



# 概率论、数理统计与随机过程

## 速通版复习资料

作者：2025 级 AI 学组

时间：2026/03-2026/06

版本：期中 ver.(概率论部分)



本资料中大部分内容来源于授课幻灯片和课本

# 目录

|                          |           |  |  |
|--------------------------|-----------|--|--|
| <b>第一章 随机事件与概率</b>       | <b>1</b>  |  |  |
| 1.1 随机事件                 | 1         |  |  |
| 1.1.1 随机现象与随机试验          | 1         |  |  |
| 1.1.2 样本空间与随机事件          | 1         |  |  |
| 1.1.3 随机事件的关系和运算         | 1         |  |  |
| 1.2 概率                   | 3         |  |  |
| 1.2.1 概率的古典定义            | 3         |  |  |
| 1.2.2 概率的统计定义            | 5         |  |  |
| 1.2.3 概率的公理化定义           | 6         |  |  |
| 1.2.4 概率的性质              | 6         |  |  |
| 1.3 条件概率                 | 7         |  |  |
| 1.3.1 条件概率与乘法公式          | 7         |  |  |
| 1.3.2 全概率公式与 Bayes 公式    | 8         |  |  |
| 1.4 事件的独立性               | 9         |  |  |
| <b>第二章 一维随机变量及其分布</b>    | <b>11</b> |  |  |
| 2.1 随机变量与分布函数            | 11        |  |  |
| 2.1.1 随机变量               | 11        |  |  |
| 2.1.2 分布函数               | 11        |  |  |
| 2.2 离散型随机变量              | 12        |  |  |
| 2.2.1 随机变量的分布律           | 12        |  |  |
| 2.2.2 常见的离散型随机变量分布       | 12        |  |  |
| 2.2.2.1 单点分布             | 13        |  |  |
| 2.2.2.2 两点分布             | 13        |  |  |
| 2.2.2.3 二项分布             | 13        |  |  |
| 2.2.2.4 泊松分布             | 13        |  |  |
| 2.2.2.5 其他常见分布           | 14        |  |  |
| 2.3 连续型随机变量              | 14        |  |  |
| 2.3.1 概率密度               | 14        |  |  |
| 2.3.2 常见的连续型随机变量分布       | 15        |  |  |
| 2.3.2.1 正态分布             | 15        |  |  |
| 2.3.2.2 均匀分布             | 17        |  |  |
| 2.3.2.3 指数分布             | 17        |  |  |
| 2.4 随机变量的函数及其概率分布        | 18        |  |  |
| 2.4.1 随机变量的函数            | 18        |  |  |
| 2.4.2 离散型随机变量的函数的概率分布    | 18        |  |  |
| 2.4.3 连续型随机变量的函数的概率分布    | 19        |  |  |
| <b>第三章 随机向量及其概率分布</b>    | <b>21</b> |  |  |
| 3.1 $n$ 维随机向量            | 21        |  |  |
| 3.1.1 随机向量的概念            | 21        |  |  |
| 3.1.2 分布函数与边缘分布函数        | 21        |  |  |
| 3.1.3 二维离散型随机向量          | 22        |  |  |
| 3.1.4 二维连续型随机向量          | 23        |  |  |
| 3.1.4.1 随机向量的概率密度和边缘概率密度 | 23        |  |  |
| 3.1.4.2 二维正态分布           | 24        |  |  |
| 3.1.4.3 二维均匀分布           | 25        |  |  |
| 3.2 条件分布                 | 25        |  |  |
| 3.2.1 离散型随机向量的条件分布律      | 25        |  |  |
| 3.2.2 连续型随机向量的条件概率密度     | 26        |  |  |
| 3.3 随机变量的相互独立性           | 27        |  |  |
| 3.3.1 两个随机向量的相互独立性       | 27        |  |  |
| 3.3.2 多个随机变量的相互独立性       | 28        |  |  |
| 3.3.3 随机向量的相互独立性         | 29        |  |  |
| 3.4 随机向量的函数及其概率分布        | 30        |  |  |
| 3.4.1 随机向量的函数            | 30        |  |  |
| 3.4.2 二维离散型随机向量的函数的概率分布  | 30        |  |  |
| 3.4.3 二维连续型随机向量的函数的概率分布  | 30        |  |  |
| 3.4.3.1 $aX + bY$ 的概率密度  | 31        |  |  |
| 3.4.3.2 $XY$ 的概率密度       | 33        |  |  |
| 3.4.3.3 $X/Y$ 的概率密度      | 33        |  |  |
| 3.4.3.4 $\max$ 和 $\min$  | 34        |  |  |
| <b>第四章 随机变量的数字特征</b>     | <b>35</b> |  |  |
| 4.1 数学期望                 | 35        |  |  |
| 4.1.1 随机变量的数学期望          | 35        |  |  |
| 4.1.2 随机变量的函数的数学期望       | 36        |  |  |
| 4.1.3 数学期望的性质            | 37        |  |  |
| 4.2 方差                   | 37        |  |  |
| 4.2.1 方差与标准差             | 37        |  |  |
| 4.2.2 方差的性质              | 37        |  |  |
| 4.3 协方差、相关系数和矩           | 39        |  |  |
| 4.3.1 协方差                | 39        |  |  |
| 4.3.2 相关系数               | 39        |  |  |
| 4.3.3 矩                  | 40        |  |  |
| 4.3.4 协方差矩阵和正态向量         | 41        |  |  |
| 4.4 条件期望与条件方差            | 42        |  |  |
| <b>第五章 大数定律与中心极限定理</b>   | <b>44</b> |  |  |
| 5.1 随机变量的收敛性             | 44        |  |  |
| 5.1.1 依概率收敛              | 44        |  |  |

---

|                                |    |                                  |    |
|--------------------------------|----|----------------------------------|----|
| 5.1.2 依分布收敛 . . . . .          | 45 | 5.3.4 Khinchin 大数定律 . . . . .    | 48 |
| 5.1.3 其他类型的收敛 . . . . .        | 45 | 5.4 中心极限定理 . . . . .             | 48 |
| 5.2 Chebyshev 不等式 . . . . .    | 45 | 5.4.1 中心极限定律的概念 . . . . .        | 48 |
| 5.3 大数定律 . . . . .             | 46 | 5.4.2 Levy-Lindberg 中心极限定理 . . . | 49 |
| 5.3.1 大数定律的概念 . . . . .        | 46 | 5.4.3 De Moivre-Laplace 中心极限定理   | 50 |
| 5.3.2 Chebyshev 大数定律 . . . . . | 46 | 5.4.4 其他形式的中心极限定理 . . . .        | 51 |
| 5.3.3 Bernoulli 大数定律 . . . . . | 47 |                                  |    |

# 第一章 随机事件与概率

## 内容提要

- 随机事件
- 概率
- 条件概率

- Bayes 公式
- 事件的独立性

## 1.1 随机事件

### 1.1.1 随机现象与随机试验

有些现象的结果有一个确定的范围 (投骰子一定会得到 1 到 6 中的一个点数), 但我们不能或很难预先知道该结果 (究竟会投出哪个点数), 那么这种现象就可认为具有随机性, 称为**随机现象**。相应地, 能预先确定结果的现象称为确定性现象。我们广义上认为确定性现象是随机现象的一种特例。对随机现象的结果的观察、记录称为**随机试验**, 简称为试验, 表示为  $E$ 。

### 1.1.2 样本空间与随机事件

#### 定义 1.1 (样本空间)

对于一个试验  $E$ , 虽然在一次试验之前不能肯定哪个结果会发生, 但试验的一切可能结果是已知的, 称  $E$  的所有可能的试验结果组成的集合称为  $E$  的样本空间, 记为  $\Omega$ ; 样本空间的元素 (亦即  $E$  的每个可能结果) 称为样本点, 记为  $\omega$ 。

举例而言, 设  $E$  表示 “掷一枚正常的硬币, 观察落地时朝上的一面是正面或反面”, 那么  $\Omega = \{\omega_0, \omega_1\}$ , 其中  $\omega_0$  表示 “正面朝上”,  $\omega_1$  表示 “反面朝上”。

在随机试验中, 其样本空间的子集叫做该试验的**随机事件**, 简称事件。以后用  $A, B, C$  等大写字母表示事件。对于一个试验  $E$ , 它的样本空间  $\Omega$  是由  $E$  的全部可能结果组成的集合, 而它的一个事件  $A$  只是由  $E$  的一部分可能结果组成的集合, 因而  $A \subset \Omega$ 。称事件  $A$  发生, 就是当且仅当属于  $A$  的某一个样本点在一次试验中出现。

#### 定义 1.2 (必然事件与不可能事件)

对于一个试验  $E$ , 在每次试验中必然发生的事件, 称为必然事件; 在每次试验中都不发生的事件, 称为不可能事件。一般地, 对于试验  $E$ , 包含它的所有可能的试验结果的样本空间  $\Omega$  是必然事件; 不包含它的任何一个试验结果的事件  $\emptyset$  是不可能事件。今后用  $\Omega$  表示必然事件, 用  $\emptyset$  表示不可能事件。

为了运算方便, 把必然事件  $\Omega$  和不可能事件  $\emptyset$  当作随机事件的两个极端情况。

### 1.1.3 随机事件的关系和运算

随机事件的关系通常可以借助集合论的概念表示。

#### 1. 事件的包含

设有两个事件  $A, B$ , 若事件  $A$  发生必然导致事件  $B$  发生, 则称事件  $B$  包含事件  $A$ , 或者事件  $A$  含于事件  $B$ , 记作  $B \supset A$  或者  $A \subset B$ 。用集合论的术语来表述, 即

$$\forall \omega \in A \Rightarrow \omega \in B$$

## 2. 事件的相等

若事件  $A, B$  相互包含, 即  $A \subset B$  且  $B \subset A$ , 则称事件  $A, B$  相等并记  $A = B$ 。

## 3. 事件的和

设有两事件  $A, B$ , 称  $\{A, B$  中至少有一个发生 $\}$  的事件为  $A$  与  $B$  的和事件, 记为  $A \cup B$ 。事件  $A \cup B$  发生意味着或者仅事件  $A$  发生, 或者仅事件  $B$  发生, 或者两者都发生。用集合论的术语来表达, 即

$$A \cup B = \{\omega \mid \omega \in A \text{ 或 } \omega \in B\}$$

设有可数个事件  $A_1, A_2, \dots, A_n, \dots$ , 定义它们的和事件为  $\{A_1, A_2, \dots, A_n, \dots$  中至少有一个发生 $\}$ , 记为

$$\bigcup_{k=1}^{\infty} A_k = \{A_1, A_2, \dots, A_n, \dots \text{ 中至少有一个发生}\}$$

## 4. 事件的积

设有两事件  $A, B$ , 称  $\{A, B$  都发生 $\}$  的事件为  $A$  与  $B$  的积事件, 记为  $A \cap B$ , 简记为  $AB$ 。用集合论的术语来表达, 即

$$A \cap B = AB = \{\omega \mid \omega \in A \text{ 且 } \omega \in B\}$$

类似地, 可以定义可数个事件  $A_1, A_2, \dots, A_n, \dots$  的积事件如下。

$$\bigcap_{k=1}^{\infty} A_k = \{A_1, A_2, \dots, A_n, \dots \text{ 都发生}\}$$

## 5. 事件的互斥和对立

设有两事件  $A, B$ , 若事件  $A, B$  不能同时发生, 即  $AB = \emptyset$ , 则称  $A, B$  是互斥的, 或称它们是互不相容的。若事件  $A_1, A_2, \dots, A_n$  中的任意两个都互斥, 则称这些事件是两两互斥的, 或称它们是互斥事件组。

互斥事件的特殊情况是对立事件: 设有事件  $A$ , 则事件  $B = \{A \text{ 不发生}\}$  称为  $A$  的对立事件 (逆事件), 记

$$B = \bar{A} = \{\omega \mid \omega \in \Omega \text{ 且 } \omega \notin A\}$$

## 6. 事件的差

设有两事件  $A, B$ , 称事件  $\{A \text{ 发生}, B \text{ 不发生}\}$  为事件  $A$  与事件  $B$  的差事件, 记为  $A - B$ 。用集合论的术语来表达, 即

$$A - B = \{\omega \mid \omega \in A \text{ 且 } \omega \notin B\}$$

根据上述定义易知

$$A - B = A\bar{B} \quad (1.1)$$

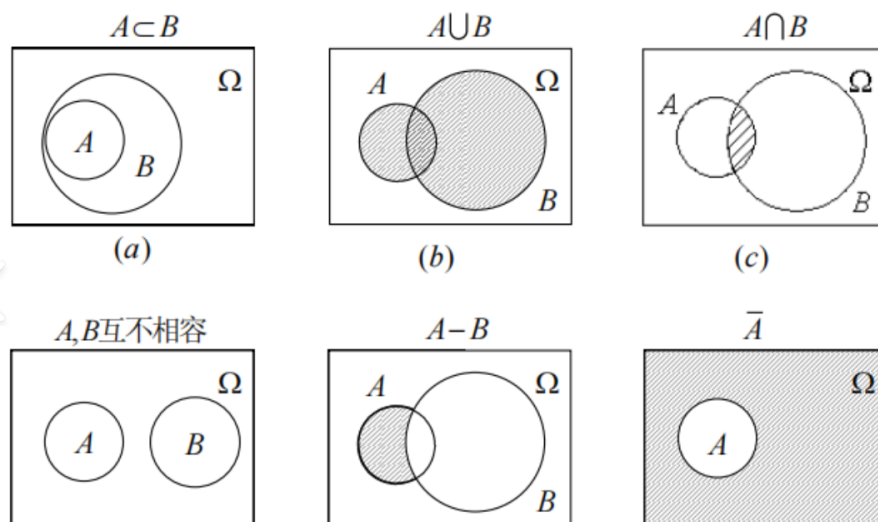


图 1.1: 随机事件的关系示意图

在进行事件的运算时，经常要用到一些定律如下。

### 定理 1.1 (事件运算的定律)

设有事件  $A, B, C$ ，则有：

#### 1. 交换律

$$A \cup B = B \cup A, \quad AB = BA$$

#### 2. 结合律

$$(A \cup B) \cup C = A \cup (B \cup C), \quad (AB)C = A(BC)$$

#### 3. 分配律

$$(A \cup B)C = (AC) \cup (BC), \quad (AB) \cup C = (A \cup C)(B \cup C)$$

#### 4. 对偶律 (De Morgan 律)

$$\overline{A \cup B} = \overline{A} \overline{B}, \quad \overline{AB} = \overline{A} \cup \overline{B}$$



### 例题 1.1 化简式子

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

解 根据式 1.1 和定理 1.1，有

$$\begin{aligned} (A \cup B) - (A - B) &= (A \cup B) \overline{(A - B)} \\ &= (A \cup B) \overline{A \overline{B}} \\ &= (A \cup B) (\overline{A} \cup B) \\ &= AA \cup AB \cup B\overline{A} \cup BB \\ &= B \end{aligned}$$

对于多个事件，通常有 De Morgan 定律的推广形式。

### 命题 1.1 (De Morgan 定律)

设有事件  $A_i (i = 1, 2, \dots, n)$ ，则满足

$$\begin{aligned} \overline{\left(\bigcup_{i=1}^n A_i\right)} &= \bigcap_{i=1}^n \overline{A_i} \\ \overline{\left(\bigcap_{i=1}^n A_i\right)} &= \bigcup_{i=1}^n \overline{A_i} \end{aligned}$$



## 1.2 概率

### 1.2.1 概率的古典定义

如果某随机试验只有有限个试验结果  $\omega_1, \omega_2, \dots, \omega_n$ ，而且从该试验的条件及实施方法分析，又没有理由认为其中的某个结果要比任何其他的结果更容易发生，那么我们只能认为每个试验结果在试验中具有同等可能的出现机会，也即  $\frac{1}{n}$  的出现机会。一般地，若试验满足

- 只有有限个试验结果  $\omega_1, \omega_2, \dots, \omega_n$ ；
- 每个试验结果在一次试验中发生的可能性相等。

则称在此试验下构建的概率模型为古典概型 (等概率概型)，相应地称这种试验为古典概型试验。进一步有



**定义 1.3 (古典概率)**

设古典概型试验的所有结果为  $\Omega = \{\omega_1, \omega_2, \dots, \omega_n\}$ 。若某事件  $A$  恰包含  $\Omega$  中的  $m$  个结果, 则  $A$  的概率  $P(A)$  定义为:

$$P(A) = \frac{A \text{ 所包含的试验结果的个数}}{\text{试验结果的总数}} = \frac{m}{n} \quad (1.2)$$

由式 1.2 算得的事件  $A$  的概率称为古典概率。



**性质** 古典概率满足以下性质。

1.  $P(A) \geq 0$
2.  $P(\Omega) = 1$
3. 若事件  $A_1, A_2, \dots, A_k$  两两互斥, 则

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k)$$

计算古典概率的关键在于求出试验结果的总数和某个特定事件包含的试验结果个数。

**例题 1.2** 一批产品共  $N$  件, 其中次品有  $M$  件, 从中任取  $n$  件, 求事件  $A = \{\text{其中恰有 } m (m \leq M) \text{ 件次品}\}$  的概率。

**解** 在  $N$  件产品中取出  $n$  件, 所有可能的取法共有  $C_N^n$  种, 每一种取法对应一个试验结果。从  $M$  件次品中取  $m$  件, 取法有  $C_M^m$  种, 从其余  $N-M$  件正品中取  $n-m$  件, 取法有  $C_{N-M}^{n-m}$  种, 由组合的乘法原理知有利于事件  $A$  (事件  $A$  包含的试验结果的个数) 的取法共有  $C_M^m C_{N-M}^{n-m}$  种。故  $A$  的概率为

$$P(A) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}$$

在有些实际应用中, 如果试验的结果有无限多个, 那么就不能按照古典概型来计算概率了。然而, 只要试验结果仍具有**等可能性**, 就可以用几何的方法来解决概率的计算问题。

