

# 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

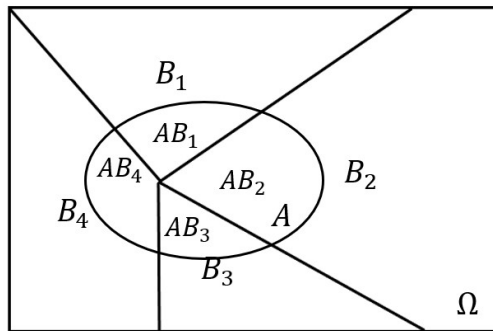


图 2.2: 全概率公式

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

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

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：

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

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  $\forall k \geq 1$ ,

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

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。

□

## 6.1 随机向量的定义

**$n$  维随机向量：**称  $n$  个随机变量  $X_1, \dots, X_n$  的整体  $\xi = (X_1, \dots, X_n)$  为  $n$  维随机向量（或者  $n$  维随机变量），一维随机向量简称随机变量。

**$n$  维随机变量数学上的精确定义：**设  $X_1 = X_1(\omega), \dots, X_n = X_n(\omega)$  都是概率空间  $(\Omega, \mathcal{F}, P)$  上的随机变量，则称

$$\xi = \xi(\Omega) \triangleq (X_1(\omega), \dots, X_n(\omega))$$

为概率空间  $(\Omega, \mathcal{F}, P)$  上的  $n$  维随机向（变）量。

例如，用炮弹向远处目标攻击，炮弹的落点用平面坐标系中的坐标表示为  $(X, Y)$ ，是一个二维随机向量。

**随机向量的函数：**设  $X_1 = X_1(\omega), \dots, X_n = X_n(\omega)$  是  $n$  个随机变量， $f(x_1, \dots, x_n)$  是  $n$  元实值函数，则称随机变量  $Y \triangleq f(x_1, \dots, x_n)$  为随机变量  $X_1, \dots, X_n$  的函数（即随机向量  $(X_1, \dots, X_n)$  的函数）。

## 6.2 二维随机变量的联合分布和边缘分布

### 6.2.1 离散情形

**离散型二维随机向量：**称二维随机向量  $\xi = (X, Y)$  是离散型的，若它只取至多可列个不同的值，即  $\xi$  可能取的值可以排成一个（有限或无穷序列）。

**二维离散型随机向量的概率分布：**设  $\xi = (X, Y)$  是二维离散型随机向量，其可能值是  $a_1, a_2, \dots$ （有限个或者无穷可列个）， $p_i \triangleq P(\xi = a_i) (i = 1, 2, \dots)$ ，则称

$$\{p_i : i = 1, 2, \dots\}$$

为  $\xi$  的概率分布，也称为  $\xi$  的概率函数或概率分布律。 $\xi = (X, Y)$  的概率分布也叫做  $(X, Y)$  的联合概率分布（简称联合分布）。

令

$$p_{ij} = P(X = x_i, Y = y_j) \quad (i, j = 1, 2, \dots),$$

$\{p_{ij}\}$  就是  $\xi = (X, Y)$  的概率分布，可以用表6.3来表示，表6.3也称  $\xi = (X, Y)$  的概率分布表。

联合分布满足性质：

表 6.3:  $(X, Y)$  的概率分布表。

$X \backslash Y$					
	$y_1$	$y_2$	$\cdots$	$y_j$	$\cdots$
$x_1$	$p_{11}$	$p_{12}$	$\cdots$	$p_{1j}$	$\cdots$
$x_2$	$p_{21}$	$p_{22}$	$\cdots$	$p_{2j}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$x_i$	$p_{i1}$	$p_{i2}$	$\cdots$	$p_{ij}$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

(1) 非负性:  $p_{ij} \geq 0, i, j = 1, 2, \cdots$  ;

(2) 规范性:  $\sum_{i,j} p_{ij} = 1$ 。

**例:** (三项分布) 设二维随机向量  $\xi = (X, Y)$  取值于集合  $E = \{(k_1, k_2) : k_1 \text{ 和 } k_2 \text{ 都是非负整数且 } k_1 + k_2 \leq n\}$ ,  $\xi$  的概率分布是:

$$P((X, Y) = (k_1, k_2)) = \frac{n!}{k_1!k_2!(n - k_1 - k_2)!} p_1^{k_1} p_2^{k_2} (1 - p_1 - p_2)^{n - k_1 - k_2},$$

其中  $n \geq 1, 0 < p_1, 0 < p_2, p_1 + p_2 < 1, (k_1, k_2) \in E$ , 这时称  $\xi$  服从三项分布。

**例:** 有一大批量粉笔, 其中 60% 是白的, 25% 是黄的, 15% 是红的, 现从中随机的依次取出 6 支, 问: 其中恰有 3 支白色, 1 支黄色, 2 支红色的概率是多少?

**解:** 令  $X =$  “6 支中白粉笔的个数”,  $Y =$  “6 支中黄粉笔的个数”, 则事件 “6 支中恰有 3 支白色, 1 支黄色, 2 支红色” 就是事件

$$\{X = 3, Y = 1\}, \text{ 即 } \{(X, Y) = (3, 1)\}.$$

由三项分布, 概率可表示为

$$P((X, Y) = (3, 1)) = \frac{6!}{3!1!2!} 0.6^3 \times 0.25 \times 0.15^2.$$

用组合数方法同样可以得到上述结果。

一般的, 对于满足  $k_1 \geq 0, k_2 \geq 0$  及  $k_1 + k_2 \leq 6$  的  $k_1, k_2$ , 由三项分布有

$$P((X, Y) = (k_1, k_2)) = \frac{6!}{k_1!k_2!(6 - k_1 - k_2)!} 0.6^{k_1} \times 0.25^{k_2} \times 0.15^{6 - k_1 - k_2}.$$

□

**二维随机向量的边缘分布:** 对于二维随机向量  $\xi = (X, Y)$ , 分量  $X$  的概率分布称为  $\xi$  关于  $X$  的**边缘分布**, 分量  $Y$  的概率分布称为  $\xi$  关于  $Y$  的**边缘分布**。

二维随机向量  $\xi = (X, Y)$  的两个边缘分布均由  $\xi$  的概率分布完全确定。

**例:** 从  $1, 2, 3, 4$  中任取一数记为  $X$ , 再从  $1, \dots, X$  中任取一数记为  $Y$ , 求  $(X, Y)$  的联合分布列及  $P(X = Y)$ 。

**解:** 易知  $X$  的分布列为:

$$P(X = i) = \frac{1}{4}, \quad i = 1, 2, 3, 4.$$

显然,  $P(X = i, Y = j) = 0, j > i, i = 1, 2, 3, 4$ , 当  $1 \leq j \leq i \leq 4$  时, 由乘法公式得

$$P(X = i, Y = j) = P(X = i)P(Y = j|X = i) = \frac{1}{4} \times \frac{1}{i} = \frac{1}{4i}.$$

从而  $(X, Y)$  的分布列为 由此可算得

$Y \backslash X$	1	2	3	4
1	$\frac{1}{4}$	0	0	0
2	$\frac{1}{8}$	$\frac{1}{8}$	0	0
3	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	0
4	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{16}$

$$P(X = Y) = \sum_{i=1}^4 P(X = Y = i) = \sum_{i=1}^4 \frac{1}{4i} = \frac{25}{48}.$$

□

**例:** (对应郑书例 2.5) 设随机变量  $X$  取值是 0 或 1, 随机变量  $Y$  取值也是 0 或 1, 且二维随机向量  $(X, Y)$  的概率分布是

$$P((X, Y) = (0, 0)) = \frac{1}{4} + \varepsilon, \quad P((X, Y) = (0, 1)) = \frac{1}{4} - \varepsilon,$$

$$P((X, Y) = (1, 0)) = \frac{1}{4} - \varepsilon, \quad P((X, Y) = (1, 1)) = \frac{1}{4} + \varepsilon,$$

其中  $0 \leq \varepsilon \leq \frac{1}{4}$ 。

易知不同的  $\varepsilon$  对应不同的联合分布, 但是

$$P(X = 0) = P((X, Y) = (0, 0)) + P((X, Y) = (0, 1)) = \frac{1}{2},$$

$$P(X=1) = P((X,Y)=(1,0)) + P((X,Y)=(1,1)) = \frac{1}{2}.$$

同理,

$$P(Y=0) = P(Y=1) = \frac{1}{2},$$

由此可见, 两个边缘分布均与  $\varepsilon$  无关, 表明有无穷多个不同的联合分布具有相同的边缘分布。

### 6.2.2 连续情形

**连续型随机向量及其联合密度函数:** 设  $\xi = (X, Y)$  为二维随机向量, 若存在非负函数  $p(x, y)$  使得

$$P(\xi \in D) = \iint_D p(x, y) dx dy,$$

对任意开矩形  $D$  成立, 则称  $\xi$  为**连续型随机向量**, 称  $p(x, y)$  为  $\xi$  的**联合密度** (函数), 也称**概率分布密度函数**, 记为  $p_{X,Y}(x, y)$ .

对于二维连续型随机向量  $\xi = (X, Y)$ , 对于平面上任意的集合  $A$ , 有

$$P(\xi \in A) = \iint_A p(x, y) dx dy, \quad (6.2.1)$$

联合密度满足归一性:

$$p(x, y) \geq 0; \quad \iint_{\mathbb{R}^2} p(x, y) dx dy = 1.$$

**例:** (对应郑书例 2.6) 设二维随机向量  $\xi = (X, Y)$  的联合密度为

$$p(x, y) = \begin{cases} ce^{-(x+y)}, & x \geq 0 \text{ 且 } y \geq 0, \\ 0, & \text{其他,} \end{cases}$$

其中  $c$  是一个常数, 求:

(1)  $c$  的值; (2)  $P(0 < X < 1, 0 < Y < 1)$ .

**解:** (1) 由归一性知

$$1 = \int_0^{+\infty} \int_0^{+\infty} ce^{-(x+y)} dx dy = c \int_0^{+\infty} e^{-x} dx \cdot \int_0^{+\infty} e^{-y} dy$$

于是  $c = 1$ .

(2) 取  $D = \{(x, y) : 0 < x < 1, 0 < y < 1\}$ , 由定义知

$$P(0 < X < 1, 0 < Y < 1) = P((X, Y) \in D) = \int_0^1 \int_0^1 e^{-(x+y)} dx dy$$

$$= \int_0^1 e^{-x} dx \cdot \int_0^1 e^{-y} dy = (1 - e^{-1})^2.$$

□

**定义：**设  $G$  是平面上面积为  $a(0 < a < +\infty)$  的区域，称二维随机向量  $\xi = (X, Y)$  服从  $G$  上的**均匀分布**，若  $P((X, Y) \in G) = 1$ ，且  $(X, Y)$  取值属于  $G$  的任何部分  $A$  ( $A$  是  $G$  的子区域) 的概率与  $A$  的面积成正比。容易推知二维随机向量  $\xi = (X, Y)$  有联合密度为

$$p(x, y) = \begin{cases} \frac{1}{a}, & (x, y) \in G, \\ 0, & \text{其他}, \end{cases} \quad (6.2.2)$$

**连续型随机向量的边缘分布：**设  $p(x, y)$  是二维随机向量  $\xi = (X, Y)$  的联合密度，则

$$p_X(x) \triangleq \int_{-\infty}^{+\infty} p(x, y) dy, \quad p_Y(y) \triangleq \int_{-\infty}^{+\infty} p(x, y) dx$$

分别是  $X, Y$  的分布密度。

**证明：**对任何  $a < b$ ，令  $A = \{(x, y) : a < x < b, -\infty < y < +\infty\}$ ，由式6.2.1知

$$\begin{aligned} P((X, Y) \in A) &= P(a < X < b) = \int_a^b \int_{-\infty}^{+\infty} p(x, y) dx dy \\ &= \int_a^b \left( \int_{-\infty}^{+\infty} p(x, y) dy \right) dx = \int_a^b p_X(x) dx. \end{aligned}$$

这表明  $p_X(x)$  是  $X$  的分布密度，同理知  $p_Y(y)$  是  $Y$  的分布密度。 □

**例：**(对应郑书例 2.7) 设  $G$  是由抛物线  $y = x^2$  和直线  $y = x$  所围成的区域 (图6.3) 若二维随机向量  $\xi = (X, Y)$  服从  $G$  上的均匀分布，试求  $\xi$  的联合分布和两个边缘分布密度。

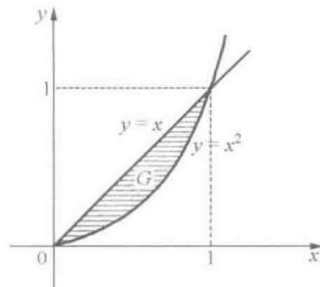


图 6.3: 区域  $G$  的示意图

解：由于  $G$  的面积为

$$\int_0^1 (x - x^2) dx = \frac{1}{6},$$

由式6.2.2知联合密度为

$$p(x, y) = \begin{cases} 6, & (x, y) \in G, \\ 0, & \text{其他,} \end{cases}$$

得  $X$  的分布密度  $p_X(x)$  和  $Y$  的分布密度  $p_Y(y)$  分别如下：

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{+\infty} p(x, y) dy \\ &= \int_{x^2}^x 6 dy = 6(x - x^2) \quad (0 \leq x \leq 1), \\ p_X(x) &= 0 \quad (x \notin [0, 1]), \\ p_Y(y) &= \int_{-\infty}^{+\infty} p(x, y) dx \\ &= \int_y^{\sqrt{y}} 6 dx = 6(\sqrt{y} - y) \quad (0 \leq y \leq 1), \\ p_Y(y) &= 0 \quad (y \notin [0, 1]). \end{aligned}$$

□

由定义知边缘密度函数由联合密度确定，但是不同的联合密度可能有相同的边缘分布密度，即联合密度不能由两个边缘分布密度完全确定。

**例：**设二维随机向量  $\xi = (X, Y)$  有联合密度

$$p_1(x, y) = \frac{1}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\},$$

二维随机向量  $\eta = (U, V)$  有联合密度

$$p_2(x, y) = \begin{cases} 2p_1(x, y), & xy \geq 0, \\ 0, & \text{其他.} \end{cases}$$

则  $X$  与  $U$  有相同的分布密度， $Y$  与  $V$  有相同的分布密度。

一方面，当  $x \leq 0$  时，

$$\begin{aligned} \int_{-\infty}^{+\infty} p_2(x, y) dy &= \int_{-\infty}^0 2p_1(x, y) dy = \frac{1}{\pi} \int_{-\infty}^0 e^{-(x^2+y^2)/2} dy \\ &= \frac{1}{\pi} e^{-x^2/2} \int_{-\infty}^0 e^{-y^2/2} dy = \frac{1}{\pi} e^{-x^2/2} \frac{1}{2} \int_{-\infty}^{+\infty} e^{-y^2/2} dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}. \end{aligned}$$



类似的, 当  $x > 0$  时,

$$\begin{aligned}\int_{-\infty}^{+\infty} p_2(x, y) dy &= \int_0^{+\infty} 2p_1(x, y) dy \\ &= \frac{1}{\pi} \int_0^{+\infty} e^{-(x^2+y^2)/2} dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.\end{aligned}$$

即, 对一切  $x$ ,  $\int_{-\infty}^{+\infty} p_2(x, y) dy = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ .

同理, 对一切  $y$ ,  $\int_{-\infty}^{+\infty} p_2(x, y) dx = \frac{1}{\sqrt{2\pi}} e^{-y^2/2}$ .

### 6.2.3 一般情形

**一般二维随机向量及其联合分布函数:** 设  $\xi = (X, Y)$  是二维随机向量, 则称

$$F(x, y) = P(X \leq x, Y \leq y) \quad (x, y \in \mathbb{R})$$

为  $\xi$  的分布函数。也称为  $(X, Y)$  的**联合分布函数**。

分布函数  $F(x, y)$  有以下性质:

- (1)  $0 \leq F(x, y) \leq 1$ ;
- (2)  $F(x, y)$  是  $x$  的右连续增函数, 也是  $y$  的右连续增函数;
- (3)  $\lim_{x \rightarrow -\infty} F(x, y) = 0, \lim_{y \rightarrow -\infty} F(x, y) = 0$ ;
- (4)  $\lim_{x \rightarrow +\infty} F(x, y) = P(Y \leq y), \lim_{y \rightarrow +\infty} F(x, y) = P(X \leq x)$ ;
- (5) 对任何  $x_1 \leq x_2, y_1 \leq y_2$ , 有

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

对性质 (1) - (4) 可以效仿一维随机变量的证明。

现在证明性质 (5), 我们指出, 对一切  $x_1 \leq x_2, y_1 \leq y_2$ , 有

$$\begin{aligned}P(x_1 < X \leq x_2, y_1 < Y \leq y_2) &= P(x_1 < X \leq x_2, Y \leq y_2) - P(x_1 < X \leq x_2, Y \leq y_1) \\ &= P(X \leq x_2, Y \leq y_2) - P(X \leq x_1, Y \leq y_2) \\ &\quad - [P(X \leq x_2, Y \leq y_1) - P(X \leq x_1, Y \leq y_1)] \\ &= F(x_2, y_2) - F(x_1, y_2) - [F(x_2, y_1) - F(x_1, y_1)]\end{aligned}$$

由  $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) \geq 0$  知性质 (5) 成立。

□

若二维随机向量  $\xi = (X, Y)$  有联合密度  $p(x, y)$ , 则  $\xi$  的联合分布函数  $F(x, y)$  与联合密度  $p(x, y)$  有关系式

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y p(u, v) du dv. \quad (6.2.3)$$

**例:** 设二维随机向量  $(X, Y)$  有密度函数

$$p(x, y) = \begin{cases} Ce^{-(2x+y)}, & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases}$$

求 (1) 常数  $C$  的值; (2) 联合分布函数  $F(x, y)$ ; (3) 概率  $P(X \leq Y)$ 。

**解:** (1) 由于

$$1 = C \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y) dx dy = C \int_0^{+\infty} e^{-2x} dx \int_0^{+\infty} e^{-y} dy = \frac{C}{2}$$

得  $C = 2$ 。

(2) 利用公式

$$\begin{aligned} F(x, y) &= \int_{-\infty}^y \int_{-\infty}^x p(t, r) dt dr \\ &= \begin{cases} \int_0^y \int_0^x 2e^{-2t-r} dt dr, & x > 0, y > 0, \\ 0, & \text{其他,} \end{cases} \\ &= \begin{cases} (1 - e^{-2x})(1 - e^{-y}), & x > 0, y > 0, \\ 0, & \text{其他.} \end{cases} \end{aligned}$$

(3) 设区域  $G = \{(x, y) | x \leq y\}$ , 则

$$P(X \leq Y) = P((X, Y) \in G) = \iint_G p(x, y) dy dx = \int_0^{+\infty} \int_x^{+\infty} 2e^{-2x-y} dy dx = \frac{2}{3}.$$

□

**例:** 设  $(X, Y)$  的联合密度函数为

$$p(x, y) = \begin{cases} 1, & 0 < x < 1, |y| < x, \\ 0, & \text{其他.} \end{cases}$$

求边际密度函数。

**解:** 根据定义

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \begin{cases} \int_{-x}^x 1 dy = 2x & x \in (0, 1), \\ 0, & x \notin (0, 1). \end{cases}$$

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx = \begin{cases} \int_y^1 1 dx = 1 - y & y \in [0, 1), \\ \int_{-y}^1 1 dx = 1 + y, & y \in (-1, 0), \\ 0, & y \notin (-1, 1). \end{cases}$$

□

### 6.2.4 二维正态分布

**二维正态分布：**若  $\xi = (X, Y)$  的联合密度  $p(x, y)$  有如下表达式，则称  $\xi$  服从二维（元）正态分布。

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{u^2 + v^2 - 2\rho uv}{2(1-\rho^2)}\right\}, \quad (6.2.4)$$

其中,

$$u = \frac{x - \mu_1}{\sigma_1}, \quad v = \frac{y - \mu_2}{\sigma_2},$$

共有 5 个参数:  $\mu_1, \mu_2 \in \mathbb{R}, \sigma_1, \sigma_2 > 0, \rho \in (-1, 1)$

**例：**设二维随机向量  $\xi = (X, Y)$  服从二维正态分布，试求出  $X$  的分布密度和  $Y$  的分布密度。

**解：**设  $X$  的分布密度为  $p_X(x)$ ，做变量代换  $v = \frac{y - \mu_2}{\sigma_2}$ ，得

$$\begin{aligned} p_X(x) &= \int_{-\infty}^{+\infty} p(x, y) dy \\ &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right\} \\ &\quad \cdot \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{y - \mu_2}{\sigma_2}\right)^2 - 2\rho\frac{y - \mu_2}{\sigma_2}\frac{x - \mu_1}{\sigma_1}\right]\right\} dy \\ &= \frac{1}{2\pi\sigma_1\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right\} \cdot \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}[v^2 - 2\rho v\frac{x - \mu_1}{\sigma_1}]\right\} dv \end{aligned}$$

其中

$$\begin{aligned} &\int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}[v^2 - 2\rho v\frac{x - \mu_1}{\sigma_1}]\right\} dv \\ &= \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(v - \rho\frac{x - \mu_1}{\sigma_1}\right)^2 - \rho^2\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right]\right\} dv \\ &= \exp\left\{\frac{\rho^2}{2(1-\rho^2)}\left(\frac{x - \mu_1}{\sigma_1}\right)^2\right\} \cdot \int_{-\infty}^{+\infty} \exp\left\{-\frac{1}{2(1-\rho^2)}\left(v - \rho\frac{x - \mu_1}{\sigma_1}\right)^2\right\} dv \end{aligned}$$

$$= \exp \left\{ \frac{\rho^2}{2(1-\rho^2)} \left( \frac{x-\mu_1}{\sigma_1} \right)^2 \right\} \sqrt{2\pi(1-\rho^2)}.$$

于是

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp \left\{ -\frac{(x-\mu_1)^2}{2\sigma_1^2} \right\}.$$

同理知

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp \left\{ -\frac{(y-\mu_2)^2}{2\sigma_2^2} \right\}.$$

这表明  $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ .

