B_3) = \frac{r}{(b+r)} \cdot \frac{r}{b+r+d} \cdot \frac{b+2d}{b+r+2d}$$

**例：**设  $n$  件产品中有  $m$  件不合格品，从中任取两件，已知两件中有一件是合格品，求另一件也是合格品的概率。

**解：**记事件  $A$  “有一件是合格品”， $B$  “另一件也是合格品”。则

$P(A) = P(\text{取出一件合格品，一件不合格品}) + P(\text{取出两件都是合格品})$

$$= \frac{C_m^1 C_{n-m}^1}{C_n^2} + \frac{C_{n-m}^2}{C_n^2} = \frac{2m(n-m) + (n-m)(n-m-1)}{n(n-1)}$$

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

$$P(AB) = P(\text{取出两件都是合格品}) = \frac{C_{n-m}^2}{C_n^2} = \frac{(n-m)(n-m-1)}{n(n-1)}$$

于是所求概率为

$$P(B|A) = \frac{P(AB)}{P(A)} = \frac{\frac{(n-m)(n-m-1)}{n(n-1)}}{\frac{(n-m)(n+m-1)}{n(n-1)}} = \frac{n-m-1}{n+m-1}$$

□

### 2.2.2 事件的独立性

**事件的独立性：**设  $(\Omega, \mathcal{F}, P)$  为概率空间，称  $A, B \in \mathcal{F}$  相互独立（独立），若

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

**性质：**(1) 若  $A, B$  独立，且  $P(B) > 0$ ，则

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

即条件概率等于无条件概率。

(2) 若  $A, B$  独立，则  $A$  与  $\bar{B}$ ， $\bar{A}$  与  $B$ ， $\bar{A}$  与  $\bar{B}$  亦独立。

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

(3) 零概率事件及其对立的事件与任意的事件都独立。

**例：**袋中有  $a$  只黑球和  $b$  只白球，令  $A$ ：“第一次摸到黑球”， $B$ ：“第二次摸到黑球”。讨论  $A$  和  $B$  的独立性。

(1) 放回情形。因为

$$P(A) = \frac{a}{a+b}, P(AB) = \frac{a^2}{(a+b)^2}, P(\overline{A}B) = \frac{ab}{(a+b)^2},$$

所以

$$P(B) = P(AB) + P(\overline{A}B) = \frac{a^2 + ab}{(a+b)^2} = \frac{a}{a+b}$$

故

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

(2) 不放回情形。易知

$$P(A) = P(B) = \frac{a}{a+b}, P(\overline{A}B) = \frac{ab}{(a+b)(a+b-1)},$$

故

$$P(A)P(B) \neq P(AB)$$

**定义：** 设  $\{A_k\}_{k \leq n} \subset \mathcal{F}$ 。称  $A_1, A_2, \dots, A_n$  相互独立，若

$$P\left(\bigcap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}), \quad 1 \leq i_1 < i_2 < \dots < i_k \leq n, k \leq n$$

注意：独立  $\Rightarrow$  两两独立，但是反之不对：

伯恩斯坦反例：一个均匀的正四面体，其第一、二、三面分别涂上红、黄、蓝三种颜色第四面同时涂上以上三种颜色。以  $A, B, C$  分别表示投一次四面体出现红、黄、蓝颜色朝下的事件，则

$$P(A) = P(B) = P(C) = \frac{1}{2}, \quad P(AB) = P(BC) = P(AC) = \frac{1}{4}$$

从而  $A, B, C$  两两独立，但是，

$$P(ABC) = \frac{1}{4} \neq P(A)P(B)P(C)$$

**独立性与概率计算：** 设  $A_1, A_2, \dots, A_n$  相互独立，则

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = 1 - \prod_{i=1}^n P(\overline{A}_i)$$

**例：** 设有某型号的高射炮，每门炮（发射一发）击中敌机的概率为 0.6，现在若干门炮同时发射（每炮射一发），问：若要以 99% 的把握击中来犯的一架敌机，至少需要配置几门高射炮？

**解：** 设  $n$  是需要配置的高射炮的门数，记  $A_i =$  “第  $i$  门炮击中敌机” ( $i = 1, \dots, n$ )， $A =$  “敌机被击中”。由于  $A = \bigcup_{i=1}^n A_i$ ，于是找到  $n$ ，使得

$$P(A) = P\left(\bigcup_{i=1}^n A_i\right) \geq 0.99$$

由于  $P(A) = 1 - P(\bar{A}) = 1 - P(\bigcup_{i=1}^n \bar{A}_i)$ , 且  $\bar{A}_1, \dots, \bar{A}_n$  相互独立, 故

$$P(A) = 1 - P(\bar{A}_1) \cdots P(\bar{A}_n) = 1 - 0.4^n$$

为使不等式成立, 必须且只需  $1 - 0.4^n \geq 0.99$ . 由此得

$$n \geq \lg 0.01 / \lg 0.4 = 5.026$$

故至少需配置 6 门高射炮方能以 99% 的把握击中敌机。

**例:** 设  $A, B, C$  三事件相互独立, 证明  $A - B$  与  $C$  独立。

**解:** 因为

$$\begin{aligned} P((A - B)C) &= P(AC - BC) = P(AC) - P(ABC) \\ &= P(A)P(C) - P(A)P(B)P(C) \\ &= (P(A) - P(A)P(B))P(C) \\ &= (P(A) - P(AB))P(C) = P(A - B)P(C). \end{aligned}$$

所以  $A - B$  与  $C$  独立。 □

## 2.3 全概率公式和贝叶斯公式

### 2.3.1 全概率公式

**完备事件组:** 若  $\{B_n\}_{n \geq 1} \subset \mathcal{F}$  满足两两互斥且  $\sum_{n=1}^{\infty} B_n = \Omega$ , 则称  $\{B_n\}_{n \geq 1}$  为完备事件组。

**全概率公式:** 假定  $\{B_n\}_{n \geq 1}$  为完备事件组, 则

$$P(A) = \sum_{n=1}^{\infty} P(B_n)P(A|B_n), \forall A \in \mathcal{F}$$

注意: 在上式中, 若  $P(B_n) = 0$ , 则规定  $P(B_n)P(A|B_n) = 0$ 。

**例:** 一保险公司相信人群可以分为 2 类: 一类是容易出事故的; 另一类是不容易出事故的。已知前者在一年内出事故的概率为 0.4, 后者在一年内出事故的概率为 0.2。前者约占人群的 30%。今有一人前来投保, 他在一年内出事故的可能性有多大?

**解:** 设  $A =$  “他在一年内出事故”,  $B =$  “他是容易出事故的”, 则  $B, \bar{B}$  构成完备事件组, 有

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



图 2.1: 完备事件组

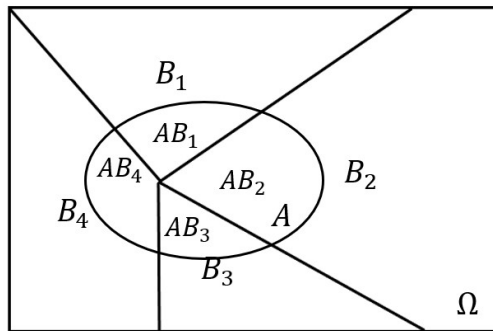


图 2.2: 全概率公式

由于  $P(B) = 0.3, P(A|B) = 0.4, P(\bar{B}) = 0.7, P(A|\bar{B}) = 0.2$ , 于是

$$P(A) = 0.3 \times 0.4 + 0.7 \times 0.2 = 0.26$$

□

**例:** 甲口袋有 1 个黑球, 2 个白球, 乙口袋有 3 个白球, 每次从两口袋中任取一球, 交换后放入另一口袋中, 求交换  $n$  次之后, 黑球仍然在甲口袋的概率。

设事件  $A_i$  为 “第  $i$  次交换后黑球仍然在甲口袋中”, 记  $p_i = P(A_i), i = 0, 1, 2, \dots$ , 则有  $p_0 = 1$ , 且

$$P(A_{i+1} | A_i) = \frac{2}{3}, \quad P(A_{i+1} | A_i^c) = \frac{1}{3}$$

由全概率公式得

$$p_n = \frac{2}{3}p_{n-1} + \frac{1}{3}(1 - p_{n-1}) = \frac{1}{3}p_{n-1} + \frac{1}{3}, \quad n \geq 1$$

得到递推公式

$$p_n - \frac{1}{2} = \left(\frac{1}{3}\right) \left(p_{n-1} - \frac{1}{2}\right), \quad n \geq 1$$

将  $p_0 = 1$  代入上式可得

$$p_n - \frac{1}{2} = \left(\frac{1}{3}\right)^n \left(\frac{1}{2}\right)$$

因此

$$p_n = \frac{1}{2} \left[ 1 + \left(\frac{1}{3}\right)^n \right]$$

□

### 2.3.2 贝叶斯公式

**贝叶斯公式：**假定  $\{B_n\}_{n \geq 1}$  为完备事件组， $A \in \mathcal{F}$  满足  $P(A) > 0$ ，则

$$P(B_n|A) = \frac{P(B_n)P(A|B_n)}{\sum_{n=1}^{\infty} P(B_n)P(A|B_n)}$$

**例：**一项血液化验有 95% 的把握诊断某种疾病，但这项化验用于健康人也会有 1% 的“假阳性”结果（即如果一个健康人接受这项化验，化验结果误诊此病人患该疾病的概率为 1%）。假定该疾病的患者事实上只占总人口的 0.5%。若某人化验结果为阳性，则此人确实患有该疾病的概率是多少？

**解：**令  $A$  表示“此人确实患该疾病”， $B$  表示“其化验结果为阳性”，则所求概率为

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} \\ &= \frac{0.95 \times 0.005}{0.95 \times 0.005 + 0.01 \times 0.995} \\ &= \frac{95}{294} \approx 0.323 \end{aligned}$$

□ **例：**一架飞机失踪了，推测它等可能的坠落在 3 个区域。令  $\alpha_i (i = 1, 2, 3)$  表示飞机在第  $i$  个区域坠落但没有被发现的概率。已知对区域 1 的搜索没有发现飞机，求在此条件下，飞机坠落在第  $i (i = 1, 2, 3)$  个区域的条件概率。

**解：**令  $B_i$  表示“飞机坠毁在第  $i$  个区域”， $i = 1, 2, 3$ ， $A$  表示“在第 1 个区域没有搜索到飞机”，则

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{\sum_{i=1}^3 P(B_i)P(A|B_i)} = \frac{\frac{\alpha_1}{3}}{\frac{\alpha_1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{\alpha_1}{\alpha_1 + 2}$$