设某试验的样本空间为有界区域  $\Omega$ , 如果样本点落入  $\Omega$  的任何子域中的机会与该子域的度量成正比, 而与其位置及形状无关, 那么样本点落入子域  $A (A \subset \Omega)$  的概率  $P(A)$  被定义为

$$P(A) = \frac{A \text{ 的度量}}{\Omega \text{ 的度量}}$$

其中  $\Omega$  的度量大于 0。上述概率模型称为**几何概型**, 它也被称作等可能概型, 因为这种概率模型同样要求具备等可能性。应用几何概型计算的概率称为**几何概率**。

**例题 1.3 (P23 第 9 题改)** 两艘船都将停靠同一码头, 它们可能在一昼夜的任意时刻到达。设两船停靠的时间分别为 1h 和 2h, 求有两艘船停靠都不需要等待的概率。

**解** 这个概率可用几何方法确定。记  $x$  和  $y$  分别为甲乙两艘轮船到达码头的时刻, 则  $(x, y)$  的可能取值形成边长为 24 的正方形  $\Omega$ , 其面积为  $S_\Omega = 24^2$ 。事件  $A = \{\text{不需要等候码头空出}\}$  有两种可能情况:

一种情况是甲船先到, 则乙船在一小时之后到达, 满足

$$y - x \geq 1$$

另一种情况是乙船先到, 则甲船在两小时之后到达, 满足

$$x - y \geq 2$$

所以事件  $A$  可表示为  $A = \{(x, y) | x - y \leq -1 \text{ 或 } x - y \geq 2\}$ 。所以事件  $A$  的区域形成了图中的阴影部分, 其面积为  $S_A = \frac{1}{2}(23^2 + 22^2)$ , 所以几何概率为

$$P(A) = \frac{S_A}{S_\Omega} = \frac{1013}{1152}$$

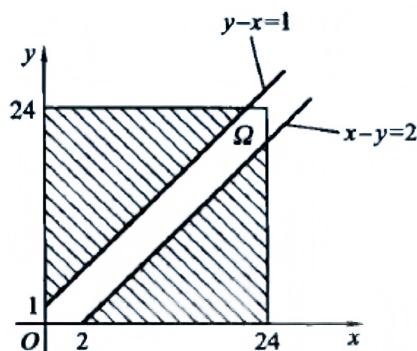


图 1.2: 两艘船停靠时间的分布图

### 1.2.2 概率的统计定义

当试验结果发生的概率不相等时，需要引入概率的统计定义。

#### 定义 1.4 (统计概率)

设有随机试验  $E$ ，在相同的条件下，重复进行了  $n$  次试验。在这  $n$  次试验中，事件  $A$  发生的次数  $n_A$  称为  $A$  发生的频数，比值  $\frac{n_A}{n}$  称为  $A$  发生的频率，记为  $f_n(A)$ ，即

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

♣

**性质** 易知事件发生的频率有以下基本性质：

1.  $f_n(A) \geq 0$
2.  $f_n(\Omega) = 1$
3. 若事件  $A_1, A_2, \dots, A_k$  两两互斥，则

$$f_n(A_1 \cup A_2 \cup \dots \cup A_k) = f_n(A_1) + f_n(A_2) + \dots + f_n(A_k)$$

假设掷一枚质量均匀的硬币，记  $A = \{\text{正面朝上}\}$ 。历史上有人做了大量的重复试验，结果如下图。

| 试验者   | $n$    | $n_A$  | $f_n(A)$ |
|-------|--------|--------|----------|
| 德摩根   | 2 048  | 1 061  | 0.518 1  |
| 蒲 丰   | 4 040  | 2 048  | 0.506 9  |
| K.皮尔逊 | 12 000 | 6 019  | 0.501 6  |
| K.皮尔逊 | 24 000 | 12 012 | 0.500 5  |

图 1.3: 历史上不同试验者对  $A$  发生的频数、频率的数据

可以看出，不管什么人抛掷一枚均匀硬币，当试验次数逐渐增多时，频率  $f_n(A)$  总是在 0.5 附近波动，并呈现出稳定于 0.5 的倾向。频率的这种“稳定性”就是通常所说的统计规律性，它揭示了隐藏在随机现象中的必然规律性，我们用频率的稳定值来刻画事件  $A$  发生的可能性大小是合适的。因此有下面的定义。

#### 定义 1.5 (事件的概率)

设有随机试验  $E$ ，若当试验的重复次数  $n$  充分大时，事件  $A$  发生的频率  $f_n(A)$  稳定地在某数  $p$  附近波动，则称数  $p$  为事件  $A$  的概率并记为

$$P(A) = p \quad (1.3)$$

♣



**性质** 概率满足以下性质：

1.  $P(A) \geq 0$
2.  $P(\Omega) = 1$
3. 若事件  $A_1, A_2, \dots, A_k$  两两互斥，则

$$P(A_1 \cup A_2 \cup \dots \cup A_k) = P(A_1) + P(A_2) + \dots + P(A_k)$$

### 1.2.3 概率的公理化定义

历史上，俄国数学家 Kolmogorov 首次给出了概率的公理化定义，叙述如下。

#### 定义 1.6 (概率的公理化定义)

设  $\mathcal{F}$  是以样本空间  $\Omega$  的一些子集为元素的集合 (谓之集类)，若它满足下面三个条件：

1.  $\Omega \in \mathcal{F}$
2. 若  $A, B \in \mathcal{F}$ ，则  $A - B \in \mathcal{F}$
3. 若  $A_n \in \mathcal{F} (n = 1, 2, \dots)$ ，则  $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$

则称  $\mathcal{F}$  为事件域，称  $\mathcal{F}$  中的元素为事件。设  $\Omega$  为样本空间， $\mathcal{F}$  为  $\Omega$  上的事件域， $P(\cdot)$  是一个定义在  $\mathcal{F}$  上的集合函数，若它满足下列条件：

1.  $P(A) \geq 0, \forall A \in \mathcal{F}$
2.  $P(\Omega) = 1$
3. 设  $A_i \in \mathcal{F} (i = 1, 2, \dots)$ ，且  $A_i A_j = \emptyset, \forall i \neq j$  有：

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

则称  $P(\cdot)$  为  $\mathcal{F}$  上的概率测度，对  $A \in \mathcal{F}$ ，称  $P(A)$  为事件  $A$  的概率。



### 1.2.4 概率的性质

概率有很多重要的性质，根据 1.2.2 小节中概率的 3 条基本性质，可以推出以下性质。

#### 定理 1.2 (概率的重要性质)

1. 不可能事件的概率是 0，即

$$P(\emptyset) = 0 \quad (1.4)$$

2. 若  $A_1, A_2, \dots, A_n$  是两两互斥的事件，则有概率的加法公理

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \quad (1.5)$$

3. 设  $A, B$  是两个事件，若  $A \subset B$ ，则有

$$P(B - A) = P(B) - P(A) \quad (1.6)$$

$$P(B) \geq P(A) \quad (1.7)$$

4. 对任一事件  $A$ ，必有

$$P(\bar{A}) = 1 - P(A) \quad (1.8)$$

5. 对任意两个事件  $A, B$ ，有

$$P(A \cup B) = P(A) + P(B) - P(AB) \quad (1.9)$$

6. 设  $\{A_n, n = 1, 2, \dots\}$  为事件列，若  $A_n \subset A_{n+1}, n = 1, 2, \dots$ ，令  $A = \bigcup_{i=1}^{\infty} A_i$ ，则

$$P(A) = \lim_{n \rightarrow \infty} P(A_n) \quad (1.10)$$

设  $\{A_n, n=1, 2, \dots\}$  为事件列, 若  $A_n \supset A_{n+1}, n=1, 2, \dots$ , 令  $A = \bigcap_{i=1}^{\infty} A_i$ , 则

$$P(A) = \lim_{n \rightarrow \infty} P(A_n) \quad (1.11)$$

7. 设  $A_1, A_2, \dots, A_n$  是任意  $n$  个事件, 则有

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j) + \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \dots + (-1)^{n+1} P(A_1 A_2 \cdots A_n) \quad (1.12)$$



**例题 1.4**(此题在课上疑问很多, 但答案是正确的) 从  $n$  双不同的手套中任取  $2k$  ( $2k < n$ ) 只, 求下列事件的概率

$A = \{\text{恰有两只手套配成一双}\}$

$B = \{\text{至少有两只手套配成一双}\}$

**解** 从  $n$  双手套中取  $2k$  只, 不同的取法总数为  $C_{2n}^{2k}$ 。有利于事件  $A$  的取法可分三步完成。

- 第一步, 从  $n$  双手套中取出一双作为配对的那两只手套, 有  $C_n^1$  种取法;
- 第二步, 从剩下的  $n-1$  双手套中取出  $2(k-1)$  双手套有  $C_{n-1}^{2k-2}$  种取法;
- 第三步, 从这不同的  $2(k-1)$  双手套中各选一只手套, 因可取左或右, 故有  $2^{2(k-1)}$  种取法。

因此, 有利于事件  $A$  的取法数为  $2^{2(k-1)} \cdot C_n^1 C_{n-1}^{2k-2}$ , 故事件  $A$  的概率为

$$P(A) = 2^{2(k-1)} \cdot \frac{C_n^1 C_{n-1}^{2k-2}}{C_{2n}^{2k}}$$

因为有利于事件  $\bar{B}$  的取法数为  $2^{2k} \cdot C_n^{2k}$ , 所以事件  $B$  的概率为

$$P(B) = 1 - P(\bar{B}) = 1 - 2^{2k} \cdot \frac{C_n^{2k}}{C_{2n}^{2k}}$$

## 1.3 条件概率

### 1.3.1 条件概率与乘法公式

#### 定义 1.7 (条件概率)

设  $A, B$  为两个事件, 且  $P(B) \neq 0$ , 记

$$P(A|B) = \frac{P(AB)}{P(B)} \quad (1.13)$$

为在事件  $B$  发生的条件下事件  $A$  的条件概率。



由条件概率的定义, 即可得到下述定理。

#### 定理 1.3 (乘法公式)

若  $P(B) > 0$ , 则:

$$P(AB) = P(B)P(A|B) \quad (1.14)$$

进一步可以推广到  $n$  ( $n > 2$ ) 个事件的情形。对任意  $n$  个事件  $A_1, A_2, \dots, A_n$ , 若  $P(A_1 A_2 \cdots A_{n-1}) > 0$ , 则

$$P(A_1 A_2 \cdots A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 A_2) \cdots P(A_n|A_1 A_2 \cdots A_{n-1}) \quad (1.15)$$



**笔记**  $P(A|B)$  表明在  $B$  已经发生的条件下  $A$  发生的概率, 因此式 1.13 中  $P(B)$  作为分母。

**例题 1.5**  $n$  个人以抽签方式决定谁将得到一张演出票,  $n$  个人依次抽签, 求第  $k$  个人 ( $1 \leq k \leq n$ ) 抽中的概率。

**解** 令  $A_k = \{\text{第 } k \text{ 个人抽到了票}\}$ ,  $\bar{A}_k = \{\text{第 } k \text{ 个人没抽到票}\}$ , 其中  $1 \leq k \leq n$ 。显然  $P(A_1) = \frac{1}{n}$ 。对于  $1 < k \leq n$  有

$$A_k \subset \bar{A}_1 \bar{A}_2 \cdots \bar{A}_{k-1}$$

故  $A_k = \bar{A}_1 \bar{A}_2 \cdots \bar{A}_{k-1} A_k$ 。由乘法公式得

$$\begin{aligned} P(A_k) &= P(\bar{A}_1 \bar{A}_2 \cdots \bar{A}_{k-1} A_k) \\ &= P(\bar{A}_1)P(\bar{A}_2|\bar{A}_1)P(\bar{A}_3|\bar{A}_1 \bar{A}_2) \cdots P(A_k|\bar{A}_1 \bar{A}_2 \cdots \bar{A}_{k-1}) \\ &= \frac{n-1}{n} \cdot \frac{n-2}{n-1} \cdot \frac{n-3}{n-2} \cdots \frac{1}{n-(k-1)} \\ &= \frac{1}{n} \quad (1 \leq k \leq n) \end{aligned}$$

### 1.3.2 全概率公式与 Bayes 公式

#### 定义 1.8 (互斥完备事件群)

设  $B_1, B_2, \cdots, B_n, \cdots$  为可数个事件, 若它们满足

$$1. B_i B_j = \emptyset \quad (i \neq j, i, j = 1, 2, \cdots)$$

$$2. B_1 \cup B_2 \cup \cdots \cup B_n \cup \cdots = \Omega$$

则称这组事件为互斥完备事件群 (或互不相容完备事件组)。

若事件  $B_1, B_2, \cdots, B_n, \cdots$  构成互斥完备事件群。现考虑任一事件  $A$ , 根据事件的运算有

$$A = A\Omega = AB_1 \cup AB_2 \cup \cdots \cup AB_n \cup \cdots$$

由  $B_1, B_2, \cdots, B_n, \cdots$  两两互斥可知  $AB_1, AB_2, \cdots, AB_n, \cdots$  亦两两互斥, 设  $P(B_i) > 0 (i = 1, 2, \cdots)$ , 结合概率的加法公理与乘法公式得

$$\begin{aligned} P(A) &= P(AB_1) + P(AB_2) + \cdots + P(AB_n) + \cdots \\ &= P(B_1)P(A|B_1) + P(B_2)P(A|B_2) + \cdots + P(B_n)P(A|B_n) + \cdots \\ &= \sum_j P(B_j)P(A|B_j) \end{aligned}$$

#### 定理 1.4 (全概率公式)

若事件  $B_1, B_2, \cdots, B_n, \cdots$  构成互斥完备事件群, 且  $P(B_i) > 0 (i = 1, 2, \cdots)$ , 则对任一事件  $A$ , 有

$$P(A) = \sum_j P(B_j)P(A|B_j) \quad (1.16)$$

**笔记** 全概率公式把事件  $A$  的“全部”概率  $P(A)$  分解成许多“部分”概率  $P(AB_j)$  之和。因此, 当概率  $P(B_j)$  与  $P(A|B_j)$  容易确定时, 通过全概率公式可以求得复杂事件  $A$  的概率  $P(A)$ 。

**例题 1.6(P25 第 29 题)** 某射击小组共有 20 名射手, 其中一级射手 4 人、二级射手 8 人、三级射手 7 人、四级射手 1 人。一、二、三、四级射手在一次射击中能中 10 环的概率分别为 0.9, 0.7, 0.5, 0.2。现从该小组中任选一名, 求他在一次射击中能中 10 环的概率。

**解** 记  $A = \{\text{该射手射击中了 10 环}\}$ ,  $B_i = \{\text{选到了第 } i \text{ 级射手}\}$ , 其中  $i = 1, 2, 3, 4$ 。由题意知  $B_i$  构成互斥完备事件群, 则由全概率公式

$$P(A) = 0.9 \times \frac{4}{20} + 0.7 \times \frac{8}{20} + 0.5 \times \frac{7}{20} + 0.2 \times \frac{1}{20} = 0.645$$

#### 定理 1.5 (Bayes 公式)

若事件  $B_1, B_2, \cdots, B_n, \cdots$  构成互斥完备事件群, 且  $P(B_i) > 0 (i = 1, 2, \cdots)$ , 则对任一事件  $A (P(A) > 0)$ , 有

$$P(B_i|A) = \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)} \quad (i = 1, 2, \cdots) \quad (1.17)$$

**证明** 由条件概率和全概率公式有

$$\begin{aligned} P(B_i|A) &= \frac{P(B_i A)}{P(A)} = \frac{P(B_i)P(A|B_i)}{P(A)} \\ &= \frac{P(B_i)P(A|B_i)}{\sum_j P(B_j)P(A|B_j)} \quad (i = 1, 2, \dots) \end{aligned}$$

在 Bayes 公式中, 通常称  $P(B_1), P(B_2), \dots$  为先验概率, 而称  $P(B_1|A), P(B_2|A), \dots$  为后验概率。因此 Bayes 公式实际上是计算后验概率的公式, 亦是概率论中的重要公式。

**例题 1.7** 某个地区患有某种疾病的人占 0.05%, 患者对某种检测结果呈阳性的概率为 0.99, 正常人对这种检测结果呈阳性的概率是 0.005, 现有一人检测结果是阳性, 问此人患有此疾病的概率是多少?

**解** 记  $A = \{\text{患有此疾病}\}, B = \{\text{检测结果呈阳性}\}$ , 则  $\bar{A} = \{\text{没有患此疾病}\}$ , 且  $P(A) = 0.0005, P(\bar{A}) = 0.9995, P(B|A) = 0.99, P(B|\bar{A}) = 0.005$ 。由 Bayes 公式, 有

$$\begin{aligned} P(A|B) &= \frac{P(A)P(B|A)}{P(A)P(B|A) + P(\bar{A})P(B|\bar{A})} \\ &= \frac{0.0005 \times 0.99}{0.0005 \times 0.99 + 0.9995 \times 0.005} \\ &\approx 0.09 \end{aligned}$$

## 1.4 事件的独立性

### 定义 1.9 (独立事件)

若两事件  $A, B$  满足:

$$P(AB) = P(A)P(B) \quad (1.18)$$

则称  $A, B$  相互独立, 简称  $A, B$  独立。进一步推广如下:

设  $A_1, A_2, \dots, A_n$  是  $n$  个事件, 若对任意  $k (1 < k \leq n)$ , 任意  $1 \leq i_1 < i_2 < \dots < i_k \leq n$  都有

$$P(A_{i_1} A_{i_2} \dots A_{i_k}) = P(A_{i_1}) P(A_{i_2}) \dots P(A_{i_k})$$

都成立, 则称事件  $A_1, A_2, \dots, A_n$  相互独立。

**注** 关于定义 1.9, 有几点注意事项。

1. 概率为零的事件与任意事件相互独立
2. 当  $P(A)P(B) \neq 0$  时, 有

$$P(A) = P(A|B) \Leftrightarrow P(B) = P(B|A) \Leftrightarrow P(AB) = P(A)P(B)$$

判断若干个事件的独立性, 只需要验证它们的积事件概率等于各自概率的乘积。

### 定理 1.6

若四对事件  $\{A, B\}, \{A, \bar{B}\}, \{\bar{A}, B\}, \{\bar{A}, \bar{B}\}$  中有一对是相互独立的, 则另外三对也是相互独立的。

类似的, 设  $n$  个事件  $A_1, A_2, \dots, A_n$  相互独立, 则把其中任意  $m (1 \leq m \leq n)$  个事件相应地换成它们的对立事件, 所得的  $n$  个事件仍相互独立。

**例题 1.8(P25 第 38 题)** 设有  $2n$  个元件, 每个元件的可靠性 (即元件能正常工作的概率) 为  $p (0 < p < 1)$ , 且各元件能否正常工作是相互独立的。若按图 1.4、图 1.5 中的方式装配成两个系统。试求两个系统的可靠性, 并比较哪个系统的可靠性大?