**例：**假定  $(\xi_1, \xi_2) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ ，试求  $(\xi_1, \xi_2)$  落在

$$D = \left\{ (x, y) \mid \left( \frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left( \frac{y-\mu_2}{\sigma_2} \right)^2 \leq \lambda^2 \right\}$$

内的概率。

**解：**所求概率

$$\begin{aligned} \iint_D p(x, y) dx dy &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \iint_D \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[ \left( \frac{x-\mu_1}{\sigma_1} \right)^2 - 2\rho \frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \left( \frac{y-\mu_2}{\sigma_2} \right)^2 \right] \right\} dx dy \end{aligned}$$

做变量代换  $u = \frac{x-\mu_1}{\sigma_1} - \rho \frac{y-\mu_2}{\sigma_2}$ ,  $v = \sqrt{1-\rho^2} \frac{y-\mu_2}{\sigma_2}$ , 则

$$\left| \frac{\partial(u, v)}{\partial(x, y)} \right| = \begin{vmatrix} \frac{1}{\sigma_1} & 0 \\ -\frac{\rho}{\sigma_2} & \frac{\sqrt{1-\rho^2}}{\sigma_2} \end{vmatrix} = \frac{\sqrt{1-\rho^2}}{\sigma_1\sigma_2}, \quad |J| = \frac{\sigma_1\sigma_2}{\sqrt{1-\rho^2}}.$$

从而

$$\begin{aligned} \iint_D p(x, y) dx dy &= \frac{1}{2\pi(1-\rho^2)} \iint_{\{u^2+v^2 \leq \lambda^2\}} \exp \left\{ -\frac{u^2+v^2}{2(1-\rho^2)} \right\} du dv \\ &= \frac{1}{2\pi(1-\rho^2)} \int_0^{2\pi} \int_0^\lambda \exp \left\{ -\frac{r^2}{2(1-\rho^2)} \right\} r dr d\theta \\ &= \int_0^{\frac{\lambda^2}{2(1-\rho^2)}} e^{-t} dt = 1 - \exp \left\{ -\frac{\lambda^2}{2(1-\rho^2)} \right\}. \end{aligned}$$

□

### 6.3 条件分布

**条件分布函数：**设  $X$  和  $Y$  是两个随机变量，给定实数  $y$ ，如果  $P(Y = y) > 0$ ，则称  $x$  的函数  $P(X \leq x|Y = y)$  为在  $Y = y$  的条件下  $X$  的**条件分布函数**，记作  $F_{X|Y}(x|y)$ ，显然，根据条件概率的定义，有

$$F_{X|Y}(x|y) = \frac{P(X \leq x, Y = y)}{P(Y = y)}.$$

#### 6.3.1 离散型情形

设  $(X, Y)$  是二维离散型随机向量，其概率分布为

$$P(X = x_i, Y = y_j) = p_{ij} \quad (i = 1, 2, \dots; j = 1, 2, \dots),$$

这里  $P(Y = y_i) > 0 \quad (j \geq 1)$ ，则在  $Y = y_i$  的条件下  $X$  的条件分布为

$$P(X = x_i|Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{ij}}{\sum_k p_{kj}} \quad (i = 1, 2, \dots).$$

**例：**设随机变量  $X$  与  $Y$  相互独立， $X$  服从参数为  $\lambda_1$  的泊松分布， $Y$  服从参数为  $\lambda_2$  的泊松分布，试求在  $X + Y = n$  条件下  $X$  的条件分布 ( $n$  为正整数)。

**解：**由于  $X + Y$  服从参数为  $\lambda_1 + \lambda_2$  的泊松分布，故对  $k = 0, 1, \dots, n$  有

$$\begin{aligned} P(X = k|X + Y = n) &= \frac{P(X = k, X + Y = n)}{P(X + Y = n)} \\ &= \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} \\ &= \frac{\lambda_1^k}{k!} e^{-\lambda_1} \frac{\lambda_2^{n-k}}{(n-k)!} e^{-\lambda_2} \bigg/ \left[ \frac{1}{n!} (\lambda_1 + \lambda_2)^n e^{-(\lambda_1 + \lambda_2)} \right] \\ &= C_n^k \left( \frac{\lambda_1}{\lambda_1 + \lambda_2} \right)^k \left( \frac{\lambda_2}{\lambda_1 + \lambda_2} \right)^{n-k}. \end{aligned}$$

这表明，在  $X + Y = n$  的条件下  $X$  的条件分布列为参数为  $n, \frac{\lambda_1}{\lambda_1 + \lambda_2}$  的二项分布。  $\square$

**例：**设随机变量  $X$  与  $Y$  相互独立，都服从参数是  $n, p$  的二项分布，试求在  $X + Y = m (0 \leq m \leq 2n)$  条件下  $X$  的条件分布。

**解：**记  $l = \min\{n, m\}$ ，易知

$$P(X + Y = m) = \sum_{i=0}^l P(X = i, Y = m - i)$$

$$\begin{aligned}
&= \sum_{i=0}^l P(X=i)P(Y=m-i) \\
&= \sum_{i=0}^l C_n^i p^i (1-p)^{n-i} C_n^{m-i} p^{m-i} (1-p)^{n-m+i} \\
&= p^m (1-p)^{2n-m} \sum_{i=0}^l C_n^i C_n^{m-i} \\
&= C_{2n}^m p^m (1-p)^{2n-m}.
\end{aligned}$$

于是, 当  $k=0, 1, \dots, l$  时,

$$\begin{aligned}
P(X=k|X+Y=m) &= \frac{P(X=k, X+Y=m)}{P(X+Y=m)} \\
&= \frac{C_n^k p^k (1-p)^{n-k} C_n^{m-k} p^{m-k} (1-p)^{n-m+k}}{C_{2n}^m p^m (1-p)^{2n-m}} \\
&= \frac{C_n^k C_n^{m-k}}{C_{2n}^m}.
\end{aligned}$$

当  $k > l$  时, 显然  $P(X=k|X+Y=m) = 0$ 。

由此可见, 在  $X+Y=m$  条件下  $X$  的条件分布是超几何分布。  $\square$

**例:** 一射手进行射击, 击中目标的概率  $p \in (0, 1)$ , 射击至击中目标两次为止。若以  $X$  表示首次击中目标所进行的射击次数, 以  $Y$  表示总共进行的射击次数。试求  $X$  和  $Y$  的联合分布列及条件分布列。

**解:**  $Y=n$  表示第  $n$  次击中目标且前  $n-1$  次恰有一次击中目标,

$$P(X=m, Y=n) = p^2(1-p)^{n-2}, \quad n=2, 3, \dots, m=1, \dots, n-1.$$

从而

$$\begin{aligned}
P(X=m) &= \sum_{n=m+1}^{\infty} P(X=m, Y=n) \\
&= \sum_{n=m+1}^{\infty} p^2(1-p)^{n-2} \\
&= p(1-p)^{m-1}, \quad m=1, 2, \dots
\end{aligned}$$

且

$$\begin{aligned}
 P(Y = n) &= \sum_{m=1}^{n-1} P(X = m, Y = n) \\
 &= \sum_{m=1}^{n-1} p^2(1-p)^{n-2} \\
 &= (n-1)p^2(1-p)^{n-2}, \quad n = 2, 3, \dots
 \end{aligned}$$

于是当  $n = 2, 3, \dots$  时,

$$P(X = m|Y = n) = \frac{p^2(1-p)^{n-2}}{(n-1)p^2(1-p)^{n-2}} = \frac{1}{n-1}, \quad m = 1, \dots, n-1.$$

当  $m = 1, 2, \dots$  时,

$$P(Y = n|X = m) = \frac{p^2(1-p)^{n-2}}{p(1-p)^{m-1}} = p(1-p)^{n-m-1}, \quad n = m+1, m+2, \dots$$

□

### 6.3.2 连续型情形

设二维随机向量  $(X, Y)$  有联合分布函数  $F(x, y)$ , 联合密度  $p(x, y)$ , 若  $p_Y(y) > 0$ , 则在  $Y = y$  条件下  $X$  的条件分布函数为

$$F_{X|Y}(x|y) = \int_{-\infty}^x \frac{p(u, y)}{p_Y(y)} du.$$

自然, 在  $Y = y$  条件下  $X$  的条件分布密度为

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}.$$

**连续场合的全概率公式:** 由基本公式

$$p(x, y) = p_Y(y)p(x|y) = p_X(x)p(y|x),$$

连续场合的全概率公式为:

$$\begin{aligned}
 p_X(x) &= \int_{-\infty}^{+\infty} p(x, y) dy = \int_{-\infty}^{+\infty} p_Y(y)p(x|y) dy, \\
 p_Y(y) &= \int_{-\infty}^{+\infty} p(x, y) dx = \int_{-\infty}^{+\infty} p_X(x)p(y|x) dx.
 \end{aligned}$$

连续场合的贝叶斯公式:

$$p(y|x) = \frac{p_Y(y)p(x|y)}{\int_{-\infty}^{+\infty} p_Y(y)p(x|y)dy},$$

$$p(x|y) = \frac{p_X(x)p(y|x)}{\int_{-\infty}^{+\infty} p_X(x)p(y|x)dx}.$$

**例:** 设二维随机向量  $(X, Y)$  满足二维正态分布, 易知  $Y$  的分布密度为

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}.$$

则在  $Y = y$  条件下  $X$  的分布密度为

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}$$

$$= \frac{1}{\sqrt{2\pi(1-\rho^2)}\sigma_1} \exp\left\{-\frac{(x-m)^2}{2(1-\rho^2)\sigma_1^2}\right\},$$

其中  $m = \mu_1 + \rho\frac{\sigma_1}{\sigma_2}(y - \mu_2)$ . □

**例:** 设二维随机向量  $(X, Y)$  的联合密度为

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

给定  $y > 0$ , 试求出条件概率  $P(X > 1|Y = y)$ 。

**解:** 在  $Y = y$  条件下  $X$  的条件分布密度是

$$p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)}.$$

其中  $p_Y(y) = \int_{-\infty}^{+\infty} p(x, y)dx = \int_0^{+\infty} \frac{1}{y}e^{-y} \cdot e^{-\frac{x}{y}}dx = e^{-y}$ , 于是

$$p_{X|Y}(x|y) = \begin{cases} \frac{1}{y}e^{-\frac{x}{y}}, & x > 0, y > 0 \\ 0, & x \leq 0, y > 0 \end{cases}$$

因此

$$P(X > 1|Y = y) = \int_1^{+\infty} \frac{1}{y}e^{-\frac{x}{y}}dx = e^{-\frac{1}{y}}.$$

□

**例:** 设随机变量  $X$  在区间  $(0, 1)$  上随机取值, 当观察到  $X = x(0 < x < 1)$  时, 随机变量  $Y$  在区间  $(x, 1)$  上随机取值, 求  $Y$  的概率密度函数  $p_Y(y)$ 。



**解：** $X$  服从区间  $(0, 1)$  上的均匀分布，对任意的  $x \in (0, 1)$ ，在  $X = x$  条件下， $Y$  的条件概率密度为

$$p(y|x) = \begin{cases} \frac{1}{1-x}, & y \in (x, 1), \\ 0, & \text{其他.} \end{cases}$$

从而，

$$p(x, y) = p(y|x)p_X(x) = \begin{cases} \frac{1}{1-x}, & 0 < x < y < 1, \\ 0, & \text{其他.} \end{cases}$$

故

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y)dx = \begin{cases} \int_0^y \frac{1}{1-x} dx = -\ln(1-y), & 0 < y < 1, \\ 0, & y \notin (0, 1). \end{cases}$$

## 6.4 随机变量的独立性

**随机变量的独立性：**设  $X$  和  $Y$  都是随机变量，如果对任何  $a < b, c < d$ ，事件  $\{a < X < b\}$  和事件  $\{c < Y < d\}$  相互独立，则称  $X$  与  $Y$  **相互独立**。

**定理：**设随机变量  $X$  的可能值是  $x_1, x_2, \dots$ （有限个或无穷可列个），随机变量  $Y$  的可能值是  $y_1, y_2, \dots$ （有限个或无穷可列个），则  $X$  与  $Y$  相互独立的充分必要条件是，对一切  $i, j$  下式成立：

$$P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_j).$$

**定理：**设随机变量  $X, Y$  分别有分布密度  $p_X(x), p_Y(y)$ ，则  $X$  与  $Y$  相互独立的充分必要条件是二元函数  $p(x, y) = p_X(x)p_Y(y)$  是二维随机向量  $(X, Y)$  的联合密度。

**证明：**充分性：设  $p_X(x)p_Y(y)$  是  $(X, Y)$  的联合密度，则对于任何  $a < b, c < d$  有

$$\begin{aligned} P(a < X < b, c < Y < d) &= \int_a^b \int_c^d p_X(x)p_Y(y)dx dy \\ &= \int_a^b p_X(x)dx \cdot \int_c^d p_Y(y)dy = P(a < X < b)P(c < Y < d). \end{aligned}$$

表明  $X$  与  $Y$  相互独立

必要性：设  $X$  与  $Y$  相互独立，则对任何  $a < b, c < d$  有

$$\begin{aligned} P(a < X < b, c < Y < d) &= P(a < X < b)P(c < Y < d) \\ &= \int_a^b p_X(x)dx \cdot \int_c^d p_Y(y)dy = \int_a^b \int_c^d p_X(x)p_Y(y)dx dy \end{aligned}$$

表明  $p_X(x)p_Y(y)$  是  $(X, Y)$  的联合密度。 □

**推论：**设二维随机向量  $(X, Y)$  的联合密度  $p(x, y)$  可以表示为

$$p(x, y) = f(x)g(y),$$

其中  $f(x) \geq 0$ ,  $g(y) \geq 0$ , 且  $\int_{-\infty}^{+\infty} f(x)dx$  收敛, 则  $X$  与  $Y$  相互独立。

**证明：**由于  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(x, y)dxdy = 1$ , 记  $c \triangleq \int_{-\infty}^{+\infty} f(x)dx > 0$ , 推知  $X$  的分布密度是  $p_X(x) = \frac{1}{c}f(x)$ ,  $Y$  的分布密度是  $p_Y(y) = cg(y)$ , 则  $p(x, y) = f(x)g(y) = p_X(x)p_Y(y)$ , 因此  $X$  与  $Y$  相互独立。  $\square$

**定理：**设  $\xi = (X, Y)$  是二维随机向量,  $X$  的分布函数是  $F_X(x)$ ,  $Y$  的分布函数是  $F_Y(y)$ , 则  $X$  和  $Y$  相互独立的充分必要条件是  $\xi$  的分布函数  $F(x, y)$  等于  $F_X(x)$  与  $F_Y(y)$  之积, 即

$$F(x, y) = F_X(x)F_Y(y). \quad (6.4.1)$$

**证明：**必要性：设  $X$  与  $Y$  相互独立, 则对任何  $n \geq 1$ , 事件  $\{-n < X \leq x\}$  与事件  $\{-n < Y \leq y\}$  相互独立, 于是

$$P(-n < X \leq x, -n < Y \leq y) = P(-n < X \leq x)P(-n < Y \leq y).$$

令  $n \rightarrow \infty$ , 即知6.4.1式成立。

充分性：设6.4.1式成立, 对任何  $a < b$ ,  $c < d$ , 有

$$\begin{aligned} P(a < X \leq b, c < Y \leq d) &= F(b, d) - F(a, d) - F(b, c) + F(a, c) \\ &= F_X(b)F_Y(d) - F_X(a)F_Y(d) - F_X(b)F_Y(c) + F_X(a)F_Y(c) \\ &= (F_X(b) - F_X(a))(F_Y(d) - F_Y(c)) \\ &= P(a < X \leq b)P(c < Y \leq d). \end{aligned}$$

由此知  $X$  与  $Y$  相互独立。  $\square$

**定理：**若随机变量  $X$  和  $Y$  相互独立, 且方差  $D(X)$  和  $D(Y)$  存在, 则  $D(X+Y) = D(X) + D(Y)$ 。

**证明：**由式5.3.2知

$$D(X+Y) = E(X+Y)^2 - (E(X+Y))^2 = (E(X^2) + E(Y^2) - 2E(XY)) - ((E(X))^2 + (E(Y))^2 + 2E(X)E(Y))$$

由独立的性质知  $E(XY) = E(X)E(Y)$ , 则

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

**例：**设二维随机向量  $(X, Y)$  服从参数为  $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$  的二维正态分布，则  $X$  与  $Y$  相互独立的充分必要条件是  $\rho = 0$ 。

**证明：**已求出  $X$  和  $Y$  的分布密度：

$$p_X(x) = \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right\}.$$

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}.$$

于是

$$p_X(x)p_Y(y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(y-\mu_2)^2}{2\sigma_2^2}\right\}.$$

结合联合密度  $p(x, y)$  (式6.2.4)，知当  $\rho = 0$  时，

$$p(x, y) = p_X(x)p_Y(y).$$

故  $X$  与  $Y$  相互独立。

若  $X$  与  $Y$  相互独立，则  $p_X(x)p_Y(y)$  是  $(X, Y)$  的联合密度，由于  $p_X(x), p_Y(y), p(x, y)$  均为连续函数，故

$$p(x, y) \equiv p_X(x)p_Y(y).$$

特别地  $p(\mu_1, \mu_2) = p_X(\mu_1)p_Y(\mu_2)$ ，于是

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} = \frac{1}{2\pi\sigma_1\sigma_2},$$

从而  $\rho = 0$ 。 □

**例：**设  $(X, Y)$  联合密度为

$$p(x, y) = \begin{cases} 8xy, & 0 \leq x \leq y \leq 1, \\ 0, & \text{其他.} \end{cases}$$

问  $X$  与  $Y$  是否独立？

**解：**易得

$$p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy = \begin{cases} \int_x^1 8xy dy = 4x(1-x^2), & x \in [0, 1], \\ 0, & \text{其他.} \end{cases}$$

$$p_Y(y) = \int_{-\infty}^{+\infty} p(x, y) dx = \begin{cases} \int_0^y 8xy dx = 4y^3, & y \in [0, 1], \\ 0, & \text{其他.} \end{cases}$$

从而，

$$p(x, y) \neq p_X(x)p_Y(y) = \begin{cases} 16x(1-x^2)y^3, & x, y \in [0, 1], \\ 0, & \text{其他.} \end{cases}$$

故  $X, Y$  不独立。  $\square$

**例：**假定一天内进入邮局的人数为服从参数  $\lambda$  的泊松分布的随机变量，如果每个进入邮局的人为男性的概率为  $p$ ，为女性的概率为  $1 - p$ ，证明进入邮局的男人数和女人数是相互独立的泊松随机变量，且参数分别为  $\lambda p$  和  $\lambda(1 - p)$ 。

**解：**设  $X$  和  $Y$  分别是进入邮局的男人数和女人数，则对任意的自然数  $i$  和  $j$ ，

$$P(X = i, Y = j) = P(X = i, Y = j | X + Y = i + j)P(X + Y = i + j).$$

注意到

$$P(X + Y = i + j) = e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!}.$$

且在给定  $i + j$  人进入邮局的条件下，恰有  $i$  个男人和  $j$  个女人的概率是  $C_{i+j}^i p^i (1-p)^j$ ，从而

$$P(X = i, Y = j) = C_{i+j}^i p^i (1-p)^j e^{-\lambda} \frac{\lambda^{i+j}}{(i+j)!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}.$$

故

$$P(X = i) = e^{-\lambda p} \frac{(\lambda p)^i}{i!} \sum_{j=0}^{+\infty} e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} = e^{-\lambda p} \frac{(\lambda p)^i}{i!}, \quad i \in \mathbb{N}.$$

且

$$P(X = j) = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!} \sum_{i=0}^{+\infty} e^{-\lambda p} \frac{(\lambda p)^i}{i!} = e^{-\lambda(1-p)} \frac{(\lambda(1-p))^j}{j!}, \quad j \in \mathbb{N}.$$

$\square$

## 6.5 两个随机变量的函数

### 6.5.1 随机向量函数的概率分布

**随机向量函数的概率分布：**假设二维随机向量  $(X, Y)$  有联合密度  $p(x, y)$ （对于离散型情形，有类似的结论），随机变量  $Z = f(X, Y)$ ，对于任何实数  $z$ ，令  $A = \{(x, y) : f(x, y) \leq z\}$ ，则  $Z$  的分布函数的计算公式为

$$P(Z \leq z) = P(Z \in A) = \iint_A p(x, y) dx dy. \quad (6.5.1)$$

**定理：**设二维随机向量  $(X, Y)$  有联合密度  $p(x, y)$ ，随机变量  $Z = X + Y$ ，则  $Z$  的分布密度为

$$p_Z(z) = \int_{-\infty}^{+\infty} p(x, z - x) dx,$$

证明：令

$$A = \{(x, y) : x + y \leq z\}$$

由式6.5.1知

$$P(Z \leq z) = P((X, Y) \in A) = \iint_{\{x+y \leq z\}} p(x, y) dx dy.$$

利用变量替换  $u = x + y$  有

$$\begin{aligned} \iint_{\{x+y \leq z\}} p(x, y) dx dy &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{z-x} p(x, y) dy \right) dx \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^z p(x, u-x) du \right) dx \\ &= \int_{-\infty}^z \left( \int_{-\infty}^{+\infty} p(x, u-x) dx \right) du. \end{aligned}$$

因此

$$P(Z \leq z) = \int_{-\infty}^z \left( \int_{-\infty}^{+\infty} p(x, u-x) dx \right) du.$$

因此  $Z$  的分布函数为  $p_Z(z) = \int_{-\infty}^{+\infty} p(x, z-x) dx$ . □

**推论：**设随机变量  $X$  和  $Y$  分别有分布密度  $p_X(x)$  和  $p_Y(y)$ ，且  $X$  和  $Y$  相互独立，则随机变量  $Z = X + Y$  有分布密度

$$p_Z(z) = \int_{-\infty}^{+\infty} p_X(x) p_Y(z-x) dx.$$

**例：**设  $(X, Y)$  服从二维正态分布，联合密度  $p(x, y)$  为  $p(x, y) = \hat{C} \exp \left\{ -\frac{u^2 - 2\rho uv + v^2}{2(1-\rho^2)} \right\}$ ，其中  $u = \frac{x-\mu_1}{\sigma_1}$ ,  $v = \frac{y-\mu_2}{\sigma_2}$ ,  $\hat{C} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}$ . 求  $Z = X + Y$  的密度。

**解：**由定理知  $Z$  的分布密度为  $p_Z(z) = \int_{-\infty}^{\infty} p(x, z-x) dx$ . 当  $y$  取  $z-x$  时，

$$v = \frac{y-\mu_2}{\sigma_2} = \frac{z-(\mu_1+\sigma_1 u)-\mu_2}{\sigma_2} = C - \frac{\sigma_1}{\sigma_2} u,$$

其中,  $C = (z - \mu_1 - \mu_2) / \sigma_2$ .

此时,

$$\begin{aligned} u^2 - 2\rho uv + v^2 &= u^2 - 2\rho u \left( C - \frac{\sigma_1 u}{\sigma_2} \right) + \left( C - \frac{\sigma_1 u}{\sigma_2} \right)^2 \\ &= \left( 1 + 2\rho \frac{\sigma_1}{\sigma_2} + \left( \frac{\sigma_1}{\sigma_2} \right)^2 \right) u^2 - 2 \left( \rho + \frac{\sigma_1}{\sigma_2} \right) C u + C^2. \end{aligned}$$

现在计算  $p_Z(z) = \int_{-\infty}^{\infty} p(x, z-x) dx$ ，已知：

$$\begin{aligned} p(x, z-x) &= \hat{C} \left\{ -\frac{Au^2 - 2Bu + C^2}{2(1-\rho^2)} \right\}, \quad \text{其中, } u = \frac{x-\mu_1}{\sigma_1}, \\ A &= 1 + 2\rho \frac{\sigma_1}{\sigma_2} + \left( \frac{\sigma_1}{\sigma_2} \right)^2, \quad B = \left( \rho + \frac{\sigma_1}{\sigma_2} \right) C, \quad C = \frac{z - (\mu_1 + \mu_2)}{\sigma_2}. \end{aligned}$$

配方:

$$Au^2 - 2Bu + C^2 = A \left( u - \frac{B}{A} \right)^2 - \left( \frac{B^2}{A} - C^2 \right)$$

于是,

$$\begin{aligned} p_Z(z) &= \hat{C} \exp \left\{ \frac{\frac{B^2}{A} - C^2}{2(1-\rho^2)} \right\} \times \int_{-\infty}^{\infty} \exp \left\{ -\frac{A \left( u - \frac{B}{A} \right)^2}{2(1-\rho^2)} \right\} \sigma_1 du \\ &= \tilde{C} \exp \left\{ \frac{B^2 - AC^2}{2(1-\rho^2)A} \right\}. \quad \tilde{C} = \hat{C} \sigma_1 \sqrt{2\pi \frac{1-\rho^2}{A}} = \frac{1}{\sqrt{2\pi\sigma_2^2 A}} \end{aligned}$$

已有:  $p_Z(z) = \tilde{C} \exp \left\{ \frac{B^2 - AC^2}{2(1-\rho^2)A} \right\}$ , 其中  $\tilde{C}$  是常数,

$$A = 1 + 2\rho \frac{\sigma_1}{\sigma_2} + \left( \frac{\sigma_1}{\sigma_2} \right)^2, B = \left( \rho + \frac{\sigma_1}{\sigma_2} \right) C, C = \frac{z - (\mu_1 + \mu_2)}{\sigma_2}.$$

$$B^2 - AC^2 = \left( \left( \rho + \frac{\sigma_1}{\sigma_2} \right)^2 - A \right) C^2 = (\rho^2 - 1) \frac{(z - (\mu_1 + \mu_2))^2}{\sigma_2^2}.$$

因此,

$$p_Z(z) = \tilde{C} \exp \left\{ -\frac{(z - \mu)^2}{2\sigma^2} \right\} = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(z - \mu)^2}{2\sigma^2} \right\}.$$

其中,  $\mu = \mu_1 + \mu_2, \sigma^2 = \sigma_2^2 A = \sigma_1^2 + \sigma_2^2 + 2\rho\sigma_1\sigma_2$ .