对  $j = 2, 3$ ,

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^3 P(B_i)P(A|B_i)} = \frac{\frac{1}{3}}{\frac{\alpha_1}{3} + 1 \times \frac{1}{3} + 1 \times \frac{1}{3}} = \frac{1}{\alpha_1 + 2}$$

□

□



**随机游走：**考虑数轴上一质点，假定它只在整数点上运动。当前时刻它处于位置  $a$ （整数），下一时刻（单位间隔时间）以概率  $p$  向正向，概率  $1-p$  向负向运动一个单位，称这样的质点运动为随机游动，当  $p = q = \frac{1}{2}$  时，称为对称随机游走。

(1) 无限制随机游走：对随机游走，以  $S_n$  表示  $n$  时刻质点的位置，假定  $S_0 = 0$ 。我们计算经过  $n$  次运动后到达位置  $k$  的概率。

由于质点在  $n$  时刻位于  $k$ ，在  $n$  次游动中，质点向右移动次数  $x$  比向左运动  $y$  多  $k$  次：

$$\begin{aligned} x - y &= k, & x + y &= n \\ x &= \frac{n+k}{2}, & y &= \frac{n-k}{2} \end{aligned}$$

为使  $x$  为整数， $k$  和  $n$  的奇偶性需要相同，即

$$P(S_n = k) = \begin{cases} C_n^{\frac{n+k}{2}} p^{\frac{n+k}{2}} (1-p)^{\frac{n-k}{2}}, & n, k \text{ 奇偶性相同} \\ 0, & n, k \text{ 奇偶性不同} \end{cases}$$

(2) 两端带有吸收壁的随机游走：设  $a, b$  为正整数。假定质点初始位置为  $a$ ，在位置 0 和  $a+b$  均有一个吸收壁，求质点被吸收的概率。

记  $q_n$  为质点初始位置是  $n$  而最终在  $a+b$  被吸收的概率，显然，

$$q_0 = 0, \quad q_{a+b} = 1$$

若质点某时刻位于  $n$ ， $n = 1, \dots, a+b-1$ 。则其在位置  $a+b$  被吸收有两种可能：(1) 运动到  $n-1$  位置被  $a+b$  吸收，(2) 运动到  $n+1$  位置被  $a+b$  吸收，由全概率公式得

$$q_n = q_{n-1}q + q_{n+1}p, \quad n = 1, \dots, a+b-1$$

由于  $p+q=1$ ，上式可以写为

$$p(q_{n+1} - q_n) = q(q_n - q_{n-1}), \quad n = 1, \dots, a+b-1$$

记  $r = \frac{q}{p}$ ，则

$$q_{n+1} - q_n = r(q_n - q_{n-1}), \quad n = 1, \dots, a+b-1$$

可以分两种情况讨论：(i) 若  $r = 1$ ，即  $p = q = \frac{1}{2}$ 。则

$$q_{n+1} - q_n = q_n - q_{n-1} = \dots = q_1 - q_0$$

$$q_{n+1} = q_0 + (n+1)(q_1 - q_0), \quad n = 1, \dots, a+b-1$$

结合边值条件, 有

$$q_n = \frac{n}{a+b}, n = 1, \cdots, a+b-1$$

(ii) 若  $r \neq 1$ , 即  $p \neq q$ :

$$q_{n+1} - q_n = r(q_n - q_{n-1}) = \cdots = r^n(q_1 - q_0)$$

即

$$q_n - q_0 = \sum_{i=0}^{n-1} (q_{i+1} - q_i) = \sum_{i=0}^{n-1} r^i (q_1 - q_0) = \frac{1-r^n}{1-r} (q_1 - q_0), \quad n = 1, \cdots, a+b-1$$

结合边值条件, 得

$$q_1 = \frac{1-r}{1-r^{a+b}}$$

则

$$q_n = \frac{1-r^n}{1-r^{a+b}}$$

### 3.1 随机变量

为了进一步研究随机现象，我们需要引入随机变量的概念。

**定义：**（随机变量的直观描述）如果条件  $S$  下的结果可以用某个变量  $X$  来描述， $X$  的值不能预先确定，而随着条件  $S$  的不同可能变化，但是对任何实数  $c$ ，事件“ $X$  取值不超过  $c$ ”是有概率的，将这样一种变量  $X$  称为随机变量。

**定义：**（随机变量的数学描述）如果条件  $S$  下的所有可能结果组成了集合  $\Omega = \{\omega\}$ ， $X = X(\omega)$  是在  $\Omega$  上有定义的实值函数，而且对任何实数  $c$ ，事件“ $\{\omega : X(\omega) \leq c\}$ ”是有概率的，将  $X$  称为随机变量。

**例：**（对应郑书例 1.2）盒中有 5 个球，其中有 2 个白球，3 个黑球。从中任取 3 个球，将其中所含的白球的数记为  $X$ 。

建模：将球编号，1~3 表示黑球，4,5 表示白球。

记摸到球的编号为  $\omega = (i, j, k)$ ，其中  $1 \leq i < j < k \leq 5$ 。  $|\Omega| = C_5^3 = 10$ 。

其中满足  $X = 0$  的  $\omega$  有  $C_2^0 C_3^3 = 1$  个；满足  $X = 1$  的  $\omega$  有  $C_2^1 C_3^2 = 6$  个；满足  $X = 2$  的  $\omega$  有  $C_2^2 C_3^1 = 3$  个。

设事件：  $\{X = 1\} = \{\omega : X(\omega) = 1\}$ ，  $\{X \leq 1\} = \{\omega : X(\omega) \leq 1\}$ 。

将  $P(\{X = 1\})$  简记为  $P(X = 1)$ 。

$$P(X = 1) = \frac{6}{10}, \quad P(X \leq 1) = \frac{7}{10}.$$

**例：**（对应郑书例 1.6）某公共汽车站每隔 10 min 会有一两某路公交车到达。某乘客随机在任意时刻到达车站。

显然，他的候车时间  $X$ （单位：min）为随机变量。 $X$  的取值范围  $0 \leq X \leq 10$ 。事件  $\{X \leq c\}$  是有概率的，这是一种几何概型，我们会在后面给出计算过程，例如：

$$P(X \leq 3) = \frac{3}{10}, \quad P(2 \leq X \leq 6) = \frac{4}{10}.$$

### 3.2 离散型随机变量

**定义：**  $X$  是离散型随机变量指：  $X$  取有限个值  $x_1, \dots, x_n$ , 或可列无穷个值  $x_1, x_2, \dots$ .  $X$  的概率分布 (列) 指:

$$p_k = P(X = x_k), \quad k = 1, \dots, n \text{ 或 } k = 1, 2, \dots.$$

将  $X$  的可能值以及相应的概率列为表3.1。表3.1称为  $X$  的**概率分布表**，它能够清楚完整的表示  $X$

$X$	$x_1$	$x_2$	$\cdots$	$x_k$	$\cdots$
$p$	$p_1$	$p_2$	$\cdots$	$p_k$	$\cdots$

表 3.1: 概率分布表

的取值以及概率的分布情况。

**定义：** 设  $X$  的可能取值是  $x_1, x_2, \dots$  (有限个或者可列无穷个)，则称

$$p_k = P(X = x_k) \quad (k = 1, 2, \dots)$$

为  $X$  的**概率分布**，这时也称为  $X$  的**概率函数**或者**概率分布律**

关于  $\{p_k\}$ ，有以下性质：

$$(1) p_k \geq 0 \quad (k = 1, 2, \dots) \quad (2) \sum_k p_k = 1$$

回忆本讲例 1 的  $X$  (抽到的白球数) 它的概率分布表如表3.2所示：

$X$	0	1	2
$p$	0.1	0.6	0.3

表 3.2:  $X$  的概率分布表

对离散型随机变量，有以下几种常见的概率分布：

### 3.2.1 两点分布 (伯努利分布)

定义随机变量  $X$  的可能值是 0 和 1 且概率分布为:

$$P(X = 1) = p, \quad P(X = 0) = 1 - p.$$

称  $X$  服从**两点分布** (也称伯努利分布), 记为  $X \sim B(1, p)$  (参数  $0 \leq p \leq 1$ )

我们定义示性函数  $1_A$ : 事件  $A$  发生则取 1;  $A$  不发生则取 0.

**例:** (对应郑书例 2.1) 100 件产品中有 3 件次品. 从中任取一件.

设事件  $A =$  “取到合格品”, 随机变量  $X = 1_A$ ,  $X$  的可能取值为 0 和 1. 取到每件产品的概率均等, 概率分布为

$$P(X = 1) = \frac{97}{100}, P(X = 0) = \frac{3}{100}$$

$X$  服从参数  $p = 0.97$  的两点分布。

### 3.2.2 二项分布

设随机变量所有可能值为  $0, 1, \dots, n$ , 且

$$P(X = k) = C_n^k p^k (1 - p)^{n-k}, k = 0, 1, \dots, n$$

称  $X$  服从参数为  $n, p$  的二项分布, 记作  $X \sim B(n, p)$  (参数  $n \geq 1, 0 \leq p \leq 1$ )

二项分布有明显的实际背景, 例如在单次实验中事件  $A$  发生的概率是  $p$ , 进行独立重复实验  $n$  次, 记事件  $A$  发生的次数为  $X$ , 则  $X \sim B(n, p)$ 。

**定理 2.1:** 对于二项分布, 分布列  $P(X = k)$  的最大值点  $k_0$  如下:

若  $(n + 1)p \notin \mathbb{Z}$ , 则  $k_0 = [(n + 1)p]$ ;

若  $(n + 1)p \in \mathbb{Z}$ , 则  $k_0 = (n + 1)p$  或  $(n + 1)p - 1$ .