图 1.4: 系统 I 工作示意图

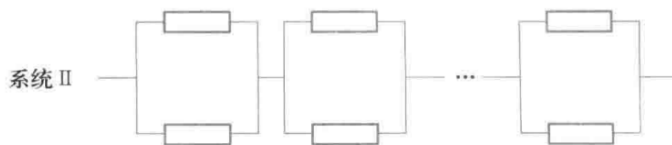


图 1.5: 系统 II 工作示意图

**解** 根据串并联特点，可以分别写出系统 I 和 II 的可靠性（正常工作的概率）

$$\begin{aligned}
 P(\text{系统 I 正常运行}) &= 1 - P(\text{两条并联线路均故障}) \\
 &= 1 - [1 - P(n \text{ 个串联元件都正常工作})]^2 \\
 &= 1 - (1 - p^n)^2 \\
 &= p^n(2 - p^n)
 \end{aligned}$$

$$\begin{aligned}
 P(\text{系统 II 正常运行}) &= [P(\text{一组并联元件中至少有一个运行})]^n \\
 &= [1 - P(\text{一组并联元件都不运行})]^n \\
 &= [1 - (1 - p)^2]^n \\
 &= p^n(2 - p)^n
 \end{aligned}$$

不难比较，当  $n \geq 2$  时系统 II 的可靠性更高。

## 第二章 一维随机变量及其分布

### 内容提要

- 分布函数
- 离散型随机变量
- 概率密度
- 连续型随机变量
- 随机变量的函数及其概率分布

## 2.1 随机变量与分布函数

### 2.1.1 随机变量

为了把随机的事件用数学语言表述, 需要考虑把试验结果数量化: 无论随机试验的结果本身与数量有无联系, 我们都能把试验的每个结果与实数对应起来。由于这样的数量依赖试验结果, 而对随机试验来说, 在每次试验之前无法断言会出现何种结果, 因而也就无法确定它会取什么值, 即它的取值具有随机性, 我们称这样的变量为随机变量。

随机变量是随着试验结果的不同而变化的量, 因此随机变量是试验结果的函数。一般地有以下定义。

#### 定义 2.1 (随机变量)

设  $E$  为一随机试验,  $\Omega$  为其样本空间, 若  $X = X(\omega), \omega \in \Omega$  为单值实函数, 且对于任意实数  $x$ , 集合  $\{\omega \mid X(\omega) \leq x\}$  都是随机事件, 则称  $X$  为随机变量。

 **笔记** 在研究随机现象的理论中引入随机变量概念, 其意义在于: 一方面借助随机变量可以更灵活便捷地描述随机现象; 另一方面也为运用微积分等近现代数学工具来研究随机现象的统计规律性奠定良好的基础。

### 2.1.2 分布函数

对于随机变量  $X$ , 我们不只是看它取哪些值, 更重要的是看它以多大的概率取那些值。由随机变量的定义可知, 对于每一个实数  $x$ , 集合  $\{X \leq x\}$  都是一个随机事件, 因而有一个确定的概率  $P\{X \leq x\}$  与  $x$  相对应, 因此, 概率  $P\{X \leq x\}$  是  $x$  的函数, 这个函数在理论和应用中都很重要, 为此, 引入以下定义。

#### 定义 2.2 (分布函数)

设  $X$  是一随机变量, 记

$$F(x) = P\{X \leq x\}, \quad x \in \mathbb{R} \quad (2.1)$$

称  $F(x)$  为  $X$  的分布函数。

关于  $X$  的随机事件的概率都可以方便地利用分布函数  $F(x)$  算出来。以  $P\{a < X \leq b\}$  为例, 其中  $a < b$ 。因为  $\{a < X \leq b\} = \{X \leq b\} - \{X \leq a\}$ , 而  $a < b$ , 所以  $\{X \leq a\} \subset \{X \leq b\}$ , 故由概率的可减性得

$$P\{a < X \leq b\} = P\{X \leq b\} - P\{X \leq a\} = F(b) - F(a) \quad (2.2)$$

可见, 如果已知  $X$  的分布函数, 就能知道  $X$  落在任意区间  $(a, b]$  的概率。注意这个区间必须左开右闭。

**性质** 分布函数  $F(x)$  也具有以下基本性质:

1.  $F(x)$  是  $x$  的单调增函数
2.  $F(x)$  是  $x$  的右连续函数



3. 对任意实数  $x$  都有  $0 \leq F(x) \leq 1$ , 且定义:

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

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

需要指出, 如果一个函数满足上述三条性质, 那么这函数一定是某个随机变量的分布函数。因而可用这三条性质来判别一个函数是否为分布函数。

**例题 2.1(P47 第 4 题)** 设随机变量  $X$  的分布函数为:

$$F(x) = A + B \arctan x \quad (-\infty < x < +\infty)$$

求常数  $A$  及  $B$  以及随机变量  $X$  落在区间  $(-1, 1]$  内的概率。

**解** 根据分布函数的性质, 需要满足  $F(-\infty) = 0$  且  $F(+\infty) = 1$ , 根据极限的运算可以求得:

$$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = A - \frac{\pi}{2} B = 0$$

$$F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = A + \frac{\pi}{2} B = 1$$

$$\text{解得 } A = \frac{1}{2}, B = \frac{1}{\pi}, \text{ 因此 } P\{-1 < X \leq 1\} = F(1) - F(-1) = \frac{1}{2}.$$

## 2.2 离散型随机变量

### 2.2.1 随机变量的分布律

若随机变量  $X$  的所有可能取值为有限个或者可数个, 则称  $X$  为离散型随机变量。

**定义 2.3 (随机变量的分布律)**

设  $X$  为离散型随机变量, 其所有可能的取值为  $\{x_1, x_2, \dots\}$ , 则称:

$$p_i = P\{X = x_i\}, \quad (i = 1, 2, \dots) \quad (2.3)$$

为随机变量  $X$  的分布律 (或概率函数)。



**性质** 随机变量  $X$  的分布律具有以下性质:

$$1. p_i \geq 0 \quad (i = 1, 2, \dots)$$

$$2. p_1 + p_2 + \dots = 1$$

随机变量  $X$  的分布律可以用一个表格表示, 也可以表示为:

$$X \sim \begin{pmatrix} x_1 & x_2 & \cdots & x_n & \cdots \\ p_1 & p_2 & \cdots & p_n & \cdots \end{pmatrix}$$

$X$  的分布律 2.3 指出了全部概率 1 在  $X$  的可能值的集合  $\{x_1, x_2, \dots\}$  上的分布情况, 因而 2.3 式又叫做随机变量  $X$  的概率分布。若已知离散型随机变量  $X$  的分布律  $p_i = P\{X = x_i\}, i = 1, 2, \dots$ , 则分布函数:

$$F(x) = P\{X \leq x\} = \sum_{x_i \leq x} P\{X = x_i\} = \sum_{x_i \leq x} p_i \quad (2.4)$$

反之, 若已知离散型随机变量  $X$  的分布函数  $F(x)$ , 欲求其分布律, 记  $\{x_1, x_2, \dots\}$  为  $F(x)$  的间断点集合, 令  $p_i = F(x_i) - F(x_i - 0), x_i \in \{x_1, x_2, \dots\}$ , 则  $X$  的分布律为:

$$P\{X = x_i\} = p_i, \quad x_i \in \{x_1, x_2, \dots\} \quad (2.5)$$

### 2.2.2 常见的离散型随机变量分布

下面介绍几种常见的离散型随机变量分布。

### 2.2.2.1 单点分布

若随机变量  $X$  的分布律为:

$$P\{X = a\} = 1$$

则称  $X$  服从单点分布, 记为:

$$X \sim \begin{pmatrix} a \\ 1 \end{pmatrix}$$

### 2.2.2.2 两点分布

若随机变量  $X$  的分布律为  $P\{X = a_0\} = 1 - p, P\{X = a_1\} = p$ , 或者统一地表示为:

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

其中  $0 < p < 1$ , 则称  $X$  服从两点分布。当  $a_0 = 0, a_1 = 1$  时, 称  $X$  服从 (0-1) 分布, 记为  $X \sim B(1, p)$  或:

$$X \sim \begin{pmatrix} 0 & 1 \\ 1-p & p \end{pmatrix}$$

若试验  $E$  如果只有两种结果, 称其是单次 Bernoulli 试验。

### 2.2.2.3 二项分布

设某事件  $A$  在一次试验中发生的概率为  $p$ , 现将此试验独立地重复进行  $n$  次。若以  $X$  表示在这  $n$  次试验中事件  $A$  发生的次数, 则  $X$  的所有可能取值为  $0, 1, 2, \dots, n$ 。上述过程也可看做独立、重复地进行了  $n$  次 Bernoulli 试验, 记为  $n$  重 Bernoulli 试验。

#### 定义 2.4 (二项分布)

若随机变量  $X$  的分布律为:

$$P\{X = i\} = C_n^i p^i (1-p)^{n-i}, \quad (i = 0, 1, 2, \dots, n) \quad (2.6)$$

其中  $0 < p < 1$ , 则称  $X$  服从参数为  $n, p$  的二项分布, 记为  $X \sim B(n, p)$ 。

**例题 2.2** 设每次射击击中目标的概率为 0.001, 现独立射击 1000 次, 求至少击中 2 次的概率。

**解** 以  $X$  表示这 1000 次独立射击中击中目标的次数, 则  $X \sim B(1000, 0.001)$ , 故所求概率为:

$$\begin{aligned} P\{X \geq 2\} &= 1 - P\{X < 2\} \\ &= 1 - P\{X = 0\} - P\{X = 1\} \\ &= 1 - 0.999^{1000} - 1000 \times 0.001 \times 0.999^{999} = 0.264 \end{aligned}$$

### 2.2.2.4 泊松分布

若把例题 2.2 中的 2 次改成更大的数字, 计算会相当麻烦, 因此引入下述近似算法。

#### 定理 2.1 (泊松定理近似)

设随机变量  $X_n \sim B(n, p_n)$ , ( $n = 1, 2, \dots$ ), 其中  $p_n$  与  $n$  有关且满足  $\lim_{n \rightarrow \infty} np_n = \lambda > 0$ , 则:

$$\lim_{n \rightarrow \infty} C_n^k p_n^k (1-p_n)^{n-k} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (k = 0, 1, 2, \dots) \quad (2.7)$$

在实际应用中, 当  $n \geq 10$  且  $p \leq 0.1$ , 且  $\lambda = np$  大小适中时, 有以下近似式:

$$C_n^k p^k (1-p)^{n-k} \approx \frac{\lambda^k}{k!} e^{-\lambda}, \quad (k = 0, 1, 2, \dots) \quad (2.8)$$

由泊松定理, 若二项分布的试验次数  $n$  很大, 二项分布的概率  $p$  很小, 且乘积  $\lambda = np$  比较适中, 则事件出现的次数  $X$  可以用下述随机变量分布来逼近。

### 定义 2.5 (泊松分布)

若随机变量  $X$  的分布律为:

$$P\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad (k = 0, 1, 2, \dots) \quad (2.9)$$

其中  $\lambda > 0$  为常数, 则称  $X$  服从参数为  $\lambda$  的泊松分布, 记为  $X \sim P(\lambda)$ 。

### 2.2.2.5 其他常见分布

在 Bernoulli 试验的条件下, 若设  $X$  为事件  $A$  首次发生时所需的试验次数, 令  $\{X = k\}$  表示第  $k$  次试验时  $A$  发生且前  $k-1$  次试验  $A$  都不发生, 由 Bernoulli 试验为独立重复试验的特点得到下述分布。

### 定义 2.6 (几何分布)

若随机变量  $X$  的分布律为:

$$P\{X = k\} = (1-p)^{k-1} p, \quad (k = 1, 2, 3, \dots) \quad (2.10)$$

其中  $0 < p < 1$  为常数, 则称  $X$  服从参数为  $p$  的几何分布, 记为  $X \sim Ge(p)$ 。

**性质** 若  $X \sim Ge(p)$ , 则  $P\{X = k+n | X > k\} = P\{X = n\} (k, n = 1, 2, 3, \dots)$ , 称为几何分布的无记忆性。

接着讨论如下问题: 从有限  $N$  个物件 (其中包含  $M$  个指定种类的物件) 中不放回地抽出  $n$  个物件, 成功抽出该指定种类的物件的次数记为  $k (k \leq \min\{n, M\})$ 。设  $X$  为抽出该指定物件的次数, 则有:

### 定义 2.7 (超几何分布)

若随机变量  $X$  的分布律为:

$$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, \quad (k = 0, 1, \dots, \min\{n, M\}) \quad (2.11)$$

则称  $X$  服从超几何分布, 记作  $X \sim H(n, M, N)$ 。

## 2.3 连续型随机变量

### 2.3.1 概率密度

除了离散型随机变量之外, 还有一类重要的随机变量——连续型随机变量, 这种随机变量  $X$  可以取某个区间如  $[a, b]$  或  $(-\infty, +\infty)$  中的一切值。在理论和实践中, 刻画连续型随机变量的概率分布常使用概率密度的定义。

### 定义 2.8 (概率密度)

设  $X$  为随机变量, 若存在非负可积函数  $f(x)$ , 使得对任意区间  $A \subset \mathbb{R}$  都有:

$$P\{X \in A\} = \int_A f(x) dx \quad (2.12)$$

则称  $X$  为连续型随机变量, 并称  $f(x)$  为  $X$  的概率密度。

**性质** 概率密度  $f(x)$  具有下述性质:

1.  $f(x) \geq 0 (-\infty < x < +\infty)$
2.  $\int_{-\infty}^{+\infty} f(x) dx = 1$
3. 若连续型随机变量  $X$  的分布函数为  $F(x)$ , 则对任意实数  $x$ , 有

$$F(x) = \int_{-\infty}^x f(t) dt \quad (2.13)$$

4. 若  $f(x)$  在  $x$  点处连续, 则有  $F'(x) = f(x)$

由高等数学的知识结合 2.13 式可知——连续型随机变量的分布函数  $F(x)$  必为连续函数。这能够推导出另一个重要事实, 即对于任意实数  $x$ , 随机变量  $X$  取到  $x$  的概率为 0。

$$P\{X = x\} = 0$$

事实上, 设  $X$  的分布函数为  $F(x)$ , 对任意  $\Delta x > 0$ , 由  $\{X = x\} \subset \{x - \Delta x < X \leq x\}$  得:

$$0 \leq P\{X = x\} \leq P\{x - \Delta x < X \leq x\} = F(x) - F(x - \Delta x)$$

因为  $X$  为连续型随机变量, 其分布函数  $F(x)$  是连续函数, 所以在上述不等式两端, 令  $\Delta x \rightarrow 0^+$  即得证。由连续型随机变量的上述特性可知:

$$P\{a < X < b\} = P\{a \leq X < b\} = P\{a < X \leq b\} = P\{a \leq X \leq b\} = \int_a^b f(x) dx \quad (2.14)$$

这表明只要区间端点  $a, b$  确定, 则  $X$  落入区间  $(a, b), [a, b), (a, b], [a, b]$  的概率都等于式 2.14 的右端积分。当  $x$  为  $X$  的某个可能值时, 事件  $\{X = x\}$  并非不可能事件, 但有  $P\{X = x\} = 0$ 。若事件  $A = \emptyset$ , 则  $P(A) = 0$ ; 反之, 若事件  $A$  的概率  $P(A) = 0$ , 则  $A$  未必是不可能事件。

**例题 2.3(P49 第 17 题)** 设随机变量  $X$  的概率密度为  $f(x) = \begin{cases} \frac{A}{\sqrt{1-x^2}}, & |x| < 1 \\ 0, & |x| \geq 1 \end{cases}$ , 求:

1. 常数  $A$
2.  $X$  落在区间  $\left[-\frac{1}{2}, \frac{1}{2}\right]$  上的概率
3.  $X$  的分布函数

**解** 求解这个问题需要用到概率密度的定义和性质。

1. 概率密度需要满足归一化条件, 即概率密度函数性质 2。

$$\int_{-\infty}^{+\infty} f(x) dx = \int_{-1}^1 \frac{A}{\sqrt{1-x^2}} dx = A\pi = 1 \implies A = \frac{1}{\pi}$$

2. 根据式 2.14 知:

$$P\left\{-\frac{1}{2} \leq X \leq \frac{1}{2}\right\} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \frac{1}{\pi\sqrt{1-x^2}} dx = \frac{1}{3}$$

3. 根据概率密度的性质 3 (式 2.13) 可以计算  $X$  的分布函数。

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < -1 \\ \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{1}{x}, & -1 \leq x < 1 \\ 1, & x \geq 1 \end{cases}$$

## 2.3.2 常见的连续型随机变量分布

### 2.3.2.1 正态分布

#### 定义 2.9 (正态分布)

若连续型随机变量  $X$  的概率密度为:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (-\infty < x < +\infty) \quad (2.15)$$

其中  $\mu, \sigma (\sigma > 0)$  为常数, 则称  $X$  服从参数为  $\mu, \sigma$  的正态分布, 记为  $X \sim N(\mu, \sigma^2)$ 。习惯上称服从正态分布的随机变量为正态变量

正态分布概率密度曲线的这种“中间高, 两头低”的特点反映了在正常状态下一般事物所遵循的客观规律。例如一大群人的身高, 特别高大者和殊为矮小者占少数, 而处中间状态者居多; 各种职业的人的收入、大批同类

产品检验时的误差等亦如此，它们都不同程度地符合正态分布。因此，正态分布无论在理论上还是在实际应用中都占有十分重要的地位。

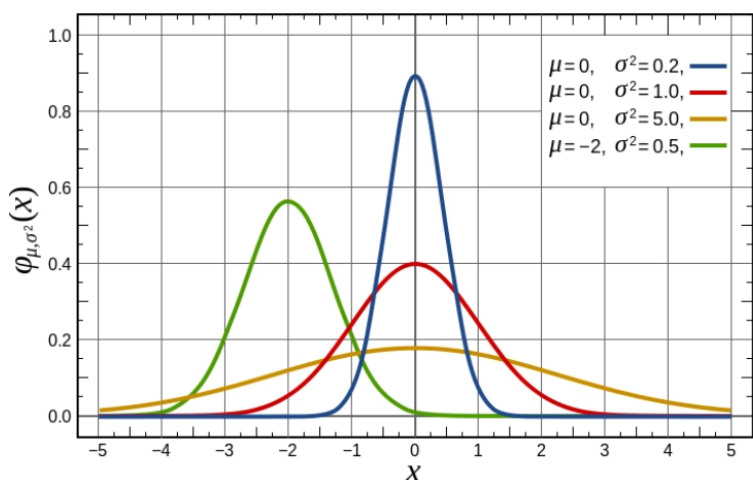


图 2.1: 不同参数  $\mu, \sigma$  下的正态变量的概率密度图像

当  $\mu = 0, \sigma = 1$  时称  $X$  服从标准正态分布，其概率密度、分布函数分别用专用符号  $\varphi(x), \Phi(x)$  表示。

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad (-\infty < x < +\infty) \quad (2.16)$$

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_{-\infty}^x \varphi(t) dt \quad (2.17)$$

#### 命题 2.1

对于标准正态分布的分布函数  $\Phi(x)$ ，有恒等式

$$\Phi(-x) = 1 - \Phi(x) \quad (2.18)$$

特别地，有  $\Phi(0) = 0.5$

标准正态分布的分布函数  $\Phi(x)$  在正态分布的概率计算中起着重要作用，人们已编制了  $\Phi(x)$  的函数值表供查用。利用  $\Phi(x)$  可进行任何正态分布的概率计算，这是由下面的命题确定的。

#### 命题 2.2

若随机变量  $X \sim N(\mu, \sigma^2)$ ，则有  $Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$ 。

**例题 2.4** 设随机变量  $X \sim N(1.4, 4)$ ，求  $P\{|X - 1.4| > 2\}$ 。

**解** 可知  $\mu = 1.4, \sigma = 2$ ，根据命题 2.1 和 2.2 可以计算：

$$\begin{aligned} P\{|X - 1.4| > 2\} &= P\left\{\left|\frac{X - 1.4}{2}\right| > 1\right\} \\ &= 1 - P\left\{\left|\frac{X - 1.4}{2}\right| \leq 1\right\} \\ &= 1 - [\Phi(1) - \Phi(-1)] \\ &= 2 - 2\Phi(1) = 2(1 - 0.8413) = 0.3174 \end{aligned}$$

## 2.3.2.2 均匀分布

## 定义 2.10 (均匀分布)

若随机变量  $X$  具有概率密度

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ 或 } x > b \end{cases} \quad (2.19)$$

则称  $X$  在区间  $[a, b]$  上服从均匀分布, 记为  $X \sim U(a, b)$ , 其中  $a, b$  为常数。

分布的“均匀”性是指  $X$  具有下述意义的等可能性, 即它落在  $[a, b]$  内任意等长度的子区间内的概率是相同的。如果随机变量  $X \sim U(a, b)$ , 那么  $X$  的分布函数

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 0, & x < a \\ \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}, & a \leq x < b \\ \int_a^b \frac{1}{b-a} dt = 1, & x \geq b \end{cases}$$

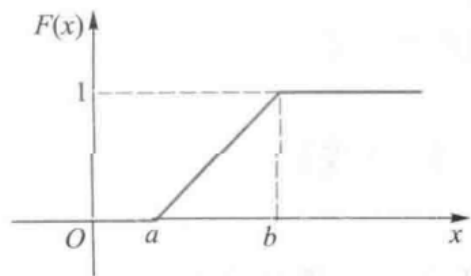
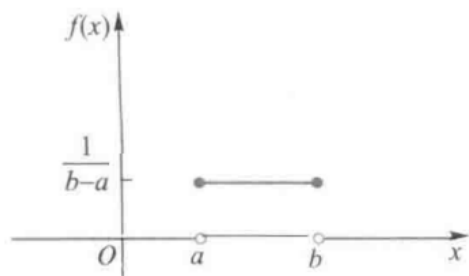


图 2.2:  $X \sim U(a, b)$  时的概率密度  $f(x)$  和分布函数  $F(x)$

## 2.3.2.3 指数分布

## 定义 2.11 (指数分布)

若随机变量  $X$  具有概率密度

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad (2.20)$$

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

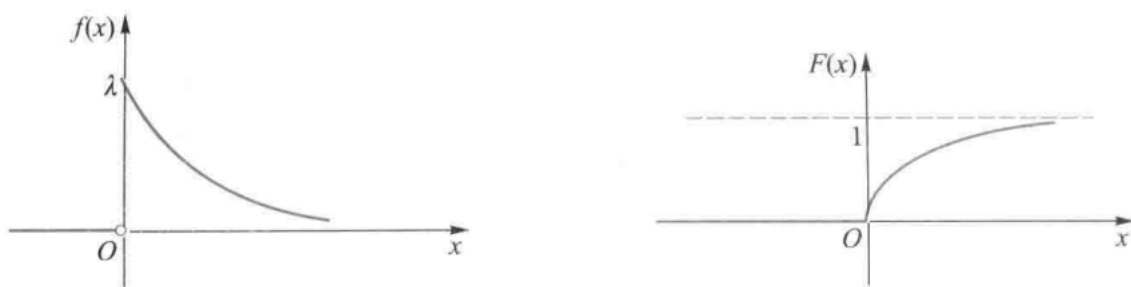
如果随机变量  $X \sim \text{Exp}(\lambda)$ , 那么  $X$  的分布函数

$$F(x) = \int_{-\infty}^x f(t) dt = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

此时, 对任意的  $x \geq 0$  有

$$P\{X > x\} = 1 - F(x) = e^{-\lambda x}$$

**性质** 若  $X \sim \text{Exp}(\lambda)$ , 则对任意的  $s > 0, t > 0$  有  $P\{X > s+t | X > s\} = P\{X > t\}$ , 这被称作指数分布的无记忆性。也有人把它比喻为“永远年轻”, 意思是若以  $X$  表示某人的寿命, 则在已知其寿命超过  $s$  岁的条件下还能多活  $t$  年的概率与其出生时能活过  $t$  年的概率是相同的。

图 2.3:  $X \sim \text{Exp}(\lambda)$  时的概率密度  $f(x)$  和分布函数  $F(x)$ 

## 2.4 随机变量的函数及其概率分布

### 2.4.1 随机变量的函数

当一个随机变量变换成另一个随机变量时, 如何由原来随机变量的概率分布来确定新随机变量的概率分布? 因此需要研究相关随机变量之间的关系, 并用已知随机变量的分布来求出与之有关的另一个随机变量的分布。这引出了随机变量的函数的定义。

#### 定义 2.12 (随机变量的函数)

记随机变量  $X$  的一切可能值集合为  $D$ , 设  $g(x)$  是定义在  $D$  上的单值实函数。若对于  $X$  的每一个可能值  $x \in D$ , 随机变量  $Y$  相应地取值  $y = g(x)$ , 则称  $Y$  为  $X$  的函数, 记为  $Y = g(X)$ 。

如果  $X$  为随机变量, 那么对一个普通的函数 (映射)  $g: D \rightarrow \mathbb{R}$ , 那么  $g(X)$  仍为随机变量。

### 2.4.2 离散型随机变量的函数的概率分布

当已知离散型随机变量  $X$  的分布律时, 要确定  $Y = g(X)$  的分布律是比较容易的。

**例题 2.5** 设离散型随机变量  $X$  的分布律为 求随机变量  $Y = \cos\left(\frac{\pi}{2}X\right)$  的分布律。

|     |               |                              |          |                              |          |
|-----|---------------|------------------------------|----------|------------------------------|----------|
| $X$ | 1             | 2                            | $\cdots$ | $n$                          | $\cdots$ |
| $p$ | $\frac{1}{2}$ | $\left(\frac{1}{2}\right)^2$ | $\cdots$ | $\left(\frac{1}{2}\right)^n$ | $\cdots$ |

**解** 因为  $\cos\left(\frac{n\pi}{2}\right) = \begin{cases} -1, & n = 2(2k-1) \\ 0, & n = 2k-1, \quad k = 1, 2, \cdots \\ 1, & n = 2(2k) \end{cases}$  所以  $Y = \cos\left(\frac{\pi}{2}X\right)$  的可能取值为  $-1, 0, 1$ 。

由于  $X$  取值  $2, 6, 10, \cdots$  都使对应的  $Y$  取  $-1$ , 可得

$$P\{Y = -1\} = \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^6 + \left(\frac{1}{2}\right)^{10} + \cdots = \frac{\frac{1}{4}}{1 - \frac{1}{16}} = \frac{4}{15}$$

同理有

$$P\{Y = 0\} = \frac{1}{2} + \left(\frac{1}{2}\right)^3 + \left(\frac{1}{2}\right)^5 + \cdots = \frac{1}{2(1 - \frac{1}{4})} = \frac{2}{3}$$

$$P\{Y = 1\} = \left(\frac{1}{2}\right)^4 + \left(\frac{1}{2}\right)^8 + \left(\frac{1}{2}\right)^{12} + \cdots = \frac{\frac{1}{16}}{1 - \frac{1}{16}} = \frac{1}{15}$$

故  $Y$  的分布律为



|     |                |               |                |
|-----|----------------|---------------|----------------|
| $Y$ | -1             | 0             | 1              |
| $p$ | $\frac{4}{15}$ | $\frac{2}{3}$ | $\frac{1}{15}$ |

 **笔记** 一般情况下, 确定离散型随机变量的函数  $Y = g(X)$  的分布律只需要把  $Y = g(X)$  可能取的不同值找出来, 再把与  $Y$  的某个可能值对应的所有  $X$  值的概率相加, 即得  $Y$  取这个值的概率。

### 2.4.3 连续型随机变量的函数的概率分布

在应用中, 更常见的是连续型随机变量的函数, 一般情况下需要根据原随机变量的概率密度确定其函数的概率密度。为区分不同随机变量的概率密度和分布函数, 将随机变量的记号写在下标处。

#### 定理 2.2 (单调函数的概率密度)

设随机变量  $X$  具有概率密度  $f_X(x)$ , 若  $g(x)$  为  $(-\infty, +\infty)$  上的可导函数, 且恒有  $g'(x) > 0$  (或恒有  $g'(x) < 0$ )。记  $h(y)$  为  $g(x)$  的反函数, 则  $Y = g(X)$  是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_X(h(y))|h'(y)|, & y \in g(\mathbb{R}) \\ 0, & \text{其他} \end{cases} \quad (2.21)$$

其中  $g(\mathbb{R}) = \{g(x) \mid x \in (-\infty, +\infty)\}$  为  $g(x)$  的值域。

**注** 如果  $f_X(x)$  在区间  $[a, b]$  以外恒为零, 那么定理 2.2 中的条件可改为: 在  $[a, b]$  上恒有  $g'(x) > 0$  (或恒有  $g'(x) < 0$ ), 定理的结论依然成立, 但要把  $g(\mathbb{R})$  换成  $g([a, b]) = \{g(x) \mid x \in [a, b]\}$ 。

定理 2.2 描述了部分连续型随机变量的函数的概率密度求法。下面用一个例子说明定理 2.2 的应用: 设随机变量  $X \sim N(\mu, \sigma^2)$ , 令  $Y = g(X) = e^X$ 。因  $g(x) = e^x$  满足定理 2.2 中的条件, 当  $y > 0$  时,  $y = e^x$  的反函数为  $x = \ln y$ , 所以由定理 2.2 可知

$$\begin{aligned} f_Y(y) &= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}} \left| \frac{1}{y} \right|, & y > 0 \\ 0, & y \leq 0 \end{cases} \\ &= \begin{cases} \frac{1}{y\sqrt{2\pi}\sigma} e^{-\frac{(\ln y - \mu)^2}{2\sigma^2}}, & y > 0 \\ 0, & y \leq 0 \end{cases} \end{aligned}$$

通常, 称具有上述概率密度的随机变量  $Y$  服从对数正态分布。求连续型随机变量的线性函数的概率密度时则可以使用下述定理。

#### 定理 2.3 (线性函数的概率密度)

设连续型随机变量  $X$  有概率密度  $f_X(x)$ , 则随机变量  $Y = kX + b$  (其中  $k, b$  为常数且  $k \neq 0$ ) 的概率密度为:

$$f_Y(y) = f_X\left(\frac{y-b}{k}\right) \frac{1}{|k|} \quad (2.22)$$

特别地, 若随机变量  $X \sim N(\mu, \sigma^2)$ , 则  $Y = kX + b \sim N(k\mu + b, k^2\sigma^2)$  ( $k \neq 0$ ), 即正态随机变量的线性函数仍然服从正态分布。

求  $Y = g(X)$  的概率密度的一般步骤是

1. 求出  $Y$  的分布函数  $F_Y(y) = P\{g(X) \leq y\}$
2. 求  $F_Y(y)$  的导函数  $F'_Y(y)$
3. 令  $f_Y(y) = F'_Y(y)$ , 并在  $F_Y(y)$  的不可导点处规定  $f_Y(y)$  的函数值为零。于是  $f_Y(y)$  就是  $Y$  的概率密度。

**例题 2.6** 设随机变量  $X \sim N(0, 1)$ , 求  $Y = X^2$  的概率密度。

**解** 因  $g(x) = x^2$  在  $(-\infty, +\infty)$  上不是单调函数, 故不能用定理。注意到  $Y = X^2$  总取非负值, 因此当  $y < 0$  时有

$$F_Y(y) = P\{Y \leq y\} = P\{X^2 \leq y\} = P(\emptyset) = 0$$

当  $y \geq 0$  时

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = P\{X^2 \leq y\} = P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x) dx = \frac{1}{\sqrt{2\pi}} \int_{-\sqrt{y}}^{\sqrt{y}} e^{-\frac{x^2}{2}} dx \end{aligned}$$

因而在  $y \neq 0$  处,  $F_Y(y)$  为可导函数, 其导数为:

$$\begin{aligned} F'_Y(y) &= \begin{cases} f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} - f_X(-\sqrt{y}) \left(-\frac{1}{2\sqrt{y}}\right), & y > 0, \\ 0, & y < 0 \end{cases} \\ &= \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0, \\ 0, & y < 0 \end{cases} \end{aligned}$$

令  $f_Y(y) = F'_Y(y)$ , 补充  $y = 0$  处  $F'_Y(y) = 0$  有:

$$f_Y(y) = \begin{cases} \frac{1}{\sqrt{2\pi}} y^{-\frac{1}{2}} e^{-\frac{y}{2}}, & y > 0 \\ 0, & y \leq 0 \end{cases}$$

此时称  $Y$  服从自由度为 1 的卡方分布, 记为  $Y \sim \chi^2(1)$ 。

有时候, 连续型随机变量的函数并不一定是连续型随机变量, 此时只需要求出其分布律即可。

**例题 2.7(P50 第 31 题)** 设  $X$  服从参数为  $\lambda$  的指数分布, 令  $Y = [X] + 1$ , 求  $Y$  的分布。

**解** 因为  $X \sim \text{Exp}(\lambda)$ , 则  $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$ , 根据取整函数的性质知  $Y = [X] + 1$  是离散型随机变量, 现欲求出其分布律。

$$\begin{aligned} P\{Y = k\} &= P\{[X] + 1 = k\} \quad (k = 1, 2, \dots) \\ &= P\{k-1 \leq X < k\} \\ &= \int_{k-1}^k f_X(x) dx \\ &= [1 - (1 - e^{-\lambda})]^{k-1} (1 - e^{-\lambda}) \end{aligned}$$