特别地, 若  $\rho = 0$  (即  $X, Y$  相互独立), 则

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

□

**例:** 设随机变量  $X$  与  $Y$  相互独立, 且  $X, Y$  分别有分布密度:

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

$$p_Y(y) = \begin{cases} \mu e^{-\mu y}, & y > 0, \\ 0, & y \leq 0, \end{cases} \quad (\mu > 0),$$

试求随机变量  $X + Y$  的分布密度。

**解:** 随机变量  $Z = X + Y$  的分布密度为

$$p(z) = \int_{-\infty}^{\infty} p_X(x) p_Y(z - x) dx.$$

易知, 当  $z \leq 0$  时,  $p(z) = 0$ , 设  $z > 0$ , 则

$$p(z) = \int_0^z \lambda e^{-\lambda x} \mu e^{-\mu(z-x)} dx = \lambda \mu e^{-\mu z} \int_0^z e^{-(\lambda-\mu)x} dx$$

$$= \begin{cases} \lambda^2 e^{-\lambda z} z, & \lambda = \mu, \\ \frac{\lambda \mu}{\lambda - \mu} (e^{-\mu z} - e^{-\lambda z}), & \lambda \neq \mu. \end{cases}$$

□

**定理:** 设二维随机向量  $(X, Y)$  有联合密度  $p(x, y)$ . 令  $Z = X/Y$  (当  $Y = 0$  时, 规定  $Z = 0$ ). 则  $Z$  为连续型, 且

$$p_Z(z) = \int_{-\infty}^{\infty} |y| p(zy, y) dy.$$

**证明:** 首先,  $\frac{x}{y} \leq z$  当且仅当 “ $y > 0$  且  $x \leq yz$ ” 或者 “ $y < 0$  且  $x \geq yz$ .” 于是,

$$F_Z(z) = P(Y > 0, X \leq Yz) + P(Y < 0, X \geq Yz).$$

其中,

$$\begin{aligned} P(Y > 0, X \leq Yz) &= \int_0^{\infty} \int_{-\infty}^{yz} p(x, y) dx dy = \int_0^{\infty} \int_{-\infty}^z p(yu, y) y du dy \\ &= \int_{-\infty}^z \left( \int_0^{\infty} y p(yu, y) dy \right) du \end{aligned}$$

类似的,

$$\begin{aligned} P(Y < 0, X \geq Yz) &= \int_{-\infty}^0 \int_{yz}^{\infty} p(x, y) dx dy = \int_{-\infty}^0 \int_{-\infty}^z p(yu, y) |y| du dy \\ &= \int_{-\infty}^z \left( \int_{-\infty}^0 |y| p(yu, y) dy \right) du \end{aligned}$$

于是,

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^z \left( \int_{-\infty}^{\infty} |y| p(yu, y) dy \right) du \\ p_Z(z) &= \int_{-\infty}^{\infty} |y| p(zy, y) dy. \end{aligned}$$

□

**例:** 随机变量  $X, Y$  相互独立, 都服从  $N(0, 1)$ . 求随机变量  $Z = X/Y$  的概率密度.

**解:** 联合密度为:

$$p(x, y) = \frac{1}{2\pi} \exp \left\{ -\frac{x^2 + y^2}{2} \right\}.$$

因此,

$$\begin{aligned} p_Z(z) &= \int_{-\infty}^{\infty} |y| p(zy, y) dy = \int_{-\infty}^{\infty} |y| \frac{1}{2\pi} \exp \left\{ -\frac{(zy)^2 + y^2}{2} \right\} dy \\ &= \frac{2}{2\pi} \int_0^{\infty} y \exp \left\{ -\frac{(z^2 + 1)y^2}{2} \right\} dy \\ &= \frac{1}{\pi} \int_0^{\infty} e^{-(z^2 + 1)u} du = \frac{1}{\pi(z^2 + 1)}. \end{aligned}$$

□

**例：**设随机变量  $X$  与  $Y$  独立同分布，共同分布是  $N(0, 1)$ ，试求随机变量  $Z = \sqrt{X^2 + Y^2}$  的概率分布。

**解：**对任何  $z \leq 0$ ，易知  $P(Z \leq z) = 0$ ，设  $z > 0$ ，则

$$P(Z \leq z) = \iint_{\{x^2+y^2 \leq z^2\}} \frac{1}{2\pi} \exp\left\{-\frac{x^2+y^2}{2}\right\} dx dy$$

做极坐标变换  $x = r \cos \theta$ ,  $y = r \sin \theta$  ( $0 \leq \theta < 2\pi, r \geq 0$ )，于是

$$P(Z \leq z) = \int_0^{2\pi} \left( \int_0^z \frac{1}{2\pi} e^{-r^2/2} r dr \right) d\theta = \int_0^z r e^{-r^2/2} dr.$$

可见， $Z = \sqrt{X^2 + Y^2}$  有分布密度

$$p(z) = \begin{cases} 0, & z \leq 0, \\ z e^{-z^2/2}, & z > 0. \end{cases}$$

这样的概率分布也称为瑞利分布。

□

**定理：**假设  $\xi = (X, Y)$  为连续型，有密度  $p(x, y)$ ，区域  $A$  满足  $P((X, Y) \in A) = 1$ ，假设

$$\eta = (U, V), \quad \text{其中 } U = f(X, Y), \quad V = g(X, Y).$$

如果：(1)  $P(\xi \in A) = 1$  且  $(f, g) : A \rightarrow G$  是一对一的；

(2)  $f, g \in C^1(A)$ ，且  $\frac{\partial(u, v)}{\partial(x, y)} \neq 0, \forall (x, y) \in A$ ，

那么， $\eta$  是连续型，且

$$p_{U, V}(u, v) = p(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|, (u, v) \in G.$$

**证明：**对于  $\forall D \subseteq G$ ，设  $D^* = \{(x, y) : (f(x, y), g(x, y)) \in D\}$ ，易知  $D^* \subseteq A$ ， $(f(x, y), g(x, y))$  是  $D^*$  到  $D$  上的一一映射，其逆映射是  $(x(u, v), y(u, v))$ ，根据重积分的变量替换公式，

$$\iint_{D^*} p(x, y) dx dy = \iint_D p(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv.$$

于是，

$$P((U, V) \in D) = P((X, Y) \in D^*) = \iint_{D^*} p(x, y) dx dy = \iint_D p(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv.$$

因此

$$p_{U, V}(u, v) = p(x(u, v), y(u, v)) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|, (u, v) \in G.$$

□



### 6.5.2 两个随机变量函数的数学期望

我们首先考虑一个特殊情形： $f(x, y) = xy$ 。

**定理：**设随机变量  $X$  与  $Y$  相互独立，且  $E(X)$  与  $E(Y)$  都存在，则

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

连续情形的证明：

$$E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp_X(x)p_Y(y)dxdy = (E(X))(E(Y)).$$

**定理：**若随机变量  $X$  与  $Y$  相互独立，则

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

**证明：**由于  $E(X + Y) = E(X) + E(Y)$ ，得

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

由  $X$  与  $Y$  相互独立得

$$E(X - E(X))(Y - E(Y)) = E(X - E(X))E(Y - E(Y)) = 0$$

因此等式成立。 □

**均值公式：**(1) 设二维随机向量  $(X, Y)$  的可能值是  $a_1, a_2, \dots$  (有限个或可列无穷个)， $f(x, y)$  是任何二元函数，则

$$E(f(X, Y)) = \sum_i f(a_i)P((X, Y) = a_i).$$

(当  $a_i$  有无穷个时，要求此级数绝对收敛)。

(2) 设二维随机向量  $(X, Y)$  有联合分布密度  $p(x, y)$ ，二元函数  $p(x, y)$  满足积分

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |f(x, y)|p(x, y)dxdy$$

收敛，则

$$E(f(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y)p(x, y)dxdy.$$

**例：**在长为  $a$  的线段上，任取两个点  $X$  和  $Y$ ，求此两点间的平均距离。

**解：**显然  $X$  和  $Y$  服从区间  $(0, a)$  上的均匀分布，且相互独立，从而  $X$  和  $Y$  的联合密度为

$$p(x, y) = \begin{cases} \frac{1}{a^2} & x, y \in (0, a), \\ 0, & \text{其他.} \end{cases}$$

从而，两点间的平均长度为

$$\begin{aligned} E|X - Y| &= \iint_{\mathbb{R}^2} |x - y| p(x, y) dx dy \\ &= \int_0^a \int_0^a |x - y| \frac{1}{a^2} dx dy \\ &= \frac{1}{a^2} \int_0^a \left( \int_0^x (x - y) dy + \int_x^a (y - x) dy \right) dx \\ &= \frac{1}{a^2} \int_0^a \left( x^2 - ax + \frac{a^2}{2} \right) dx = \frac{a}{3}. \end{aligned}$$

□

**例：**  $x, y, z$  为相互独立的随机变量， $h, l, f, g$  为任意确定性映射。判断

- (1) 令  $a = f(x, y)$ ,  $b = g(x, z)$ ,  $a$  与  $b$  是否独立,  $a, b | x$  是否独立?
- (2) 令  $c = x + y$ , 则  $x, y | c$  是否独立?
- (3)  $h(l(x, y), z)$  与  $x$  是否独立?  $h(l(x, y), z)$  与  $x$  在  $l(x, y)$  给定条件下是否独立?

**解：**

- (1)  $a$  与  $b$  不独立，都依赖于  $x$ ,  $a, b | x$  独立。
- (2) 不独立
- (3)  $h(l(x, y), z)$  与  $x$  不独立，给定  $l(x, y)$  则独立。

## 6.6 二维随机向量的数字特征

**两个随机变量的协方差：**假设随机变量  $X, Y$  的期望和方差存在，则称

$$E(X - E(X))(Y - E(Y))$$

为  $X$  与  $Y$  的**协方差**，记为  $\text{cov}(X, Y)$  或  $\sigma_{XY}$ 。

若  $\sigma_{XY} = 0$ ，则称  $X$  与  $Y$  不相关。

注: 协方差存在, 因为

$$2(X - EX)(Y - EY) \leq (X - EX)^2 + (Y - EY)^2.$$

协方差的计算公式为:

$$\text{cov}(X, Y) = E(XY) - (EX)(EY).$$

注意: 协方差为 0 不等价于随机变量  $X$  和  $Y$  独立。

例如, 令随机变量  $X \sim U(0, 2\pi)$ , 设  $Y = \sin X$ ,  $Z = \cos X$ ,  $Y$  和  $Z$  的协方差为

$$\text{cov}(Y, Z) = E(YZ) - E(Y)E(Z) = \frac{1}{2}E(\sin 2X) - E(\sin X)E(\cos X) = 0.$$

而  $Y$  和  $Z$  显然是不独立的。

**定理:** 假设  $X, Y$  的方差存在, 则

$$(\text{cov}(X, Y))^2 \leq \text{var}(X) \cdot \text{var}(Y). \quad (6.6.1)$$

**证明:** 若  $\text{var}(X) = 0$ , 则  $X \equiv c$ , 于是  $\text{cov}(X, Y) = 0$ . 若  $\text{var}(X) > 0$ , 则设

$$\begin{aligned} g(t) &:= E(t(X - EX) + (Y - EY))^2 \\ &= t^2 \text{var}(X) + 2t \text{cov}(X, Y) + \text{var}(Y) \geq 0 \end{aligned}$$

由于不等式恒成立, 故  $g(t)$  的判别式  $\leq 0$ , 即  $(\text{cov}(X, Y))^2 \leq \text{var}(X) \cdot \text{var}(Y)$ . □

**随机变量的相关系数:** 设  $0 < \text{var}(X), \text{var}(Y) < \infty$ , 则称

$$\frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X)}\sqrt{\text{var}(Y)}}$$

为  $X$  与  $Y$  的**相关系数**, 记为  $\rho_{XY}$ , 简记为  $\rho$ 。

**定理:** 设  $\rho$  是随机变量  $X$  与  $Y$  的相关系数, 则有

- (1)  $|\rho| \leq 1$ ;
- (2)  $X$  与  $Y$  独立, 则不相关, 从而  $\rho = 0$ ;
- (3)  $|\rho| = 1$  当且仅当存在  $a, b$  以概率 1 使得  $Y = a + bX$ .

**证明:** (1) 可以直接由式 6.6.1 推知成立。

(2) 若  $X$  与  $Y$  相互独立, 则

$$E(X - E(X))(Y - E(Y)) = E(X - E(X))E(Y - E(Y)) = 0,$$

从而  $\rho = 0$ 。

(3) 设

$$\begin{aligned} g(t) &:= E(t(X - EX) + (Y - EY))^2 \\ &= t^2 \text{var}(X) + 2t \text{cov}(X, Y) + \text{var}(Y) \end{aligned}$$

则  $|\rho| = 1$  当且仅当  $g(t)$  的判别式为 0, 即存在  $t_0$  使得

$$\begin{aligned} g(t_0) &= E(t_0(X - EX) + (Y - EY))^2 = 0 \\ &\Leftrightarrow Y = -t_0 X + EX + EY. \end{aligned}$$

□

**例:** 设  $(X, Y)$  服从二维正态分布, 联合密度为

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} \quad u = \frac{x-\mu_1}{\sigma_1}, v = \frac{y-\mu_2}{\sigma_2}.$$

求  $\rho_{XY}$ 。

**解:** 由之前的结论,  $\mu_1 = E(X), \mu_2 = E(Y), \sigma_1^2 = \text{var}(X), \sigma_2^2 = \text{var}(Y)$ 。

故

$$\begin{aligned} \rho_{XY} &= \frac{E(X - \mu_1)(Y - \mu_2)}{\sigma_1\sigma_2} = E\left(\frac{X - \mu_1}{\sigma_1} \frac{Y - \mu_2}{\sigma_2}\right) \\ &= \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \cdot e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} dv du. \end{aligned}$$

先对  $v$  积分,  $v^2 - 2\rho uv + u^2 = (v - \rho u)^2 + (1 - \rho^2)u^2$ ,

$$\begin{aligned} \int_{-\infty}^{\infty} uv \cdot e^{-\frac{1}{2(1-\rho^2)}(u^2+v^2-2\rho uv)} dv &= ue^{-\frac{u^2}{2}} \times \int_{-\infty}^{\infty} ve^{-\frac{(v-\rho u)^2}{2(1-\rho^2)}} dv \\ &= ue^{-\frac{u^2}{2}} \times \sqrt{2\pi(1-\rho^2)} \cdot \rho u. \end{aligned}$$

代入积分式, 再对  $u$  积分,

$$\rho_{XY} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \rho u^2 e^{-\frac{u^2}{2}} du = \rho.$$

□

**例:** 设二维随机向量  $(X, Y)$  的联合密度是

$$p(x, y) = \begin{cases} \frac{2}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\}, & xy > 0, \\ 0, & \text{其他.} \end{cases}$$

求  $\rho_{XY}$ 。

解：上一讲第二节已经指出  $X \sim N(0, 1)$ ,  $Y \sim N(0, 1)$ , 故

$$E(X) = E(Y) = 0, \quad \text{var}(X) = \text{var}(Y) = 1.$$

因此

$$\begin{aligned} \text{cov}(X, Y) &= E(XY) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xyp(x, y)dx dy \\ &= \iint_{\{(x, y): xy > 0\}} xy \frac{2}{2\pi} \exp\left\{-\frac{1}{2}(x^2 + y^2)\right\} dx dy \\ &= \int_0^{\infty} ye^{-y^2/2} \left( \int_0^{\infty} \frac{2}{2\pi} xe^{-x^2/2} dx \right) dy \\ &\quad + \int_{-\infty}^0 ye^{-y^2/2} \left( \int_{-\infty}^0 \frac{2}{2\pi} xe^{-x^2/2} dx \right) dy \\ &= \int_0^{+\infty} \frac{2}{2\pi} ye^{-y^2/2} dy - \int_{-\infty}^0 \frac{2}{2\pi} ye^{-y^2/2} dy \\ &= \int_0^{+\infty} \frac{2}{\pi} ye^{-y^2/2} dy = \frac{2}{\pi} \end{aligned}$$

故相关系数为  $\rho_{XY} = \frac{2}{\pi}$ 。

□

## 6.7 条件期望

**条件期望的定义：** 设  $X$  和  $Y$  是两个随机变量。

(1) 若在  $Y = y$  的条件下  $X$  的可能值是  $x_1, x_2, \dots$  (有限个或无穷可列个), 条件概率分布是  $P(X = x_i | Y = y_i) (i = 1, 2, \dots)$  则称

$$\sum_i x_i P(X = x_i | Y = y)$$

为在  $Y = y$  条件下  $X$  的**条件期望**, 记为  $E(X | Y = y)$ 。

(2) 若在  $Y = y$  的条件下  $X$  有条件分布密度  $p_{X|Y}(x|y)$ , 则称积分

$$\int_{-\infty}^{\infty} xp_{X|Y}(x|y)dx$$

为在  $Y = y$  的条件下  $X$  的**条件期望**, 记为  $E(X | Y = y)$ 。

设二维随机向量  $(X, Y)$  有联合密度  $p(x, y)$ , 有

$$E(X | Y = y) = \frac{1}{p_Y(y)} \int_{-\infty}^{\infty} xp(x, y)dx.$$

**定理：** 设二维随机向量  $(X, Y)$  有联合密度  $p(x, y)$ ，则

$$E(X) = \int_{\{y: p_Y(y) > 0\}} E(X|Y = y) p_Y(y) dy.$$

**证明：** 首先，若  $p_Y(y) = 0$ ，则对任何  $A > 0$  有

$$\left| \int_{-A}^A xp(x, y) dx \right| \leq A \int_{-A}^A p(x, y) dx \leq A \int_{-\infty}^{+\infty} p(x, y) dx = A p_Y(y) = 0,$$

于是

$$\int_{-\infty}^{+\infty} xp(x, y) dx = \lim_{A \rightarrow \infty} \int_{-A}^A xp(x, y) dx = 0$$

可见

$$\begin{aligned} \int_{\{y: p_Y(y) > 0\}} E(X|Y = y) p_Y(y) dy &= \int_{\{y: p_Y(y) > 0\}} \left( \int_{-\infty}^{+\infty} xp(x, y) dx \right) dy \\ &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} xp(x, y) dy \right) dx = \int_{-\infty}^{+\infty} x \left( \int_{-\infty}^{+\infty} p(x, y) dy \right) dx \\ &= \int_{-\infty}^{+\infty} xp_X(x) dx = E(X). \end{aligned}$$

□

对于离散情形，有类似的定理。

**定理：** 设  $(X, Y)$  是二维随机向量， $Y$  的可能值是  $y_1, y_2, \dots$ （有限个或可列无穷个）， $P(Y = y_i) > 0$  ( $i = 1, 2, \dots$ )， $X$  的可能值是  $x_1, x_2, \dots$ （有限个或可列无穷个），且  $E(X)$  存在，则

$$E(X) = \sum_i E(X|Y = y_i) P(Y = y_i).$$

**证明：** 由于  $P(X = x_k, Y = y_i) = P(X = x_k|Y = y_i) P(Y = y_i)$ ，知

$$\begin{aligned} E(X) &= \sum_k x_k P(X = x_k) = \sum_k x_k \sum_i P(X = x_k, Y = y_i) \\ &= \sum_k \sum_i x_k P(X = x_k|Y = y_i) P(Y = y_i) \\ &= \sum_i E(X|Y = y_i) P(Y = y_i). \end{aligned}$$

□

**例：**设  $(X, Y)$  的联合密度函数为

$$p(x, y) = \begin{cases} \frac{e^{-x/y} e^{-y}}{y}, & x, y \in (0, +\infty), \\ 0, & \text{其他.} \end{cases}$$

求  $E(X|Y = y)$ 。

**解：**对给定的  $y \in (0, +\infty)$ ，在  $Y = y$  条件下  $X$  的条件密度函数为

$$p(x|y) = \begin{cases} \frac{p(x, y)}{p_Y(y)} = \frac{\frac{1}{y} e^{-x/y} e^{-y}}{\int_0^{+\infty} \frac{1}{y} e^{-x/y} e^{-y} dx} = \frac{e^{-x/y}}{y} & x \in (0, +\infty), \\ 0, & x \notin (0, +\infty), \end{cases}$$

因此， $X$  在给定  $Y = y$  条件下的条件分布恰好是参数为  $\frac{1}{y}$  的指数分布。从而

$$E(X|Y = y) = \int_0^{+\infty} x \frac{e^{-x/y}}{y} dx = y.$$

**例：**(对应郑书例 7.7) 一矿工在有三个门的矿井中迷了路，第 1 个门通到一个通道，走 2 个小时可到达地面；第 2 个门通到另一个通道，走 3 个小时又回到原处；第 3 个门通到第 3 个通道，沿它走 5 个小时也回到原处，假定该矿工总是等可能从 3 个门选择任意一个进入通道，试问，该矿工到达地面平均需要多长时间。

**解：**设矿工到达地面所需时间为  $X$ ，选择门的编号为  $Y$ ，则  $P(Y = 1) = P(Y = 2) = P(Y = 3) = \frac{1}{3}$ ，于是

$$E(X) = \sum_{i=1}^3 P(Y = i) E(X|Y = i) = \frac{1}{3} \sum_{i=1}^3 E(X|Y = i).$$

易知， $E(X|Y = 1) = 2$ ， $E(X|Y = 2) = E(X) + 3$ ， $E(X|Y = 3) = E(X) + 5$ ，于是

$$E(X) = \frac{1}{3}(2 + E(X) + 3 + E(X) + 5)$$

推知  $E(X) = 10$ ，即矿工到达地面平均要 10 小时。

## 7.1 $n$ 维随机向量

**$n$  维随机向量及其联合分布函数:** 设  $\xi = (X_1, \dots, X_n)$  是  $n$  维向量, 称

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n), \quad x_1, \dots, x_n \in \mathbb{R}^n$$

为  $\xi$  的联合分布函数, 也记为  $F_\xi$  或  $F_{X_1, \dots, X_n}$ .

**离散型  $n$  维随机向量:** 若  $\xi$  取有限个或可列个值 ( $n$  维向量), 则称  $\xi$  为离散型.

**连续型  $n$  维随机向量:** 若存在非负可积函数  $p(x_1, \dots, x_n)$  使得对任意  $n$  维矩形  $D$  都有

$$P(\xi \in D) = \int \cdots \int_D p(x_1, \dots, x_n) dx_1 \cdots dx_n,$$

则称  $\xi$  为连续型随机向量, 称  $p(x_1, \dots, x_n)$  为  $\xi$  的联合密度, 也记为  $P_{X_1, \dots, X_n}$ . (注: 上式对一般集合  $D$  都成立).

**$n$  维随机向量的边缘分布:** 设  $\xi = (X_1, \dots, X_n)$  是  $n$  维向量, 对任意  $1 \leq k < n, 1 \leq i_1 < \cdots < i_k \leq n$ , 则称  $(X_{i_1}, \dots, X_{i_k})$  为  $\xi$  的 (一个  $k$  维) 边缘, 其分布被称为  $\xi$  的边缘分布.

**例:** (多项分布) 设  $U_1, \dots, U_n$  是取值  $1, \dots, t$  的随机变量, 且相互独立, 都服从如下分布:

$$P(U_i = k) = p_k, \quad k = 1, \dots, t,$$

其中  $t \geq 2, p_k > 0, \forall k$  且  $p_1 + \cdots + p_t = 1$ .

记

$$X_k = |\{1 \leq i \leq n : U_i = k\}| = \sum_{i=1}^n 1_{\{U_i=k\}}.$$

$\xi = (X_1, \dots, X_t)$  的联合分布列:

$$P(\xi = (i_1, \dots, i_t)) = \frac{n!}{i_1! \cdots i_t!} p_1^{i_1} \cdots p_t^{i_t}.$$

因为  $X_t = n - \sum_{s=1}^{t-1} X_s$ ,  $p_t = 1 - \sum_{s=1}^{t-1} p_s$ , 所以  $\xi$  与  $(X_1, \dots, X_{t-1})$  等价.

本例的背景模型为:  $n$  次独立重复试验 (投掷一枚  $t$  面骰子).

**例:** 口袋中有 5 个白球, 8 个黑球, 从中不放回的依次取出 3 个, 若第  $i$  次取出白球, 则  $X_i = 1$ , 否则令  $X_i = 0$ ,  $i = 1, 2, 3$ , 求  $(X_1, X_2, X_3)$  的联合分布列.

**解:**

$$P(X_1 = 0, X_2 = 0, X_3 = 0) = \frac{8 \times 7 \times 6}{13 \times 12 \times 11} = 0.1958$$



$$\begin{aligned}
P(X_1 = 1, X_2 = 0, X_3 = 0) &= P(X_1 = 0, X_2 = 1, X_3 = 0) \\
&= P(X_1 = 0, X_2 = 0, X_3 = 1) = \frac{8 \times 7 \times 5}{13 \times 12 \times 11} = 0.1632 \\
P(X_1 = 1, X_2 = 1, X_3 = 0) &= P(X_1 = 0, X_2 = 1, X_3 = 1) \\
&= P(X_1 = 1, X_2 = 0, X_3 = 1) = \frac{5 \times 4 \times 8}{13 \times 12 \times 11} = 0.0932 \\
P(X_1 = 1, X_2 = 1, X_3 = 1) &= \frac{5 \times 4 \times 3}{13 \times 12 \times 11} = 0.0035
\end{aligned}$$

**独立性:** 若对任意  $a_i < b_i, i = 1, \dots, n$  都有

$$\begin{aligned}
&P(a_1 < X_1 < b_1, \dots, a_n < X_n < b_n) \\
&= P(a_1 < X_1 < b_1) \cdots P(a_n < X_n < b_n)
\end{aligned}$$

则称  $n$  个随机变量  $X_1, \dots, X_n$  相互独立.

若  $X_1, \dots, X_n$  相互独立, 且  $F_{X_i} = F_{X_1}, i = 2, \dots, n$ , 则称  $X_1, \dots, X_n$  独立同分布.

若相互独立, 则上式中的  $a_i < X_i < b_i$  可以改为  $X_i \in B_i$ , 其中  $B_1, \dots, B_n$  为任意一维 Borel 集.

**定理:** 设  $X_1, \dots, X_n (n \geq 2)$  都是随机变量, 分别有分布密度  $p_1(x_1), \dots, p_n(x_n)$ , 则  $X_1, \dots, X_n$  相互独立的充分必要条件是  $n$  元函数

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n) = p_1(x_1) \cdots p_n(x_n).$$

为  $n$  维随机向量  $(X_1, \dots, X_n)$  的联合密度.

对于离散型随机向量, 有类似的结论, 设  $n$  个随机变量的取值分别为  $X_1 = x_1^{(1)}, x_2^{(1)}, \dots; \dots; X_n = x_1^{(n)}, x_2^{(n)}, \dots$ , 则  $X_1, \dots, X_n$  相互独立的充分必要条件是

$$\begin{aligned}
&P(X_1 = x_{i_1}^{(1)}, \dots, X_n = x_{i_n}^{(n)}) \\
&= P(X_1 = x_{i_1}^{(1)}) \cdots P(X_n = x_{i_n}^{(n)}) = p_{i_1}^{(1)} \cdots p_{i_n}^{(n)}
\end{aligned}$$

**定义:** 若  $X_i$  与  $X_j$  相互独立,  $\forall i \neq j$ , 则称  $X_1, \dots, X_n$  两两独立.

**例:** 甲、乙玩石头剪刀布. 甲出  $X$ , 乙出  $Y$ , 结局为  $Z$ . 则  $X, Y, Z$  两两独立, 但不相互独立.

**例:** 设随机向量  $(X, Y, Z)$  在矩形区域  $a < x < b, c < y < d, e < z < f$  内服从均匀分布, 求  $X, Y, Z$  的分布密度函数, 以及  $X, Y, Z$  是否相互独立.