证明: 显然

$$\frac{p_n(k+1)}{p_n(k)} = \frac{n-k}{k+1} \cdot \frac{p}{1-p}$$

由于  $\frac{n-k}{k+1} \cdot \frac{p}{1-p} > 1$  等价于  $k < (n + 1)p - 1$ , 于是有:

(a) 当  $k < (n + 1)p - 1$  时,  $p_n(k + 1) > p_n(k)$

(b) 当  $k > (n + 1)p - 1$  时,  $p_n(k + 1) < p_n(k)$

(c) 当  $k = (n + 1)p - 1$  时,  $p_n(k + 1) = p_n(k)$

(i) 若  $(n+1)p \notin \mathbb{Z}$ , 设  $k_0 = [(n+1)p] < (n+1)p < k_0 + 1$ , 当  $k < m$  时,  $k \leq k_0 - 1 < (n+1)p - 1$ , 因此  $p_n(k) < p_n(k+1)$ ; 当  $k \geq k_0$  时,  $k > (n+1)p - 1$ , 因此  $p_n(k) > p_n(k+1)$ , 所以  $k_0$  为最大值。

(ii) 若  $(n+1)p \in \mathbb{Z}$ , 设  $k_0 = (n+1)p$ , 有  $p_n(k_0) = p_n(k_0 + 1)$ , 进而利用性质 (a) 和性质 (b) 知  $k_0$  为最大值。□

### 3.2.3 泊松分布

**定义:** 设随机变量  $X$  的所有可能取值是全体非负整数, 且

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

则称  $X$  服从参数为  $\lambda$  的泊松分布, 记为  $X \sim \mathcal{P}(\lambda)$  (参数:  $\lambda > 0$ )。

泊松分布常见于生物学, 物理学, 工业的应用中, 例如电话交换台收到的电话呼唤次数, 放射性物质在一定时间内放出的粒子数。

**定理:** 泊松分布的分布列最大值点  $k_0 = [\lambda]$ 。

证明: 注意到  $p_{k+1} = \frac{\lambda}{k+1} p_k$ , 故由分布函数知

若  $k+1 \leq \lambda$ , 则  $p_{k+1} \geq p_k$

若  $k+1 \geq \lambda$ , 则  $p_{k+1} \leq p_k$

因此当  $k_0 = [\lambda]$  时, 分布列取最大值。□

**例:** 已知某商场一天来的顾客服从参数为  $\lambda$  的泊松分布, 而每个来商场的顾客购物概率为  $p$ , 证明此商场一天内购物的顾客数服从参数为  $\lambda p$  的泊松分布。

**解:** 用  $Y$  表示商场内一天购物的顾客数, 则由全概率公式知, 对任意正整数  $k$  有

$$\begin{aligned} P(Y = k) &= \sum_{i=k}^{\infty} P(X = i) P(Y = k | X = i) = \sum_{i=k}^{\infty} \frac{\lambda^i e^{-\lambda}}{i!} C_i^k p^k (1-p)^{i-k} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda} \sum_{i=k}^{\infty} \frac{[\lambda(1-p)]^{i-k}}{(i-k)!} = \frac{(\lambda p)^k}{k!} e^{-\lambda} e^{\lambda(1-p)} = \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned}$$

□

### 3.2.4 超几何分布

**定义：**若随机变量  $X$  的概率分布满足：

$$P(X = k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}, \quad k = 0, 1, \dots, n.$$

则称  $X$  服从超几何分布，记为  $X \sim H(N, D, n)$ （参数  $N, D, n$  满足  $N \geq D \geq 0$ ）

设一批产品有  $N$  个产品， $D$  个次品，任取  $n$  个，抽到的次品数为  $X$ 。如果进行放回抽样则  $X$  服从二项分布，如果进行不放回抽样则  $X$  服从超几何分布。

**定理 2.3：**给定  $n$ 。当  $N \rightarrow \infty, \frac{D}{N} \rightarrow p$  时，

$$\lim_{N \rightarrow \infty} \frac{C_D^k C_{N-D}^{n-k}}{C_N^n} = C_n^k p^k (1-p)^{n-k}, \quad k \geq 0$$

证明：由于  $0 < p < 1$ ，当  $N$  充分大时， $n < D < N$ ，且  $n$  是固定的，易知

$$\begin{aligned} \frac{C_D^k C_{N-D}^{n-k}}{C_N^n} &= \frac{D!}{k!(D-k)!} \cdot \frac{(N-D)!}{(n-k)!(N-D-n+k)!} \cdot \frac{n!(N-n)!}{N!} \\ &= \frac{n!}{k!(n-k)!} \cdot \frac{D(D-1) \cdots (D-k+1)}{N^k} \\ &\quad \cdot \frac{(N-D)(N-D-1) \cdots (N-D-n+k+1)}{N^{n-k}} \\ &\quad \cdot \frac{N^n}{N(N-1) \cdots (N-n+1)} \\ &= C_n^k \left( \prod_{i=1}^k \frac{D-i+1}{N} \right) \left( \prod_{i=1}^{n-k} \frac{N-D-i+1}{N} \right) \left( \prod_{i=1}^n \frac{N}{N-i+1} \right) \\ &\rightarrow C_n^k p^k (1-p)^{n-k} \quad (N \rightarrow \infty) \end{aligned}$$

□

该定理的直观解释是，如果一批产品的总量  $N$  很大，其中次品占比为  $p$ ，则从整批产品随机抽取  $n$  个，抽到次品的个数  $k$  近似服从参数为  $p, n$  的二项分布。

### 3.2.5 几何分布

**定义：**若随机变量  $X$  的所有可能值是全体整数，且概率分布满足：

$$P(X = k) = (1-p)^{k-1} p, \quad k = 1, 2, \dots$$

则称  $X$  服从几何分布，记为  $X \sim G(p)$ ，参数  $0 < p < 1$ 。

例如，某个射手向目标连续射击，如果他单次射中目标的概率为  $p$ ，则他首次射中目标所需要的射击次数  $X$  是一个随机变量，且满足几何分布。

几何分布具备无记忆性:  $P(X - n = k \mid X > n) = P(X = k)$ .

**例：**设  $X$  是只取自然数的离散随机变量，若  $X$  的分布具有无记忆性，证明  $X$  的分布一定为几何分布。

**证明：**由无记忆性知

$$P(X > n + m \mid X > m) = \frac{P(X > n + m)}{P(X > m)} = P(X > n),$$

将  $n$  换为  $n - 1$  仍有

$$P(X > n + m - 1) = P(X > n - 1)P(X > m).$$

两式相减有

$$P(X = n + m) = P(X = n)P(X > m).$$

设  $P(X = 1) = p$ ，若取  $n = m = 1$  有

$$P(X = 2) = p(1 - p).$$

若取  $n = 2, m = 1$  则有

$$P(X = 3) = P(X = 2)P(X > 1) = p(1 - p)^2.$$

若令  $P(X = k) = p(1 - p)^{k-1}$ ，则用数学归纳法得

$$P(X = k + 1) = P(X = k)P(X > 1) = p(1 - p)^k, \quad k = 0, 1, \dots.$$

这表明  $X$  的分布为几何分布。 □

### 3.2.6 离散均匀分布

**定义：**若随机变量  $X$  的概率分布满足：

$$P(X = k) = \frac{1}{N}, \quad k = 1, \dots, N.$$

则称  $X$  服从离散均匀分布。



### 3.3 连续随机变量

**定义：**连续型随机变量指：存在  $p(x)$  使得

$$P(a \leq X \leq b) = \int_a^b p(x)dx, \quad \forall a < b.$$

称  $p(\cdot)$  为  $X$  的概率密度 (函数), 也记为  $p_X(\cdot)$ .

连续随机变量有以下性质：

- (1) 非负:  $p(x) \geq 0$
- (2) 规范:  $\int_{-\infty}^{\infty} p(x)dx = 1$
- (3)  $P(X = x) = 0$  在任意一点选中的概率都为 0.
- (4)  $p(\cdot)$  在  $x$  连续, 即  $P(X \in [x, x + \Delta x]) = p(x)\Delta x + o(\Delta x)$ ,

以下是常见的连续随机变量：

#### 3.3.1 均匀分布

**定义：**如果随机变量  $X$  的分布密度为：

$$p(x) = \begin{cases} \frac{1}{b-a}; & \text{若 } a \leq x \leq b; \\ 0, & \text{其他.} \end{cases}$$

则称  $X$  服从区间  $[a, b]$  (或  $(a, b)$ ) 上的均匀分布, 记为  $X \sim U(a, b)$  (参数  $a < b$ ) :

均匀分布的分布函数也可以写为  $p(x) = \frac{1}{b-a} 1_{\{a \leq x \leq b\}}$ .

例如, 某公共汽车站每隔 10 分钟会有一班公交车到达, 一位搭乘该车的乘客在任意时刻到达车站是等可能的, 则他的候车时间  $X$  是一个随机变量, 且满足  $[0, 10]$  上的均匀分布。

#### 3.3.2 指数分布

**定义：**如果随机变量  $X$  的分布密度为：

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0.$$

则称  $X$  服从参数为  $\lambda$  的指数分布, 记为  $X \sim \text{Exp}(\lambda)$  (参数  $\lambda > 0$ )

若  $X$  服从参数为  $\lambda$  的指数分布, 则对任何  $0 \leq a < b$  有:

$$P(a < X < b) = \lambda \int_a^b e^{-\lambda x} dx = e^{-\lambda a} - e^{-\lambda b}$$

$$P(X > a) = e^{-\lambda a}$$

**定理:** (无记忆性):  $P(X - s > t \mid X > s) = e^{-\lambda t}, \forall t, s \geq 0$ .

不难看出,  $P(X - s > t \mid X > s) = \frac{P(X-s>t)}{P(X>s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = P(X > t)$

注意到, 无记忆性是指数分布独有的, 即设  $X$  是非负的随机变量,  $P(X-s > t \mid X > s) = P(X > s)$  对  $\forall t, s \geq 0$  恒成立的充分必要条件是  $X$  服从指数分布。

证明: 之前已经证明了充分性, 现只需证明必要性: 设  $X$  是非负随机变量满足  $P(X - s > t \mid X > s) = e^{-\lambda t}$ , 则

$$P(X > s) > 0, \quad P(X > s + t) = P(X > s)P(X > t)$$

令  $f(u) = P(X > u)$ , 则  $f(s + t) = f(s)f(t)$

于是  $f(1) = f(\frac{1}{n} \times n) = (f(\frac{1}{n}))^n$

从而  $f(\frac{m}{n}) = f(\frac{1}{n} \times m) = (f(\frac{1}{n}))^m = (f(1))^{\frac{m}{n}}$

故对任意正有理数  $r$ , 有  $f(r) = (f(1))^r$ 。由于  $0 < f(1) < 1$  且  $f(u)$  是关于  $u$  的减函数, 因此对任意  $u \geq 0$ , 有  $f(u) = (f(1))^u$ 。

令  $\lambda = -\ln f(1)$ , 则  $f(u) = e^{-\lambda u}$ , 即

$$P(X > u) = e^{-\lambda u} = \int_{+\infty}^u e^{-\lambda x} dx$$

$$P(a < X < b) = \int_a^b \lambda e^{-\lambda x} dx \quad (0 \leq a < b)$$

说明  $X$  服从指数分布。 □

### 3.3.3 正态分布

**定义:** 如果随机变量  $X$  的分布密度为:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

则称  $X$  服从参数为  $\mu, \sigma$  的正态分布, 记为  $X \sim N(\mu, \sigma^2)$  (参数  $\mu \in \mathbb{R}, \sigma > 0$ )

参数  $\mu = 0$ ,  $\sigma^2 = 1$  时的正态分布称为标准正态分布  $N(0, 1)$ , 分布密度是:

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

**归一性:**  $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$ :

设  $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ , 将积分的平方写为二重积分:

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{2\pi} \iint_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy.$$

做极坐标变换:

$$x = r \cos \theta, y = r \sin \theta \Rightarrow \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial y}{\partial r} \\ \frac{\partial x}{\partial \theta} & \frac{\partial y}{\partial \theta} \end{vmatrix} = r.$$

因此, 二重积分可以写为

$$\frac{1}{2\pi} \int_0^{2\pi} \left( \int_0^{\infty} e^{-\frac{r^2}{2}} r dr \right) d\theta = \int_0^{\infty} e^{-R} dR = 1$$

对于其他正态分布的密度函数  $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ : 令  $y = \frac{x-\mu}{\sigma}$ , 则

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1.$$

定义函数  $\Phi$ :

$$\Phi(x) = \int_{-\infty}^x \phi(x) dx.$$

容易看出  $\Phi(-x) = 1 - \Phi(x)$ .