令  $p = 1 - e^{-\lambda}$ , 不难看出  $Y$  服从参数为  $p$  的几何分布, 亦即  $Y \sim \text{Ge}(1 - e^{-\lambda})$ 。

## 第三章 随机向量及其概率分布

### 内容提要

- 随机向量
- 随机向量的分布律与概率密度
- 条件分布
- 条件概率密度
- 随机变量的相互独立性
- 随机向量的函数及其概率分布

### 3.1 $n$ 维随机向量

#### 3.1.1 随机向量的概念

在实际问题中,有些试验的结果需要同时用两个或两个以上的随机变量来描述。这类例子可以举出很多,而对于同一个试验结果的各个随机变量之间,一般有某种联系,因而需要把它们作为一整体来研究。

##### 定义 3.1 (随机向量)

设  $\Omega$  是随机试验  $E$  的样本空间,  $\omega$  为样本点, 而  $X_1(\omega), X_2(\omega), \dots, X_n(\omega)$  是定义在  $\Omega$  上的  $n$  个随机变量, 则  $n$  维向量  $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  称为  $n$  维随机向量或  $n$  维随机变量。

随机向量  $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  是从样本空间  $\Omega$  到  $n$  维欧氏空间  $\mathbb{R}^n$  的一个映射:

$$\omega \in \Omega \mapsto (X_1(\omega), X_2(\omega), \dots, X_n(\omega)) \in \mathbb{R}^n$$

其中的每一个分量  $X_i(\omega) (1 \leq i \leq n)$  是一维随机变量。随机向量的所有分量中包含的  $\omega$  是同一个  $\omega$ , 因此在不强调样本点  $\omega$  时, 常把  $(X_1(\omega), X_2(\omega), \dots, X_n(\omega))$  简记为  $(X_1, X_2, \dots, X_n)$ 。

由于高维随机向量的性质和讨论是相似的, 下面主要以二维向量  $(X, Y)$  进行研究。

#### 3.1.2 分布函数与边缘分布函数

##### 定义 3.2 ((联合) 分布函数)

设  $(X, Y)$  是二维随机向量, 对任意实数  $x, y$ , 二元函数

$$F(x, y) = P\{X \leq x, Y \leq y\} = \{ \omega \mid X(\omega) \leq x, Y(\omega) \leq y \} \quad (3.1)$$

称为二维随机向量  $(X, Y)$  的分布函数, 或  $X$  与  $Y$  的联合分布函数。分布函数  $F(x, y)$  也表示事件  $\{X \leq x\}$  与  $\{Y \leq y\}$  同时发生的概率。

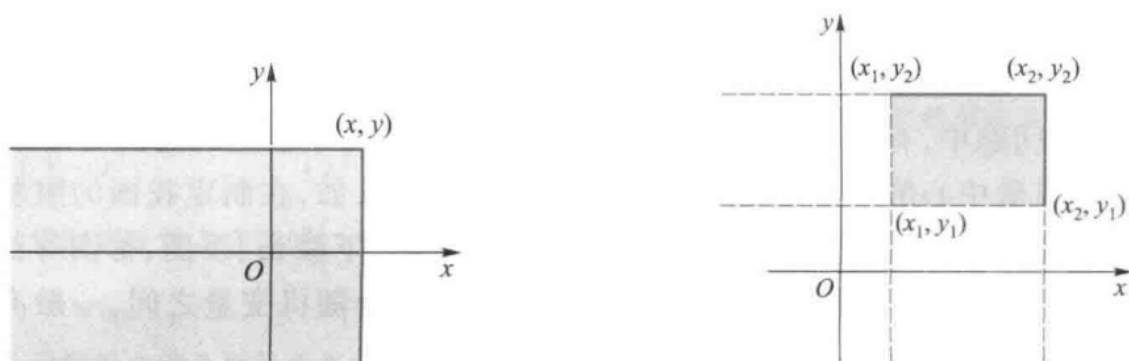
若把二维随机向量  $(X, Y)$  看作是平面上的随机点的坐标, 则分布函数  $F(x, y)$  在  $(x, y)$  处的函数值, 就是随机点  $(X, Y)$  落入以点  $(x, y)$  为顶点而位于该点左下方的无穷矩形域内的概率。由上述几何解释并借助图 3.1, 不难算出事件  $\{x_1 < X \leq x_2, y_1 < Y \leq y_2\}$  的概率

$$P\{x_1 < X \leq x_2, y_1 < Y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1) \quad (3.2)$$

**性质** 分布函数  $F(x, y)$  具有以下性质:

1.  $F(x, y)$  对每个变量是单调增函数, 即当  $x_1 < x_2$  时,  $F(x_1, y) \leq F(x_2, y)$ ; 当  $y_1 < y_2$  时,  $F(x, y_1) \leq F(x, y_2)$
2.  $F(x, y)$  对每个变量是右连续的, 即  $F(x+0, y) = F(x, y)$ ,  $F(x, y+0) = F(x, y)$
3.  $F(-\infty, y) = F(x, -\infty) = F(-\infty, -\infty) = 0$ ,  $F(+\infty, +\infty) = 1$
4. 对于任意两点  $(x_1, y_1)$ ,  $(x_2, y_2)$ , 若  $x_1 \leq x_2$ ,  $y_1 \leq y_2$ , 则

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

图 3.1: 随机向量  $(X, Y)$  的分布函数的几何直观

若给定二维随机向量  $(X, Y)$  的分布函数  $F(x, y)$ , 则它的两个分量  $X, Y$  的分布函数  $F_X(x), F_Y(y)$  与分布函数  $F(x, y)$  之间必然存在联系。事实上有

$$\begin{aligned} F_X(x) &= P\{X \leq x\} = P\{X \leq x, Y < +\infty\} = \lim_{y \rightarrow +\infty} P\{X \leq x, Y \leq y\} \\ &= \lim_{y \rightarrow +\infty} F(x, y) = F(x, +\infty) \end{aligned}$$

因此有边缘分布函数的定义。

#### 定义 3.3 (边缘分布函数)

称  $F_X(x), F_Y(y)$  分别为二维随机向量  $(X, Y)$  关于  $X, Y$  的边缘分布函数, 即

$$F_X(x) = F(x, +\infty), \quad F_Y(y) = F(+\infty, y)$$

边缘分布函数  $F_X(x), F_Y(y)$  完全由分布函数  $F(x, y)$  唯一确定, 但反过来并不一定成立。

### 3.1.3 二维离散型随机向量

#### 定义 3.4 (二维离散型随机向量的分布律)

若二维随机向量  $(X, Y)$  的所有可能取的值是有限对或可数对  $(x_i, y_j)$  ( $i, j = 1, 2, \dots$ ), 则称  $(X, Y)$  是二维离散型随机向量, 并称  $P\{X = x_i, Y = y_j\} = p_{ij}$  ( $i, j = 1, 2, \dots$ ) 为二维离散型随机向量  $(X, Y)$  的分布律或  $X$  与  $Y$  的联合分布律。

**性质** 由  $p_{ij}$  的定义结合概率的性质可知, 它应满足

1.  $0 \leq p_{ij} \leq 1 \quad (i, j = 1, 2, \dots)$

2.  $\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} p_{ij} = 1$

$(X, Y)$  的分布律也可以用下述表格表示。

| $p_{ij}$      |          | $Y$           |               |         |               |         | $p_{i \cdot}$ |
|---------------|----------|---------------|---------------|---------|---------------|---------|---------------|
|               |          | $y_1$         | $y_2$         | $\dots$ | $y_m$         | $\dots$ |               |
| $X$           | $x_1$    | $p_{11}$      | $p_{12}$      | $\dots$ | $p_{1m}$      | $\dots$ | $p_{1 \cdot}$ |
|               | $x_2$    | $p_{21}$      | $p_{22}$      | $\dots$ | $p_{2m}$      | $\dots$ | $p_{2 \cdot}$ |
|               | $\vdots$ | $\vdots$      | $\vdots$      | $\dots$ | $\vdots$      | $\dots$ | $\vdots$      |
|               | $x_n$    | $p_{n1}$      | $p_{n2}$      | $\dots$ | $p_{nm}$      | $\dots$ | $p_{n \cdot}$ |
|               | $\vdots$ | $\vdots$      | $\vdots$      | $\dots$ | $\vdots$      | $\dots$ | $\vdots$      |
| $p_{\cdot j}$ |          | $p_{\cdot 1}$ | $p_{\cdot 2}$ | $\dots$ | $p_{\cdot m}$ | $\dots$ | 1             |

若已知二维离散型随机向量  $(X, Y)$  的分布律  $P\{X = x_i, Y = y_j\} = p_{ij} (i, j = 1, 2, \dots)$ , 则  $(X, Y)$  的分布函数可按下列式确定, 即对满足  $x_i \leq x, y_j \leq y$  的下表的  $i, j$  求和。

$$F(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} p_{ij} \quad (3.3)$$

从  $(X, Y)$  的联合分布律还可以求出每个分量的分布律。

#### 定义 3.5 (边缘分布律)

称  $p_{i\cdot} (i = 1, 2, \dots)$  和  $p_{\cdot j} (j = 1, 2, \dots)$  为  $(X, Y)$  关于  $X$  和关于  $Y$  的边缘分布律。亦即

$$p_{i\cdot} = P\{X = x_i\} = \sum_{j=1}^{\infty} p_{ij}, \quad (i = 1, 2, \dots) \quad (3.4)$$

$$p_{\cdot j} = P\{Y = y_j\} = \sum_{i=1}^{\infty} p_{ij}, \quad (j = 1, 2, \dots) \quad (3.5)$$

由  $(X, Y)$  的分布律可以唯一地确定  $X$  和  $Y$  的边缘分布律  $p_{i\cdot} (i = 1, 2, \dots)$  和  $p_{\cdot j} (j = 1, 2, \dots)$ 。

**注** 随机向量的分布律可以唯一确定其边缘分布律, 但反之不一定成立。

### 3.1.4 二维连续型随机向量

#### 3.1.4.1 随机向量的概率密度和边缘概率密度

连续型随机变量的分布通常用概率密度去描述, 类似的可以定义连续型随机向量的概率密度。

#### 定义 3.6 (二维随机向量的概率密度)

设  $(X, Y)$  是二维随机向量, 若存在一个非负可积函数  $f(x, y) (-\infty < x < +\infty, -\infty < y < +\infty)$ , 使对任意区域  $A \subset \mathbb{R}^2$  都有:

$$P\{(X, Y) \in A\} = \iint_A f(x, y) dx dy$$

则称  $(X, Y)$  为二维连续型随机向量, 称  $f(x, y)$  为  $(X, Y)$  的概率密度, 或  $X$  与  $Y$  的联合概率密度。

**性质** 随机向量  $(X, Y)$  的概率密度具有以下性质

1.  $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2$
2.  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = 1$
3. 设二维连续型随机向量  $(X, Y)$  的分布函数为  $F(x, y)$ , 则对任意实数  $x, y$ , 都有

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

4. 若  $f(x, y)$  在点  $(x, y)$  的某邻域内连续, 则

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y)$$

当给定二维连续型随机向量  $(X, Y)$  的概率密度  $f(x, y)$ , 那么两个分量  $X$  和  $Y$  的概率密度  $f_X(x), f_Y(y)$  也就随之确定。因为对任意区间  $[a, b] \subset \mathbb{R}$ , 有

$$\begin{aligned} P\{X \in [a, b]\} &= P\{X \in [a, b], -\infty < Y < +\infty\} \\ &= \int_a^b \left[ \int_{-\infty}^{+\infty} f(u, v) dv \right] du \end{aligned}$$

其中  $\int_{-\infty}^{+\infty} f(x, y) dy$  为  $x$  的非负可积函数, 所以  $X$  是一维连续型随机变量, 进而有边缘概率密度的定义。

**定义 3.7 (边缘概率密度)**

称下面两式所确定的  $f_X(x), f_Y(y)$  分别为  $(X, Y)$  关于  $X, Y$  的边缘概率密度。

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy \quad (3.6)$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx \quad (3.7)$$

**例题 3.1(P77 第 10 题)** 设二维随机向量  $(X, Y)$  的概率密度为

$$f(x, y) = \begin{cases} Cx^2y, & x^2 \leq y \leq 1 \\ 0, & \text{其他} \end{cases}$$

1. 确定常数  $C$
2. 求概率  $P\{|X| \leq Y\}$
3. 求关于  $X, Y$  的边缘概率密度

**解** 求解该问题需要用到随机向量概率密度的定义和性质。

1. 随机向量的概率密度函数需要满足归一化条件。

$$\iint_{x^2 \leq y \leq 1} Cx^2y dx dy = \int_{-1}^1 dx \int_{x^2}^1 Cx^2y dy = \frac{4}{21}C = 1$$

解得  $C = \frac{21}{4}$ 。

2. 划定区域  $A = \{(x, y) | x^2 \leq y \leq 1, |x| \leq y\}$ , 由概率密度的定义知

$$P\{|X| \leq Y\} = \iint_A f(x, y) dx dy = \int_{-1}^0 dx \int_{-x}^1 f(x, y) dy + \int_0^1 dx \int_x^1 f(x, y) dy = 0.7$$

3. 根据定义 3.7 可以分别计算关于  $X, Y$  的边缘概率密度。

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \begin{cases} \frac{21}{8}(x^2 - x^6), & -1 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}$$

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} \frac{7}{2}y^{\frac{3}{2}}, & 0 \leq y \leq 1 \\ 0, & \text{其他} \end{cases}$$

**3.1.4.2 二维正态分布****定义 3.8 (二维正态分布)**

若二维随机向量  $(X, Y)$  的概率密度为

$$f(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \quad (3.8)$$

其中  $-\infty < x < +\infty, -\infty < y < +\infty$ ,  $\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$  都是常数, 且  $\sigma_1 > 0, \sigma_2 > 0, |\rho| < 1$ , 则称  $(X, Y)$  服从二维正态分布, 记为  $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ , 并称  $(X, Y)$  为二维正态随机向量。

二维正态随机向量  $(X, Y)$  关于  $X$  的边缘概率密度可以根据定义 3.4 计算。

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}, \quad (-\infty < x < +\infty)$$

同理可得  $(X, Y)$  关于  $Y$  的边缘概率密度。

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \frac{1}{\sqrt{2\pi}\sigma_2} e^{-\frac{(y-\mu_2)^2}{2\sigma_2^2}}, \quad (-\infty < y < +\infty)$$

**定理 3.1**

当随机向量  $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$  时, 它的每个分量为正态变量。且  $X \sim N(\mu_1, \sigma_1^2), Y \sim N(\mu_2, \sigma_2^2)$ 。

**3.1.4.3 二维均匀分布**

对于平面上的几何概型问题, 可以用服从二维均匀分布的随机向量  $(X, Y)$  来建模求解。

**定义 3.9 (二维均匀分布)**

若二维随机向量  $(X, Y)$  的概率密度为

$$f(x, y) = \begin{cases} \frac{1}{A}, & \text{当 } (x, y) \in G \\ 0, & \text{当 } (x, y) \notin G \end{cases} \quad (3.9)$$

其中  $G$  是平面上的一个有界区域, 且  $S(G) = A$ , 则称二维随机向量  $(X, Y)$  在  $G$  上服从二维均匀分布。

**3.2 条件分布****3.2.1 离散型随机向量的条件分布律**

当同时研究多个随机变量时, 变量间的相互影响、相互依赖关系是一个值得关注的问题。本节讨论条件分布概念, 即在已知一随机向量的部分分量的值的条件下, 研究其余分量的条件概率分布。对于二维随机向量  $(X, Y)$ , 现欲求在  $Y = y_j$  的条件下  $X = x_i$  的概率。当  $P\{Y = y_j\} > 0$  时, 利用条件概率的计算公式可知

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

不难验证, 上述这组条件概率 (注意: 这里  $j$  是固定的,  $i$  取遍  $1, 2, \dots$ ) 符合分布律的两条基本性质:

1.  $P\{X = x_i | Y = y_j\} \geq 0 \quad (i = 1, 2, \dots)$
2.  $\sum_{i=1}^{\infty} P\{X = x_i | Y = y_j\} = 1$

**定义 3.10 (二维离散型随机向量的条件分布律)**

设  $(X, Y)$  是二维离散型随机向量, 其分布律以及  $(X, Y)$  关于  $X$  和关于  $Y$  的边缘分布律分别由定义 3.4 和 3.5 给出。对于固定的  $j$ , 若  $P\{Y = y_j\} > 0$ , 则称

$$P\{X = x_i | Y = y_j\} = \frac{p_{ij}}{p_{\cdot j}}, \quad (i = 1, 2, \dots) \quad (3.10)$$

为在  $Y = y_j$  的条件下随机变量  $X$  的条件分布律。类似地, 对于固定的  $i$ , 若  $P\{X = x_i\} > 0$ , 则称

$$P\{Y = y_j | X = x_i\} = \frac{p_{ij}}{p_{i\cdot}}, \quad (j = 1, 2, \dots) \quad (3.11)$$