**解:** 由均匀分布定义

$$p(x, y, z) = \frac{1}{(b-a)(d-c)(f-e)} \quad a < x < b, c < y < d, e < z < f.$$

当  $x, y, z$  所在边界矩形是独立的, 且在矩形内时有:

$$p_X(x) = \int_e^f \int_c^d \frac{1}{(b-a)(d-c)(f-e)} dydz = \frac{1}{b-a}$$

$$p_Y(y) = \int_e^f \int_a^b \frac{1}{(b-a)(d-c)(f-e)} dx dz = \frac{1}{d-c}$$

$$p_Z(z) = \int_e^f \int_a^b \frac{1}{(b-a)(d-c)(f-e)} dx dy = \frac{1}{f-e}.$$

由于  $p(x, y, z) = p_X(x)p_Y(y)p_Z(z)$ , 因此  $X, Y, Z$  之间相互独立。

**定义:** 设  $\mathbf{X} = (X_1, \dots, X_m)$  和  $\mathbf{Y} = (Y_1, \dots, Y_n)$  分别是  $m$  维和  $n$  维随机向量, 给定  $\mathbf{y} = (y_1, \dots, y_n)$ , 若  $P(\mathbf{Y} = \mathbf{y}) > 0$ , 则  $x_1, \dots, x_m$  的函数

$$P(X_1 \leq x_1, \dots, X_m \leq x_m | \mathbf{Y} = \mathbf{y})$$

称为在  $\mathbf{Y} = \mathbf{y}$  条件下  $\mathbf{X}$  的**条件分布函数**, 记为  $F_{\mathbf{X}|\mathbf{Y}}(x_1, \dots, x_m | \mathbf{y})$ .

若  $\mathbf{X} = (X_1, \dots, X_m)$  和  $\mathbf{Y} = (Y_1, \dots, Y_n)$  有联合密度  $p(x_1, \dots, x_m, y_1, \dots, y_n)$ , 则

$$F_{\mathbf{X}|\mathbf{Y}}(x_1, \dots, x_m | y_1, \dots, y_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_m} \frac{p(u_1, \dots, u_m, y_1, \dots, y_n)}{p_Y(y_1, \dots, y_n)} du_1 \dots du_m,$$

这里  $p_Y(y_1, \dots, y_n)$  是  $\mathbf{Y} = (Y_1, \dots, Y_n)$  的联合密度, 称这里的被积函数为在  $\mathbf{Y} = (y_1, \dots, y_n)$  条件下  $\mathbf{X}$  的**条件分布密度**。

**例:** 设  $X_1, X_2, X_3$  为独立同分布的连续型随机变量, 求  $P(X_3 < X_1 | X_1 = \min \{X_1, X_2\})$ .

**解:**

$$\begin{aligned} P(X_3 < X_1 | X_1 = \min \{X_1, X_2\}) &= \frac{P(X_3 < X_1, X_1 = \min \{X_1, X_2\})}{P(X_1 = \min \{X_1, X_2\})} \\ &= \frac{\int_{-\infty}^{+\infty} \int_{x_3}^{+\infty} \int_{x_1}^{+\infty} p(x_2) dx_2 p(x_1) dx_1 p(x_3) dx_3}{\int_{-\infty}^{+\infty} \int_{x_1}^{+\infty} p(x_2) dx_2 p(x_1) dx_1} \\ &= \frac{\int_{-\infty}^{+\infty} \int_{x_3}^{+\infty} (1 - F(x_1)) dF(x_1) dF(x_3)}{\int_{-\infty}^{+\infty} (1 - F(x_1)) dF(x_1)} \\ &= \frac{\int_{-\infty}^{+\infty} \frac{1}{2} - F(x_3) + \frac{1}{2} F^2(x_3) dF(x_3)}{1/2} = \frac{1/6}{1/2} = \frac{1}{3} \end{aligned}$$

### 7.1.1 $n$ 维随机向量的数字特征

设  $\xi = (X_1, \dots, X_n)$  是  $n$  维随机向量, 每个  $X_i$  都有期望和方差, 易知协方差

$$\sigma_{X_i X_j} = E((X_i - E(X_i))(X_j - E(X_j))) \quad (i \neq j)$$

必然存在。

**期望:** 称  $(E(X_1), \dots, E(X_n))$  为  $\xi$  的期望, 记为  $E(\xi)$ .

**协方差阵:** 记  $\sigma_{ij} = \text{cov}(X_i, X_j)$ ,  $\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}\sigma_{jj}}}$ . 称  $\Sigma = (\sigma_{ij})_{n \times n}$  为  $\xi$  的协方差阵,  $\mathbf{R} = (\rho_{ij})_{n \times n}$  为  $\xi$  的相关系数阵。

**例:** 设随机变量  $X_1, X_2, X_3$  满足

$$aX_1 + bX_2 + cX_3 = 0,$$

$$E(X_1) = E(X_2) = E(X_3) = d,$$

$$\text{var}(X_1) = \text{var}(X_2) = \text{var}(X_3) = \sigma^2.$$

求相关系数  $\rho_{12}, \rho_{23}, \rho_{31}$ .

**解:** 对等式  $aX_1 + bX_2 = -cX_3$  两侧求方差得  $a^2\sigma^2 + b^2\sigma^2 + 2ab\sigma^2\rho_{12} = c^2\sigma^2$ , 由此解得

$$\rho_{12} = \frac{c^2 - a^2 - b^2}{2ab},$$

同理, 对等式  $aX_1 + cX_3 = -bX_2$  两侧求方差得

$$\rho_{13} = \frac{b^2 - a^2 - c^2}{2ac},$$

同理, 对等式  $bX_2 + cX_3 = -aX_1$  两侧求方差得

$$\rho_{23} = \frac{a^2 - b^2 - c^2}{2bc}.$$

特别的, 当  $d \neq 0$  时, 有  $(a + b + c)d = 0$ , 因此  $a + b + c = 0$ , 由此可得

$$c^2 = a^2 + b^2 + 2ab, \quad b^2 = a^2 + c^2 + 2ac, \quad a^2 = b^2 + c^2 + 2bc,$$

代入  $\rho_{12}, \rho_{23}, \rho_{31}$  表达式得  $\rho_{12} = \rho_{23} = \rho_{31} = 1$ 。

### 7.1.2 $n$ 个随机变量的函数

**定理：** 设  $Y = f(X_1, \cdots, X_n)$  的分布函数是  $F(y)$ ，令

$$A(y) = \{(x_1, \cdots, x_n) : f(x_1, \cdots, x_n) \leq y\}$$

其中  $y$  是任意实数，则

$$F_Y(y) = P(f(\xi) \leq y) = \int \cdots \int_{A(y)} p(x_1, \cdots, x_n) dx_1 \cdots dx_n$$

**定理：** (均值公式) 设  $Y = f(X_1, \cdots, X_n)$ ， $n$  维随机向量  $(X_1, \cdots, X_n)$  有联合密度  $p(x_1, \cdots, x_n)$ ，则

$$EY = \int \cdots \int f(x_1, \cdots, x_n) p(x_1, \cdots, x_n) dx_1 \cdots dx_n.$$

**例：** ( $\chi^2$  分布) 假设  $X_1, \cdots, X_n$  独立同分布，都服从  $N(0, 1)$ . 于是， $Y_n := X_1^2 + \cdots + X_n^2 \sim \Gamma(\frac{n}{2}, \frac{1}{2})$ ，密度为

$$p_n(x) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} x^{n/2-1} e^{-x/2}, \quad x > 0.$$

其中

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx$$

**证明：** 利用数学归纳法，已经证明 (郑书例 5.2)  $Y_1 = X_1^2$  的分布密度是

$$p_1(x) = \frac{1}{\sqrt{2\pi}} x^{-1/2} e^{-x/2}, \quad x > 0.$$

设  $n = k$  时结论成立，考虑  $n = k + 1$  的情形，由于  $Y_{k+1} = Y_k + X_{k+1}^2$ ， $Y_k$  与  $X_{k+1}^2$  相互独立，则  $Y_{k+1}$  的分布密度为

$$\begin{aligned} p_{k+1}(x) &= \int_{-\infty}^{+\infty} p_k(u) p_1(x-u) du \\ &= \int_0^x \frac{1}{2^{k/2} \Gamma(\frac{k}{2})} u^{k/2-1} e^{-u/2} \frac{1}{\sqrt{2\pi}} (x-u)^{-1/2} e^{-(x-u)/2} du \\ &= \int_0^x \frac{e^{-x/2}}{2^{(k+1)/2} \Gamma(\frac{k}{2}) \sqrt{\pi}} u^{k/2-1} (x-u)^{-1/2} du \\ &= \frac{x^{(k+1)/2-1} e^{-x/2}}{2^{(k+1)/2} \Gamma(\frac{k}{2}) \sqrt{\pi}} \int_0^1 v^{k/2-1} (1-v)^{-1/2} dv \quad (\text{做变量替换 } u = xv) \\ &= C x^{(k+1)/2-1} e^{-x/2} \quad (C \text{ 是与 } x \text{ 无关的常数}). \end{aligned}$$

由归一性

$$1 = \int_0^{+\infty} C x^{(k+1)/2-1} e^{-x/2} dx = C 2^{(k+1)/2} \int_0^{+\infty} t^{(k+1)/2-1} e^{-t} dt$$

故

$$C = \frac{1}{2^{(k+1)/2} \Gamma\left(\frac{k+1}{2}\right)}.$$

因此  $n = k + 1$  时结论成立。对一切  $n \geq 1$ ,  $Y_n$  均服从  $\Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$ 。称  $Y_n$  服从  $n$  个自由度的  $\chi^2$  (卡方) 分布。□

**例:** 假设  $X_1, \dots, X_n$  独立同分布, 都服从参数为  $\lambda$  的指数分布, 则  $Y_n := \sum_{i=1}^n X_i$  ( $n \geq 1$ ) 的分布密度是

$$p_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}, \quad (x > 0).$$

**证明:** 利用数学归纳法, 当  $n = 1$  时, 结论显然成立。设  $n = k$  时结论成立, 考虑  $n = k + 1$  的情形, 由于  $Y_{k+1} = Y_k + X_{k+1}$ ,  $Y_k$  与  $X_{k+1}$  相互独立, 则  $Y_{k+1}$  的分布密度为

$$\begin{aligned} p_{k+1}(x) &= \int_{-\infty}^{+\infty} p_k(u) p_1(x-u) du \\ &= \int_0^x \frac{\lambda^k}{(k-1)!} u^{k-1} e^{-\lambda u} \lambda e^{-\lambda(x-u)} du \\ &= \frac{\lambda^{k+1}}{(k-1)!} e^{-\lambda x} \int_0^x u^{k-1} du = \frac{\lambda^{k+1}}{k!} x^k e^{-\lambda x}. \end{aligned}$$

因此  $n = k + 1$  时结论成立。对一切  $n \geq 1$ ,  $Y_n$  的概率密度均为  $p_n(x) = \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x}$ , ( $x > 0$ )。并且随机变量  $Z_n = 2\lambda Y_n$  服从  $2n$  个自由度的  $\chi^2$  分布。□

**例:**  $N$  件产品中有  $D$  件次品. 随机抽取  $n$  件, 设包含  $X$  件次品. 可以利用期望的性质, 求  $E(X)$  与  $\text{var}(X)$ . (其中,  $N \geq n \geq 2$ ).

**解:** 随机数目的分解:  $X = X_1 + \dots + X_n$ , 其中

$$X_i = \begin{cases} 1, & \text{若第 } i \text{ 件是次品;} \\ 0, & \text{若第 } i \text{ 件是合格品.} \end{cases}$$

由期望的线性、伯努利分布的期望,

$$E(X) = \sum_{i=1}^n E(X_i) = \sum_{i=1}^n P(\text{第 } i \text{ 件是次品}) = n \frac{D}{N}.$$

由于  $\text{var}(X) = EX^2 - (EX)^2$ . 根据对称性,

$$E(X^2) = \sum_{i=1}^n E(X_i^2) + \sum_{i \neq j} E(X_i X_j) = nE(X_1^2) + n(n-1)E(X_1 X_2)$$

由乘法公式,

$$E(X_1 X_2) = P(\text{前两件都是次品}) = \frac{D}{N} \cdot \frac{D-1}{N-1}.$$

因此,

$$\begin{aligned} \text{var}(X) &= n \frac{D}{N} + n(n-1) \frac{D}{N} \cdot \frac{D-1}{N-1} - \left( n \frac{D}{N} \right)^2 \\ &= \frac{n(N-n)D(N-D)}{N^2(N-1)} \quad (N > 1) \end{aligned}$$

**例:** 随机向量  $\mathbf{X} = (X_i)_{n \times 1}$ ,  $E(\mathbf{X}) = \mu$ ,  $\text{var}(\mathbf{X}) = \Sigma$ , 矩阵  $\mathbf{A}_{n \times n}$ , 证明:  $E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{A} \Sigma) + \mu^\top \mathbf{A} \mu$ .

**证明:**

$$\begin{aligned} E(\mathbf{X}^\top \mathbf{A} \mathbf{X}) &= E(\text{tr}(\mathbf{X}^\top \mathbf{A} \mathbf{X})) = E(\text{tr}(\mathbf{A} \mathbf{X} \mathbf{X}^\top)) \\ &= \text{tr}(E(\mathbf{A} \mathbf{X} \mathbf{X}^\top)) = \text{tr}(\mathbf{A} E(\mathbf{X} \mathbf{X}^\top)) \\ &= \text{tr}(\mathbf{A} (\text{var}(\mathbf{X}) + \mu \mu^\top)) = \text{tr}(\mathbf{A} \Sigma) + \text{tr}(\mathbf{A} \mu \mu^\top) \\ &= \text{tr}(\mathbf{A} \Sigma) + \mu^\top \mathbf{A} \mu. \end{aligned}$$

□

### 7.1.3 $n$ 个随机变量的多个函数

**定理:** 设  $\xi = (X_1, \dots, X_n)$  为连续型随机向量, 且  $\mathbb{R}^n$  中的区域  $A$  满足  $P(\xi \in A) = 1$ , 函数  $f_1(x_1, \dots, x_n), \dots, f_n(x_1, \dots, x_n)$  满足下列条件:

(1) 对任何实数  $u_1, \dots, u_n$ , 方程组

$$f_k(x_1, \dots, x_n) = u_k, \quad (k = 1, \dots, n)$$

在  $A$  中至多有一个解  $x_i = x_i(u_1, \dots, u_n)$ ,  $i = 1, \dots, n$ ;

(2) 对一切  $k = 1, \dots, n$ ,  $f_k$  在  $A$  中有连续偏导数;

(3) 雅可比行列式

$$J = \frac{\partial(y_1, \dots, y_n)}{\partial(x_1, \dots, x_n)} \neq 0.$$

设  $Y_k = f_k(X_1, \dots, X_n)$  ( $k = 1, \dots, n$ ),  $G = \{(u_1, \dots, u_n) : \text{方程组 } f_k(x_1, \dots, x_n) = u_k, (k = 1, \dots, n) \text{ 在 } A \text{ 中有解}\}$ , 则  $\eta = (Y_1, \dots, Y_n)$  是连续型, 且联合密度

$$p_\eta(y_1, \dots, y_n) = p_\xi(x_1, \dots, x_n) |J^{-1}|, (y_1, \dots, y_n) \in G.$$

**定理:** 设  $\xi = (X_1, \dots, X_n)$  的协方差阵为  $\Sigma$ , 且

$$Y_i = \sum_{j=1}^n a_{ij} X_j, j = 1, \dots, m.$$

记  $\mathbf{A} = (a_{ij})_{m \times n}$ ,  $\eta = (Y_1, \dots, Y_m)$ , 则

$$(E(\eta))^{\top} = \mathbf{A}(E(\xi))^{\top},$$

$$\text{cov}(\eta, \eta) = \mathbf{A}\Sigma\mathbf{A}^{\top}.$$

**证明:** 由于  $E(Y_i) = \sum_{j=1}^n a_{ij} E(X_j)$ , 故  $(E(\eta))^{\top} = \mathbf{A}(E(\xi))^{\top}$  成立, 又由于  $Y_i - E(Y_i) = \sum_{j=1}^n a_{ij} (X_j - E(X_j))$ , 知

$$(Y_i - E(Y_i))(Y_k - E(Y_k)) = \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} (X_j - E(X_j))(X_l - E(X_l)).$$

于是

$$\begin{aligned} \text{cov}(Y_i, Y_k) &= \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} E(X_j - E(X_j))(X_l - E(X_l)) \\ &= \sum_{j=1}^n \sum_{l=1}^n a_{ij} a_{kl} \sigma_{jl}, \end{aligned}$$

这里  $\sigma_{jl} = \text{cov}(X_j, X_l)$ .

由于  $\Sigma = (\sigma_{jl})_{n \times n}$ , 知  $\text{cov}(\eta, \eta) = \mathbf{A}\Sigma\mathbf{A}^{\top}$  成立。 □

**次序统计量:** 设  $n$  个随机变量  $X_1, \dots, X_n$ , 将它们从小到大排列:

$$X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)},$$

称  $X_{(k)}$  为第  $k$  个次序统计量.

**例:** 设  $X_1, \dots, X_n$  独立同分布, 都服从  $U(0, 1)$ . 已知对于  $\forall 0 < x < 1$ ,

$$P(X_{(k)} \leq x) = \sum_{i=k}^n \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i}.$$

求  $E(X_{(k)})$  与  $\text{var}(X_{(k)})$ .

**解:** 由于对于  $\forall 0 < x < 1$ ,

$$P(X_{(k)} \leq x) = \sum_{i=k}^n \frac{n!}{i!(n-i)!} x^i (1-x)^{n-i}.$$

$k \leq i \leq n-1$ , 上式单项的导数是

$$\begin{aligned} & \frac{n!}{i!(n-i)!} (ix^{i-1}(1-x)^{n-i} - x^i(n-i)(1-x)^{n-i-1}) \\ &= \frac{n!}{(i-1)!(n-i)!} x^{i-1}(1-x)^{n-i} - \frac{n!}{i!(n-i-1)!} x^i(1-x)^{n-i-1} \\ &= a_{i-1} - a_i, \end{aligned}$$

$i = n$  时,  $(x^n)' = a_{n-1}$ , 于是,  $\forall 0 < x < 1$ ,

$$p_{X_{(k)}}(x) = \sum_{i=k}^{n-1} (a_{i-1} - a_i) + a_{n-1} = a_{k-1}.$$

已有  $q_k(x) := p_{X_{(k)}}(x) = \frac{n!}{(k-1)!(n-k)!} x^{k-1}(1-x)^{n-k}$ , 且对于  $\forall \ell, m \geq 1$ , 由分部积分

$$\begin{aligned} \int_0^1 x^\ell (1-x)^m dx &= \frac{m}{\ell+1} \int_0^1 x^{\ell+1} (1-x)^{m-1} dx \\ &= \cdots = \frac{m!}{(\ell+1) \cdots (\ell+m)} \int_0^1 x^{\ell+m} dx = \frac{\ell!m!}{(\ell+m+1)!} \end{aligned}$$

期望: 取  $\ell = k, m = n-k$ , 知

$$\begin{aligned} EX_{(k)} &= \int_0^1 x q_k(x) dx = \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^k (1-x)^{n-k} dx \\ &= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{k!(n-k)!}{(n+1)!} = \frac{k}{n+1}. \end{aligned}$$

二阶矩: 取  $\ell = k+1, m = n-k$ ,

$$\begin{aligned} EX_{(k)}^2 &= \int_0^1 x^2 q_k(x) dx = \frac{n!}{(k-1)!(n-k)!} \int_0^1 x^{k+1} (1-x)^{n-k} dx \\ &= \frac{n!}{(k-1)!(n-k)!} \cdot \frac{(k+1)!(n-k)!}{(n+2)!} = \frac{k(k+1)}{(n+1)(n+2)}. \end{aligned}$$

方差:

$$\begin{aligned} \text{var}(X_{(k)}) &= EX_{(k)}^2 - (EX_{(k)})^2 = \frac{k(k+1)}{(n+1)(n+2)} - \frac{k^2}{(n+1)^2} \\ &= \frac{k^2(n+1) + k(n+1) - k^2(n+2)}{(n+1)^2(n+2)} = \frac{k(n+1-k)}{(n+1)^2(n+2)}. \end{aligned}$$

#### 7.1.4 $n$ 维正态分布

我们已经定义过  $n$  维正态分布。



**$n$  维正态分布:** 假设  $n$  维随机向量  $\xi$  有如下的联合密度, 则称  $\xi$  服从  $n$  维正态分布, 记为  $\xi \sim N(\mu, \Sigma)$ .

$$p(\mathbf{x}) = \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu) \right\}.$$

**定理 8.1:** 设  $(X_1, \dots, X_n)^\top \sim N(\mu, \Sigma)$ ,  $\mathbf{A} = (a_{ij})_{n \times n}$ ,  $|\mathbf{A}| \neq 0$ ,  $Y_i = \sum_{j=1}^n a_{ij} X_j (i = 1, \dots, n)$ , 则

$$(Y_1, \dots, Y_n)^\top \sim N(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^\top). \quad (7.1.1)$$

**证明:** 设  $\mathbf{y} = (y_1, \dots, y_n)^\top$ ,  $\mathbf{x} = (x_1, \dots, x_n)^\top$ , 对于任意  $n$  维矩形  $D$ , 记

$$D^* = \{\mathbf{x} = (x_1, \dots, x_n)^\top : \mathbf{A}\mathbf{x} \in D\},$$

则

$$P((Y_1, \dots, Y_n)^\top \in D) = P((X_1, \dots, X_n)^\top \in D^*) = \iint_{D^*} \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mu)^\top \Sigma^{-1}(\mathbf{x} - \mu) \right\} d\mathbf{x}$$

做变量替换  $\mathbf{x} = \mathbf{A}^{-1}\mathbf{y}$ , 雅可比行列式为

$$\frac{\partial(x_1, \dots, x_n)}{\partial(y_1, \dots, y_n)} = |\mathbf{A}^{-1}| = |\mathbf{A}|^{-1}.$$

于是

$$\begin{aligned} P((Y_1, \dots, Y_n)^\top \in D) &= \int \cdots \int_D \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma|}} \exp \left\{ -\frac{1}{2}(\mathbf{A}^{-1}\mathbf{y} - \mu)^\top \Sigma^{-1}(\mathbf{A}^{-1}\mathbf{y} - \mu) \right\} \|\mathbf{A}\|^{-1} d\mathbf{y} \\ &= \int \cdots \int_D \frac{1}{\sqrt{2\pi}^n \sqrt{|\mathbf{A}\Sigma\mathbf{A}^\top|}} \exp \left\{ -\frac{1}{2}(\mathbf{y} - \mathbf{A}\mu)^\top (\mathbf{A}\Sigma\mathbf{A}^\top)^{-1}(\mathbf{y} - \mathbf{A}\mu) \right\} d\mathbf{y}. \end{aligned}$$

这表明  $(Y_1, \dots, Y_n)^\top \sim N(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^\top)$ . □

**推论:** 若  $\xi$  服从  $n$  元正态分布  $N(\mu, \Sigma)$ , 则存在一个正交变换  $\mathbf{U}$ , 使得  $\eta = \mathbf{U}\xi$  是一个具有独立正态分布分量的随机向量, 它的数学期望为  $\mathbf{U}\mu$ , 方差分量是  $\Sigma$  的特征值。

**证明:** 对实对称矩阵  $\Sigma$ , 存在正交矩阵  $\mathbf{U}$ , 使得  $\mathbf{U}\Sigma\mathbf{U}^\top = \mathbf{D}$ , 其中  $\mathbf{D}$  为对角矩阵, 对角元是  $\Sigma$  的特征值, 若  $\Sigma$  的秩为  $r$ , 则有  $r$  个特征值不为零。

将这里的  $\mathbf{U}$  作为定理 8.1 的变换矩阵, 则可得推论结果。 □

**推论:** 正交变换下, 多维标准正态变量保持其独立性, 同方差性不变。

**证明:** 设  $\mathbf{X} = (X_1, \dots, X_n)^\top$  服从  $n$  元正态分布, 且  $X_i$  相互独立有相同的方差  $\sigma^2$ , 则协方差矩阵  $D(\mathbf{X}) = \sigma^2 \mathbf{I}$ , 若  $\mathbf{U}$  是正交阵,  $\mathbf{Y} = \mathbf{U}\mathbf{X}$ , 由定理 8.1 知  $\mathbf{Y}$  服从正态分布, 协方差为

$$\mathbf{U}\sigma^2\mathbf{I}\mathbf{U}^\top = \sigma^2\mathbf{I}$$

因此  $\eta$  仍然是相互独立且具有相同方差。 □

**推论：**若  $\xi \sim N(\mu, \Sigma)$ ，其中  $\Sigma$  是  $n$  阶正定阵，则

$$(\xi - \mu)^\top \Sigma^{-1} (\xi - \mu) \sim \chi_n^2$$

**证明：**设正定阵  $\Sigma = \mathbf{L}\mathbf{L}^\top$ ，则

$$\begin{aligned} (\xi - \mu)^\top \Sigma^{-1} (\xi - \mu) &= (\xi - \mu)^\top (\mathbf{L}\mathbf{L}^\top)^{-1} (\xi - \mu) \\ &= [\mathbf{L}^{-1}(\xi - \mu)]^\top [\mathbf{L}^{-1}(\xi - \mu)] = \eta^\top \eta \end{aligned}$$