**定理:** 令  $x^* = \frac{x-\mu}{\sigma}$ , 则

$$P(a < X < b) = \int_a^b \frac{1}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) dx = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right).$$

**推论:** 设随机变量  $X \sim N(\mu, \sigma^2)$ , 则对一切正数  $k$ , 有

$$P(\mu - k\sigma < X < \mu + k\sigma) = \Phi(k) - \Phi(-k) = 2\Phi(k) - 1$$

例如查表得  $\Phi(3) = 0.9987$ , 因此

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = \Phi(3) - \Phi(-3) = 0.9974$$

该结果说明正态随机变量  $X$  的取值基本落在区间  $(\mu - 3\sigma, \mu + 3\sigma)$  内。

### 3.3.4 伽马分布

**定义：**如果随机变量  $X$  的分布密度为：

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0.$$

则称随机变量  $X$  服从伽马分布，记为  $X \sim \Gamma(\alpha, \beta)$  (参数  $\alpha, \beta > 0$ )

其中，称  $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$  为  $\Gamma$  函数。

若  $\Gamma(\alpha)$  为  $\Gamma$  函数，则函数具备以下性质：

$$(1) \Gamma(\alpha + 1) = \alpha \Gamma(\alpha)$$

证明：

$$\int_0^\infty y^\alpha e^{-y} dy = -y^\alpha e^{-y} \Big|_0^\infty + \int_0^\infty \alpha y^{\alpha-1} e^{-y} dy = \alpha \int_0^\infty y^{\alpha-1} e^{-y} dy$$

□

$$(2) \Gamma(1) = 1; \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$$

证明：

$$\Gamma\left(\frac{1}{2}\right) = \int_0^\infty \frac{1}{\sqrt{y}} e^{-y} dy = \sqrt{2} \int_0^\infty e^{-\frac{x^2}{2}} dx = \sqrt{\pi}.$$

□

$$(3) \alpha = 1 \text{ 时就是指数分布参数为 } \beta.$$

### 3.4 随机变量的严格定义

**定义：**假设  $(\Omega, \mathcal{F}, P)$  是概率空间， $X: \Omega \rightarrow \mathbb{R}$  满足：

$$\text{对任意 } x \in \mathbb{R} \text{ 都有 } \{X \leq x\} \in \mathcal{F},$$

则称  $X$  是一个随机变量。

**定义：**令  $F(x) = P(X \leq x), x \in \mathbb{R}$ . 称  $F$  为随机变量  $X$  的分布函数，也记为  $F_X$ .

**定理：**分布函数  $F = F_X$  的三条性质：

- (1) 单调性：若  $x \leq y$ , 则  $F(x) \leq F(y)$ .
- (2) 规范性： $\lim_{x \rightarrow -\infty} F(x) = 0; \lim_{x \rightarrow \infty} F(x) = 1$ .
- (3) 右连续性： $\lim_{y \rightarrow x+} F(y) = F(x)$ .

- 离散型：  $P(X = x_i) = p_i$ .  $x_i$  为  $F_X$  的跳点,  $p_i$  为跳跃幅度.
- 连续型：  $F_X(x) = \int_{-\infty}^x p(z)dz$ , 且

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

反过来, 若  $F_X$  “几乎” 连续可导, 则为连续型 (定理 4.3, 4.4).

- 尾分布函数：  $G(x) = P(X > x) = 1 - F(x)$ .

连续型：  $p(x) = -G'(x)$ .

- 例.  $X \sim \text{Exp}(\lambda)$ .

$$\begin{aligned} G(x) &= e^{-\lambda x}, \quad \forall x > 0, \\ \Rightarrow G'(x) &= -\lambda G(x). \quad \lambda: \text{速率}. \end{aligned}$$

- 由  $F_X(x)$  可求出  $P(X \in B), \forall B$ .
- 若  $F_X = F_Y$ , 则称  $X$  与  $Y$  同分布, 记为  $X \stackrel{d}{=} Y$ .
- $X = Y$ , 即  $P(X = Y) = 1$ , 则  $F_X = F_Y$ . 反之不然.

## 4.1 随机变量的函数

**随机变量的函数：**设  $y = g(x)$  是定义在  $\mathbb{R}$  上的一个函数， $X$  是一个随机变量，那么  $Y = g(X)$  作为  $X$  的函数，同样也是一个随机变量。

在实际问题中，如果已知随机变量  $X$  的分布，我们可以求出另一个随机变量  $Y = g(X)$  的分布。我们将从离散和连续两种场合分别讨论随机变量函数的分布。

注：为了让  $Y$  是数学意义上严格定义的随机变量，必须对函数  $f(x)$  有所假定才能使得  $\{Y \leq c\}$  是有概率的事件，通常假定  $f(x)$  是 Borel 函数，即对于任何实数  $c$ ， $\{x : f(x) \leq c\}$  是 Borel 集，有以下定理：

**定理：**设  $X = X(\omega)$  是概率空间  $(\Omega, \mathcal{F}, P)$  上的随机变量，则对任何 Borel 函数  $f(x)$ ， $Y = f(X(\omega))$  也是这个概率空间上的随机变量。

**证明：**给定任意实数  $c$ ，令

$$B = \{x : f(x) \leq c\}$$

则  $\{\omega : Y \leq c\} = \{\omega : f(X(\omega)) \leq c\} = \{\omega : X(\omega) \in B\}$ ，由于  $B$  是 Borel 集，则由定理 [?] 知  $\{\omega : X(\omega) \in B\} \in \mathcal{F}$ ，所以  $\{Y \leq c\} \in \mathcal{F}$ ， $Y$  是随机变量。

我们遇到的随机函数一般都是 Borel 函数，所以  $Y = X(\omega)$  一般都是随机变量。

### 4.1.1 离散随机变量函数的分布

设  $X$  是离散随机变量， $X$  的分布列为：则  $Y = g(X)$  也是一个离散随机变量，此时  $Y$  的分布列可

$X$	$x_1$	$x_2$	$\cdots$	$x_k$	$\cdots$
$p$	$p_{x_1}$	$p_{x_2}$	$\cdots$	$p_{x_k}$	$\cdots$

以简单表示为：若  $p_{x_1}, p_{x_2}, \cdots, p_{x_k}, \cdots$  中有某些值相等时，把那些相等的值分别合并，并将对应概

$Y$	$g(x_1)$	$g(x_2)$	$\cdots$	$g(x_k)$	$\cdots$
$p$	$p_{x_1}$	$p_{x_2}$	$\cdots$	$p_{x_k}$	$\cdots$

率相加。

**例：**已知随机变量  $X$  的分布如下，求  $Y = X^2 + X$  的分布列。

$X$	-2	-1	0	1	2
$p$	0.2	0.1	0.1	0.3	0.3

解:  $Y = X^2 + X$  的分布列为

$Y$	2	0	0	2	6
$p$	0.2	0.1	0.1	0.3	0.3

合并得到

$Y$	0	2	6
$p$	0.2	0.5	0.3

**定理:** (离散卷积公式) 若  $\xi, \eta$  是相互独立的随机变量, 且取非负整数值, 分布列分别为  $\{k; a_k\}$  和  $\{k; b_k\}$ 。则随机变量  $\zeta = \xi + \eta$  的分布列为  $P(\zeta = k) = \sum_{i=0}^k a_i b_{k-i}$ , 称为**卷积公式**。

证明: 注意到  $P(\zeta = k) = P(\xi = 0, \eta = k) + P(\xi = 1, \eta = k-1) + \cdots + P(\xi = k, \eta = 0)$ 。其中  $= P(\xi = i, \eta = k-i) = a_i b_{k-i}$ , 因此  $P(\zeta = k) = \sum_{i=0}^k a_i b_{k-i}$ 。  $\square$

**例:** (泊松分布可加性) 设  $X \sim \mathcal{P}(\lambda_1), Y \sim \mathcal{P}(\lambda_2)$ , 且  $X, Y$  相互独立, 证明  $X + Y \sim \mathcal{P}(\lambda_1 + \lambda_2)$ 。

解: 泊松分布函数  $P(X = k) = \frac{\lambda_1^k}{k!} e^{-\lambda_1}$ ,  $P(Y = k) = \frac{\lambda_2^k}{k!} e^{-\lambda_2}$ , 由卷积公式,

$$P(X + Y = k) = \sum_{i=0}^k P(X = i) P(Y = k-i) = \sum_{i=0}^k \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2}$$

由二项式展开, 上式整理为

$$\frac{e^{-(\lambda_1 + \lambda_2)}}{k!} \sum_{i=0}^k \frac{\lambda_1^i \lambda_2^{k-i} k!}{i!(k-i)!} = \frac{e^{-(\lambda_1 + \lambda_2)}}{k!} (\lambda_1 + \lambda_2)^k$$

$\square$

#### 4.1.2 连续随机变量函数的分布

对于连续随机变量, 一般先求分布函数, 如果能写出分布密度就写出分布密度。

**例:** 设  $X \sim N(\mu, \sigma^2)$ , 试求  $Y = \frac{1}{\sigma}(X - \mu)$  的概率分布。

解: 对任何实数  $y$ , 由于  $\{Y \leq y\} = \{X \leq \sigma y + \mu\}$ , 于是

$$P(Y \leq y) = P(X \leq \sigma y + \mu)$$

$$= \int_{-\infty}^{\mu+\sigma y} \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} dx$$