为在  $X = x_i$  的条件下随机变量  $Y$  的条件分布律。

**例题 3.2(P77 第 13 题)** 设某医院一天中出生的婴儿个数  $X \sim P(\lambda)$ ,  $Y$  是其中的男婴个数。假设每出生一个婴儿, 其是男婴的概率为 0.5, 试求

1.  $(X, Y)$  的分布律
2.  $Y$  的分布律
3. 给定  $Y$  时  $X$  的条件分布律

**解** 求解该问题需要用到分布律, 条件分布律的相关定义。



1. 当  $X = n$  时,  $Y \sim B\left(n, \frac{1}{2}\right)$ 。对于  $(X, Y)$  的分布律只需求

$$\begin{aligned} P\{X = n, Y = k\} &= P\{X = n\}P\{Y = k|X = n\} \quad (n = 0, 1, 2, \dots; k = 0, 1, 2, \dots, n) \\ &= \left(\frac{\lambda^n}{n!} e^{-\lambda}\right) \cdot \left[\left(\frac{1}{2}\right)^n C_n^k\right] \end{aligned}$$

2. 对于  $Y$  的分布律只需求级数

$$\begin{aligned} P\{Y = k\} &= \sum_{n=k}^{\infty} P\{X = n, Y = k\} \quad (k = 0, 1, 2, \dots, n) \\ &= \sum_{n=k}^{\infty} \frac{e^{-\lambda}}{k!(n-k)!} \cdot \left(\frac{\lambda}{2}\right)^n \\ &= \frac{1}{k!} \left(\frac{\lambda}{2}\right)^k e^{-\lambda} \sum_{n=k}^{\infty} \frac{1}{(n-k)!} \cdot \left(\frac{\lambda}{2}\right)^{n-k} \\ &= \frac{1}{k!} \left(\frac{\lambda}{2}\right)^k e^{-\lambda} \sum_{m=0}^{\infty} \frac{1}{m!} \cdot \left(\frac{\lambda}{2}\right)^m \\ &= \frac{1}{k!} \left(\frac{\lambda}{2}\right)^k e^{-\lambda} \cdot e^{\frac{\lambda}{2}} \\ &= \frac{1}{k!} \left(\frac{\lambda}{2}\right)^k e^{-\frac{\lambda}{2}} \end{aligned}$$

不难看出  $Y \sim P\left(\frac{\lambda}{2}\right)$ 。

3. 给定  $Y = k$  时, 求  $X$  的条件分布律为

$$\begin{aligned} P\{X = n|Y = k\} &= \frac{P\{X = n, Y = k\}}{P\{Y = k\}} \quad (n = k, k+1, k+2, \dots) \\ &= \frac{1}{(n-k)!} \left(\frac{\lambda}{2}\right)^{n-k} e^{-\frac{\lambda}{2}} \end{aligned}$$

### 3.2.2 连续型随机向量的条件概率密度

设  $(X, Y)$  为二维连续型随机向量, 其概率密度为  $f(x, y)$ ,  $(X, Y)$  关于  $Y$  的边缘概率密度为  $f_Y(y)$ 。此时由于对任意实数  $y$  有  $P\{Y = y\} = 0$ , 因而直接由条件概率公式去计算  $P\{a \leq X \leq b | Y = y\}$  是行不通的。为此假定对  $\forall \varepsilon > 0$ ,  $P\{|y - \varepsilon < Y \leq y + \varepsilon\} > 0$ , 并通过极限手段, 把概率  $P\{a \leq X \leq b | Y = y\}$  定义为

$$P\{a \leq X \leq b | Y = y\} = \lim_{\varepsilon \rightarrow 0^+} P\{a \leq X \leq b | y - \varepsilon < y \leq y + \varepsilon\}$$

根据条件概率的计算公式, 有

$$\begin{aligned} P\{a \leq X \leq b | Y = y\} &= \lim_{\varepsilon \rightarrow 0^+} P\{a \leq X \leq b | y - \varepsilon < y \leq y + \varepsilon\} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{P\{a \leq X \leq b, y - \varepsilon < y \leq y + \varepsilon\}}{P\{y - \varepsilon < y \leq y + \varepsilon\}} \\ &= \lim_{\varepsilon \rightarrow 0^+} \frac{\int_a^b \int_{y-\varepsilon}^{y+\varepsilon} f(u, v) dv du}{\int_{-\infty}^{+\infty} \int_{y-\varepsilon}^{y+\varepsilon} f(u, v) dv du} \end{aligned}$$

若  $f(x, y)$  在点  $(x, y)$  处连续, 则上式的右端的极限值为

$$\frac{\int_a^b f(u, y) du}{\int_{-\infty}^{+\infty} f(u, y) du} = \int_a^b \frac{f(u, y)}{f_Y(y)} du$$

亦即有条件概率

$$P\{a \leq X \leq b \mid Y = y\} = \int_a^b \frac{f(u, y)}{f_Y(y)} du$$

### 定义 3.11 (条件概率密度和条件分布函数)

设  $(X, Y)$  为二维连续型随机向量, 其概率密度为  $f(x, y)$ ,  $(X, Y)$  关于  $Y$  的边缘概率密度为  $f_Y(y)$ 。若对于固定的  $y$ ,  $f_Y(y) > 0$ , 则称

$$f_{X|Y}(x \mid y) = \frac{f(x, y)}{f_Y(y)}, \quad F_{X|Y}(x \mid y) = \int_{-\infty}^x f_{X|Y}(u \mid y) du \quad (3.12)$$

为在  $Y = y$  的条件下  $X$  的条件概率密度和条件分布函数。类似地, 当  $f_X(x) > 0$  时, 可定义在  $X = x$  的条件下  $Y$  的条件概率密度和条件分布函数为

$$f_{Y|X}(y \mid x) = \frac{f(x, y)}{f_X(x)}, \quad F_{Y|X}(y \mid x) = \int_{-\infty}^y f_{Y|X}(v \mid x) dv \quad (3.13)$$

类似于条件概率的乘法公式, 条件概率密度同样满足乘法公式。

### 定理 3.2 (条件概率密度的乘法公式)

设有随机向量  $(X, Y)$ , 则有恒等式:

$$f(x, y) = f_X(x)f_{Y|X}(y \mid x) = f_Y(y)f_{X|Y}(x \mid y) \quad (3.14)$$

**例题 3.3** 设数  $X$  在区间  $(0, 1)$  内随机地取值, 当观察到  $X = x$  ( $0 < x < 1$ ) 时, 数  $Y$  在区间  $(x, 1)$  内随机地取值, 求  $Y$  的概率密度  $f_Y(y)$ 。

**解** 由题意知  $X \sim U(0, 1)$ , 且  $X$  具有概率密度

$$f_X(x) = \begin{cases} 1, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

当给定条件  $X = x$  ( $0 < x < 1$ ) 时,  $Y$  的条件分布为区间  $(x, 1)$  上的均匀分布, 即:

$$f_{Y|X}(y \mid x) = \begin{cases} \frac{1}{1-x}, & x < y < 1 \\ 0, & \text{其他} \end{cases}$$

其中  $x$  固定,  $0 < x < 1$ 。由式 3.13 得:

$$f(x, y) = f_{Y|X}(y \mid x)f_X(x) = \begin{cases} \frac{1}{1-x}, & 0 < x < 1, x < y < 1 \\ 0, & \text{其他} \end{cases}$$

此时对概率密度  $f(x, y)$  的  $x$  积分得到  $y$  的概率密度。

$$f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx = \begin{cases} \int_0^y \frac{1}{1-x} dx = -\ln(1-y), & 0 < y < 1 \\ 0, & \text{其他} \end{cases}$$

## 3.3 随机变量的相互独立性

### 3.3.1 两个随机向量的相互独立性

随机变量的相互独立性是概率论与数理统计中的一个很重要的概念, 它是由随机事件的相互独立性引申而来的。根据第一章中事件的独立性可引出随机变量的相互独立性。

**定义 3.12 (两个随机变量的相互独立性)**

设  $(X, Y)$  是二维随机向量,  $F(x, y)$  及  $F_X(x), F_Y(y)$  分别是  $(X, Y)$  的分布函数及边缘分布函数, 若对任意实数  $x, y$  恒有:

$$P\{X \leq x, Y \leq y\} = P\{X \leq x\}P\{Y \leq y\} \quad (3.15)$$

亦即:

$$F(x, y) = F_X(x)F_Y(y) \quad (3.16)$$

则称  $X$  和  $Y$  是相互独立的。

**笔记**

- 当  $(X, Y)$  是二维离散型随机向量时,  $X$  与  $Y$  相互独立的条件式等价于对  $(X, Y)$  的所有可能取值  $(x_i, y_j) (i, j = 1, 2, \dots)$ , 都有

$$P\{X = x_i, Y = y_j\} = P\{X = x_i\}P\{Y = y_j\}, \quad (i, j = 1, 2, \dots) \quad (3.17)$$

- 当  $(X, Y)$  是二维连续型随机向量时, 设  $f(x, y)$  及  $f_X(x), f_Y(y)$  分别是  $(X, Y)$  的概率密度及边缘概率密度, 则  $X$  与  $Y$  相互独立的条件式等价于

$$f(x, y) = f_X(x)f_Y(y) \quad (3.18)$$

几乎处处成立 (在平面上挖去一个“面积”为 0 的点集后, 处处成立)。

**命题 3.1**

设二维随机向量  $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ , 则  $X$  与  $Y$  相互独立的充要条件是  $\rho = 0$ 。

**证明** 可知  $(X, Y)$  的两个边缘概率密度的乘积为

$$f_X(x)f_Y(y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

先证充分性: 设  $\rho = 0$ , 把  $\rho = 0$  代入到  $(X, Y)$  的概率密度  $f(x, y)$  中, 因为对一切实数  $x, y$  都有:

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\} \\ &= f_X(x)f_Y(y) \end{aligned}$$

故  $X$  与  $Y$  相互独立。

再证必要性: 设  $X$  与  $Y$  相互独立, 由于  $f(x, y), f_X(x)$  及  $f_Y(y)$  都是连续函数, 故对一切实数  $x, y$ , 都有

$$f(x, y) = f_X(x)f_Y(y)$$

在上式中令  $x = \mu_1, y = \mu_2$ , 得到

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

从而  $\rho = 0$ 。

**3.3.2 多个随机变量的相互独立性**

相互独立的概念可以推广到  $n(n > 2)$  个随机变量的场合。设  $n$  维随机向量  $(X_1, X_2, \dots, X_n)$  的分布函数为

$$F(x_1, x_2, \dots, x_n) = P\{X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n\}$$

若对任意实数  $x_1, x_2, \dots, x_n$ , 都有

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1)F_{X_2}(x_2) \cdots F_{X_n}(x_n) \quad (3.19)$$

则称随机变量  $X_1, X_2, \dots, X_n$  是相互独立的。相互独立的概念还可以做进一步的推广。设  $X_1, X_2, \dots, X_n, \dots$  为一串随机变量 (或称为随机变量序列), 若对任意正整数  $n > 1$ , 都有  $X_1, X_2, \dots, X_n$  相互独立, 则称随机变量序列  $X_1, X_2, \dots, X_n, \dots$  是相互独立的。

 **笔记** 判断若干个随机变量是否相互独立, 关键在于利用 3.19 等式判断这些随机变量组成的随机向量的分布函数, 是否可以写成几个随机变量分布函数之积的形式。有时候也可以借助分布律, 概率密度等判断。

**例题 3.4(P79 第 20 题)** 设三维随机向量  $(X, Y, Z)$  的概率密度为

$$f(x, y, z) = \begin{cases} \frac{1}{8}(1 - xyz), & -1 \leq x, y, z \leq 1 \\ 0, & \text{其他} \end{cases}$$

证明  $X, Y, Z$  两两独立, 但不相互独立。

**证明** 根据概率密度中  $x, y, z$  的轮换性, 以  $X, Y$  为例子讨论。对于随机变量  $X, Y$ , 其概率密度

$$f_X(x) = \begin{cases} \int_{-1}^1 \int_{-1}^1 f(x, y, z) dy dz = \frac{1}{2}, & -1 \leq x \leq 1 \\ 0, & \text{其他} \end{cases}, \quad f_Y(y) = \begin{cases} \frac{1}{2}, & -1 \leq y \leq 1 \\ 0, & \text{其他} \end{cases}$$

接着求  $(X, Y)$  的概率密度

$$f_{X,Y}(x, y) = \begin{cases} \int_{-1}^1 f(x, y, z) dz = \frac{1}{4}, & -1 \leq x, y \leq 1 \\ 0, & \text{其他} \end{cases}$$

此时有  $f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y)$ , 因此  $X, Y$  相互独立。同理有  $Y, Z$  和  $X, Z$  相互独立。但显然  $f(x, y, z) \neq f_X(x) \cdot f_Y(y) \cdot f_Z(z)$ , 因此  $X, Y, Z$  不相互独立。

### 3.3.3 随机向量的相互独立性

设  $m$  维随机向量  $(X_1, X_2, \dots, X_m)$  的分布函数为  $F_1(x_1, x_2, \dots, x_m)$ ,  $n$  维随机向量  $(Y_1, Y_2, \dots, Y_n)$  的分布函数为  $F_2(y_1, y_2, \dots, y_n)$ ,  $m+n$  维随机向量  $(X_1, X_2, \dots, X_m, Y_1, Y_2, \dots, Y_n)$  的分布函数为  $F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n)$ 。若对任意实数  $x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n$ , 都有

$$F(x_1, x_2, \dots, x_m, y_1, y_2, \dots, y_n) = F_1(x_1, x_2, \dots, x_m)F_2(y_1, y_2, \dots, y_n) \quad (3.20)$$

则称  $(X_1, X_2, \dots, X_m)$  与  $(Y_1, Y_2, \dots, Y_n)$  是相互独立的。

#### 定理 3.3 (随机向量的分量独立性)

从  $(X_1, X_2, \dots, X_m)$  中任选  $i$  个 ( $1 \leq i \leq m$ ) 变量  $(X_{m_1}, X_{m_2}, \dots, X_{m_i})$ , 且下标  $1 \leq m_1 < m_2 < \dots < m_i \leq m$ ; 从  $(Y_1, Y_2, \dots, Y_n)$  中任选  $j$  个 ( $1 \leq j \leq n$ ) 变量  $(Y_{n_1}, Y_{n_2}, \dots, Y_{n_j})$ , 且下标  $1 \leq n_1 < n_2 < \dots < n_j \leq n$ , 则  $(X_{m_1}, X_{m_2}, \dots, X_{m_i})$  与  $(Y_{n_1}, Y_{n_2}, \dots, Y_{n_j})$  相互独立。反之亦然。

下面还有两个重要定理, 其中有关随机向量的函数的定义见 3.4 小节。

#### 定理 3.4

设  $(X_1, X_2, \dots, X_m)$  与  $(Y_1, Y_2, \dots, Y_n)$  相互独立, 若  $h(x_1, x_2, \dots, x_m), g(y_1, y_2, \dots, y_n)$  是连续函数, 则  $h(X_1, X_2, \dots, X_m)$  与  $g(Y_1, Y_2, \dots, Y_n)$  也相互独立。

#### 定理 3.5

若随机变量  $X_1, X_2, \dots, X_n$  相互独立, 把它们分为不相交的  $k$  组, 每组中的所有变量由一个连续函数复合而生成一个新的随机变量, 则这  $k$  个随机变量仍相互独立。

## 3.4 随机向量的函数及其概率分布

### 3.4.1 随机向量的函数

在上一章节中我们讨论了随机变量的函数及其概率分布。而对有些应用问题来说, 需要将高维的随机数据讲行压缩或降维, 这就需要用到随机变量的函数的定义。

#### 定义 3.13 (随机向量的函数和向量值函数)

设  $(X_1, X_2, \dots, X_n)$  是  $n$  维随机向量,  $D \subset \mathbb{R}^n$  是  $(X_1, X_2, \dots, X_n)$  的所有可能值集合。

- 若  $g(x_1, x_2, \dots, x_n)$  为定义在  $D$  上的  $n$  元函数,  $Y$  为随机变量, 当  $(X_1, X_2, \dots, X_n)$  取得可能值  $(x_1, x_2, \dots, x_n) \in D$  时,  $Y$  相应地取值  $g(x_1, x_2, \dots, x_n)$ , 则称  $Y$  为随机向量  $(X_1, X_2, \dots, X_n)$  的函数, 记作  $Y = g(X_1, X_2, \dots, X_n)$ 。
- 若  $g(x_1, x_2, \dots, x_n) (j = 1, 2, \dots, k)$  为定义在  $D$  上的  $n$  元函数,  $Y_j (j = 1, 2, \dots, k)$  为随机变量, 其中正整数  $k \leq n$ , 当  $(X_1, X_2, \dots, X_n)$  取得可能值  $(x_1, x_2, \dots, x_n) \in D$  时,  $Y_j$  相应地取值  $g_j(x_1, x_2, \dots, x_n) (j = 1, 2, \dots, k)$ , 则称  $(Y_1, Y_2, \dots, Y_k)$  为随机向量  $(X_1, X_2, \dots, X_n)$  的向量值函数, 记作

$$(Y_1, Y_2, \dots, Y_k) = (g_1(X_1, X_2, \dots, X_n), g_2(X_1, X_2, \dots, X_n), \dots, g_k(X_1, X_2, \dots, X_n))$$

### 3.4.2 二维离散型随机向量的函数的概率分布

确定离散型随机向量的函数的概率分布是比较简单的。总的思路是设法将新随机变量 (或随机向量) 所表示的事件转化为旧随机向量所表示的等价事件, 然后根据旧随机向量的分布律求出新随机变量的分布律。