其中  $\eta = \mathbf{L}^{-1}(\xi - \mu)$ ，由定理 8.1 知它是均值为  $\mathbf{0}$  的  $n$  维正态变量，协方差矩阵为

$$\mathbf{L}^{-1} \Sigma (\mathbf{L}^{-1})^\top = \mathbf{I}$$

从而  $\eta$  的各个分量是相互独立的标准状态变量，因此

$$\eta^\top \eta = \chi_1^2 + \cdots + \chi_n^2 \sim \chi_n^2.$$

□

**定理 8.2：**设  $(X_1, \cdots, X_m, X_{m+1}, \cdots, X_n)^\top \sim N(\mu, \Sigma) (1 \leq m < n)$ ，且

$$\mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma^{(1)} & \mathbf{0} \\ \mathbf{0} & \Sigma^{(2)} \end{bmatrix},$$

其中  $\mu^{(1)}$  是  $m$  维列向量， $\mu^{(2)}$  是  $n-m$  维列向量， $\Sigma^{(1)}$  是  $m$  阶矩阵， $\Sigma^{(2)}$  是  $n-m$  阶矩阵，则

$$\mathbf{X}^{(1)} = (X_1, \cdots, X_m)^\top \sim N(\mu^{(1)}, \Sigma^{(1)}), \quad \mathbf{X}^{(2)} = (X_{m+1}, \cdots, X_n)^\top \sim N(\mu^{(2)}, \Sigma^{(2)}).$$

**证明：**记  $\mathbf{x}^{(1)} = (x_1, \cdots, x_m)^\top$ ， $\mathbf{x}^{(2)} = (x_{m+1}, \cdots, x_n)^\top$ ，易知  $(X_1, \cdots, X_m, X_{m+1}, \cdots, X_n)^\top$  的联合密度为

$$\begin{aligned} p(x_1, \cdots, x_m, x_{m+1}, \cdots, x_n) &= \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma^{(1)}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(1)} - \mu^{(1)})^\top (\Sigma^{(1)})^{-1} (\mathbf{x}^{(1)} - \mu^{(1)}) \right\} \\ &\quad \cdot \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma^{(2)}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(2)} - \mu^{(2)})^\top (\Sigma^{(2)})^{-1} (\mathbf{x}^{(2)} - \mu^{(2)}) \right\}. \end{aligned}$$

于是

$$P((X_1, \cdots, X_m)^\top \in D) = \int \cdots \int_D \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma^{(1)}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(1)} - \mu^{(1)})^\top (\Sigma^{(1)})^{-1} (\mathbf{x}^{(1)} - \mu^{(1)}) \right\} d\mathbf{x}^{(1)}$$

$$\begin{aligned}
& \cdot \int \cdots \int_{\mathcal{R}^{n-m}} \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma^{(2)}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(2)} - \mu^{(2)})^\top (\Sigma^{(2)})^{-1} (\mathbf{x}^{(2)} - \mu^{(2)}) \right\} d\mathbf{x}^{(2)} \\
& = \int \cdots \int_D \frac{1}{\sqrt{2\pi}^n \sqrt{|\Sigma^{(1)}|}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^{(1)} - \mu^{(1)})^\top (\Sigma^{(1)})^{-1} (\mathbf{x}^{(1)} - \mu^{(1)}) \right\} d\mathbf{x}^{(1)}.
\end{aligned}$$

这表明  $(X_1, \dots, X_m)^\top \sim N(\mu^{(1)}, \Sigma^{(1)})$ , 同理知  $(X_{m+1}, \dots, X_n)^\top \sim N(\mu^{(2)}, \Sigma^{(2)})$ .  $\square$

在上述定理的假设条件下,  $(X_1, \dots, X_m)^\top$  与  $(X_{m+1}, \dots, X_n)^\top$  相互独立. 进而推知多元正态分布  $(X_1, \dots, X_n)$  两两独立的充分必要条件是两两不相关.

**定理 8.3:**  $(X_1, \dots, X_m, \dots, X_n)^\top \sim N(\mu, \Sigma) (1 \leq m < n)$ , 则

$$(X_1, \dots, X_m) \sim N(\mu^{(1)}, \Sigma_{11})$$

其中

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \quad \mu = \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}.$$

**证明:** 令

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \mathbf{I}_m & \mathbf{0} \\ -\Sigma_{21}\Sigma_{11}^{-1} & \mathbf{I}_{n-m} \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} = \mathbf{B} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$$

则由式7.1.1知

$$(Y_1, \dots, Y_n)^\top \sim N(\mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}^\top).$$

易知

$$\mathbf{B}\mu = \mathbf{B} \begin{bmatrix} \mu^{(1)} \\ \mu^{(2)} \end{bmatrix}, \quad \mathbf{B}\Sigma\mathbf{B}^\top = \begin{bmatrix} \Sigma_{11} & \mathbf{0} \\ \mathbf{0} & \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} \end{bmatrix}.$$

根据定理知

$$(X_1, \dots, X_m) \sim N(\mu^{(1)}, \Sigma_{11}).$$

$\square$

**定理 8.4:** 设  $(X_1, \dots, X_n)^\top \sim N(\mu, \Sigma)$ ,  $\mathbf{A}$  是  $m \times n$  的矩阵且  $\mathbf{A}$  的秩等于  $m$ ,  $(Y_1, \dots, Y_m)^\top = \mathbf{A}(X_1, \dots, X_m)^\top$ , 则

$$(Y_1, \dots, Y_m)^\top \sim N(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A}^\top).$$

**证明:** 若  $m = n$ , 则结论与式7.1.1相同; 若  $m < n$ , 则在  $\mathbf{A}$  下方添加  $n - m$  行使得到的矩阵

$$\mathbf{B} = \begin{bmatrix} \mathbf{A} \\ \mathbf{C} \end{bmatrix}$$

非奇异, 令

$$(Z_1, \dots, Z_n)^\top = \mathbf{B}(X_1, \dots, X_n)^\top.$$

由式 7.1.1 知

$$(Z_1, \dots, Z_n)^\top \sim N(\mathbf{B}\mu, \mathbf{B}\Sigma\mathbf{B}^\top).$$

注意到

$$\mathbf{B}\mu = \begin{bmatrix} \mathbf{A}\mu \\ \mathbf{C}\mu \end{bmatrix}, \quad \mathbf{B}\Sigma\mathbf{B}^\top = \begin{bmatrix} \mathbf{A}\Sigma\mathbf{A}^\top & \mathbf{A}\Sigma\mathbf{C}^\top \\ \mathbf{C}\Sigma\mathbf{A}^\top & \mathbf{C}\Sigma\mathbf{C}^\top \end{bmatrix},$$

$$(Z_1, \dots, Z_m)^\top = (Y_1, \dots, Y_m)^\top.$$

由定理 8.3 知定理成立。 □

**定理 8.5:** 设  $(X_1, \dots, X_n)^\top \sim N(\mu, \Sigma)$ , 则有

$$(1) \ E(\mathbf{X}) \triangleq (E(X_1), \dots, E(X_n))^\top = \mu;$$

$$(2) \ \text{cov}(\mathbf{X}, \mathbf{X}) = \Sigma.$$

**证明:** 先考虑  $\Sigma = \mathbf{I}$  的情形, 此时  $X_1, \dots, X_n$  独立同分布, 且  $X_i \sim N(\mu_i, 1)$ , 于是

$$(E(X_1), \dots, E(X_n))^\top = \mu, \quad \text{cov}(\mathbf{X}, \mathbf{X}) = \mathbf{I},$$

故  $\Sigma = \mathbf{I}$  时定理成立。

现考虑一般情形, 设  $\Sigma$  是任何  $n$  阶正定矩阵, 存在方阵  $\mathbf{A}$ , 使得  $\mathbf{A}\Sigma\mathbf{A} = \mathbf{I}$ , 令  $\mathbf{Y} = \mathbf{A}\mathbf{X}$ , 由定理 8.1 知  $\mathbf{Y} \sim N(\mathbf{A}\mu, \mathbf{A}\Sigma\mathbf{A})$ , 即

$$\mathbf{Y} \sim N(\mathbf{A}\mu, \mathbf{I}),$$

因此

$$E(\mathbf{Y}) = \mathbf{A}\mu, \quad \text{cov}(\mathbf{Y}, \mathbf{Y}) = \mathbf{I}.$$

由于  $\mathbf{X} = \mathbf{A}^{-1}\mathbf{Y}$ , 利用期望的线性性质得到

$$E(\mathbf{X}) = \mathbf{A}^{-1}E(\mathbf{Y}) = \mu,$$

$$\text{cov}(\mathbf{X}, \mathbf{X}) = \mathbf{A}^{-1} \text{cov}(\mathbf{Y}, \mathbf{Y})(\mathbf{A}^{-1})^\top = \Sigma.$$

□

**例:** 若  $\xi \sim N(0, I_d)$ , 试证明  $\frac{\xi}{\|\xi\|} (\xi \neq 0)$  为  $\|x\|_2 = 1$  上的均匀分布。

**证明:** 只需说明  $\forall \|x\|_2 = 1, R^\top R = I_d$ , 有  $p_{\frac{\xi}{\|\xi\|}}(x) = p_{\frac{\xi}{\|\xi\|}}(Rx)$ .

由于概率密度函数  $p_{\frac{\xi}{\|\xi\|}}(Rx)$  等价于  $\frac{\xi}{\|\xi\|}$  经过线性变换  $R^\top$  后, 得到的变量  $Z = \frac{R^\top \xi}{\|\xi\|}$  的概率密度函数,

$$p_{\frac{\xi}{\|\xi\|}}(Rx) = p_{\frac{R^\top \xi}{\|\xi\|}}(x),$$

注意到  $\|R^\top \xi\| = \|\xi\|$ , 故

$$p_{\frac{\xi}{\|\xi\|}}(Rx) = p_{\frac{R^\top \xi}{\|R^\top \xi\|}}(x)$$

由定理 8.1 有  $R^\top \xi \sim N(0, I_d)$ , 因此

$$p_{\frac{\xi}{\|\xi\|}}(Rx) = p_{\frac{R^\top \xi}{\|R^\top \xi\|}}(x) = p_{\frac{\xi}{\|\xi\|}}(x).$$

**例:** 若  $\xi_1, \xi_2$  是相互独立的随机变量, 均服从标准正态分布, 而

$$\eta_1 = a\xi_1 + b\xi_2, \quad \eta_2 = c\xi_1 + d\xi_2,$$

则由于

$$E(\eta_1) = 0, \quad D(\eta_1) = a^2 D(\xi_1) + b^2 D(\xi_2) = a^2 + b^2$$

$$E(\eta_2) = 0, \quad D(\eta_2) = c^2 D(\xi_1) + d^2 D(\xi_2) = c^2 + d^2$$

$$\text{cov}(\eta_1, \eta_2) = ac + bd, \quad \rho_{\eta_1, \eta_2} = \frac{ac + bd}{\sqrt{a^2 + b^2} \sqrt{c^2 + d^2}}$$

因此  $\eta_1 \sim N(0, a^2 + b^2)$ ,  $\eta_2 \sim N(0, c^2 + d^2)$ , 且

$$(\eta_1, \eta_2) \sim N(0, 0, a^2 + b^2, c^2 + d^2, \frac{ac + bd}{\sqrt{a^2 + b^2} \sqrt{c^2 + d^2}})$$

当  $ac + bd = 0$  时,  $\rho_{\eta_1, \eta_2} = 0$ ,  $\eta_1$  与  $\eta_2$  独立。

当  $\rho_{\eta_1, \eta_2} = \pm 1$ , 即  $(ac + bd)^2 = (a^2 + b^2)(c^2 + d^2)$  时,  $(\eta_1, \eta_2)$  退化为一个点。退化为一维分布, 而当  $a = b = c = d = 0$  时,  $(\eta_1, \eta_2)$  退化为一个点。

**条件分布:** 若  $\xi = \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix}$  服从  $n$  元正态分布  $N(\mu, \Sigma)$ ,  $E(\xi_1) = \mu_1, E(\xi_2) = \mu_2, \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$ ,

则在给定  $\xi_1 = x_1$  下,  $\xi_2$  的分布仍然为正态分布, 条件数学期望

$$\mu_{2|1} = E(\xi_2 | \xi_1 = x_1) = \mu_2 + \Sigma_{21} \Sigma_{11}^{-1} (x_1 - \mu_1)$$

条件方差

$$\Sigma_{22|1} = \Sigma_{22} - \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12}$$

这里  $E(\xi_2 | \xi_1 = x_1)$  称为  $\xi_2$  关于  $\xi_1$  的回归, 注意到它是  $x_1$  的线性函数。又有条件方差与  $x_1$  无关。

**证明：**考虑

$$\begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix} = \mathbf{I}.$$

其中,  $\Sigma_{12} = \Sigma_{21}^\top$  和  $\mathbf{A}_{12} = \mathbf{A}_{21}^\top$ 。则有

$$\Sigma_{11}\mathbf{A}_{12} + \Sigma_{12}\mathbf{A}_{22} = \mathbf{0}.$$

则  $\mathbf{A}_{22}^{-1}\mathbf{A}_{12}^\top = -\Sigma_{12}^\top\Sigma_{11}^{-1}$ . 所以  $\mathbf{A}_{22}^{-1}\mathbf{A}_{21} = -\Sigma_{21}\Sigma_{11}^{-1}$ . 另外有

$$\Sigma_{12}\mathbf{A}_{12} + \Sigma_{22}\mathbf{A}_{22} = \mathbf{I}.$$

则有

$$(\Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12})\mathbf{A}_{22} = \mathbf{I}.$$

而配方有

$$\begin{aligned} & (x_2 - \mu_2)^\top \mathbf{A}_{22}(x_2 - \mu_2)^\top + 2(x_2 - \mu_2)^\top \mathbf{A}_{21}(x_1 - \mu_1) \\ &= [x_2 - \mu_2 + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}(x_1 - \mu_1)]^\top \mathbf{A}_{22} [x_2 - \mu_2 + \mathbf{A}_{22}^{-1}\mathbf{A}_{21}(x_1 - \mu_1)] \\ & \quad + f(x_1). \end{aligned}$$

得到证明。

**例：**二元场合, 若  $(\xi_1, \xi_2)^\top$  服从正态分布  $N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ , 则

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

在给定  $\xi_1 = x_1$  条件下,  $\xi_2$  的条件分布还是正态分布而且其条件期望由定理可以推知

$$E(\xi_2|\xi_1 = x_1) = \mu_2\rho\frac{\sigma_2}{\sigma_1}(x_1 - \mu_1)$$

条件方差可以推知为

$$\sigma_2^2 - \frac{(\rho\sigma_1\sigma_2)^2}{\sigma_1^2} = \sigma_2^2(1 - \rho^2).$$

## 8.1 随机序列的收敛性

设随机变量  $\eta = \eta(\omega), \xi_1 = \xi_1(\omega), \xi_2 = \xi_2(\omega), \dots$  都是概率空间  $(\Omega, \mathcal{F}, P)$  上的实值函数, 我们在表述上常常省去  $\omega$ 。

**定义:** 称随机变量  $\xi_1, \xi_2, \dots$  **依概率收敛** 于  $\eta$ , 若对任何  $\varepsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} P(\{\omega : |\xi_n(\omega) - \eta(\omega)| \geq \varepsilon\}) = 0.$$

此时记作  $\xi_n \xrightarrow{P} \eta$ .

**定义:** 称随机变量  $\xi_1, \xi_2, \dots$  **概率为 1 (或几乎必然) 的收敛** 于  $\eta$ , 若

$$P(\{\omega : \lim_{n \rightarrow \infty} \xi_n(\omega) = \eta(\omega)\}) = 1.$$

此时记作  $\xi_n \xrightarrow{a.s.} \eta$ , 其中 a.s. 是 almost surely 的缩写。

**定义:** 称随机变量  $\xi_1, \xi_2, \dots$  **弱收敛** 于  $\eta$ , 若对于  $\eta$  的分布函数  $F(x)$  的任何连续点  $x$ , 下式皆成立:

$$\lim_{n \rightarrow \infty} P(\{\omega : \xi_n(\omega) \leq x\}) = P(\{\omega : \eta(\omega) \leq x\}).$$

此时记作  $\xi_n \xrightarrow{w} \eta$ 。弱收敛也称为依分布收敛。

**定理:** 设  $\xi_n \xrightarrow{a.s.} \eta$ , 则  $\xi_n \xrightarrow{P} \eta$ 。

**证明 (不做要求):** 研究集合  $A = \{\omega : \xi_1(\omega), \xi_2(\omega), \dots \text{不收敛于 } \eta(\omega)\}$ , 假设  $\xi_n \xrightarrow{a.s.} \eta$ , 知  $P(A) = 0$ , 对任何  $\varepsilon > 0$ , 令

$$B = \{\omega : \text{有无穷多个 } n, \text{ 使得 } |\xi_n(\omega) - \eta(\omega)| \geq \varepsilon\},$$

$$B_m = \{\omega : \text{有 } n \geq m, \text{ 使得 } |\xi_n(\omega) - \eta(\omega)| \geq \varepsilon\},$$

则  $B_m \supset B_{m+1}$ ,  $B = \bigcap_{m=1}^{\infty} B_m$ , 于是

$$\lim_{m \rightarrow \infty} P(B_m) = P(B) \leq P(A) = 0,$$

因为  $P(|\xi_n - \eta| \geq \varepsilon) \leq P(B_m)$ , 所以

$$\lim_{m \rightarrow \infty} P(|\xi_n - \eta| \geq \varepsilon) = 0.$$

表明  $\xi_n \xrightarrow{P} \eta$ 。

□

注意，逆定理不成立：

**例 (不做要求)：** 设  $\Omega = (0, 1)$ ,  $\mathcal{F}$  由  $(0, 1)$  中所有 Borel 子集组成,  $P$  是这样的概率测度: 对任何区间  $(a, b) (0 \leq a < b \leq 1)$ ,  $P((a, b)) = b - a$ , 在概率空间  $(\Omega, \mathcal{F}, P)$  上考虑下列随机变量序列:

对任何正整数  $k$  及  $j = 1, \dots, 2^k$ , 令

$$X_{k1} = \begin{cases} 1, & 0 < \omega < \frac{1}{2^k}, \\ 0, & \text{其他;} \end{cases}, \quad X_{kj} = \begin{cases} 1, & \frac{j-1}{2^k} < \omega < \frac{j}{2^k}, \\ 0, & \text{其他;} \end{cases} \quad (j > 1).$$

这些  $X_{kj} : k \geq 1, j = 1, \dots, 2^k$  可排成一个序列:  $X_{11}, X_{12}, X_{21}, X_{22}, X_{23}, X_{24}, \dots$  (按照字典排列法, 将第一个足标从小到大排, 若相同则按第二个足标从小到大排), 将该序列记为  $\xi_1, \xi_2, \dots$ , 其中  $\xi_n = X_{k_n j_n}$ , 则对任何  $\varepsilon \in (0, 1)$  有:

$$P(|\xi_n| \geq \varepsilon) = P(\xi_n = 1) = \frac{1}{2^{k_n}}$$

在  $n \rightarrow \infty$  时  $k_n \rightarrow \infty$ , 故有  $\lim_{n \rightarrow \infty} P(|\xi_n| \geq \varepsilon) = 0$ , 这表明  $\xi_n \xrightarrow{P} 0$ .

而对于任何  $\omega \in (0, 1)$ ,  $\lim_{n \rightarrow \infty} \xi_n(\omega)$  不存在。实际上对任何  $\omega$  和  $k$ , 存在唯一的  $j_k$  使得  $X_{kj_k}(\omega) = 1$ , 而  $j \neq j_k$  时  $X_{kj}(\omega) = 0$ , 由此可见,  $\lim_{n \rightarrow \infty} \xi_n(\omega)$  不存在。即  $\xi_n \xrightarrow{a.s.} \eta$  不成立。

**定理 (不做要求)：** 设  $\xi_n \xrightarrow{P} \eta$ , 则  $\xi_n \xrightarrow{w} \eta$ .

**证明：** 设  $x_0$  是  $\eta$  的分布函数  $F(x)$  的连续点, 记

$$F_n(x) = P(\xi_n \leq x) \quad (n = 1, 2, \dots).$$

易知, 对任何  $\varepsilon > 0$ , 有

$$\{\xi_n \leq x_0\} \subset \{\xi_n - \eta \leq -\varepsilon\} \cup \{\eta \leq x_0 + \varepsilon\},$$

于是  $P(\xi_n \leq x_0) \leq P(\xi_n - \eta \leq -\varepsilon) + P(\eta \leq x_0 + \varepsilon)$ .

故

$$F_n(x_0) - F(x_0) \leq P(|\xi_n - \eta| \geq \varepsilon) + F(x_0 + \varepsilon) - F(x_0).$$

类似地, 有

$$\{\xi_n \leq x_0\} \supset \{\xi_n - \eta \leq \varepsilon, \eta \leq x_0 - \varepsilon\},$$

于是

$$P(\xi_n \leq x_0) \geq P(\xi_n - \eta \leq \varepsilon, \eta \leq x_0 - \varepsilon) \geq P(\eta \leq x_0 - \varepsilon) - P(\xi_n - \eta > \varepsilon)$$

故

$$F_n(x_0) - F(x_0) \geq F(x_0 - \varepsilon) - F(x_0) - P(|\xi_n - \eta| \geq \varepsilon).$$



因此

$$|F_n(x_0) - F(x_0)| \leq F(x_0 + \varepsilon) - F(x_0 - \varepsilon) + P(|\xi_n - \eta| \geq \varepsilon).$$

由于  $x_0$  是  $F(x)$  的连续点, 因此对任何  $\delta > 0$ , 有  $\varepsilon > 0$ , 满足

$$F(x_0 + \varepsilon) - F(x_0 - \varepsilon) < \frac{\delta}{2}.$$

取  $n_0$ , 当  $n \geq n_0$  时,

$$P(|\xi_n - \eta| \geq \varepsilon) < \frac{\delta}{2}.$$

于是对一切  $n \geq n_0$ , 有

$$|F_n(x_0) - F(x_0)| < \delta.$$

这表明  $F_n(x_0) \rightarrow F(x_0) (n \rightarrow \infty)$ , 故  $\xi_n \xrightarrow{\omega} \eta$ . □

注意, 逆定理不真:

**例:** 设随机变量  $X \sim N(0, 1)$ , 令

$$\xi_{2n-1} = X, \quad \xi_{2n} = -X \quad (n = 1, 2, \dots).$$

易知所有的  $\xi_n$  有相同的分布函数  $\phi(x)$ , 为标准正态分布函数, 显然  $\xi_n \xrightarrow{\omega} \eta$ . 但是对  $\varepsilon > 0$ , 有

$$P(|\xi_n - X| \geq \varepsilon) = \begin{cases} 0, & n \text{ 是奇数,} \\ P(|X| \geq \frac{\varepsilon}{2}), & n \text{ 是偶数.} \end{cases}$$

可见  $\xi_1, \xi_2, \dots$  并不依概率收敛于  $X$ .

设  $X_1, X_2, \dots$  是随机变量序列, 令

$$S_n = X_1 + \dots + X_n.$$

**定义:** 若  $E(X_n), n = 1, 2, \dots$  都存在, 且

$$\frac{1}{n} (S_n - E(S_n)) \xrightarrow{P} 0$$

则称  $X_1, X_2, \dots$  服从 (弱) 大数律 (Weak Law of Large Numbers, WLLN)。

**定义:** 若  $E(X_n), n = 1, 2, \dots$  都存在, 且

$$\frac{1}{n} (S_n - E(S_n)) \xrightarrow{\text{a.s.}} 0,$$

则称  $X_1, X_2, \dots$  服从强大数律 (SLLN)。

**定义:** 若对任意  $n \geq 2$  都有  $X_1, \dots, X_n$  相互独立, 则称  $X_1, X_2, \dots$  是相互独立的随机变量序列。

若  $X_1, X_2, \dots$  相互独立, 且  $X_n \stackrel{d}{=} X_1, \forall n \geq 2$ , 则称  $X_1, X_2, \dots$  独立同分布, 记为 i.i.d. (independent and identically distributed).

**例:** 设随机变量序列  $\{X_n\}$  独立同分布, 其密度函数为

$$p(x) = \begin{cases} \frac{1}{\beta}, & 0 < x < \beta, \\ 0, & \text{其他.} \end{cases}$$

其中常数  $\beta > 0$ , 令  $Y_n = \max\{X_1, X_2, \dots, X_n\}$ , 证明  $Y_n \xrightarrow{P} \beta$ .