变量替换  $\frac{x-\mu}{\sigma} = t$  得

$$P(Y \leq y) = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

说明  $Y \sim N(0, 1)$  □

**定理:** 设随机变量  $X$  有分布密度  $p(x)$ , 且在区间  $(a, b)$  ( $-\infty \leq a < b \leq +\infty$ ) 上满足  $P(a < X < b) = 1$ 。又  $Y = f(X)$ , 其中  $f(x)$  是  $(a, b)$  上严格单调的连续函数,  $g(y)$  是  $f(x)$  的反函数, 且  $g'(y)$  处处存在, 令

$$q(y) = \begin{cases} p(g(y))|g'(y)|, & y \in (\alpha, \beta), \\ 0, & \text{其他} \end{cases}$$

其中  $(\alpha, \beta)$  是反函数  $g(y)$  的存在区间, 即  $\alpha = \min\{A, B\}$ ,  $\beta = \max\{A, B\}$ ,  $A \triangleq \lim_{x \rightarrow a+} f(x)$ ,  $B \triangleq \lim_{x \rightarrow b-} f(x)$ , 则  $q(y)$  是  $Y$  的分布密度。

**证明:** 设  $f(x)$  是严格增函数 (当  $f(x)$  是严格减函数时, 可以类似的证明)。那么对于  $u \in (\alpha, \beta)$  有

$$\begin{aligned} P(Y \leq u) &= P(f(X) \leq u) = P(X \leq g(u)) \\ &= \int_{-\infty}^{g(u)} p(x) dx = \int_a^{g(u)} p(x) dx \end{aligned}$$

做变量替换  $x = g(y)$ , 则

$$P(Y \leq u) = \int_a^u p(g(y)) |g'(y)| dy = \int_{-\infty}^u q(y) dy$$

当  $u \leq \alpha$  时,

$$P(Y \leq u) = P(X \leq a) = 0 = \int_{-\infty}^u q(y) dy$$

当  $u \geq \beta$  时,

$$\begin{aligned} P(Y \leq u) &= P(X \leq b) = 1 = \int_a^b p(x) dx \\ &= \int_{\alpha}^{\beta} p(g(y)) |g'(y)| dy = \int_{-\infty}^u q(y) dy \end{aligned}$$

综上, 对于一切实数  $u$ , 有  $P(Y \leq u) = \int_{-\infty}^u q(y) dy$ , 故  $q(y)$  是  $Y = f(X)$  的密度函数。 □

**例:** (对应郑书例 5.3) 研究水箱内某种微生物的增长情况。设在时 0 微生物的总数是  $v$  ( $v > 0$ ), 增长率是  $X$ , 在时刻  $t$  微生物总数是  $Y = ve^{Xt}$  ( $t > 0$ )。若  $X$  有分布密度

$$p(x) = \begin{cases} 3(1-x)^2, & 0 < x < 1, \\ 0, & \text{其他}, \end{cases}$$



试求  $Y$  的概率分布。

**解：**反函数的求解需要注意函数和区间的变化。

令  $f(x) = ve^{Xt} (0 < x < 1)$ ，则其反函数为：

$$g(y) = \frac{1}{t} \ln \frac{y}{v} \quad (v < y < ve^t)$$

易知  $g'(y) = \frac{1}{ty}$ ，根据定理知， $Y = ve^{xt}$  的分布密度是：

$$q(y) = \begin{cases} 3(1 - \frac{1}{t} \ln \frac{y}{v})^2 \frac{1}{ty}, & v < y < ve^t, \\ 0, & \text{其他}, \end{cases}$$

□

**例：**(对应郑书例 5.4, 对数正态分布) 设  $X$  是只取正值的随机变量，使得  $Y = \ln X$  服从正态分布  $N(\mu, \sigma^2)$ ，试求出  $X$  的分布函数和分布密度。

**解：**对任何  $x > 0$ ，有

$$\begin{aligned} F_X(x) &= P(X \leq x) = P(\ln X \leq \ln x) = P(Y \leq \ln x) \\ &= \int_{-\infty}^{\ln x} \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} (y - \mu)^2 \right\} dy \end{aligned}$$

做变量替换  $y = \ln u$ ，得

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}\sigma u} \exp \left\{ -\frac{1}{2\sigma^2} (\ln u - \mu)^2 \right\} du$$

当  $x \leq 0$  时，称变量  $X$  服从对数正态分布。不难看出， $X$  的分布密度  $p(u)$  为：当  $u \leq 0$  时， $p(u) = 0$ ，当  $u > 0$  时， $p(u)$  是上式中的被积函数。 □

**例：**设  $\theta \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$ ， $\psi = \tan \theta$ ，求  $\psi$  的密度函数。

**解：**设  $\psi = \tan \theta$  的反函数为  $g(y)$ ，则  $g(y) = \arctan y$ 。由定理得  $p_\psi(y) = p_{U(-\frac{\pi}{2}, \frac{\pi}{2})}(g(y))g'(y) = p_{U(-\frac{\pi}{2}, \frac{\pi}{2})}(\arctan y) \frac{1}{1+y^2} = \frac{1}{\pi(1+y^2)}$ ， $y \in \mathbb{R}$ ，称该变量  $\psi$  符合 **Cauchy 分布**。 □

## 4.2 随机变量的反函数

**随机变量的反函数：** 设  $F(x)$  是任何分布函数（即  $F(x)$  非减，右连续，且  $\lim_{x \rightarrow -\infty} F(x) = 0$ ,  $\lim_{x \rightarrow +\infty} F(x) = 1$ ），令

$$F^{-1}(p) \triangleq \min \{x : F(x) \geq p\} \quad (0 < p < 1)$$

则称  $F^{-1}(p)$  是  $F(x)$  的**广义反函数**。

注意， $F(x)$  是右连续增函数，满足不等式  $F(x) \geq p$  的  $x$  中必有最小者，当  $F(x)$  是严格增的连续函数时， $F^{-1}(p)$  正好是方程  $F(x) = p$  的唯一根，此时  $F^{-1}(p)$  是  $F(x)$  的普通反函数。

**引理：**  $F^{-1}(p)$  ( $0 < p < 1$ ) 有如下性质：

(1)  $F^{-1}(p)$  是  $p$  的增函数。

(2)  $F(F^{-1}(p)) \geq p$ ，若  $F(x)$  在点  $x = F^{-1}(p)$  处连续，则

$$F(F^{-1}(p)) = p.$$

(3)  $F^{-1}(p) \leq x$  的充分必要条件是  $p \leq F(x)$ 。

**证明：** (2) 由于  $F(F^{-1}(p) + \varepsilon) \geq p$  ( $\forall \varepsilon > 0$ )，令  $\varepsilon \rightarrow 0$ ，利用  $F(x)$  的右连续性知  $F(F^{-1}(p)) \geq p$ 。若  $F(x)$  在点  $F^{-1}(p)$  处连续，从  $F(F^{-1}(p) - \varepsilon) < p$  ( $\varepsilon > 0$ ) 推知  $F(F^{-1}(p)) = \lim_{\varepsilon \rightarrow 0} F(F^{-1}(p) - \varepsilon) \leq p$ ，从而  $F(F^{-1}(p)) = p$ 。

(3) 若  $F(x) \geq p$ ，从非减性质知  $x \geq F^{-1}(p)$ ；反之若  $x \geq F^{-1}(p)$ ，则  $F(x) \geq F(F^{-1}(p)) \geq p$ ，故性质 (3) 成立。  $\square$

**定理：** 设  $F(x)$  是任何分布函数，若  $U$  是服从区间  $[0, 1]$  上的均匀分布的随机变量，且

$$X = F^{-1}(U)$$

则  $X$  的分布函数恰好是  $F(x)$ 。

**证明：** 对任何  $y \in (0, 1)$ ，从性质 (3) 知  $x \geq F^{-1}(y)$  的充分必要条件是  $F(x) \geq y$ ，于是

$$P(X \leq x) = P(F^{-1}(U) \leq x) = P(U \leq F(x)) = F(x).$$

这表明  $X$  的分布函数是  $F(x)$ 。  $\square$

## 5.1 随机变量的数学期望

实际问题的概率分布比较难以确定,有时只需掌握随机变量的数学特征就足够了。随机变量的数学期望(expectation)的含义是,随机变量平均取值(mean)的大小。

- $X$  的大量独立观测值(记为  $a_1, a_2, \dots, a_n$ ) 的算术平均,当样本数无穷大时,算术平均收敛于期望值:

$$\bar{a} = \frac{1}{n} (a_1 + \dots + a_n).$$

- $X$  的所有可能值的加权平均(总和).

### 5.1.1 离散型随机变量的数学期望

**离散型随机变量的数学期望:** 假设  $X$  是离散型随机变量,分布列为

$$P(X = x_k) = p_k, \quad k = 1, \dots, n \text{ 或 } k = 1, 2, \dots.$$

其中  $X$  的可能值是  $x_1, x_2, \dots$ , 如果  $\sum_k |x_k| p_k < \infty$ , 那么,称  $X$  的期望存在,称  $\sum_k x_k p_k$  为  $X$  的数学期望,记为  $EX$ .

注意,级数  $\sum_k |x_k| p_k$  收敛可以保证和数  $\sum_k x_k p_k$  与加项的先后次序无关。更一般的假定是级数  $\sum_k x_k^+ p_k$  和  $\sum_k x_k^- p_k$  中至少一个收敛(这里  $x_k^+ = \max\{x_k, 0\}$ ,  $x_k^- = \max\{-x_k, 0\}$ ) 这时和数  $\sum_k x_k p_k$  与加项的先后次序无关。

注意到  $E(X)$  完全由  $X$  的概率分布确定,因此  $E(X)$  也称为相应概率分布的期望,下面计算几个常见的概率分布的期望:

#### (1) 两点分布

设随机变量  $X$  服从两点分布,  $P(X = 1) = p, P(X = 0) = 1 - p$ . 则,

$$E(X) = 1 \cdot p + 0 \cdot (1 - p) = p.$$

#### (2) 二项分布

设随机变量  $X$  服从二项分布:  $P(X = k) = C_n^k p^k q^{n-k} := b(n; k), k = 0, 1, \dots, n, (q = 1 - p)$ .