#### 命题 3.2 (泊松分布的可加性)

设随机变量  $X \sim P(\lambda_1), Y \sim P(\lambda_2)$ , 且  $X$  与  $Y$  相互独立, 定义  $(X, Y)$  的函数  $Z = X + Y$ , 则  $Z$  是随机变量且  $Z \sim P(\lambda_1 + \lambda_2)$ 。

**证明**  $Z = X + Y$  显然是随机变量, 且  $Z$  的取值为所有非负整数  $0, 1, 2, \dots$ , 则事件  $\{Z = k\}$  是一组互斥事件  $\{X = i, Y = k - i\}, i = 0, 1, \dots, k$  的并。因此, 对  $k = 0, 1, 2, \dots$  可以求出分布律:

$$\begin{aligned} P\{Z = k\} &= \sum_{i=0}^k P\{X = i, Y = k - i\} \\ &= \sum_{i=0}^k P\{X = i\}P\{Y = k - i\} \\ &= \sum_{i=0}^k \frac{\lambda_1^i}{i!} e^{-\lambda_1} \frac{\lambda_2^{k-i}}{(k-i)!} e^{-\lambda_2} \\ &= e^{-(\lambda_1 + \lambda_2)} \sum_{i=0}^k \frac{1}{i!(k-i)!} \lambda_1^i \lambda_2^{k-i} \\ &= \frac{(\lambda_1 + \lambda_2)^k}{k!} e^{-(\lambda_1 + \lambda_2)} \end{aligned}$$

即  $Z = X + Y$  服从参数为  $\lambda_1 + \lambda_2$  的泊松分布。

根据数学归纳法, 若随机变量  $X_1, X_2, \dots, X_n$  相互独立, 且  $X_i \sim P(\lambda_i), i = 1, 2, \dots, n$ , 则这些随机变量的和变量  $X_1 + X_2 + \dots + X_n \sim P(\lambda_1 + \lambda_2 + \dots + \lambda_n)$ 。即泊松分布具有可加性。

### 3.4.3 二维连续型随机向量的函数的概率分布

想要求出  $Z = g(X, Y)$  的概率密度, 可以先求出  $Z = g(X, Y)$  的分布函数  $F_Z(z) = P\{Z \leq z\} = P\{g(X, Y) \leq z\}$ , 再求导得到概率密度。我们还可以依靠下述定理。

**定理 3.6**

设二维连续型随机向量  $(X, Y)$  的概率密度为  $f(x, y)$ , 若  $U = g_1(X, Y), V = g_2(X, Y)$  是  $(X, Y)$  到  $(U, V)$  的一一对应变换, 即存在逆变换  $X = h_1(U, V), Y = h_2(U, V)$ , 并且  $g_1, g_2, h_1, h_2$  都存在连续的一阶偏导数, 则  $(U, V)$  仍为连续型随机向量, 且其概率密度为

$$f_{U,V}(u, v) = \begin{cases} f(h_1(u, v), h_2(u, v)) |J|, & (u, v) \in \mathcal{G} \\ 0, & \text{其他} \end{cases} \quad (3.21)$$

其中  $\mathcal{G}$  为向量值函数  $(u, v) = (g_1(x, y), g_2(x, y))$  的值域,  $J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial h_1}{\partial u} & \frac{\partial h_1}{\partial v} \\ \frac{\partial h_2}{\partial u} & \frac{\partial h_2}{\partial v} \end{vmatrix}$  是 Jacobi 行列式。

定理 3.6 的叙述类似于定理 2.2, 可以看做是多维情况下的推广。下面用具体的例题加以介绍。

**例题 3.5** 设随机变量  $X$  与  $Y$  相互独立且  $X, Y \sim N(0, \sigma^2)$ , 令  $U = X + Y, V = X - Y$ , 则  $U$  与  $V$  是否相互独立?

**解** 由独立性知,  $(X, Y)$  的概率密度为

$$f(x, y) = f_X(x) \cdot f_Y(y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}, \quad -(\infty < x, y < +\infty)$$

因  $(U, V) = (X + Y, X - Y) = (g_1(X, Y), g_2(X, Y))$  存在唯一的逆变换

$$(X, Y) = \left( \frac{U+V}{2}, \frac{U-V}{2} \right) = (h_1(U, V), h_2(U, V))$$

且上述正、逆变换中的所有函数都存在连续的一阶偏导数, 其 Jacobi 行列式

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

故由定理 3.6 直接求出

$$f_{U,V}(u, v) = f\left(\frac{u+v}{2}, \frac{u-v}{2}\right) \cdot \left|-\frac{1}{2}\right| = \frac{1}{4\pi\sigma^2} e^{-\frac{u^2+v^2}{4\sigma^2}}, \quad (-\infty < u, v < +\infty)$$

对比二维正态分布的定义式, 可知  $(U, V) \sim N(0, 0; 2\sigma^2, 2\sigma^2; 0)$ , 根据命题 3.1 可知  $U$  和  $V$  相互独立。

有时, 我们只需求二维连续型随机向量  $(X, Y)$  的一个函数  $g(X, Y)$  的概率密度, 下面就几种常见情况来探讨。

### 3.4.3.1 $aX + bY$ 的概率密度

我们先来讨论两个相互独立的随机变量的和是如何分布的。

#### 定理 3.7 (卷积公式)

设  $X$  和  $Y$  是两个随机变量, 其概率密度函数分别为  $f_X(x)$  和  $f_Y(y)$ 。则  $Z = X + Y$  的概率密度

$$f_Z(z) = \int_{-\infty}^{+\infty} f(x, z-x) dx = \int_{-\infty}^{+\infty} f(z-y, y) dy \quad (3.22)$$

这被称作卷积公式, 其中  $f_Z$  的定义式也被称作  $f_X$  和  $f_Y$  的卷积。

**证明** 我们用两种求解此类问题的通用证明法予以说明。

- 证法一 (分布函数法) 根据分布函数的定义可以写出  $F_Z(z)$

$$F_Z(z) = P\{Z \leq z\} = P\{X + Y \leq z\}$$

因为  $X$  和  $Y$  是连续型随机变量且相互独立, 因此

$$\begin{aligned}
 F_Z(z) &= \iint_{x+y \leq z} f(x, y) \, dx \, dy \\
 &= \iint_{x+y \leq z} f_X(x) f_Y(y) \, dx \, dy \\
 &= \int_{-\infty}^{+\infty} \int_{-\infty}^{z-x} f_X(x) f_Y(y) \, dy \, dx \\
 &= \int_{-\infty}^{+\infty} f_X(x) \left[ \int_{-\infty}^{z-x} f_Y(y) \, dy \right] \, dx
 \end{aligned}$$

内层积分表示为  $F_Y(z-x)$

$$F_Z(z) = \int_{-\infty}^{+\infty} f_X(x) F_Y(z-x) \, dx$$

对  $z$  求导, 得到概率密度函数

$$\begin{aligned}
 f_Z(z) &= \frac{d}{dz} F_Z(z) = \int_{-\infty}^{+\infty} f_X(x) \frac{d}{dz} F_Y(z-x) \, dx \\
 &= \int_{-\infty}^{+\infty} f_X(x) f_Y(z-x) \, dx
 \end{aligned}$$

类似的可以推出对  $y$  积分的形式。

- 证法二 (逆变换法) 设  $U = X + Y$ , 为求  $U$  的概率密度, 补一个函数

$$V = Y$$

这样便构成  $(X, Y)$  到  $(U, V)$  的一一对应变换  $\begin{cases} U = X + Y \\ V = Y \end{cases}$ , 其逆变换为  $\begin{cases} X = U - V \\ Y = V \end{cases}$ , 且 Jacobi 行列式

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1, \text{ 于是}$$

$$f_{U,V}(u, v) = f(u-v, v) \cdot |1| = f(u-v, v)$$

因此求得  $U = X + Y$  的概率密度

$$f_U(u) = \int_{-\infty}^{+\infty} f_{U,V}(u, v) \, dv = \int_{-\infty}^{+\infty} f(u-v, v) \, dv$$

综上便得求  $Z = U = X + Y$  的概率密度的如下公式

$$f_{X+Y}(t) = \int_{-\infty}^{+\infty} f(x, t-x) \, dx = \int_{-\infty}^{+\infty} f_X(x) f_Y(t-x) \, dx$$

若在最初令  $V = X$ , 便得到

$$f_{X+Y}(t) = \int_{-\infty}^{+\infty} f(t-y, y) \, dy = \int_{-\infty}^{+\infty} f_X(t-y) f_Y(y) \, dy$$

### 命题 3.3 (正态分布的可加性)

若  $X_k \sim N(\mu_k, \sigma_k^2)$  ( $k = 1, 2, \dots, n$ ), 并且它们相互独立, 则有

$$X_1 + X_2 + \dots + X_n \sim N\left(\sum_{k=1}^n \mu_k, \sum_{k=1}^n \sigma_k^2\right)$$

更一般地,  $n$  个相互独立的正态随机变量的线性组合仍然是正态随机变量。

$$a_1 X_1 + a_2 X_2 + \dots + a_n X_n \sim N\left(\sum_{k=1}^n a_k \mu_k, \sum_{k=1}^n a_k^2 \sigma_k^2\right)$$

类比卷积公式定理, 我们可以给出  $Z = aX + bY$  ( $a, b \neq 0$ ) 的概率分布。



**定理 3.8 (随机变量的线性组合的概率密度)**

设  $X$  和  $Y$  是两个相互独立的随机变量, 其概率密度函数分别为  $f_X(x)$  和  $f_Y(y)$ 。则  $Z = aX + bY (a, b \neq 0)$  的概率密度

$$f_Z(z) = \frac{1}{|b|} \int_{-\infty}^{+\infty} f_X(x) f_Y\left(\frac{z-ax}{b}\right) dx = \frac{1}{|a|} \int_{-\infty}^{+\infty} f_X\left(\frac{z-by}{a}\right) f_Y(y) dy \quad (3.23)$$

**3.4.3.2  $XY$  的概率密度****定理 3.9 (随机变量的乘积的概率密度)**

设  $X$  和  $Y$  是两个相互独立的随机变量, 其概率密度函数分别为  $f_X(x)$  和  $f_Y(y)$ 。则  $Z = XY$  的概率密度

$$f_Z(z) = \int_{-\infty}^{+\infty} \frac{1}{|y|} f_X\left(\frac{z}{y}\right) f_Y(y) dy \quad (3.24)$$

**3.4.3.3  $X/Y$  的概率密度****定理 3.10 (随机变量的商的概率密度)**

设  $X$  和  $Y$  是两个相互独立的随机变量, 其概率密度函数分别为  $f_X(x)$  和  $f_Y(y)$ 。则  $Z = X/Y$  的概率密度

$$f_Z(z) = \int_{-\infty}^{+\infty} |y| f_X(yz) f_Y(y) dy \quad (3.25)$$



举例而言, 对于相互独立的随机变量  $X, Y \sim N(0, 1)$ , 要求出  $Z = \frac{X}{Y}$  的概率密度, 可以直接代入式 3.25

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} |y| f_X(yz) f_Y(y) dy \\ &= 2 \int_0^{+\infty} y \frac{1}{\sqrt{2\pi}} e^{-\frac{(yz)^2}{2}} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy \\ &= \frac{1}{\pi} \int_0^{+\infty} y e^{-\frac{(1+z^2)y^2}{2}} dy \\ &= \frac{1}{\pi(1+z^2)} \quad (-\infty < z < +\infty) \end{aligned}$$

具有上述概率密度的概率分布称为标准 Cauchy 分布。

**例题 3.6(P79 第 27 题)** 向一目标射击, 目标中心为坐标原点, 已知命中点的横坐标  $X$  和纵坐标  $Y$  相互独立, 且均服从正态分布  $N(0, \sigma^2)$ 。令  $\rho = \sqrt{X^2 + Y^2}$ ,  $\theta = \arctan \frac{Y}{X}$ , 求  $(\rho, \theta)$  的概率密度并判断  $\rho$  与  $\theta$  是否相互独立。

**解** 根据题意构造  $(X, Y)$  到  $(\rho, \theta)$  的一一对应变换  $\begin{cases} \rho = \sqrt{X^2 + Y^2} \\ \theta = \arctan \frac{Y}{X} \end{cases}$ , 其逆变换为  $\begin{cases} X = \rho \cos \theta \\ Y = \rho \sin \theta \end{cases}$ , 且 Jacobi 行列

式  $J = \frac{\partial(x, y)}{\partial(\rho, \theta)} = \begin{vmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{vmatrix} = \rho$ , 于是

$$\begin{aligned} f_{\rho, \theta}(\rho, \theta) &= \begin{cases} f(\rho \cos \theta, \rho \sin \theta) \cdot |\rho|, & \rho \geq 0, 0 \leq \theta < 2\pi \\ 0, & \text{其他} \end{cases} \\ &= \begin{cases} \rho f_X(\rho \cos \theta) f_Y(\rho \sin \theta), & \rho \geq 0, 0 \leq \theta < 2\pi \\ 0, & \text{其他} \end{cases} \\ &= \begin{cases} \frac{\rho}{2\pi\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}} & \rho \geq 0, 0 \leq \theta < 2\pi \\ 0, & \text{其他} \end{cases} \end{aligned}$$

对概率密度的每个变量积分得到边缘概率密度

$$f_{\rho}(\rho) = \int_{-\infty}^{+\infty} f_{\rho, \theta}(\rho, \theta) d\theta = \begin{cases} \frac{\rho}{\sigma^2} e^{-\frac{\rho^2}{2\sigma^2}}, & \rho \geq 0 \\ 0, & \text{其他} \end{cases} \quad (3.26)$$

$$f_{\theta}(\theta) = \int_{-\infty}^{+\infty} f_{\rho, \theta}(\rho, \theta) d\rho = \begin{cases} \frac{1}{2\pi}, & 0 \leq \theta < 2\pi \\ 0, & \text{其他} \end{cases}$$

因为  $f_{\rho, \theta}(\rho, \theta) = f_{\rho}(\rho)f_{\theta}(\theta)$ , 故两随机变量相互独立。

其中, 式子 3.26 也被称作服从参数  $\sigma$  的瑞利 (Rayleigh) 分布。

### 3.4.3.4 max 和 min

如果随机变量  $X$  与  $Y$  相互独立, 它们的分布函数分别为  $F_X(x)$  和  $F_Y(y)$ , 设  $M = \max\{X, Y\}$ ,  $N = \min\{X, Y\}$ , 下面来寻求  $M, N$  的概率分布。对任意实数  $z$ , 有  $\{\max\{X, Y\} \leq z\} = \{X \leq z, Y \leq z\}$ , 由  $X$  与  $Y$  相互独立, 可知

$$\begin{aligned} F_M(z) &= P\{M \leq z\} = P\{\max\{X, Y\} \leq z\} \\ &= P\{X \leq z, Y \leq z\} = P\{X \leq z\}P\{Y \leq z\} \\ &= F_X(z)F_Y(z) \end{aligned}$$

类似地, 由  $\{\min\{X, Y\} > z\} = \{X > z, Y > z\}$  及  $X$  与  $Y$  的相互独立性, 可得

$$\begin{aligned} F_N(z) &= P\{N \leq z\} = P\{\min\{X, Y\} \leq z\} \\ &= 1 - P\{\min\{X, Y\} > z\} = 1 - P\{X > z, Y > z\} \\ &= 1 - P\{X > z\}P\{Y > z\} \\ &= 1 - [1 - F_X(z)][1 - F_Y(z)] \end{aligned}$$

只需再对分布函数求导就可以得到概率密度。再推广到  $n$  个随机变量的情况, 有以下定理。

#### 定理 3.11 (max 和 min 函数的分布)

若  $X_1, X_2, \dots, X_n$  是  $n$  个相互独立的随机变量,  $F_{X_k}(x_k)$  为  $X_k (k = 1, 2, \dots, n)$  的分布函数, 设  $M = \max\{X_1, X_2, \dots, X_n\}$ ,  $N = \min\{X_1, X_2, \dots, X_n\}$ , 则有其分布函数

$$F_M(z) = F_{X_1}(z)F_{X_2}(z) \cdots F_{X_n}(z) \quad (3.27)$$

$$F_N(z) = 1 - [1 - F_{X_1}(z)][1 - F_{X_2}(z)] \cdots [1 - F_{X_n}(z)] \quad (3.28)$$

例如, 若  $X_1, X_2, \dots, X_n$  是  $n$  个相互独立且服从同一分布的随机变量且  $X_1 \sim U(0, 1)$ , 已知均匀分布  $U(0, 1)$  的分布函数为

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

又有服从同一分布的这一条件, 则

$$f_M(z) = F'_M(z) = (F^n(x))' = nF^{n-1}(x)f(x) = \begin{cases} nx^{n-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

同理可得

$$f_N(z) = \begin{cases} n(1-x)^{n-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

## 第四章 随机变量的数字特征

### 内容提要

- 数学期望
- 方差
- 协方差与相关系数
- 矩
- 协方差矩阵
- 条件期望与条件方差

### 4.1 数学期望

#### 4.1.1 随机变量的数学期望

中学已经学过数学期望的基本概念, 本节就离散型随机变量和连续性随机变量介绍其标准定义。

##### 定义 4.1 (离散型随机变量的数学期望)

设离散型随机变量  $X$  的分布律为  $P\{X = x_i\} = p_i (i = 1, 2, \cdots)$ 。若级数  $\sum_{i=1}^{\infty} x_i p_i$  绝对收敛, 则称  $\sum_{i=1}^{\infty} x_i p_i$  为随机变量  $X$  的数学期望, 简称期望或均值, 记作  $E(X)$ , 即有

$$E(X) = \sum_{i=1}^{\infty} x_i p_i \quad (4.1)$$

 **笔记** 数学期望应与  $X$  的取值  $x_1, x_2, \cdots, x_n, \cdots$  的排列次序无关, 因此要求级数  $\sum_{i=1}^{\infty} x_i p_i$  无论以何种方式重排各项的求和顺序都能收敛到同一数值, 根据高等数学知识, 该级数应该是绝对收敛的。