**证明:** 因为当  $x < 0$  时, 有  $P(Y_n \leq x) = 0$ , 当  $x \geq \beta$  时, 有  $P(Y_n \leq x) = 1$ , 当  $0 \leq x < \beta$  时, 有

$$P(Y_n \leq x) = \prod_{i=1}^n P(X_i \leq x) = \prod_{i=1}^n \int_0^x \frac{1}{\beta} dx = \left(\frac{x}{\beta}\right)^n,$$

所以对任意的  $\varepsilon > 0 (\varepsilon < \beta)$ , 当  $n \rightarrow \infty$  时, 有

$$P(|Y_n - \beta| \geq \varepsilon) = P(Y_n \leq \beta - \varepsilon) = \left(\frac{\beta - \varepsilon}{\beta}\right)^n \rightarrow 0,$$

所以有  $Y_n \xrightarrow{P} \beta$ . □

## 8.2 大数定律

**切比雪夫大数定律:** 假设  $X_1, X_2, \dots$  相互独立, 且存在  $M$  使得  $\text{var}(X_i) \leq M, \forall i$ . 设  $S_n = X_1 + \dots + X_n$ , 那么,

$$\frac{1}{n} (S_n - ES_n) \xrightarrow{P} 0 \quad (n \rightarrow \infty).$$

**证明:** 令  $A_n = \left\{ \left| \frac{1}{n} (S_n - ES_n) \right| \geq \varepsilon \right\}$ . 需验证  $P(A_n) \rightarrow 0$ .

由切比雪夫不等式,

$$P(A_n) = P(|S_n - ES_n| \geq n\varepsilon) \leq \frac{1}{(n\varepsilon)^2} \text{var}(S_n)$$

由于  $X_1, X_2, \dots$  两两不相关, 所以  $\text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) \leq nM$ , 于是

$$P(A_n) \leq \frac{nM}{n^2\varepsilon^2} = \frac{M}{\varepsilon^2} \cdot \frac{1}{n} \rightarrow 0.$$

由此知定律成立.

其中定律里的条件“相互独立”可减弱为“两两不相关”. □

**推论:** 设  $X_1, X_2, \dots$  独立同分布,  $\text{var}(X_1) < \infty$ , 则当  $n \rightarrow \infty$  时,

$$\frac{S_n}{n} \xrightarrow{P} E(X_1).$$

**推论：**（伯努利大数律）单次试验中  $A$  发生的概率为  $p$ ，设在  $n$  次试验中事件  $A$  发生了  $\nu_n$  次，则当  $n \rightarrow \infty$  时，

$$\frac{\nu_n}{n} \xrightarrow{P} p.$$

**证明：** 令

$$X_i = \begin{cases} 1, & \text{第 } i \text{ 次试验中 } A \text{ 发生,} \\ 0, & \text{第 } i \text{ 次试验中 } A \text{ 不发生,} \end{cases} \quad (i = 1, 2, \dots),$$

则  $\frac{\nu_n}{n} = \frac{1}{n} \sum_{i=1}^n X_i$  由于  $X_1, X_2, \dots$  是独立同分布的随机变量序列， $E(X_i) = p$ ， $\text{var}(X_i) = p(1-p)$  ( $i = 1, 2, \dots$ )，故由上一推论知本推论成立。  $\square$

如果不假定  $E(X_i)$  存在，上述推论是否成立？

**例：** 设  $X_1, X_2, \dots$  独立同分布，密度为  $p(x) = \frac{1}{\pi(x^2+1)}$ 。可以证明， $\frac{S_n}{n}$  与  $X_1$  有相同的密度。于是，对任何  $a$  和  $\varepsilon > 0$ ，有

$$P\left(\left|\frac{S_n}{n} - a\right| > \varepsilon\right) = P(|X_1 - a| > \varepsilon) \text{ 不趋于 } 0.$$

故  $\frac{S_n}{n}$  不能以概率收敛于  $a$ 。

**定理：**（Cantelli 强大数定律）假设  $X_1, X_2, \dots$  相互独立， $E(X_i)$  存在，且  $E(X_i - E(X_i))^4 \leq M$ ， $\forall i$ 。那么

$$\frac{1}{n} (S_n - ES_n) \xrightarrow{\text{a.s.}} 0.$$

**推论：** 设  $X_1, X_2, \dots$  独立同分布， $E(X_1^4)$  存在，则  $\frac{S_n}{n} \xrightarrow{\text{a.s.}} E(X_1)$ 。

本推论可以由 Cantelli 强大数定律直接推出。

**推论：**（Borel 强大数律）单次小试验中事件  $A$  发生的概率为  $p$ 。在独立重复试验中，前  $n$  次试验中  $A$  发生的次数为  $\nu_n$ ，则

$$\frac{\nu_n}{n} \xrightarrow{\text{a.s.}} p.$$

**定理 2.4.** (Kolmogorov's SLLN). 假设  $X_1, X_2, \dots$  独立同分布，期望存在，则  $\frac{1}{n} S_n \xrightarrow{\text{a.s.}} EX_1$ 。

**例：** 设  $X_1, X_2, \dots$  是相互独立的随机变量序列， $X_1 \equiv 0$ ，对一切  $n \geq 2$ ， $X_n$  只取三个可能值  $n, -n, 0$ ，且

$$P(X_n) = n = P(X_n = -n) = \frac{1}{2n \ln n}, \quad P(X_n = 0) = 1 - \frac{1}{n \ln n}.$$

证明  $X_1, X_2, \dots$  服从切比雪夫大数定律。

**证明：** 易知

$$E(X_n) = 0, \quad \text{var}(X_1) = 0, \quad \text{var}(X_n) = \frac{n}{\ln n} \quad (n = 2, 3, \dots).$$

令  $S_n = \sum_{i=1}^n X_i$ , 则  $\text{var}(S_n) = \sum_{i=1}^n \text{var}(X_i) = \sum_{i=2}^n \frac{i}{\ln i}$ . 由于  $x \geq 3$  时  $\frac{x}{\ln x}$  是  $x$  的增函数, 故  $\text{var}(S_n) \leq \frac{2}{\ln 2} + \frac{n^2}{\ln n}$ , 利用切比雪夫不等式, 有

$$P\left(\left|\frac{S_n - E(S_n)}{n}\right| \geq \varepsilon\right) \leq \frac{1}{n^2 \varepsilon} \text{var}(S_n) \rightarrow 0 \quad (\varepsilon > 0, n \rightarrow \infty).$$

这表明  $X_1, X_2, \dots$  服从切比雪夫大数定律。

大数定律和强大数定律有广泛的应用:

(1): 统计方法的理论依据. 假设采集到数据数据:  $X_1, \dots, X_n$  为  $X$  的  $n$  次独立观测值, 它们独立同分布.

常用  $\frac{1}{n}(X_1 + \dots + X_n)$  估计期望:

$$\bar{X} := \frac{1}{n}(X_1 + \dots + X_n) \xrightarrow{\text{a.s.}} E(X).$$

常用  $\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  估计方差:

$$\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - (\bar{X})^2 \xrightarrow{\text{a.s.}} E(X^2) - (E(X))^2 = \text{var}(X).$$

(2): 用于计算机模拟期望、概率.

**例:** 设有  $m$  枚炮弹同时射击, 第  $i$  枚炮弹落点为  $(x_i, y_i)$ ,

$$\varphi(x_1, y_1; \dots; x_m, y_m) = \begin{cases} 1, & \text{若落点造成有效毁伤;} \\ 0, & \text{否则.} \end{cases}$$

设第  $i$  枚炮弹的瞄准点为  $(a_i, b_i)$ , 实际落点  $(X_i, Y_i)$ . 模型假设:  $X_1, \dots, X_m; Y_1, \dots, Y_m$  相互独立, 且

$$X_i \sim N(a_i, \sigma_1^2), \quad Y_i \sim N(b_i, \sigma_2^2).$$

根据 SLLN:

$$\begin{aligned} & P(\varphi(X_1, Y_1; \dots; X_m, Y_m) = 1) \\ & \approx \frac{1}{n} \sum_{k=1}^n \varphi(X_1^{(k)}, Y_1^{(k)}; \dots; X_m^{(k)}, Y_m^{(k)}). \end{aligned}$$

利用数据  $X_1^{(k)}, Y_1^{(k)}; \dots; X_m^{(k)}$  即可用计算机计算概率的估计值。

(3): 估计积分  $I = \int_a^b f(x)dx$ .

利用变量替换  $x = a + (b-a)u$  得,  $I = \int_0^1 f(a + (b-a)u)(b-a)du$ , 因此不妨假设  $a = 0, b = 1$ .

得到  $I = \int_0^1 f(x) \cdot 1dx = E(f(U))$ . 其中  $U$  为服从区间  $(0, 1)$  上均匀分布的随机变量。

根据 SLLN:

$$I \approx \frac{1}{n} \sum_{i=1}^n (f(U_1) + \cdots + f(U_n))$$

因此只需得到服从区间  $(0, 1)$  上均匀分布的随机数  $u_1, \cdots, u_n$ , 即可得到积分的近似值。该方法还可以推广到高维的数值积分。

### 8.3 中心极限定理

**中心极限定理:** 设  $X_1, X_2, \cdots$  为随机变量序列, 若  $E(X_n), \text{var}(X_n), n = 1, 2, \cdots$  都存在,  $\text{var}(X_n)$  不全为 0, 令  $S_n = \sum_{i=1}^n X_i$ , 且

$$S_n^* = \frac{S_n - E(S_n)}{\sqrt{\text{var}(S_n)}} \xrightarrow{\omega} Z \sim N(0, 1),$$

则称  $X_1, X_2, \cdots$  服从中心极限定理 (Central Limit Theorem, CLT)(或适合中心极限定理)。

**定理:** (Linderberg-Levy 中心极限定理) 假设  $X_1, X_2, \cdots$  独立同分布,  $E(X_1)$  存在且  $0 < \text{var}(X_1) < \infty$ . 那么,

$$S_n^* \xrightarrow{\omega} Z \sim N(0, 1).$$

**例:** 加法器同时收到 20 个噪声电压  $V_k, k = 1, \cdots, 20$ , 它们独立同分布,  $V_1 \sim U(0, 10)$ . 记  $V = \sum_{k=1}^{20} V_k$ , 求  $P(V > 105)$ .

**解:** 易知  $E(V_1) = 5, \text{var}(V_1) = \frac{10^2}{12}$ .

设

$$V^* = \frac{V - 20 \times 5}{\sqrt{20 \times \frac{100}{12}}}$$

根据中心极限定理,

$$P(V > 105) = P\left(V^* > \frac{105 - 20 \times 5}{\sqrt{20 \times \frac{100}{12}}}\right) \approx 1 - \Phi(0.387)$$

查表得  $\Phi(x^*) = 0.652$ , 从而所求的  $p = 1 - 0.652 = 0.348$ .

**例:** 旅馆有 500 间客房, 每间有一台 2 千瓦的空调. 入住率为 80%. 问: 需多少千瓦的电力能有 99% 的把握保证电力足够?

**解:** 假设提供  $x$  千瓦.

设事件  $A_i =$  “第  $i$  间房开空调”,  $P(A_i) = 80\%$ ,  $X_i = 2 \times 1_{A_i}$ .

则易知  $E(X_1) = 1.6$ ,  $\text{var}(X_1) = 4 \times 0.8 - 1.6^2 = 0.64$ .

要求  $x$  满足:  $P(S_n \leq x) \geq 99\%$ . 设

$$S_n^* = \frac{S_n - 500 \times 1.6}{\sqrt{500 \times 0.64}}$$

根据中心极限定理, 要求

$$P(S_n \leq x) = P(S_n^* \leq x^*) \approx \Phi(x^*) \geq 0.99$$

其中

$$x^* = \frac{x - 500 \times 1.6}{\sqrt{500 \times 0.64}}.$$

查表得  $\Phi(2.33) = 0.99$ . 即, 要求  $x^* \geq 2.33$ .

即, 要求  $x \geq 800 + 2.33 \times \sqrt{320} = 841.68$ , 从而需 842 千瓦.

## 9.1 统计学若干基本概念

**定义：**所考察的对象的总和称为**总体**，在统计学中可以归结为随机变量或其他形式的随机量。

例如，考察电子产品的使用寿命，于是将所有电子产品的使用寿命作为总体。所谓总体特性，就是使用寿命的特性，或者是刻画使用寿命的随机变量  $X$  的特性，该随机变量的分布称为**总体分布**。可以假定  $X$  的分布为指数分布，其分布密度有下列形式：

$$p(x, \theta) = \frac{1}{\theta} \exp \left\{ -\frac{x}{\theta} \right\} \quad (x > 0, \theta > 0),$$

式中  $\theta$  是分布的**参数**。

设用  $F(x, \theta)$  表示随机变量  $X$  分布密度相应的分布函数，用  $F_\theta$  表示相应的分布。为获取分布  $F_\theta$  的信息，我们假定  $F_\theta$  属于一个**分布族**，用  $\mathcal{F} = \{F_\theta, \theta \in \Theta\}$  表示这个分布族。在分布族  $\mathcal{F}$  的表达式中  $\theta$  称为参数， $\Theta$  称为**参数空间**。在统计学中，随机变量  $X$  称为总体，它的分布  $F_\theta$  称为**总体分布**。这样， $X \sim F_\theta \in \mathcal{F}$  形成了这个统计问题的**模型**，称为总体模型。

例如，电子产品的使用寿命  $X$  的分布  $F_\theta$  由分布密度确定，其中参数  $\theta \in (0, +\infty)$ 。当  $\theta$  确定后，我们获得了电子产品使用寿命的全部信息。

总体模型只涉及  $X$  这个随机变量，而没有涉及数据。观察数据  $\mathbf{x} = (x_1, \dots, x_n)$  是  $\mathbf{X} = (X_1, \dots, X_n)$  的观察值。其中  $X_1, \dots, X_n$  是独立同分布的，这由样本的产生所确定，其共同的分布为  $F_\theta$ 。在统计学中，我们称  $\mathbf{X} = (X_1, \dots, X_n)$  为**样本**，称  $n$  为**样本量**，称  $\mathbf{X}$  的取值  $\mathbf{x} = (x_1, \dots, x_n)$  为**样本值**。称  $\mathbf{X}$  的所有可能取值的集合为样本空间  $\mathcal{X}$ ，在样本空间上的分布为  $P_\theta$ ，我们称  $\mathbf{X} \sim P_\theta (\theta \in \Theta)$  为**统计模型**。

模型的参数  $\theta$  可以是常数向量或者其他的量，主要特征是：一旦参数的值确定后，统计模型中的分布就完全确定了。在某些统计问题中，我们需要了解与参数有关的量，即  $\theta$  的函数  $g(\theta)$ ，为了简便，将  $g(\theta)$  也称为参数。

**例：**(测量问题) 对某待估量  $a$  重复独立测量  $n$  次，得到测量值  $x_1, \dots, x_n$ 。

测量值带有误差，总体分布  $X = a + e$ ，其中  $e \sim N(0, \sigma^2)$ 。即， $X \sim N(a, \sigma^2)$ 。相应的参数空间为  $\Theta = \{\theta = (a, \sigma^2) : a \in \mathbb{R}, \sigma^2 > 0\}$ 。

参数  $\theta = (a, \sigma^2)$ 。其中， $\sigma^2$  不是所关心的，称为讨厌参数。

$P_\theta: \vec{X}$  的联合密度为

$$f_\theta(x) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x_i - a)^2}{2\sigma^2} \right\}.$$

这样,  $\mathbf{X} \sim P_\theta$ ,  $\theta = (a, \sigma^2) \in \Theta$  形成了统计模型。

研究对象  $\theta$  或  $g(\theta)$ . 例,  $g(a, \sigma^2) = a$ .

**定义:** 设  $\mathbf{X} \sim P_\theta(\theta \in \Theta)$  是一个统计模型, 则定义在样本空间  $\mathcal{X}$  上的任何函数  $T(\mathbf{x})(\mathbf{x} \in \mathcal{X})$  都称为**统计量**。

在统计学中统计量通常是指具体的函数, 不能泛指, 尤其不能含有未知参数。从数学上, 统计量是一个只依赖数据的函数, 当  $(x_1, \dots, x_n)$  的值给定后根据函数关系可以算出  $T(\mathbf{x})$  的值。

然而, 如果将统计量看成样本  $\mathbf{X}$  的函数,  $T(\mathbf{X})$  还是一个随机变量, 具有分布, 且在不同参数值下具有不同的分布。严格意义下, 统计量具有分布族。

例如在测量问题中, 最常见的统计量为样本均值  $T = \frac{1}{n}(X_1 + \dots + X_n)$ , 当观察值为  $x_1, \dots, x_n$  时,  $T = \frac{1}{n}(x_1 + \dots + x_n)$  为一个数值。当  $\mathbf{X} = (X_1, \dots, X_n)$  时, 统计量是样本的函数, 为随机变量。我们可以计算  $T$  的分布,  $T \sim N\left(a, \frac{\sigma^2}{n}\right)$ 。注意统计量的分布含有未知参数  $(a, \sigma^2)$ 。

## 9.2 若干统计问题

**估计问题:** 依赖于样本的统计量就可以作为参数  $a$  的估计, 在估计问题中, 估计参数的统计量也称为估计量。

**例:** (测量问题续) 测量问题中待测量  $a$  的一个估计为  $T_1 = \frac{1}{n} \sum_{i=1}^n x_i$ 。当  $(X_1, \dots, X_n)$  服从多元正态分布时, 其常数线性组合的分布也是正态分布, 利用  $X_i (i = 1, \dots, n)$  独立同分布的特性, 计算  $T_1$  的期望和方差, 可得

$$E(T_1) = \frac{1}{n} \left[ \sum_{i=1}^n E(X_i) \right] = a,$$

$$\text{var}(T_1) = \frac{1}{n^2} \left[ \sum_{i=1}^n \text{var}(X_i) \right] = \frac{\sigma^2}{n},$$

这样我们得到  $T_1 \sim N\left(a, \frac{\sigma^2}{n}\right)$ 。

**假设检验:** 对假设  $H_0$  回答“是”或“否”。

例如, 规定不合格率不能超过 3%。现有 200 件产品, 从中任意抽取 10 件, 发现 2 件不合格。问: 是否可以出厂?



**线性回归：**研究变量  $Y$  对  $x$  的线性依赖关系，

$$Y = b_0 + b_1x + e, \quad e \sim N(0, \sigma^2).$$

### 9.3 极大似然

我们引入两类常用的统计模型：

**离散统计模型：**设  $(X_1, \dots, X_n)$  为独立重复观察得到的样本，其中  $X_i (i = 1, \dots, n)$  为离散型随机变量，样本分布列具有下列一般性质：

$$P_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n)) = \prod_{i=1}^n P_\theta(X_i = x_i) \quad (\theta \in \Theta),$$

此时  $\theta$  为参数，对于固定的样本值  $(x_1, \dots, x_n)$ ，作为  $\theta$  的函数

$$L(\theta) = \prod_{i=1}^n P_\theta(X_i = x_i)$$

称为**似然函数**。

**连续统计模型：**此时  $X_i (i = 1, \dots, n)$  为连续型随机变量，样本  $(X_1, \dots, X_n)$  具有联合密度

$$\prod_{i=1}^n p(x_i, \theta) \quad (\theta \in \Theta).$$

对于固定的样本值  $(x_1, \dots, x_n)$ ， $\theta$  的函数

$$L(\theta) = \prod_{i=1}^n p(x_i, \theta) \tag{9.3.1}$$

也称为**似然函数**。

我们可以用似然函数刻画  $\theta$  与数据的匹配程度。例如，在离散情况下，似然函数  $L(\theta)$  就是总体参数为  $\theta$  的情况下，事件  $\{(X_1, \dots, X_n) = (x_1, \dots, x_n)\}$  的概率。极大似然估计就是挑选使  $P_\theta((X_1, \dots, X_n) = (x_1, \dots, x_n))$  达到最大的  $\theta$  值作为真值的估计。这个思路可以写为定义：

**定义：**设  $\theta \in \Theta$  为统计模型  $(X_1, \dots, X_n) \sim P_\theta$  的参数，统计模型可为连续型或者离散型，又设  $x_1, \dots, x_n$  为总体的样本值，若存在  $\hat{\theta}(x_1, \dots, x_n)$ ，使得

$$L(\hat{\theta}(x_1, \dots, x_n)) = \max_{\theta \in \Theta} L(\theta),$$

其中  $L(\cdot)$  为离散统计模型或者连续统计模型的似然函数，则称  $\hat{\theta}(x_1, \dots, x_n)$  为  $\theta$  的**最大似然估计** (简称 ML 估计)。若  $\hat{\theta}$  为参数  $\theta$  的 ML 估计，则  $\theta$  的函数  $g(\theta)$  的 ML 估计定义为  $g(\hat{\theta})$ 。

注 1：为记号统一，离散和连续模型下的似然函数都用式 9.3.1 的表达式表示。

注 2：与统计量的表示法一样，可用  $\hat{\theta}(x_1, \dots, x_n)$  或  $\hat{\theta}(X_1, \dots, X_n)$  表示参数  $\theta$  的估计量  $\hat{\theta}$ ，当用  $\hat{\theta}(x_1, \dots, x_n)$  表示时，强调  $\hat{\theta}$  的计算，当用  $\hat{\theta}(X_1, \dots, X_n)$  表示时，强调  $\hat{\theta}$  的统计特性。

注 3：本定义中， $g(\theta)$  的 ML 估计定义为  $g(\hat{\theta})$ ，此处  $\hat{\theta}$  为参数  $\theta$  的 ML 估计。这样定义后使得 ML 估计具有不变性。若  $g(\theta)$  是一个一一变换， $\eta = g(\theta)$  可以视为模型的一个新参数化，在新的模型中，似然函数  $\tilde{L}(\eta) = \prod_{i=1}^n p(x_i, \theta)$ 。显然， $\hat{\eta}$  使  $\tilde{L}(\eta)$  达到最大等价于原来的似然函数  $L(\theta)$  在  $\hat{\theta}$  处达到最大。这样  $\hat{\eta} = g(\hat{\theta})$ 。定义中并不要求  $g(\theta)$  是一个一一变换，可以形式上定义一个似然函数

$$L(g) = \max_{\{g: g(\theta)=g\}} L(\theta),$$

利用该似然函数，求得  $g$  的最大似然估计就是  $g(\hat{\theta})$ 。

**正态模型：**考虑随机变量  $X \sim N(\mu, \sigma^2)$ ，有独立重复观察得到的样本  $(X_1, \dots, X_n) = (x_1, \dots, x_n)$ ，希望由这些观察值求出参数  $\mu, \sigma^2$  的 ML 估计，对于样本值  $(x_1, \dots, x_n)$ ，似然函数为

$$L(\theta) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{1}{2} \sum_{i=1}^n \left( \frac{x_i - \mu}{\sigma} \right)^2 \right\},$$

其中  $\theta = (\mu, \sigma^2) \in \mathbb{R} \times \mathbb{R}^+$  为参数。为求似然函数  $L(\theta)$  的最大值点，首先固定二维向量  $\theta = (\mu, \sigma^2)$  的  $\sigma^2$ ，求  $L(\theta) = L(\mu, \sigma^2)$  相对于变量  $\mu$  的最大值点  $\mu^*(\sigma^2)$ ，随后代入  $L(\theta)$ ，得到  $L(\mu^*(\sigma^2), \sigma^2)$ ，再求  $\sigma^2$  的最大值点  $\sigma^{2*}$ ，可得

$$L(\mu^*(\sigma^{2*}), \sigma^{2*}) \geq L(\mu^*(\sigma^2), \sigma^2) \geq L(\mu, \sigma^2),$$

因此  $(\mu^*(\sigma^{2*}), \sigma^{2*})$  为  $\theta = (\mu, \sigma^2)$  的 ML 估计。

我们首先求  $\mu^*(\sigma^2)$ 。注意到，求  $\mu$  的值  $\mu^*$  使得  $L(\mu, \sigma^2)$  达到最大，等价于求  $\mu^*$  使得  $\sum_{i=1}^n (x_i - \mu)^2$  达到最小。由于

$$\sum_{i=1}^n (x_i - \mu)^2 = \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2,$$

其中  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ , 可知当  $\mu = \bar{x}$  时,  $\sum_{i=1}^n (x_i - \mu)^2$  达到最小, 将  $\mu = \bar{x}$  代入  $L(\mu, \sigma^2)$  的表达式中, 得

$$L(\bar{x}, \sigma^2) = \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^n \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{2\sigma^2} \right\}.$$

由于  $\sigma^2 \in (0, +\infty)$ ,  $L$  作为  $\sigma^2$  的函数恒取正值。现等价的求解方程  $\frac{\partial \ln L}{\partial \sigma^2} = 0$ , 即

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \bar{x})^2 = 0$$

解得  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 。由此可知,  $\mu = \bar{x}$ ,  $\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$  使得  $L(\mu, \sigma^2)$  在它的定义域上达到最大, 再由 ML 估计定义知

$$\hat{\mu} = \bar{x}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

分别为  $\mu$  和  $\sigma^2$  的 ML 估计。

**例：**设对飞机的最大飞行速度进行测试, 测得 15 个数据 (单位  $m/s$ ) 如下:

422.2, 418.7, 425.6, 420.3, 425.8, 423.1, 431.5, 428.2, 434.0, 438.3, 412.3, 417.2, 413.5, 441.3, 423.7.