对于  $\forall 1 \leq k \leq n$ ,

$$\begin{aligned} k \cdot b(n; k) &= k \cdot \frac{n!}{k!(n-k)!} p^k q^{n-k} = \frac{n!}{(k-1)!(n-k)!} p^k q^{n-k} \\ &= \frac{n \cdot (n-1)!}{(k-1)!(n-k)!} p \cdot p^{k-1} q^{n-k} = np \cdot b(n-1; k-1) \end{aligned}$$

因此,

$$\begin{aligned} E(X) &= \sum_{k=0}^n k \cdot b(n; k) = \sum_{k=1}^n np \cdot b(n-1; k-1) \\ &= np \sum_{\ell=0}^{n-1} b(n-1, \ell) = np. \end{aligned}$$

### (3) 泊松分布

设随机变量  $X$  服从泊松分布:

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} =: p_k, \quad k = 0, 1, 2, \dots$$

则对于  $\forall k \geq 1$ ,

$$x_k p_k = k \frac{\lambda^k}{k!} e^{-\lambda} = \lambda \frac{\lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda p_{k-1}.$$

因此,

$$E(X) = \sum_{k=0}^{\infty} k p_k = \sum_{k=1}^{\infty} \lambda p_{k-1} = \lambda \sum_{\ell=0}^{\infty} p_{\ell} = \lambda.$$

### (4) 几何分布

设随机变量  $X$  满足几何分布, 即

$$P(X = k) = q^{k-1} p =: p_k, \quad k = 1, 2, \dots, (q = 1 - p).$$

直接计算期望:

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

### (5) 离散均匀分布

设随机变量  $X$  的可能值是  $1, \dots, N$ , 且

$$P(X = k) = \frac{1}{N} \quad (k = 1, \dots, N).$$

直接计算

$$E(X) = \sum_{k=1}^N k P(X = k) = \frac{N+1}{2}.$$

## (6) 超几何分布

设随机变量  $X$  满足超几何分布, 即

$$P(X = k) = \frac{C_D^k C_{N-D}^{n-k}}{C_N^n}, \quad k = 0, 1, \dots, n.$$

记  $h(N, D, n; k) = A_1 \cdot A_2 \cdot A_3$

$$= \frac{D!}{k!(D-k)!} \cdot \frac{(N-D)!}{(n-k)!(N-D-(n-k))!} \cdot \frac{n!(N-n)!}{N!}.$$

记  $x' = x - 1$ . 则,  $\forall 1 \leq k \leq n$ ,

$$k \cdot A_1 = \frac{D!}{(k-1)!(D-k)!} = D \times \frac{D!}{k!(D-k')!}.$$

进一步,

$$A_2 = \frac{(N'-D')!}{(n'-k')!(N'-D'-(n'-k'))!},$$

$$A_3 = \frac{n \cdot n'! (N' - n')!}{N \cdot N'!} = \frac{n}{N} \times \frac{n'! (N' - n')!}{N'!}.$$

记  $x' = x - 1$ . 则  $\forall 1 \leq k \leq n$ ,

$$k \cdot h(N, D, n; k) = \frac{nD}{N} \times h(N', D', n'; k').$$

因此,

$$E(X) = \sum_{k=1}^n k \cdot h(N, D, n; k) = \frac{nD}{N} \sum_{k'=0}^{n'} h(N', D', n'; k') = \frac{nD}{N}$$

对于该期望, 当  $D = 1$  时, 退化为伯努利分布,  $E(X) = p = \frac{D}{N}$ .

当  $D \geq 2$  时, 不放回抽样, 仍有  $E(X) = np$ .

## 5.1.2 一般随机变量的数学期望

若  $X$  为任意随机变量. 做如下近似: 对于  $\forall n \in \mathbb{Z}$ ,

$$\text{当 } n\varepsilon < X \leq (n+1)\varepsilon \text{ 时, 令 } X^* = n\varepsilon. \quad (5.1.1)$$

该假设的直观含义是:  $X^* \leq X < X^* + \varepsilon$ , 因此  $EX^* \leq EX < EX^* + \varepsilon$ .

**一般随机变量的数学期望:** 若  $EX^*$  存在且当  $\varepsilon \rightarrow 0$  时有极限, 则称  $X$  的期望存在, 且称该极限为  $X$  的期望, 记为  $E(X)$ 。

对离散型随机变量, 离散型随机变量期望的定义和一般随机变量的数学期望的定义一致。

**例:** 对于连续性随机变量  $X$ , 且  $X \geq 0$ . 证明  $E(X) = \int_0^{+\infty} P(X > x)dx$ .

**解:** 令

$$G(x) = P(X > x) = \int_x^{\infty} p(y)dy$$

则  $G'(x) = -p(x)$ . 于是,

$$\int_0^{\infty} xp(x)dx = \int_0^{\infty} x dG(x) = \int_0^{\infty} G(x)dx.$$

接下来, 我们计算一些常见连续型随机变量的数学期望:

### (1) 均匀分布

设随机变量  $X$  服从区间  $[a, b]$  上的均匀分布, 即  $X$  有密度分布:

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

由定义知

$$E(X) = \int_a^b x \frac{1}{b-a} dx = \frac{a+b}{2}.$$

### (2) 指数分布

设随机变量  $X$  有分布密度

$$p(x) = \lambda e^{-\lambda x}, \quad x > 0, \lambda > 0.$$

由定义知  $E(X) = \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx = - \int_0^{\infty} x de^{-\lambda x} = \int_0^{\infty} e^{-\lambda x} dx = \frac{1}{\lambda}$ .

### (3) 正态分布

设随机变量  $X$  有分布密度

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

对于  $X \sim N(0, 1)$ , 由对称性直接计算得,

$$E(X) = \int_{-\infty}^{\infty} x \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0$$

同理,  $X \sim N(\mu, \sigma^2)$ , 则  $p(\mu+x) = p(\mu-x)$ , 因此  $E(X) = \int_{-\infty}^{\infty} (x+\mu)p(\mu+x)d(x+\mu) = \int_{-\infty}^{\infty} xp(x+\mu)dx + \mu \int_{-\infty}^{\infty} p(x+\mu)dx = \mu$ .

例, 对于柯西分布,

$$p(x) = \frac{1}{\pi} \cdot \frac{1}{1+x^2}.$$

但是,  $\int_{-\infty}^{\infty} |x|p(x)dx = \infty$ . 因此,  $EX$  不存在!

#### (4) 伽马分布

设随机变量  $X$  有分布密度

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

对于  $\forall x > 0$ ,

$$xp(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^\alpha e^{-\beta x} = \frac{\Gamma(\alpha+1)}{\beta \Gamma(\alpha)} \cdot \frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)} x^\alpha e^{-\beta x} = \frac{\alpha}{\beta} \cdot \hat{p}(x).$$

因此,

$$E(X) = \int_0^\infty xp(x)dx = \frac{\alpha}{\beta} \int_0^\infty \hat{p}(x)dx = \frac{\alpha}{\beta}$$

### 5.1.3 数学期望的性质

**定理:** (1) 若  $X \equiv a$ , 则  $E(X) = a$ ;

(2) 若  $X \geq 0$ , 且  $E(X)$  存在, 则  $EX \geq 0$ ;

(3) 若  $F_X = F_Y$  (或, 若  $X = Y$ ), 且  $E(X)$  存在, 则  $E(Y)$  存在, 且  $E(X) = E(Y)$ ;

(4) 线性: 假设  $E(X), E(Y)$  存在. 则,

$$E(a(X)) = aE(X), \quad E(X+Y) = E(X) + E(Y).$$

(5) 单调性: 假设  $E(X), E(Y)$  存在, 又若  $X \geq Y$ , 则  $E(X) \geq E(Y)$ ;

(6)  $E|X| \geq |E(X)|$ ;

(7) 若随机变量  $X, Y$  独立, 且期望  $E(X), E(Y)$  存在, 则

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

**推论:** (1) 线性: 假设  $E(X), E(Y)$  存在. 则,

$$E(aX + bY) = aE(X) + bE(Y).$$

(2) 和的期望: 假设  $E(X_1), \dots, E(X_n)$  都存在,  $\eta = X_1 + \dots + X_n$ . 则  $E(\eta)$  存在, 且

$$E(\eta) = \sum_{i=1}^n E(X_i)$$

推论 (1) 可以由性质 (4) 推出, 推论 (2) 可以由数学归纳法和性质 (4) 推出。

**例：**超几何分布  $\eta \sim H(N, D, n)$  的期望可以使用推论 (2) 计算：若第  $i$  个产品是次品，则令  $X_i = 1$ ；否则，令  $X_i = 0$ 。则，

$$\eta = X_1 + \cdots + X_n \Rightarrow E(\eta) = np$$

**马尔科夫不等式：**设  $X \geq 0$ ，且  $EX$  存在。则对任意  $C > 0$ ，有

$$P(X \geq C) \leq \frac{1}{C}EX.$$

**证明：**令  $A = \{X \geq C\}$ 。则  $1_A \leq \frac{X}{C}$ 。于是，

$$P(A) = E1_A \leq E\frac{X}{C} = \frac{1}{C}EX.$$

**例：**若  $X \geq 0$ ，且  $EX = 0$ ，证明  $P(X > 0) = 0$ 。

**解：**

$$\begin{aligned} P\left(X \geq \frac{1}{n}\right) &\leq nEX = 0 \\ \Rightarrow P(X > 0) &= \lim_{n \rightarrow \infty} P\left(X \geq \frac{1}{n}\right) = 0. \end{aligned}$$

□

## 5.2 随机变量函数的期望

**定理：**(1)  $X$  是离散型随机变量，且下面的级数绝对收敛，则

$$Ef(X) = \sum_k f(x_k) p_k \quad (5.2.1)$$

(2)  $X$  是连续型随机变量，且下面的积分绝对收敛，则

$$Ef(X) = \int_{-\infty}^{\infty} f(x)p(x)dx. \quad (5.2.2)$$

**例：**(对应郑书例 6.1) 设  $X \sim U(0, 2\pi)$ ，求  $E(\sin X)$ 。

**解：**令  $p(x)$  是  $x$  的分布密度，用公式：

$$E \sin X = \int_{-\infty}^{\infty} \sin x \cdot p(x)dx = \frac{1}{2\pi} \int_0^{2\pi} \sin x dx = 0$$

□



**例：**(对应郑书例 6.2) 设随机变量  $X$  服从参数为  $\lambda$  的指数分布，又  $v_0 > 0$ ,

$$Y = \begin{cases} X, & X < v_0, \\ v_0, & X \geq v_0, \end{cases}$$

求  $E(Y)$ 。

**解：**设  $f(x) = \min\{x, v_0\}$ ，则  $Y = f(X)$ ，由于  $X$  的分布密度是

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & X \leq 0, \end{cases}$$

由式 5.2.2 知

$$\begin{aligned} E(Y) &= \int_{-\infty}^{+\infty} f(x)p(x)dx = \int_0^{+\infty} f(x)\lambda e^{-\lambda x}dx \\ &= \int_0^{v_0} x\lambda e^{-\lambda x}dx + \int_{v_0}^{+\infty} v_0\lambda e^{-\lambda x}dx \\ &= \frac{1}{\lambda}(1 - e^{-\lambda v_0}) \end{aligned}$$