试估计飞机最大速度的均值。

**解：**将飞机最大飞行速度的观察值  $x_1, \dots, x_{15}$  视为随机变量  $X_1, \dots, X_{15}$  的观察值, 随机变量的共同分布为正态分布  $N(\mu, \sigma^2)$ , 该假设根据物理背景或经验确定。

将数据代入上面的结果, 得

$$\hat{\mu} = \bar{x} = 425.05$$

本例建立的统计模型称为**正态模型**, 有时也称为**测量模型**, 其数据的总体  $X$  的分布是正态分布  $N(\mu, \sigma^2)$ , 而这些数据就是来自这个总体的一组观测值, 对于正态总体, 其参数  $\mu$  和  $\sigma^2$  的 ML 估计分别是  $\hat{\mu} = \bar{x}, \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ 。

注意到在实际问题中, 正态模型的参数空间不一定是  $\Theta = \mathbb{R} \times \mathbb{R}^+$ , 参数空间不同, 其相应的模型就是不同的模型。在求解问题时, 需要特别注意参数空间的范围。

例如实际问题中有下列两种常见的正态模型: 一种是  $X \sim N(\mu_0, \sigma^2)$ , 其中  $\sigma^2 > 0$ ,  $\mu_0$  为已知的参数值, 求解方程  $\frac{\partial \ln L}{\partial \sigma^2} = 0$ , 即

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu_0)^2 = 0$$

解得  $\sigma^2$  的 ML 估计为  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_0)^2$ 。

另一种是  $X \sim N(\mu, \sigma_0^2)$ , 其中  $\mu \in \mathbb{R}$ , 而  $\sigma_0^2$  为已知的参数值, 此时求解  $L(\theta)$  的关于  $\mu$  的最大值仍等价于求解  $\mu^*$  使  $\sum_{i=1}^n (x_i - \mu)^2$  达到最小, 因此  $\mu$  的 ML 估计为  $\hat{\mu} = \bar{x}$ .

**例:** (次品率的估计). 某工人生产 20 件产品, 检查出恰有一件为次品. 估计该工人生产的次品率.

**解:** 由题意知, 总体  $X \sim B(1, p), p \in [0, 1]$ . 样本量:  $n = 20$ .

似然函数:

$$L(p) = C_n^s p^s (1-p)^{n-s}, \text{ 其中 } s = x_1 + \cdots + x_n.$$

显然  $\hat{p}$  也为  $\ln L(p) = s \ln p + (n-s) \ln(1-p)$  的最大值点, 且

$$\frac{d}{dp} \ln L(p) = \frac{s}{p} - \frac{n-s}{1-p} \Rightarrow \hat{p} = \frac{s}{n}.$$

将  $n = 20, s = 1$  代入, 因此,  $\hat{p} = \frac{1}{20}$ .

**例:** (对应郑书例 1.4) 设  $X_1, \cdots, X_n \sim \text{iid } U(0, \theta)$ , (样本量:  $n$ ). 求:  $\theta$  的最大似然估计.

**解:** 似然函数:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} 1_{\{0 \leq x_i \leq \theta\}} = \frac{1}{\theta^n} 1_{\{0 \leq x_1, \cdots, x_n \leq \theta\}}.$$

仅当  $\theta \geq \max_{1 \leq i \leq n} x_i$  时,  $L(\theta) > 0$ .

当  $\theta \geq \max_{1 \leq i \leq n} x_i$  时,  $L(\theta) = \frac{1}{\theta^n}$ , 关于  $\theta$  单调下降.

从而,  $\hat{\theta} = \max_{1 \leq i \leq n} x_i$ , 即  $\max_{1 \leq i \leq n} X_i$ .

**例:** (对应郑书例 1.5) 设  $X$  为某产品的寿命, 服从指数分布, 即  $X \sim \text{Exp}(\lambda)$ , 设  $(x_1, \cdots, x_n)$  为来自总体  $X$  的一个样本值, 样本量:  $n$ . 求:  $E(X)$  的最大似然估计.

**解:** 似然函数:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} = \exp \{ n(\ln \lambda - \lambda \bar{x}) \}.$$

$\hat{\lambda}$  是  $\ln \lambda - \bar{x}$  的最大值点:

$$\frac{d}{d\lambda} (\ln \lambda - \bar{x}) = \frac{1}{\lambda} - \bar{x} \Rightarrow \hat{\lambda} = 1/\bar{x}.$$

$E(X) = 1/\lambda$ , 因此,  $\widehat{E(X)} = 1/\hat{\lambda} = \bar{x}$ , 或  $\bar{X}$ .

**例:** (对应郑书例 1.6) 设  $X$  服从泊松分布,  $X \sim \mathcal{P}(\lambda)$ , 设  $(x_1, \cdots, x_n)$  为来自总体  $X$  的一个样本值, 样本量:  $n$ . 求:  $\lambda$  的最大似然估计.

**解:** 似然函数:

$$L(\lambda) = \prod_{i=1}^n \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{1}{\prod_{i=1}^n x_i!} \lambda^{n\bar{x}} e^{-\lambda n} = \frac{1}{\prod_{i=1}^n x_i!} e^{n(\bar{x} \ln \lambda - \lambda)}.$$

$\hat{\lambda}$  是  $\bar{x} \ln \lambda - \lambda$  的最大值点:

$$\frac{d}{d\lambda}(\bar{x} \ln \lambda - \lambda) = \frac{\bar{x}}{\lambda} - 1 \Rightarrow \hat{\lambda} = \bar{x}, \text{ 或 } \bar{X}.$$

**例:** 设总体  $X$  的分布函数为  $f(x, \theta) = \frac{2x}{\theta^2}, 0 \leq x \leq \theta$ , 设  $(x_1, \dots, x_n)$  为来自总体  $X$  的一个样本值, 求分布中位数的 ML 估计。

**解:** 设分布中位数为  $a$ , 则

$$\int_0^a f(x, \theta) dx = \int_a^\theta f(x, \theta) dx = \frac{1}{2}$$

其中

$$\int_0^a f(x, \theta) dx = \frac{a^2}{\theta^2} = \frac{1}{2} \Rightarrow a = \frac{\theta}{\sqrt{2}}$$

似然函数为

$$L(\theta) = \prod_{i=1}^n \frac{2x_i}{\theta^2} = \frac{2 \prod_{i=1}^n x_i}{\theta^{2n}}$$

由于  $\theta \geq x_i, i = 1, \dots, n$ , 上式关于  $\theta$  单调减, 故  $\theta$  的 ML 估计为  $\hat{\theta} = \max\{x_1, \dots, x_n\}$ 。中位数  $a$  的 ML 估计为  $\hat{a} = \frac{\max\{x_1, \dots, x_n\}}{\sqrt{2}}$ 。

## 9.4 矩估计

**定义** 设  $X_1, \dots, X_n$  为来自总体  $X \sim F_\theta (\theta \in \Theta)$  的一个样本, 通常  $\alpha_l \triangleq E_\theta(X^l)$  称为  $l$  阶**总体矩**, 而  $a_l \triangleq \frac{1}{n} \sum_{i=1}^n X_i^l$  称为  $l$  阶**样本矩**。所涉及的矩存在且有限。

(1)  $l$  阶总体矩  $\alpha_l = E_\theta(X^l)$  的矩估计定义为相应的样本矩, 即

$$\hat{\alpha}_l = a_l = \frac{1}{n} \sum_{i=1}^n X_i^l, \quad l = 1, 2, \dots;$$

(2) 若存在连续函数  $\phi$  使  $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$  成立, 则  $g(\theta)$  的矩估计定义为

$$\hat{g}(\theta) = \phi(a_1, \dots, a_k).$$

**例:** (对应郑书例 2.4) 设总体:  $X \sim U(0, \theta)$ . 样本量:  $n$ . 求  $\theta$  的矩估计.

**解:**  $\alpha_1 = \frac{1}{\theta} \int_0^\theta x dx = \frac{1}{2}\theta$ , 即  $\theta = 2\alpha_1$ . 故  $\hat{\theta}_1 = 2\bar{X}$  是  $\theta$  的矩估计.

$\alpha_2 = \frac{1}{\theta} \int_0^\theta x^2 dx = \frac{1}{3}\theta^2$ , 即  $\theta = \sqrt{3\alpha_2}$ . 从而  $\hat{\theta}_2 = \sqrt{3\bar{X}^2}$  也是  $\theta$  的矩估计.

比较  $\hat{\theta}_1$  与最大似然估计  $\hat{\theta} = \max_{1 \leq i \leq n} X_i$ .

(1) 当  $2\bar{X} < \max_{1 \leq i \leq n} X_i$  时,  $\hat{\theta}_1$  不合理. 但,  $\hat{\theta}$  总是合理.

(2) 期望:

$$E_{\theta}(\hat{\theta}_1) = 2E(\bar{X}) = \theta, \quad E_{\theta}(\hat{\theta}) = \frac{n}{n+1}\theta$$

(3) 方差:

$$\begin{aligned} \text{var}_{\theta}(\hat{\theta}_1) &= \frac{4}{n} \text{var}_{\theta}(X) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{1}{3n} \theta^2 \\ \text{var}_{\theta}\left(\frac{n+1}{n}\hat{\theta}\right) &= \frac{(n+1)^2}{n^2} \cdot \frac{n}{(n+1)^2(n+2)} \theta^2 = \frac{1}{n(n+2)} \theta^2. \end{aligned}$$

一些常见分布的矩估计:

(1) 如果总体  $\xi \sim B(N, p)$ ,  $N, p$  为未知参数, 因为  $E(\xi) = Np$ ,  $D(\xi) = Np(1-p)$ , 由方程组

$$\begin{cases} Np = \bar{\xi} \\ Np(1-p) = S^2 \end{cases}, \text{ 解得 } N, p \text{ 的矩估计为}$$

$$\begin{cases} \hat{N} = \frac{\bar{\xi}^2}{\bar{\xi} - S^2}, \\ \hat{p} = 1 - \frac{S^2}{\bar{\xi}}. \end{cases}$$

(2) 如果总体  $\xi \sim P(\lambda)$ ,  $\lambda$  为未知参数, 因为  $E(\xi) = \lambda$ ,  $D(\xi) = \lambda$ , 所以  $\lambda$  的矩估计为

$$\hat{\lambda} = \bar{\xi} \text{ 或 } \hat{\lambda} = S^2.$$

$E(\xi) = \lambda$ 。

(3) 如果总体服从几何分布,  $P(\xi = k) = p(1-p)^{k-1}$ ,  $p$  为未知参数, 则因为  $E(\xi) = \frac{1}{p}$ , 所以解得  $p$  的矩估计为

$$\hat{p} = \frac{1}{\bar{\xi}}.$$

(4) 如果总体  $\xi \sim U(\theta_1, \theta_2)$ ,  $\theta_1, \theta_2$  均为未知参数, 因为

$$E(\xi) = \frac{\theta_1 + \theta_2}{2}, \quad D(\xi) = \frac{(\theta_1 - \theta_2)^2}{12}.$$

所以由方程

$$\frac{\theta_1 + \theta_2}{2} = \bar{\xi}, \quad \frac{(\theta_1 - \theta_2)^2}{12} = S^2$$

解得  $\theta_1, \theta_2$  的矩估计分别为

$$\hat{\theta}_1 = \bar{\xi} - \sqrt{3}S, \quad \hat{\theta}_2 = \bar{\xi} + \sqrt{3}S.$$

**例:** 设总体  $\xi$  的分布为  $p(\xi = k) = (k-1)\theta^2(1-\theta)^{k-2}$ ,  $k = 2, 3, \dots$ ,  $0 < \theta < 1$ , 求  $\theta$  的矩估计。

**解:**

$$E(\xi) = \sum_{k=2}^{\infty} kp(\xi = k) = \sum_{k=2}^{\infty} k(k-1)\theta^2(1-\theta)^{k-2},$$

$$= 2 \sum_{k=2}^{\infty} C_k^2 \theta^2 (1-\theta)^{k-2} = 2\theta^2 \sum_{k=0}^{\infty} C_{k+2}^2 (1-\theta)^k$$

注意到由泰勒展开式  $(1-x)^{-3} = \sum_{k=0}^{\infty} \frac{(k+1)(k+2)}{2} x^k = \sum_{k=0}^{\infty} C_{k+2}^2 x^k$ , 因此

$$E(\xi) = 2\theta^2 \sum_{k=0}^{\infty} C_{k+2}^2 (1-\theta)^k = 2\theta^2 (1 - (1-\theta))^{-3} = \frac{2}{\theta}$$

因此解得

$$\hat{\theta} = \frac{2}{\xi}.$$

## 9.5 估计的无偏性

**定义:** 设  $X_1, \dots, X_n \sim \text{iid } F(x, \theta)$  是一个统计模型,  $g(\theta)$  为待估量。若统计量  $T = T(X_1, \dots, X_n)$  满足

$$E_{\theta}[T(X_1, \dots, X_n)] = g(\theta), \quad \forall \theta \in \Theta.$$

则称  $T$  为  $g(\theta)$  的**无偏估计**。

**例:** 继续讨论正态模型中的估计问题。已经求得期望和方差的 ML 估计:

$$\hat{\mu} = \bar{X}, \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

样本均值  $\bar{X}$  是期望  $\mu$  的无偏估计。

$$\begin{aligned} E_{\theta}(\bar{X}) &= E_{\theta}\left(\frac{1}{n}(X_1 + \dots + X_n)\right) \\ &= \frac{1}{n}(E_{\theta}(X_1) + \dots + E_{\theta}(X_n)) = \mu. \end{aligned}$$

样本方差  $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  不是  $\sigma^2$  的无偏估计。

注意到  $x_i - \mu = (x_i - \bar{x}) + (\bar{x} - \mu)$ :

$$\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 + (\bar{x} - \mu)^2.$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 - (\bar{X} - \mu)^2.$$

$$E_{\theta}(\hat{\sigma}^2) = \text{var}(X) - \text{var}(\bar{X}) = \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2.$$

将方差的 ML 估计适当修改, 可以得到  $\sigma^2$  的无偏估计。

**定理:** 若总体方差  $\sigma^2$  存在, 则  $S^2$  是  $\sigma^2$  的无偏估计, 其中,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

**例:** 设总体:  $X \sim \text{Exp}(\lambda)$ . 样本量:  $n$ . 寻找  $\lambda$  的无偏估计。

**解:** 最大似然估计 & 矩估计:  $\hat{\lambda} = 1/\bar{X} = \frac{n}{S_n}$ , 其中,

$$S_n = X_1 + \cdots + X_n \sim \Gamma(n, \lambda), \quad p_{S_n}(y) = \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y}, y > 0.$$

于是,

$$\begin{aligned} E(\hat{\lambda}) &= E\left(\frac{n}{S_n}\right) = n \int_0^\infty \frac{1}{y} \cdot \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy \\ &= n \frac{\lambda \Gamma(n-1)}{\Gamma(n)} \int_0^\infty \frac{\lambda^{n-1}}{\Gamma(n-1)} y^{n-2} e^{-\lambda y} dy = \frac{n}{n-1} \lambda \end{aligned}$$

$n \geq 2$  时,  $\frac{n-1}{n} \hat{\lambda}$  为  $\lambda$  的无偏估计。

$n = 1$  时,  $E\hat{\lambda} = \int_0^\infty \frac{1}{x} \times \lambda e^{-\lambda x} dx = \infty$ .  $\lambda$  的无偏估计不存在。

**例:** 设总体  $X$  服从参数为  $\lambda$  的泊松分布。现在求  $\exp\{-2\lambda\}$  的估计。

**解:**  $E(X) = \lambda$  知  $X$  是  $\lambda$  的一个无偏估计。显然  $g_1(X) = \exp\{-2X\}$  是一个可能的估计, 但不是无偏估计 ( $E[g_1(x)] = e^{-\lambda(1-e^{-2})} > e^{-2\lambda}$ )。因此考虑另一个估计  $g_2(X)$ , 令

$$g_2(x) = \begin{cases} 1, & x \text{ 是偶数,} \\ -1, & x \text{ 是奇数,} \end{cases}$$

此时

$$E(g_2(X)) = \sum_{k=0}^{\infty} (-1)^k P(X = k) = \exp\{-\lambda\} \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} = \exp\{-2\lambda\}.$$

确实是一个无偏估计, 但是这个估计是荒谬的, 当  $X$  为奇数时, 它取负值。因此, 在样本量很小的时候不能片面追求无偏性。

**例:** 设总体  $\xi \sim N(\mu, \sigma^2)$ , 求  $c$  使得  $c \sum_{i=1}^n (\xi_{i+1} - \xi_i)^2$  为  $\sigma^2$  的无偏估计。

**解:**

$$E\left(\sum_{i=1}^n (\xi_{i+1} - \xi_i)^2\right) = \sum_{i=1}^n E(\xi_{i+1}^2 - 2\xi_{i+1}\xi_i + \xi_i^2) = \sum_{i=1}^n 2\sigma^2 = 2n\sigma^2$$

因此可以取  $c = \frac{1}{2n}$ 。

**例:** 设  $\hat{\theta}$  为参数  $\theta$  的无偏估计, 且有  $\text{var}(\hat{\theta}) > 0$ , 试证明  $(\hat{\theta})^2$  不是  $\theta^2$  的无偏估计。



**证明：**由方差的定义可知，

$$\text{var}(\hat{\theta}) = E(\hat{\theta}^2) - (E(\hat{\theta}))^2 > 0$$

由于  $\hat{\theta}$  为参数  $\theta$  的无偏估计，即  $E(\hat{\theta}) = \theta$ ，因此

$$E(\hat{\theta}^2) = \text{var}(\hat{\theta}) + (E(\hat{\theta}))^2 = \text{var}(\hat{\theta}) + \theta^2 > \theta^2,$$

所以  $(\hat{\theta})^2$  不是  $\theta^2$  的无偏估计。

**例：**设  $x_1, \dots, x_n$  是来自  $N(\theta, 1)$  的样本，证明  $g(\theta) = |\theta|$  没有无偏估计。

**证明：**假设  $T(x_1, \dots, x_n)$  为  $g(\theta)$  的无偏估计，则

$$\left(\frac{1}{\sqrt{2\pi}}\right) \int_{-\infty}^{+\infty} T(x_1, \dots, x_n) \exp\left\{-\sum_{i=1}^n \frac{(x_i - \theta)^2}{2}\right\} dx_1 dx_2 \cdots dx_n = |\theta|.$$

由上式可知，等式左侧关于  $\theta$  处处可导，而等式右侧在  $\theta = 0$  不存在导数，因此假设不成立。即  $g(\theta) = |\theta|$  没有无偏估计。

**例：**设从均值  $\mu$ ，方差为  $\sigma^2 > 0$  的总体中分别抽取容量为  $n_1$  和  $n_2$  的两独立样本， $\bar{x}_1$  和  $\bar{x}_2$  分别是这两个样本的均值。试证明，对于任意常数  $a, b$  ( $a + b = 1$ )， $Y = a\bar{x}_1 + b\bar{x}_2$  都是  $\mu$  的无偏估计，并确定常数  $a, b$  使得  $\text{var}(Y)$  达到最小。

**证明：**由于  $\bar{x}_1$  和  $\bar{x}_2$  分别是容量为  $n_1$  和  $n_2$  的两独立样本的均值，故

$$E(\bar{x}_1) = \mu, \quad E(\bar{x}_2) = \mu, \quad \text{var}(\bar{x}_1) = \frac{\sigma^2}{n_1}, \quad \text{var}(\bar{x}_2) = \frac{\sigma^2}{n_2}.$$

因而

$$E(Y) = E(a\bar{x}_1 + b\bar{x}_2) = aE(\bar{x}_1) + bE(\bar{x}_2) = a\mu + b\mu = (a + b)\mu = \mu$$

因此  $Y = a\bar{x}_1 + b\bar{x}_2$  是  $\mu$  的无偏估计。

又由  $a + b = 1$  知  $Y = a\bar{x}_1 + b\bar{x}_2 = a\bar{x}_1 + (1 - a)\bar{x}_2$ ，从而

$$\text{var}(Y) = \frac{a^2\sigma^2}{n_1} + \frac{(1-a)^2\sigma^2}{n_2} = \sigma^2 \left[ \left(\frac{1}{n_1} + \frac{1}{n_2}\right)a^2 - \frac{2}{n_2}a + \frac{1}{n_2} \right],$$

求导知，当  $a = \frac{1/n_2}{1/n_1 + 1/n_2} = \frac{n_1}{n_1 + n_2}$  时， $\text{var}(Y)$  达到最小，此时  $b = \frac{n_2}{n_1 + n_2}$ 。

结果表明，来自同一总体的容量为  $n_1$  和  $n_2$  的两独立样本的合样本的均值  $\bar{\bar{x}} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n_1 + n_2}$  是线性无偏估计类  $U = \{a\bar{x}_1 + (1 - a)\bar{x}_2\}$  中方差最小的。

## 9.6 无偏估计的优良性

设  $X_1, \dots, X_n \sim \text{iid } F_\theta(x) (\theta \in \Theta)$  为统计模型,  $g(\theta)$  为待估量,  $g(\theta)$  的估计量  $T(X_1, \dots, X_n)$  的均方误差定义为

$$R(\theta, T) = E_\theta[T(X_1, \dots, X_n) - g(\theta)]^2.$$

估计的均方误差有时也称为**风险函数**。

**例:** 设  $X_1, \dots, X_n \sim \text{iid } N(\mu, \sigma^2)$ ,  $\mu \in (+\infty, -\infty)$ ,  $\sigma^2 > 0$ , 在这个统计模型中,  $\mu$  的 ML 估计为  $\hat{\mu} = \bar{X}$ , 其均方误差或风险函数为

$$\begin{aligned} E(\bar{X} - \mu)^2 &= \text{var}(\bar{X}) = \text{var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n^2} \text{var}\left(\sum_{i=1}^n X_i\right) \\ &= \frac{1}{n^2} \sum_{i=1}^n \text{var}(X_i) = \frac{\sigma^2}{n}. \end{aligned}$$

它是不随  $\mu$  的变化而变化的。

**例:** 设  $X_1, \dots, X_n \sim \text{iid } B(1, p)$ , 即  $X_i (i = 1, \dots, n)$  以概率  $p$  取 1, 以概率  $1 - p$  取 0。该统计模型在实际应用中是很常见的, 我们已经求得参数  $p$  的 ML 估计为  $\hat{p} = \bar{X}$ , 不难验证  $\hat{p}$  也是  $p$  的无偏估计, 其均方误差为

$$E(\bar{X} - p)^2 = \text{var}(\bar{X}) = \frac{1}{n} \text{var}(X_1) = \frac{p(1-p)}{n}.$$

由此可知, 估计  $\bar{X}$  的风险函数为  $R(p, \bar{X}) = \frac{p(1-p)}{n}$ 。

注意: 对于任何一个待估参数  $g(\theta)$ , 可以定义估计  $g(\theta_0)$ , 它将保证在  $\theta_0$  处, 风险函数最小, 这意味着一个估计如果要成为最优估计, 它的风险函数必须处处为 0。但是这是不可能的, 为此我们需要考虑限制估计类。现在我们只考虑无偏估计类, 希望在无偏估计类内找到最优估计。设  $T(X)$  是  $g(\theta)$  的无偏估计, 其均方误差变成方差:

$$R(\theta, T) = E_\theta[T - g(\theta)]^2 = E_\theta[T - E_\theta(T)]^2 = \text{var}_\theta(T).$$

**定义:** 设  $X_1, \dots, X_n \sim \text{iid } F_\theta(x) (\theta \in \Theta)$  为统计模型,  $g(\theta)$  为待估量,  $g(\theta)$  的一个估计量为  $T(X_1, \dots, X_n)$  如果

- (1)  $T$  是  $g(\theta)$  的无偏估计,
- (2) 对于  $g(\theta)$  的任意无偏估计  $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$ , 都有

$$\text{var}_\theta(T) \leq \text{var}_\theta(\tilde{T}), \quad \forall \theta \in \Theta.$$

则称  $T$  为  $g(\theta)$  的 (一致) **最小方差无偏估计** (Uniformly Minimum Variance Unbiased, UMVU).

我们记

$$F_{\theta}(t) = P_{\theta}(T \leq t)$$

为统计量  $T$  的分布。

**定义:** 假设统计量  $T = T(X_1, \dots, X_n)$  满足:

对任意统计量  $\tilde{T} = \tilde{T}(X_1, \dots, X_n)$  都有,

在  $T = t$  的条件下,  $\tilde{T}$  的分布与参数  $\theta$  无关,

那么, 称  $T$  为**充分统计量**。

**例:** (对应郑书例 4.3) 设总体:  $X \sim B(1, p)$ , 样本量:  $n$ . 考虑  $T = X_1 + \dots + X_n$ . 易知  $T$  的分布为二项分布, 现在讨论  $T$  的充分性。

对  $t = 0, 1, \dots, n$ , 令

$$S_t := \{(x_1, \dots, x_n) : x_i \in \{0, 1\}, \forall i; \text{ 且 } x_1 + \dots + x_n = t\}.$$

那么,  $\forall (x_1, \dots, x_n) \in S_t$ ,

$$\begin{aligned} & P_p(X_1 = x_1, \dots, X_n = x_n | T = t) \\ &= \frac{P_p(X_1 = x_1, \dots, X_n = x_n)}{P_p(T = t)} = \frac{p^t(1-p)^{n-t}}{C_n^t p^t(1-p)^{n-t}} = \frac{1}{C_n^t}. \end{aligned}$$

因此  $\forall t$ , 在  $T = t$  的条件下,  $(X_1, \dots, X_n)$  服从  $S_t$  上的均匀分布. 该分布与  $p$  无关. 因此,  $T$  是充分统计量.