□

**琴生不等式：**若  $\phi$  为凸函数，则

$$\phi(E(X)) \leq E(\phi(X)).$$

将期望等价于平均，代入琴生不等式即可证明。

**例：**连续型随机变量  $X, Y$  的概率密度函数分别为  $p(x), q(x)$  且  $p(x), q(x) \neq 0$ ， $f$  为一凸函数， $f(1) = 0$ ，证明： $E_{x \sim q} f\left(\frac{p(x)}{q(x)}\right) \geq 0$ 。

**证明：**由琴生不等式，

$$E_{x \sim q} f\left(\frac{p(x)}{q(x)}\right) \geq f\left(E_{x \sim q}\left(\frac{p(x)}{q(x)}\right)\right) = f\left(\int_{-\infty}^{+\infty} \frac{p(x)}{q(x)} q(x) dx\right) = f(1) = 0.$$

**例：**连续型随机变量  $X, Y$  的概率密度函数分别为  $p(x), q(x)$  且  $p(x), q(x) \neq 0$ ，我们定义  $X$  关于  $Y$  的 KL-divergence 为  $KL(X||Y) = E_X\left(\ln \frac{p(x)}{q(x)}\right)$ ，试证明  $KL(X||Y) \geq 0$ 。

**证明：**

$$KL(X||Y) = \int p(x) \left(-\ln \frac{q(x)}{p(x)}\right) dx,$$

由于  $-\ln x$  是凸函数，由琴生不等式知

$$\int p(x) \left(-\ln \frac{q(x)}{p(x)}\right) dx \geq -\ln \left(\int p(x) \frac{q(x)}{p(x)} dx\right) = -\ln 1 = 0.$$

□

**例：**设随机变量  $X$  服从参数为  $\lambda$  的泊松分布，证明

$$E(X^n) = \lambda E((X+1)^{n-1}).$$

利用此结果计算  $E(X^3)$ 。

**证明：**

$$E(X^n) = \sum_{k=0}^{\infty} k^n \frac{\lambda^n}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} k^{n-1} \frac{\lambda^{k-1}}{(k-1)!}$$

设  $k' = k - 1$ ，则

$$E(X^n) = \lambda e^{-\lambda} \sum_{k'=0}^{\infty} (k'+1)^{n-1} \frac{\lambda^{k'}}{k'!} = \lambda E((X+1)^{n-1}).$$

由此得

$$E(X^3) = \lambda E(X+1)^2 = \lambda(E(X^2) + 2E(X) + 1) = \lambda(\lambda E(X+1) + 2\lambda + 1) = \lambda^3 + 3\lambda^2 + \lambda.$$

**例：**设  $X$  是仅取非负整数的离散随机变量，若其数学期望存在，证明

$$(1) E(X) = \sum_{k=1}^{\infty} P(X \geq k).$$

$$(2) \sum_{k=0}^{\infty} kP(X > k) = \frac{1}{2}[E(X^2) - E(X)].$$

**证明：**(1) 由于  $E(X) = \sum_{k=1}^{\infty} kP(X = k)$  存在，所以该级数绝对收敛，从而

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} kP(X = k) = \sum_{k=1}^{\infty} \left[ \sum_{i=1}^k P(X = k) \right] \\ &= \sum_{i=1}^{\infty} \left[ \sum_{k=i}^{\infty} P(X = k) \right] = \sum_{k=1}^{\infty} P(X \geq k). \end{aligned}$$

(2)

$$\begin{aligned} \sum_{k=0}^{\infty} kP(X > k) &= \sum_{k=0}^{\infty} k \sum_{i=k+1}^{\infty} P(X = i) = \sum_{i=1}^{\infty} \sum_{k=0}^{i-1} kP(X = i) \\ &= \sum_{i=1}^{\infty} P(X = i) \frac{(i-1)i}{2} = \frac{1}{2} \sum_{i=1}^{\infty} i^2 P(X = i) - \frac{1}{2} \sum_{i=1}^{\infty} i P(X = i) \\ &= \frac{1}{2} E(X^2) - \frac{1}{2} E(X). \end{aligned}$$

**例：**甲乙两人进行象棋比赛，每局甲胜的概率为  $p$ ，乙胜的概率为  $q = 1 - p$ ，比赛进行到有一人连胜两局为止，求平均比赛局数。

**解：**设  $X$  为决定胜负所需的局数，可以取值为  $2, 3, \dots$ ，事件  $\{X \geq k\}$  表示“到  $k-1$  局时没有一人连胜两局”，所以

$$P(X \geq 1) = 1,$$

$$P(X \geq 2k) = p^k q^{k-1} + p^{k-1} q^k = (pq)^{k-1}, \quad k = 1, 2, \dots,$$

$$P(X \geq 2k+1) = 2p^k q^k, \quad k = 1, 2, \dots.$$

利用上一题第一问提供的公式，可得

$$\begin{aligned} E(X) &= \sum_{k=1}^{\infty} P(X \geq k) = 1 + \sum_{k=1}^{\infty} (pq)^{k-1} + 2 \sum_{k=1}^{\infty} (pq)^k \\ &= 1 + \frac{1}{1-pq} + \frac{2pq}{1-pq} = \frac{2+pq}{1-pq}. \end{aligned}$$

注意到对任意的  $0 < p < 1$  总有  $p(1-p) \leq \frac{1}{4}$ ，故由  $E(X)$  关于  $pq$  单调增可得

$$E(X) \leq \frac{2+\frac{1}{4}}{1-\frac{1}{4}} = 3$$

故这种比赛最终决定胜负的平均局数不超过 3 局，在  $p = \frac{1}{2}$  时达到上界。

### 5.3 随机变量的方差

**随机变量的方差和标准差：**假设  $E(X)$  存在，且  $E(X - EX)^2$  也存在。则称  $E(X - EX)^2$  为  $X$  的方差，记为  $\text{var}(X)$  或  $D(X)$ 。称  $\sqrt{\text{var}(X)}$  为标准差。

**切比雪夫不等式：**设  $X$  是随机变量，如果  $E(X)$  和  $\text{var}(X)$  都存在，则  $\forall \varepsilon > 0$ ，有

$$P(|X - EX| \geq \varepsilon) \leq \frac{1}{\varepsilon^2} \text{var}(X). \quad (5.3.1)$$

**证明：** $\{|X - EX| \geq \varepsilon\} = \{(X - EX)^2 \geq \varepsilon^2\}$ ，对  $Y = (X - EX)^2$  用马尔可夫不等式，得

$$P(|X - EX| \geq \varepsilon) = P(Y \geq \varepsilon^2) \leq \frac{1}{\varepsilon^2} E(Y).$$

□

**推论：**若  $\text{var}(X) = 0$ ，则

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

**证明：**由切比雪夫不等式知

$$P(|X - E(X)| \geq \frac{1}{n}) = 0 \quad (n = 1, 2, \dots),$$

于是

$$\begin{aligned} P(X \neq E(X)) &= P\left(\bigcup_{n=1}^{\infty} \left\{|X - E(X)| \geq \frac{1}{n}\right\}\right) \\ &\leq \sum_{n=1}^{\infty} P(|X - E(X)| \geq \frac{1}{n}) = 0. \end{aligned}$$

所以  $P(X = E(X)) = 1$ 。 □

对于方差的计算方法，有以下定理：

**定理：**  $X$  为一般随机变量，且期望  $E(X^2)$  和  $E(X)$  存在，则

$$\text{var}(X) = E(X^2) - (EX)^2. \quad (5.3.2)$$

**证明：**

$$\begin{aligned} \text{var}(X) &= E(X^2 - 2X \cdot EX + (EX)^2) \\ &= E(X^2) - 2E(X) \cdot E(X) + (E(X))^2 = E(X^2) - (EX)^2 \end{aligned}$$

具体地，离散型或连续型的公式如下：

$$\begin{aligned} \text{var}(X) &= \sum_k x_k^2 p_k - (EX)^2 \\ \text{var}(X) &= \int_{-\infty}^{\infty} x^2 p(x) dx - (EX)^2 \end{aligned}$$

□

**定理：**  $X$  的线性变换的方差：

$$\text{var}(aX + b) = a^2 \text{var}(X)$$

该定理可以利用式5.3.2计算。

$$\begin{aligned} \text{var}(aX + b) &= E((aX + b)^2) - (E(aX + b))^2 \\ &= (a^2 E(X^2) + 2abE(X) + b^2) - (a^2 (E(X))^2 + 2abE(X) + b^2) \\ &= a^2 (E(X^2) - (E(X))^2) = a^2 \text{var}(X) \end{aligned}$$

□

**定理：** 设  $X$  为随机变量，则方差  $D(X) = \inf_{c \in \mathbb{R}} E(X - c)^2$ 。

**证明：【方法一】** 利用  $E(X + c) = E(X) + c$  与  $D(X + c) = D(X)$ ，可得

$$D(X) = D(X - c) = E(X - c)^2 - (E(X - c))^2 \leq E(X - c)^2.$$

等号成立条件是  $E(X) = c$ 。

【方法二】利用

$$\begin{aligned} E(X - c)^2 &= E(X - E(X) + E(X) - c)^2 \\ &= D(X) + 2E[(X - E(X))(E(X) - c)] + (E(X) - c)^2 \\ &= D(X) + (E(X) - c)^2. \end{aligned}$$

在  $c = E(X)$  处取得最小值  $D(X)$ 。 □

下面计算常见随机变量的方差：

(1) 两点分布

设随机变量  $X$  服从两点分布，即  $X \sim B(1, p)$ ，根据之前的计算  $E(X) = p$ ， $E(X^2) = 0^2 \cdot P(X = 0) + 1^2 \cdot P(X = 1) = p$ ，由式5.3.2知

$$D(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1 - p)$$

(2) 二项分布

设随机变量  $X$  服从参数为  $n, p$  的二项分布，即

$$P(X = k) = C_n^k p^k q^{n-k} =: b(n; k), \quad k = 0, 1, \dots, n, (q = 1 - p).$$

已经计算期望  $EX = np$ ，且由分布函数知，对于  $\forall 1 \leq k \leq n$ ，

$$k \cdot b(n; k) = np \cdot b(n - 1, k - 1).$$

那么对于  $\forall 2 \leq k \leq n$ ，

$$\begin{aligned} k(k - 1) \cdot b(n; k) &= np \cdot (k - 1) \cdot b(n - 1, k - 1) \\ &= np \cdot (n - 1)p \cdot b(n - 2, k - 2) \end{aligned}$$

于是，

$$E(X(X - 1)) = \sum_{k=2}^n k(k - 1) \cdot b(n; k) = np(n - 1)p \sum_{k=2}^n b(n - 2; k - 2) = np(n - 1)p = (np)^2 - np^2,$$

从而

$$D(X) = E(X^2) - (E(X))^2 = E(X(X - 1)) + E(X) - (E(X))^2 = npq.$$

(3) 泊松分布

设随机变量  $X$  服从参数为  $\lambda$  的泊松分布, 即

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

已经计算  $X$  的期望  $E(X) = \lambda$ , 且由分布函数,  $\forall k \geq 1, kp_k = \lambda p_{k-1}$ . 因此, 对于  $\forall k \geq 2$ ,

$$k(k-1)p_k = \lambda(k-1)p_{k-1} = \lambda^2 p_{k-2}.$$

于是,

$$E(X(X-1)) = \sum_{k=2}^{\infty} k(k-1) \cdot p_k = \lambda^2 \sum_{k=2}^{\infty} p_{k-2} = \lambda^2,$$

从而

$$Dvar(X) = E(X^2) - (E(X))^2 = E(X(X-1)) + E(X) - (E(X))^2 = \lambda.$$

#### (4) 均匀分布

设随机变量  $X$  服从区间  $[a, b]$  上的均匀分布, 即  $X$  有密度分布:

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

已经计算  $X$  的期望  $E(x) = \frac{a+b}{2}$ , 且

$$\int_{-\infty}^{+\infty} x^2 p(x) dx = \frac{1}{b-a} \int_a^b x^2 dx = \frac{b^2 + ab + a^2}{3}.$$

由式5.3.2得

$$D(X) = \frac{b^2 + ab + a^2}{3} - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}.$$

#### (5) 指数分布

设随机变量  $X$  服从参数为  $\lambda$  的指数分布, 即  $X$  的分布函数为:

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & X \leq 0, \end{cases}$$

已经计算期望  $E(X) = \frac{1}{\lambda}$ , 且

$$\int_{-\infty}^{+\infty} x^2 p(x) dx = \lambda \int_0^{+\infty} x^2 e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

由式5.3.2得

$$D(X) = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda}\right)^2 = \frac{1}{\lambda^2}$$

## (6) 正态分布

设随机变量  $X$  服从正态分布, 即  $X$  的分布函数为:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

若  $\mu = E(X) = 0, \sigma^2 = 1$ , 则,

$$\begin{aligned} D(X) &= E(X^2) = \int_{-\infty}^{\infty} x^2 \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ &= -\frac{2}{\sqrt{2\pi}} \int_0^{\infty} x d e^{-\frac{x^2}{2}} = \frac{2}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} dx = 1 \end{aligned}$$

一般情形, 做变量替换  $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$ . 则

$$\begin{aligned} X - E(X) &= (\mu + \sigma Y) - (\mu + \sigma E(Y)) = \sigma(Y - E(Y)) \\ \Rightarrow D(X) &= E((X - E(X))^2) = E(\sigma(Y - E(Y))^2) = \sigma^2 D(Y) = \sigma^2 \end{aligned}$$

## (7) 伽马分布

设随机变量  $X$  服从伽马分布, 有分布密度

$$p(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x > 0$$

已经计算  $X$  的期望  $E(X) = \frac{\alpha}{\beta}$ , 由式5.3.2知

$$D(X) = \int_0^{+\infty} x^2 \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx - \left(\frac{\alpha}{\beta}\right)^2$$

做变量替换  $\beta x = t$ , 易知

$$\begin{aligned} D(X) &= \frac{1}{\Gamma(\alpha)\beta^2} \int_0^{+\infty} t^{\alpha+1} e^{-t} dt - \left(\frac{\alpha}{\beta}\right)^2 \\ &= \frac{1}{\Gamma(\alpha)\beta^2} \Gamma(\alpha+2) - \left(\frac{\alpha}{\beta}\right)^2 \\ &= \frac{(\alpha+1)\alpha\Gamma(\alpha)}{\Gamma(\alpha)\beta^2} - \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2} \end{aligned}$$

**随机变量的标准化:** 一般地, 若  $X$  的方差存在, 且  $\text{var}(X) > 0$ , 则

$$X^* = \frac{X - E(X)}{\sqrt{\text{var}(X)}}$$

满足  $E(X^*) = 0, \text{var}(X^*) = 1$ . 称  $X^*$  为  $X$  的标准化。

**例：**设随机变量  $X \sim N(0, 1)$ , 则对一切正整数  $k$ ,

$$E(X^{2k-1}) = 0, \quad E(X^{2k}) = (2k-1)(2k-3)\cdots 3 \cdot 1.$$

**证明：**对任何  $m \geq 1$ , 积分  $\int_{-\infty}^{+\infty} |x|^m e^{-x^2/2} dx$  收敛, 因此  $E(X^m)$  存在, 由于

$$x^{2k-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

是  $x$  的奇函数, 故

$$E(X^{2k-1}) = \int_{-\infty}^{+\infty} x^{2k-1} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = 0$$

$$\begin{aligned} E(X^{2k}) &= \int_{-\infty}^{+\infty} x^{2k} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^{2k-1} d(e^{-x^2/2}) \\ &= (2k-1) \int_{-\infty}^{+\infty} x^{2k-2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \\ &= (2k-1) E(X^{2k-2}). \end{aligned}$$

这是递推公式, 故

$$E(X^{2k}) = (2k-1)(2k-3)\cdots 3 \cdot 1$$

**例：**设随机变量  $X$  的概率密度函数为  $f(x) = \frac{k\theta^k}{x^{k+1}}, x > \theta > 0, k > 2$  为正整数, 求 (1)  $E(X)$ , (2)  $D(X)$ .

**解：**(1):

$$E(X) = \int_{\theta}^{+\infty} \frac{k\theta^k}{x^{k+1}} x dx = \frac{k\theta}{k-1}.$$

(2):

$$D(X) = \int_{\theta}^{+\infty} \frac{k\theta^k}{x^{k+1}} x^2 dx - (E(X))^2 = \frac{3k-2k^2}{(k-2)(k-1)^2}.$$

**例：**设连续随机变量  $X$  的分布函数为  $F(x)$ , 且数学期望存在, 证明:

$$E(X) = \int_0^{\infty} [1 - F(x)] dx - \int_{-\infty}^0 F(x) dx.$$

**证明：**

$$E(X) = \int_{-\infty}^{\infty} xp(x) dx = \int_{-\infty}^0 xp(x) dx + \int_0^{\infty} xp(x) dx.$$



将第一个积分改写为:

$$\begin{aligned}\int_{-\infty}^0 xp(x)dx &= \int_{-\infty}^0 \left( \int_x^0 dy \right) p(x)dx \\ &= - \int_{-\infty}^0 \int_{-\infty}^y p(x)dx dy \\ &= - \int_{-\infty}^0 F(y)dy.\end{aligned}$$

第二个积分同理,

$$\begin{aligned}\int_0^{\infty} xp(x)dx &= \int_0^{\infty} \left( \int_0^x dy \right) p(x)dx \\ &= \int_0^{\infty} \int_y^{\infty} p(x)dx dy \\ &= \int_0^{\infty} [1 - F(y)]dy.\end{aligned}$$

将二式加和即可得

$$E(X) = \int_0^{\infty} [1 - F(x)]dx - \int_{-\infty}^0 F(x)dx.$$

## 5.4 随机变量的其他数学特征

**原点矩和中心矩:**

设  $X$  是随机变量, 如果  $E(X^k)$  存在 ( $k$  是正整数), 则称  $E(X^k)$  是  $X$  的  $k$  阶原点矩, 常常记为  $\nu_k$ 。

设  $X$  是随机变量, 如果  $E(X)$  存在, 且  $E(X - E(X))^k$  存在 ( $k$  是正整数), 则称  $E(X - E(X))^k$  为  $X$  的  $k$  阶中心矩, 常常记为  $\mu_k$ 。

显然,  $E(X) = \nu_1$ ,  $\text{var}(X) = \mu_2$ 。

**随机变量的  $p$  分位数:** 若  $X$  是随机变量,  $0 < p < 1$ , 且

$$P(X < a) \leq p \leq P(X \leq a),$$

则称  $a$  为  $X$  的一个  $p$  分位数。

$p = 0.5$  时, 也称  $a$  为一个中位数。

**例:** 设随机变量  $X$  的可能值是 1, 2, 3 且

$$P(X = 1) = \frac{1}{3}, \quad P(X = 2) = \frac{1}{6}, \quad P(X = 3) = \frac{1}{2}.$$

则  $E(X) = \frac{13}{6}$ , 中位数有无穷个, 区间  $[2, 3]$  中的每个数都是  $X$  的中位数。

**例:** (对应郑书例 8.3) 设随机变量  $X \sim N(0, 1)$ , 则对一切正整数  $k$ ,

$$E(x^{2k-1}) = 0, \quad E(X^{2k}) = (2k-1)(2k-3) \cdots 3 \cdot 1.$$

□

**定理:** 设  $X = X(\omega)$  是随机变量, 对某个  $\alpha \geq 1$ ,  $E(|X|^\alpha)$  存在, 则  $E(X)$  存在, 且

$$E(|X|) \leq (E(|X|^\alpha))^{1/\alpha}.$$

**证明:** 首先指出, 对一切  $x \geq 0$ ,  $\alpha \geq 0$ , 如下不等式成立:

$$x^\alpha \geq a^\alpha + \alpha a^{\alpha-1}(x-a). \quad (5.4.1)$$

实际上, 令  $f(x) = x^\alpha - a^\alpha - \alpha a^{\alpha-1}(x-a)$ , 则  $f'(x) = \alpha(x^{\alpha-1} - a^{\alpha-1})$ , 从而  $f(x)$  在  $x = a$  处达到最小值, 由于  $f(a) = 0$ , 因此式 5.4.1 成立。

由于  $\alpha \geq 1$ , 有  $|X(\omega)| \leq |X(\omega)|^\alpha + 1$ , 知  $E(X)$  存在。令  $a = E(X)$ , 由式 5.4.1 知

$$|X(\omega)|^\alpha \geq (E(|X|))^\alpha + \alpha(E(|X|))^{\alpha-1}(|X(\omega)| - E(|X|)).$$

两侧取数学期望, 得  $(E(|X|))^\alpha \leq E(|X|^\alpha)$ , 表明定理成立。

□