一般情况下, 求充分统计量是麻烦的. 有如定理:

**定理:** (因子分解定理) 设  $X_1, \dots, X_n \sim \text{iid} p(x, \theta)$ , 其中  $p(x, \theta)$  为分布密度或分布列. 若  $T = T(X_1, \dots, X_n)$  满足:

$$\prod_{i=1}^n p(x_i, \theta) = q_{\theta}(T(x_1, \dots, x_n)) h(x_1, \dots, x_n)$$

则  $T$  是充分统计量。

**例:** 设总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ .

联合密度为

$$\begin{aligned} \prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \mu)^2}{2\sigma^2} \right\}. \end{aligned}$$

指数上的部分可以写为  $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$  的函数。因此,  $(\bar{X}, \sum_{i=1}^n (X_i - \bar{X})^2)$  是充分统计量。

若总体改为  $X \sim N(\mu, 1)$ , 则联合密度改为

$$\begin{aligned} \prod_{i=1}^n p(x_i, \theta) &= \frac{1}{\sqrt{2\pi}^n} \exp \left\{ -\sum_{i=1}^n \frac{(x_i - \mu)^2}{2} \right\} \\ &= \frac{1}{\sqrt{2\pi}^n} \exp \left\{ -\frac{1}{2} \sum_{i=1}^n x_i^2 \right\} \exp \left\{ \mu \sum_{i=1}^n x_i \right\} \exp \left\{ -\frac{n}{2} \mu^2 \right\}. \end{aligned}$$

因此,  $\bar{X}$  是充分统计量。

**例:** 假设二维随机向量  $(X, Y)$  服从二元正态分布, 参数  $\theta = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ , 从总体中抽取一个样本  $((X_1, Y_1), \dots, (X_n, Y_n))$ , 样本量为  $n$ , 其联合密度为

$$\begin{aligned} \prod_{i=1}^n p(x_i, y_i; \theta) &= \left( \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \right)^n \\ &\cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \sum_{i=1}^n \left[ \left( \frac{x_i - \mu_1}{\sigma_1} \right)^2 - \frac{2\rho(x_i - \mu_1)(y_i - \mu_2)}{\sigma_1\sigma_2} + \left( \frac{y_i - \mu_2}{\sigma_2} \right)^2 \right] \right\}, \end{aligned}$$

$x_i - \mu_1 = x_i - \bar{x} + \bar{x} - \mu_1$  和  $y_i - \mu_2 = y_i - \bar{y} + \bar{y} - \mu_2$  代入上式  $e$  指数上的表达式中, 化简, 表达式可以写为

$$\left( \bar{x}, \bar{y}, \sum_{i=1}^n (x_i - \bar{x})^2, \sum_{i=1}^n (y_i - \bar{y})^2, \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right)$$

的函数, 因此联合密度具有形式

$$\prod_{i=1}^n p(x_i, y_i; \theta) = q_\theta \left( \bar{x}, \bar{y}, \sum_{i=1}^n (x_i - \bar{x})^2, \sum_{i=1}^n (y_i - \bar{y})^2, \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \right).$$

由因子分解定理知

$$T = \left( \bar{X}, \bar{Y}, \sum_{i=1}^n (X_i - \bar{X})^2, \sum_{i=1}^n (Y_i - \bar{Y})^2, \sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y}) \right)$$

**例:** 设总体:  $X \sim \text{Exp}(\lambda)$ , 样本量:  $n$ .

联合密度具有形式

$$p_\lambda(x_1, \dots, x_n) = \lambda^n \exp \left\{ -\lambda \sum_{i=1}^n x_i \right\} \cdot 1_{\{x_1, \dots, x_n > 0\}}.$$

其中  $1_{\{x_1, \dots, x_n > 0\}}$  是示性函数, 当所有  $x_i$  大于 0 时取 1, 否则为 0。显然该函数与参数无关, 因此由因子分解定理有, 令  $T = T(X_1, \dots, X_n) = X_1 + \dots + X_n$ . 则  $T$  是充分统计量。

注意充分统计量不唯一，没有起到数据压缩的作用。**定义：**设  $X_1, \dots, X_n \sim \text{iid} p(x, \theta)$ ，又设  $T = T(X_1, \dots, X_n)$  为充分统计量. 若对任意  $\phi$ ,

$$E_{\theta} \phi(T) = 0, \forall \theta \in \Theta \text{ 可推出 } P_{\theta}(\phi(T) = 0) = 1, \forall \theta \in \Theta.$$

则称  $T$  为**完全充分统计量**.

**例：**设总体  $X \sim N(\theta, 1)$ ，样本量为  $n$ 。由因子分解定理知， $T_1 = \sum_{i=1}^n X_i$  是一个充分统计量， $T_2 = (T_1, X_1 - X_2)$  也是充分统计量。取  $\phi(T_2) = X_1 - X_2$ ，易知

$$E_{\theta}[\phi(T_2)] = E_{\theta}(X_1) - E_{\theta}(X_2) \equiv 0$$

但是  $P(\phi(T_2) = 0) = P(X_1 = X_2) \neq 1$ ，这说明  $T_2 = (T_1, X_1 - X_2)$  不是完全充分统计量。

**定理：**设  $T = T(X_1, \dots, X_n)$  为完全充分统计量. 若

$$E_{\theta}(\phi(T)) = g(\theta), \forall \theta,$$

则  $\phi(T)$  是  $g(\theta)$  的 UMVU 估计.

该定理说明，对于待估量  $g(\theta)$ ，只要找到依赖于完全充分统计量的函数  $\phi(T)$ ，使得  $\phi(T)$  是  $g(\theta)$  的无偏估计，则  $\phi(T)$  就是  $g(\theta)$  的 UMVU 估计。因此，要找到  $g(\theta)$  的 UMVU 估计，只需在完全充分统计量中寻找即可。

**例：**设总体  $\xi \sim U(0, \theta)$ ，已经证明样本量为  $n$  的样本的最大值为  $\theta$  的充分统计量，记为  $\xi_{(n)}$ ，证明  $\xi_{(n)}$  也为  $\theta$  的完全充分统计量。

**证明：**因为  $\xi_{(n)}$  的密度为

$$f_{\xi_{(n)}}(x; \theta) = n f(x; \theta) [F(x; \theta)]^{n-1} = \begin{cases} \frac{nx^{n-1}}{\theta^n}, & 0 \leq x \leq \theta, \\ 0, & \text{其他,} \end{cases}$$

如果有函数  $g(x)$ ，使得对一切  $0 < \theta$ ，有

$$E_{\theta}[g(\xi_{(n)})] = 0,$$

即

$$0 \equiv \int_0^{\theta} g(x) f_{\xi_{(n)}}(x; \theta) dx = \frac{n}{\theta^n} \int_0^{\theta} g(x) x^{n-1} dx,$$

故

$$\int_0^{\theta} g(x) x^{n-1} dx \equiv 0.$$

对两侧求导得

$$g(\theta) \theta^{n-1} \equiv 0.$$

从而对一切  $\theta > 0$ ,  $g(\theta) = 0$ , 因此  $\xi_{(n)}$  是完全充分统计量。

**定义:** 若密度或分布列  $p(x, \theta)$  能进行如下分解:

$$p(x, \theta) = S(\theta)h(x) \exp \left\{ \sum_{k=1}^m C_k(\theta)T_k(x) \right\}, \quad (\theta \in \Theta)$$

则称  $p(x, \theta), \theta \in \Theta$  为**指数族分布**.

注:  $x$  可为高维向量, 于是  $p(x, \theta)$  为联合密度/联合分布列.

**例:** 设总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ , 参数  $\theta = (\mu, \sigma^2)$ , 分布密度具有形式:

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

显然, 它具有指数分布的形式。

**例:** 设总体:  $X$  服从参数为  $p$  的二项分布, 样本量:  $n$ , 分布列可以改写为

$$\begin{aligned} P(X=x) &= C_n^x p^x (1-p)^{n-x} \\ &= (1-p)^n C_n^x \exp \left\{ x \ln \frac{p}{1-p} \right\} \end{aligned}$$

$P(X=x)$  所表示的三个因子的乘积符合指数族分布的要求。

**引理:** 若总体  $X$  是指数族, 则设  $(X_1, \dots, X_n)$  为  $X$  的一个样本, 将  $(X_1, \dots, X_n)$  看做随机向量, 则其联合分布也是指数族分布。

**例:** 设总体  $X \sim N(\mu, \sigma^2)$ , 可得  $(X_1, \dots, X_n)$  的联合分布密度为

$$\begin{aligned} p(x_1, \dots, x_n; \mu, \sigma^2) &= \prod_{i=1}^n p(x_i, \theta) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left\{ -\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\sigma^2} \right\} \\ &= \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp \left\{ -\frac{n\mu^2}{2\sigma^2} \right\} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n x_i^2 + \frac{\mu}{\sigma^2} \sum_{i=1}^n x_i \right\}, \end{aligned}$$

故正态分布的一个样本的联合分布也是指数族分布。直接使用引理也可以得证。

**定理:** 总体  $X$  具有指数族分布,  $\Theta \in \mathbb{R}^m$  且含内点;  $(C_1, \dots, C_m)$  是在  $\Theta$  上一对一、连续的函数; 诸  $C_i$  之间 ( $T_i$  之间) 无线性关系. 则

$$\left( \sum_{i=1}^n T_1(X_i), \dots, \sum_{i=1}^n T_k(X_i) \right)$$

是完全充分统计量.

**例：**设总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ , 参数  $\theta = (\mu, \sigma^2)$ 。

设  $T_1 = \sum_{i=1}^n X_i, T_2 = \sum_{i=1}^n X_i^2$ , 则由上面的定理知  $(T_1, T_2)$  是完全充分统计量.

$\bar{X}, S^2$  是  $\mu, \sigma^2$  的 UMVU 估计, 其中:

$$\bar{X} = \frac{1}{n}T_1, \quad S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{1}{n-1} \left( T_2 - \frac{1}{n}T_1^2 \right)$$

是  $(T_1, T_2)$  的函数, 并且是  $\mu, \sigma^2$  的无偏估计.

改为已知  $\mu$  (例如, 已知  $\mu = 1$ ). 则  $\theta = \sigma^2, m = 1$  :

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

$T_1 = \sum_{i=1}^n (X_i - \mu)^2$  是完全充分统计量.

$\hat{\sigma}^2$  是  $\sigma^2$  的 UMVU 估计, 其中

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2.$$

**例：**某工人生产 20 件产品, 其中 1 件为次品. 求: 次品率的 UMVU 估计.

**解：**由题意, 总体:  $Y \sim B(1, p)$ , 参数  $p = \theta \in [0, 1]$ , 样本量:  $n = 20$ .

分布列: (记  $k = y_1 + \cdots + y_n$ )

$$\begin{aligned} P_p(Y_1 = y_1, \cdots, Y_n = y_n) &= \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} \\ &= p^k (1-p)^{n-k} = e^{k \cdot \log p + (n-k) \log(1-p)} = e^{n \log(1-p) + (\log p - \log(1-p))k} \end{aligned}$$

可见  $Y$  的分布列具有指数族分布的形式. 因此,  $T_1 = X = Y_1 + \cdots + Y_{20}$  是完全充分统计量.

又由于  $E(X) = nE(Y) = np$ , 即  $\frac{X}{n}$  是  $p$  的无偏估计, 因此,  $\hat{p} = X/20$  是 UMVU 估计.

**例：**(对应郑书例 4.15) 总体:  $X \sim N(\mu, 1)$ , 样本量:  $n$ , 求  $\mu^2$  的 UMVU 估计.

**解：**参数  $\theta = \mu$ , 联合密度为:

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

$T_1 = \sum_{i=1}^n X_i$  是完全充分统计量, 因此  $\bar{X}$  也是完全充分统计量.

由  $\text{var}(Y) = E(Y^2) - (E(Y))^2$  知,

$$\mu^2 = (E_\mu \bar{X})^2 = E_\mu \bar{X}^2 - \text{var}_\mu(\bar{X}) = E_\mu \bar{X}^2 - \frac{1}{n} = E_\mu \left( \bar{X}^2 - \frac{1}{n} \right)$$

因此,  $\bar{X}^2 - \frac{1}{n}$  是  $\mu^2$  的 UMVU 估计.

**例:** 设总体  $Y \sim N(X\beta, \sigma^2 I_n)$ , 其中  $Y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p} (p \leq n), \beta \in \mathbb{R}^{p \times 1}$ .  $X$  已知.  $(\beta, \sigma^2)$  是参数. 设  $\hat{\beta} = (X^\top X)^{-1} X^\top Y$ , 证明  $\hat{\beta}$  为  $\beta$  的 UMVU 估计.

**证明:** 由于

$$E(\hat{\beta}) = (X^\top X)^{-1} X^\top E(Y) = (X^\top X)^{-1} X^\top X \beta = \beta,$$

因此  $(\hat{\beta})$  是  $\beta$  的无偏估计.  $Y_1, \dots, Y_n$  的联合密度函数为

$$p(y_1, \dots, y_n; \beta, \sigma^2) = (\sqrt{2\pi}\sigma)^{-n} \exp \left\{ -\frac{1}{2\sigma^2} \|Y - X\beta\|_2^2 \right\} = Q(\theta) \exp \{ \theta_1 T_1(Y) + \theta_2 T_2(Y) \}$$

其中  $\theta_1 = -\frac{1}{2\sigma^2}, \theta_2 = \frac{\beta}{\sigma^2}, T_1(Y) = Y^\top Y, T_2(Y) = X^\top Y, Q(\theta) = (\sqrt{2\pi}\sigma)^{-n} \cdot \exp \left\{ -\frac{1}{2\sigma^2} (X\beta)^\top (X\beta) \right\}$ .

由指数分布族的性质知,  $T_1(Y)$  和  $T_2(Y)$  为完全充分统计量, 而

$$\hat{\beta} = (X^\top X)^{-1} X^\top Y = (X^\top X)^{-1} T_2(Y)$$

这表明  $\hat{\beta}$  是完全充分统计量的函数, 可知  $\hat{\beta}$  是  $\beta$  的 UMVU 估计.

## 9.7 估计的相合性

**定义:** 设  $T_n = T_n(X_1, \dots, X_n)$  满足:  $\forall \varepsilon > 0$ ,

$$P_\theta(|T_n - g(\theta)| \geq \varepsilon) \xrightarrow{n \rightarrow \infty} 0, \quad \forall \theta \in \Theta.$$

则称  $T_n$  为  $g(\theta)$  的相合估计, 或估计  $T_n$  具有相合性.

**定理:** 设  $X_1, \dots, X_n \sim \text{iid } F_\theta(x) (\theta \in \Theta)$ ,  $E_\theta(X_i)$  存在且有限, 则

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{P} E_\theta(X_1) \quad (n \rightarrow \infty),$$

定理说明, 在简单随机抽样的情况下, 样本均值是总体均值的相合估计.

**推论:** 设  $X_1, \dots, X_n \sim \text{iid } F_\theta(x) (\theta \in \Theta)$ , 则  $\alpha_l = E_\theta(X_1^l)$  的矩估计  $a_l = \frac{1}{n} \sum_{j=1}^n X_j^l$  为  $\alpha_l$  的相合估计.



**定理：** 设  $X_1, \dots, X_n \sim \text{iid } F_\theta(x) (\theta \in \Theta)$ , 则  $\theta$  的函数  $g(\theta)$  的矩估计具有相合性. (注:  $g(\theta) = \phi(\alpha_1, \dots, \alpha_k)$ , 其中  $\phi$  为连续函数).

隐藏使用了定理:

**定理** 若  $T_n \xrightarrow{P} \theta$ ,  $f$  为连续函数, 则  $f(T_n) \xrightarrow{P} f(\theta)$ .

提示: 由连续性: 对于任意  $\epsilon > 0$ , 存在  $\delta$ , 有  $|T_n - \theta| \leq \delta$ , 则  $|f(T_n) - f(\theta)| \leq \epsilon$ . 故前者事件是后者子集, 有  $P(|T_n - \theta| \leq \delta) \leq P(|f(T_n) - f(\theta)| \leq \epsilon)$ .

**例：** 设总体:  $X \sim U(0, \theta)$ , 样本量:  $n$ .

由定理可以直接得到, 参数  $\theta$  的矩估计  $2\bar{X}$  具有相合性.

最大似然估计  $T_n = \max_{1 \leq i \leq n} X_i$  也具有相合性:  $\forall 0 < \epsilon < \theta$ ,

$$\begin{aligned} P_\theta(|T_n - \theta| \geq \epsilon) &= P_\theta(T_n \leq \theta - \epsilon) \\ &= P_\theta(X \leq \theta - \epsilon)^n = \left(\frac{\theta - \epsilon}{\theta}\right)^n \rightarrow 0. \end{aligned}$$

**例：** (对应郑书例 5.3) 考虑某物种三种类型的个体, 以 1, 2, 3 表示个体的三种类型, 设此三种类型出现的概率分别为:

$$p(1, \theta) = \theta^2, \quad p(2, \theta) = 2\theta(1 - \theta), \quad p(3, \theta) = (1 - \theta)^2,$$

其中  $0 < \theta < 1$ , 现有  $n$  个个体的类型  $\mathbf{x} = (x_1, \dots, x_n)$ , 设其中共有  $n_1$  个 1,  $n_2$  个 2,  $n_3$  个 3, 求 ML 估计并探究相合性.

$\mathbf{x}$  出现的概率为

$$\begin{aligned} P(X = \mathbf{x}) &= p(1, \theta)^{n_1} p(2, \theta)^{n_2} p(3, \theta)^{n_3} \\ &= \theta^{2n_1 + n_2} (1 - \theta)^{n_2 + 2n_3} 2^{n_2} \\ &= 2^{n_2} \theta^{2n_1 + n_2} (1 - \theta)^{2n - (2n_1 + n_2)}. \end{aligned}$$

这是对  $\theta$  作 ML 估计的似然函数, 与总样本量  $2n$ , 某事件出现  $2n_1 + n_2$  次的二项分布的似然函数相同. 利用二项分布的 ML 估计公式, 可得  $\theta$  的 ML 估计为

$$\hat{\theta}_n = \frac{2n_1 + n_2}{2n}.$$

为证明  $\hat{\theta}_n$  的相合性, 我们只需证明

$$\frac{n_1}{n} \xrightarrow{P} \theta^2, \quad \frac{n_2}{n} \xrightarrow{P} 2\theta(1 - \theta) (n \rightarrow \infty).$$

事实上, 利用大数定律可得  $\frac{n_1}{n} \xrightarrow{P} p(1, \theta) = \theta^2$ ,  $\frac{n_2}{n} \xrightarrow{P} p(2, \theta) = 2\theta(1 - \theta)$ , 将两个极限合并即可得到

$$\hat{\theta}_n = \frac{n_1}{n} + \frac{n_2}{2n} \xrightarrow{P} \theta^2 + \theta - \theta^2 = \theta$$

即  $\hat{\theta}_n$  为  $\theta$  的相合估计。

**例:** 设总体  $\xi \sim U(0, \theta)$ ,  $Y = (\prod_{i=1}^n \xi_i)^{\frac{1}{n}}$ , 证明  $eY$  是  $\theta$  的相合估计。

**证明:** 需要证明  $eY \xrightarrow{P} \theta$ , 等价于  $Y \xrightarrow{P} \frac{\theta}{e}$ 。注意到

$$E(\ln \xi) = \int_0^\theta \ln x \frac{1}{\theta} dx = \frac{1}{\theta} x \ln x \Big|_0^\theta - \int_0^\theta \frac{1}{\theta} dx = \ln \theta - 1.$$

因此, 由独立同分布的大数定律得,

$$Y = \left( \prod_{i=1}^n \xi_i \right)^{\frac{1}{n}} = e^{\frac{1}{n} \sum_{i=1}^n \ln \xi_i} \xrightarrow{P} e^{E(\ln \xi)} = \frac{\theta}{e}.$$

## 9.8 估计的渐近分布

**定义:** 设  $T_n = T_n(X_1, \dots, X_n)$  满足:

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow{d} Z \sim N(0, \sigma^2), \quad \forall \theta \in \Theta,$$

则称  $T_n$  是渐近正态的, 其中  $\sigma^2 = \sigma_\theta^2$  称为渐近方差。

**定理:** 若  $Y_n \xrightarrow{d} Y$ , 且随机变量序列  $A_n$  和  $B_n (n = 1, 2, \dots)$  分别依概率收敛于  $a$  和  $b$ , 则  $A_n + B_n Y_n \xrightarrow{d} a + bY$ 。

**定理:** (中心极限定理) 设  $X_i (i = 1, \dots, n)$  是独立同分布的, 且  $E(X_i) = \mu$ ,  $\text{var}(X_i) = \sigma^2 < +\infty$ , 那么  $\sqrt{n}(\bar{X} - \mu)$  弱收敛到  $N(\mu, \sigma^2)$ , 因此  $\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma}$  弱收敛到标准正态分布  $N(0, 1)$ 。

**定理:** ( $\Delta$  方法). 设  $T_n$  为  $\theta$  的估计,  $\sqrt{n}(T_n - \theta) \xrightarrow{d} Z \sim N(0, \tau^2)$ ,  $h'(\theta)$  存在且不为 0, 则

$$\sqrt{n}(h(T_n) - h(\theta)) \xrightarrow{d} W \sim N(0, h'(\theta)^2 \tau^2).$$

**例:** 设总体:  $X \sim N(\mu, \sigma^2)$ , 样本量:  $n$ .

UMVU 估计:  $\hat{\mu} = \bar{X}, \hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ .

$\hat{\mu}$  渐近正态: 事实上,

$$\sqrt{n}(\bar{X} - \mu) \sim N(0, \sigma^2).$$

**例:** (对应郑书例 6.2) 设  $X_1, \dots, X_n$  是来自总体  $X$  的一个样本, 总体  $X$  服从泊松分布, 分布列为

$$P(X_i = k) = \frac{\lambda^k}{k!} \exp\{-\lambda\} \quad (k = 0, 1, 2, \dots; \lambda > 0).$$

已经求得  $\lambda$  的 ML 估计为  $\hat{\lambda} = \bar{X}$ , 不难验证  $\lambda$  的矩估计也是  $\bar{X}$ , 利用中心极限定理, 有

$$\sqrt{n}(\bar{X} - \lambda) = \sqrt{n}[\bar{X} - E(X)] \xrightarrow{\omega} N(0, \text{var}(X)),$$

其中  $\text{var}(X) = \lambda$ , 故  $\bar{X}$  是渐近正态的, 并且渐近方差为  $\lambda$ . 现在试求  $\lambda^2$  的估计, 并讨论其渐近分布. 因  $\lambda$  的 ML 估计为  $\hat{\lambda} = \bar{X}$ , 故  $\lambda^2$  的 ML 估计为  $\bar{X}^2$ , 有

$$\sqrt{n}[h(\bar{X}) - h(\lambda)] \xrightarrow{\omega} N(0, (h'(\lambda))^2 \lambda) = N(0, 4\lambda^3) \quad (n \rightarrow \infty),$$

式中  $h(\bar{X}) = \bar{X}^2$ ,  $h(\lambda) = \lambda^2$ , 由此可知  $\lambda^2$  的 ML 估计  $\bar{X}^2$  是渐近正态的, 其渐近方差为  $4\lambda^2$ .

**例:** 总体:  $X \sim N(\mu, 1)$ , 待估量:  $g(\mu) = P_\mu(X \leq x_0)$ .

方法一、设  $g(\mu) = P_\mu(X - \mu \leq x_0 - \mu) = \Phi(x_0 - \mu)$ .

由 CLT,  $\mu$  的最大似然估计  $\hat{\mu} = \bar{X}$  渐近正态, 渐近方差 = 1.

再由  $\Delta$  方法,  $g(\mu)$  的最大似然估计  $g(\hat{\mu}) = \Phi(x_0 - \bar{X})$  渐近正态, 渐近方差为

$$\sigma_1^2 = g'(\mu)^2 \cdot 1 = \varphi(x_0 - \mu)^2.$$

方法二、注意到  $\frac{1}{n} \sum_{i=1}^n 1_{\{X_i \leq x_0\}} \xrightarrow{P_\mu} P_\mu(X_i \leq x_0) = g(\mu)$ , 渐近正态, 渐近方差为

$$\sigma_2^2 = \text{var}(1_{\{X \leq x_0\}}) = g(\mu)(1 - g(\mu)) = \Phi(x_0 - \mu)(1 - \Phi(x_0 - \mu))$$