连续型随机变量的数学期望应该与定义 4.1 保持形式一致, 因此只需要把求和改成积分即可。设连续型随机变量  $X$  的概率密度为  $f(x)$ , 若积分  $\int_{-\infty}^{+\infty} x f(x) dx$  绝对收敛, 则称  $\int_{-\infty}^{+\infty} x f(x) dx$  为随机变量  $X$  的数学期望, 记为  $E(X)$ , 即

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx \quad (4.2)$$

数学期望  $E(X)$  是由随机变量  $X$  的概率分布所确定的, 因此当  $X$  服从从某一分布时, 也把  $E(X)$  称为该分布的数学期望。然而, 一些概率分布的数学期望是不存在的。

**例题 4.1(P103 第 7 题)** 证明标准 Cauchy 分布的数学期望不存在。

**证明** 标准 Cauchy 分布的概率密度函数为

$$f(x) = \frac{1}{\pi(1+x^2)}, \quad (-\infty < x < +\infty) \quad (4.3)$$

根据连续型随机变量数学期望的定义, 若  $E(X)$  存在, 则下列反常积分必须绝对收敛:

$$E(X) = \int_{-\infty}^{+\infty} x f(x) dx = \int_{-\infty}^{+\infty} \frac{x}{\pi(1+x^2)} dx$$

判断上述积分是否绝对收敛, 则考察积分

$$\begin{aligned} \int_{-\infty}^{+\infty} |x| f(x) dx &= \frac{1}{\pi} \int_{-\infty}^{+\infty} \frac{|x|}{1+x^2} dx \\ &= \frac{2}{\pi} \int_0^{+\infty} \frac{x}{1+x^2} dx \end{aligned}$$

现在判定反常积分  $\int_0^{+\infty} \frac{x}{1+x^2} dx$  的敛散性。

$$\begin{aligned}\int_0^{+\infty} \frac{x}{1+x^2} dx &= \lim_{b \rightarrow +\infty} \int_0^b \frac{x}{1+x^2} dx \\ &= \lim_{b \rightarrow +\infty} \left[ \frac{1}{2} \ln(1+x^2) \right]_0^b \\ &= \frac{1}{2} \lim_{b \rightarrow +\infty} \ln(1+b^2)\end{aligned}$$

由于  $\lim_{b \rightarrow +\infty} \ln(1+b^2) = +\infty$ , 所以该积分发散。则标准 Cauchy 分布的数学期望不存在。

表4.1(见下一小节)给出了一些常见分布的数学期望, 不难看出这些分布的参数决定了期望的值。

### 4.1.2 随机变量的函数的数学期望

根据随机变量的数学期望的定义, 可以给出下面的定理。

#### 定理 4.1 (随机变量的函数的数学期望)

设离散型随机变量  $X$  的分布律  $P\{X=x_i\}=p_i (i=1, 2, \dots)$ , 若  $g(x)$  连续, 且  $\sum_{i=1}^{\infty} g(x_i)p_i$  绝对收敛, 则

$$E[g(X)] = \sum_{i=1}^{\infty} g(x_i)p_i \quad (4.4)$$

设连续型随机变量  $X$  的概率密度为  $f(x)$ , 若  $g(x)$  连续, 且  $\int_{-\infty}^{+\infty} g(x) \cdot f(x) dx$  绝对收敛, 则

$$E[g(X)] = \int_{-\infty}^{+\infty} g(x)f(x) dx \quad (4.5)$$

 **笔记** 对二维随机向量的函数有类似的结论, 这里一次性讲完。

- 设二维离散型随机向量  $(X, Y)$  的分布律为  $P\{X=x_i, Y=y_j\}=p_{ij} (i, j=1, 2, \dots)$ , 若  $g(x, y)$  为二元连续函数, 且  $\sum_{j=1}^{\infty} \sum_{i=1}^{\infty} g(x_i, y_j)p_{ij}$  绝对收敛, 则

$$E[g(X, Y)] = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} g(x_i, y_j)p_{ij} \quad (4.6)$$

- 设二维连续型随机向量  $(X, Y)$  的概率密度为  $f(x, y)$ , 若  $g(x, y)$  为二元连续函数, 且  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y)dxdy$  绝对收敛, 则

$$E[g(X, Y)] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y)f(x, y)dxdy \quad (4.7)$$

- 可以借助二维随机向量分布求出每个变量的数学期望。若  $E(X)$  存在, 取  $g(x, y)=x$ , 当  $(X, Y)$  为离散型随机向量时, 则有

$$E(X) = \sum_{j=1}^{\infty} \sum_{i=1}^{\infty} x_i p_{ij} = \sum_{i=1}^{\infty} x_i p_i. \quad (4.8)$$

当  $(X, Y)$  为连续型随机向量时, 则有

$$E(X) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x f(x, y) dxdy = \int_{-\infty}^{+\infty} x f_X(x) dx \quad (4.9)$$

**例题 4.2** 设  $(X, Y)$  在半圆域  $D = \{(x, y) | x^2 + y^2 \leq 1, y \geq 0\}$  上服从均匀分布, 求  $X^2Y$  的数学期望。

**解** 可以列出  $(X, Y)$  的概率密度为

$$f(x, y) = \begin{cases} \frac{2}{\pi}, & x^2 + y^2 \leq 1, y \geq 0 \\ 0, & \text{其他} \end{cases}$$

所以

$$E(X^2Y) = \iint_D x^2 y \frac{2}{\pi} dx dy = \frac{2}{\pi} \int_{-1}^1 x^2 dx \int_0^{\sqrt{1-x^2}} y dy = \frac{4}{15\pi}$$

### 4.1.3 数学期望的性质

**性质** 数学期望具有如下性质 ( $C$  为任意常数,  $X, Y$  表示随机变量且期望存在)

1.  $E(C) = C$
2.  $E(CX) = CE(X)$
3.  $E(X+Y) = E(X) + E(Y)$
4. 若  $X$  与  $Y$  相互独立, 则

$$E(XY) = E(X)E(Y) \quad (4.10)$$

5. (Cauchy-Schwarz 不等式)  $[E(XY)]^2 \leq E(X^2)E(Y^2)$

一般地, 数学期望具有如下的线性性质:

$$E\left(\sum_{k=1}^n C_k X_k\right) = \sum_{k=1}^n C_k E(X_k) \quad (4.11)$$

性质 (4) 还可以推广到  $n(n > 2)$  个相互独立的随机变量, 即若  $X_1, X_2, \dots, X_n$  是  $n$  个相互独立的随机变量, 则

$$E(X_1 X_2 \cdots X_n) = E(X_1)E(X_2) \cdots E(X_n) \quad (4.12)$$

## 4.2 方差

### 4.2.1 方差与标准差

中学还学过方差的基本概念, 在概率论里有如下具体定义。

#### 定义 4.2 (方差与标准差)

设  $X$  为随机变量, 若  $E\{[X - E(X)]^2\}$  存在, 则称

$$D(X) = E\{[X - E(X)]^2\} \quad (4.13)$$

为  $X$  的方差, 称  $\sqrt{D(X)}$  为  $X$  的标准差, 记为  $\sigma(X)$ , 即  $\sigma(X) = \sqrt{D(X)}$ 。

 **笔记** 由于  $\sigma(X)$  的量纲与  $X$  的量纲相同, 因而, 在工程技术领域中常用标准差  $\sigma(X)$ 。记  $\mu = E(X)$ , 因为

$$[X - E(X)]^2 = (X - \mu)^2 = X^2 - 2\mu X + \mu^2$$

根据数学期望的性质, 有

$$E\{[X - E(X)]^2\} = E(X^2) - 2\mu E(X) + \mu^2 = E(X^2) - \mu^2$$

所以

$$D(X) = E(X^2) - [E(X)]^2 \quad (4.14)$$

### 4.2.2 方差的性质

在讨论方差的性质之前, 我们先来介绍一个重要的不等式。

#### 定理 4.2 (Markov 不等式)

设  $X$  为随机变量, 若  $E(|X|^r) < +\infty$ , 其中  $r > 0$  为实数, 则对任意正数  $\varepsilon$  有

$$P\{|X| \geq \varepsilon\} \leq \frac{E(|X|^r)}{\varepsilon^r} \quad (4.15)$$

**证明** 以下仅就  $X$  为连续型随机变量的情形加以证明, 并设  $X$  的概率密度为  $f(x)$ , 则

$$\begin{aligned} P\{|X| \geq \varepsilon\} &= \int_{|x| \geq \varepsilon} f(x) dx \\ &\leq \int_{|x| \geq \varepsilon} \frac{|x|^r}{\varepsilon^r} f(x) dx \quad (\text{因为在积分域上, 恒有 } 1 \leq \frac{|x|^r}{\varepsilon^r}) \\ &\leq \frac{1}{\varepsilon^r} \int_{-\infty}^{+\infty} |x|^r f(x) dx = \frac{E(|X|^r)}{\varepsilon^r} \end{aligned}$$

接着, 我们可以给出方差的重要性质

**性质** 方差具有如下性质 ( $C$  为任意常数, 出现的随机变量方差均存在)

1.  $D(C) = 0$
2.  $D(CX) = C^2 D(X)$
3. 若  $X$  与  $Y$  相互独立, 则  $D(X \pm Y) = D(X) + D(Y)$
4.  $D(X) = 0$  的充要条件是  $P\{X = E(X)\} = 1$
5. 若  $X_1, X_2, \dots, X_n$  是相互独立的随机变量, 则

$$D\left(\sum_{i=1}^n C_i X_i\right) = \sum_{i=1}^n C_i^2 D(X_i) \quad (4.16)$$

**例题 4.3(P104 第 23 题)** 设随机变量  $X, Y$  相互独立且方差存在, 证明

$$D(XY) = D(X)D(Y) + [E(X)]^2 D(Y) + E[(Y)]^2 D(X) \quad (4.17)$$

**证明** 根据式 4.14, 可以得到

$$E(X^2) = D(X) + [E(X)]^2, \quad E(Y^2) = D(Y) + [E(Y)]^2$$

因为  $X, Y$  相互独立, 因此  $X^2, Y^2$  相互独立, 于是有

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

这一结果式子表明  $D(XY) \geq D(X)D(Y)$  恒成立。

根据这些性质和简单的证明, 可以得到下表。

| 分布        | 记号                 | 数学期望 $E(X)$         | 方差 $D(X)$             |
|-----------|--------------------|---------------------|-----------------------|
| (0, 1) 分布 | $B(1, p)$          | $p$                 | $p(1-p)$              |
| 二项分布      | $B(n, p)$          | $np$                | $np(1-p)$             |
| 泊松分布      | $P(\lambda)$       | $\lambda$           | $\lambda$             |
| 几何分布      | $Ge(p)$            | $\frac{1}{p}$       | $\frac{1-p}{p^2}$     |
| 正态分布      | $N(\mu, \sigma^2)$ | $\mu$               | $\sigma^2$            |
| 均匀分布      | $U(a, b)$          | $\frac{a+b}{2}$     | $\frac{(b-a)^2}{12}$  |
| 指数分布      | $Exp(\lambda)$     | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ |

**表 4.1:** 常见分布的数学期望 (聪明的你可以自己证明一遍)

## 4.3 协方差、相关系数和矩

### 4.3.1 协方差

人们常常希望了解随机向量的分量之间的相互关系，本段将针对二维随机向量  $(X, Y)$  来讨论有关这方面的问题，具体而言，可以给出协方差的定义。

#### 定义 4.3 (协方差)

设  $(X, Y)$  为二维随机向量，若  $E\{[X - E(X)][Y - E(Y)]\}$  存在，则定义  $X$  与  $Y$  的协方差为

$$\text{Cov}(X, Y) = E\{[X - E(X)][Y - E(Y)]\} \quad (4.18)$$

记作  $\text{Cov}(X, Y)$ 。

 **笔记** 读者可以自行证明下面这个更简便计算协方差的定义式。

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) \quad (4.19)$$

**性质** 若  $X_1, X_2, X, Y$  为随机变量， $a, b$  为常数，并假定所涉及的协方差均存在，则

1.  $\text{Cov}(a, X) = 0$
2.  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
3.  $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$
4.  $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$
5. 若  $X$  与  $Y$  相互独立，则  $\text{Cov}(X, Y) = 0$
6. (Cauchy-Schwarz 不等式) 若  $E(X^2), E(Y^2)$  存在，则

$$[E(XY)]^2 \leq E(X^2)E(Y^2), \quad [\text{Cov}(X, Y)]^2 \leq D(X)D(Y) \quad (4.20)$$

#### 定理 4.3

设有随机变量  $X, Y$ ，且方差存在，则有恒等式

$$D(X \pm Y) = D(X) \pm 2\text{Cov}(X, Y) + D(Y) \quad (4.21)$$

**证明** 根据方差的定义，有

$$\begin{aligned} D(X \pm Y) &= E\{[(X \pm Y) - E(X \pm Y)]^2\} \\ &= E\{[X \pm Y - E(X) \mp E(Y)]^2\} \\ &= E\{[[X - E(X)] \pm [Y - E(Y)]]^2\} \\ &= E\{[X - E(X)]^2 \pm 2[X - E(X)][Y - E(Y)] + [Y - E(Y)]^2\} \\ &= E\{[X - E(X)]^2\} \pm 2E\{[X - E(X)][Y - E(Y)]\} + E\{[Y - E(Y)]^2\} \\ &= D(X) \pm 2\text{Cov}(X, Y) + D(Y) \end{aligned}$$

### 4.3.2 相关系数

在 4.2.2 小节中方差的性质 (3) 中提到，如果  $X, Y$  相互独立，那么式子 4.21 右侧的  $\pm 2\text{Cov}(X, Y)$  项就不存在。这提醒我们——当两个随机变量相互独立时，他们的协方差为 0。由此我们再给出相关系数的定义。

#### 定义 4.4 (相关系数)

若  $D(X) > 0, D(Y) > 0$ ，则称下式为  $X$  与  $Y$  的相关系数，记为  $\rho(X, Y)$

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}} \quad (4.22)$$



**性质** 相关系数具有下列性质:

1. 若  $X$  与  $Y$  相互独立, 则  $\rho(X, Y) = 0$
2.  $|\rho(X, Y)| \leq 1$
3.  $|\rho(X, Y)| = 1$  的充要条件是: 存在常数  $a, b$ , 使得  $P\{Y = a + bX\} = 1$

**注** 一般而言,  $\rho(X, Y)$  的绝对值是介于 0 与 1 之间的常数。当  $|\rho(X, Y)| = 1$  时,  $X$  与  $Y$  之间几乎必然地存在着某种线性关系。因而当  $|\rho(X, Y)|$  接近于 1 时, 我们说  $X$  与  $Y$  之间的线性相关程度较强; 当  $|\rho(X, Y)|$  接近于 0 时, 则说  $X$  与  $Y$  之间的线性相关程度较弱。特别地, 当  $\rho(X, Y) = 0$  时, 称  $X$  与  $Y$  不线性相关 (简称  $X$  与  $Y$  不相关)。可以看出, 相关系数  $\rho(X, Y)$  是一个反映随机变量  $X$  与  $Y$  之间线性相关程度的量。

下面的例题说明了一个易错点——随机变量的相互独立性和相关性不能划等号。

**例题 4.4** 设二维连续型随机向量  $(X, Y)$  具有概率密度

$$f(x, y) = \begin{cases} \frac{1}{\pi}, & x^2 + y^2 \leq 1, \\ 0, & \text{其他} \end{cases}$$

证明:  $X$  与  $Y$  是不相关的, 但  $X$  与  $Y$  并不相互独立。

**证明** 根据概率密度的定义有

$$f_X(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-x^2}, & |x| \leq 1 \\ 0, & |x| > 1 \end{cases}, \quad f_Y(y) = \begin{cases} \frac{2}{\pi}\sqrt{1-y^2}, & |y| \leq 1 \\ 0, & |y| > 1 \end{cases}$$

因此

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

同理得  $E(Y) = 0$ , 又因为

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y) = \iint_{x^2+y^2 \leq 1} \frac{xy}{\pi} dx dy = 0$$

这表明  $X$  与  $Y$  不相关。但因为集合  $D = \{(x, y) \mid |x| > 1, |y| > 1, x^2 + y^2 > 1\}$  上有

$$f(x, y) = 0 \neq f_X(x)f_Y(y) = \frac{4}{\pi^2}\sqrt{1-x^2}\sqrt{1-y^2}$$

而点集  $D$  的面积不为零, 所以  $X$  与  $Y$  不是相互独立的。

 **笔记** 随机变量  $X, Y$  相互独立则必然不相关, 反之不一定成立。

**命题 4.1 (正态分布向量的相关系数)**

设二维随机向量  $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$ , 则  $\rho(X, Y) = \rho$ 。

### 4.3.3 矩

现在介绍几种反映随机变量 (向量) 数字特征的普遍量。

**定义 4.5 (随机变量的原点矩和中心矩)**

设  $X$  为随机变量,  $n$  为非负整数, 若

$$\alpha_n = E(X^n) \quad (4.23)$$

存在, 则称它为  $X$  的  $n$  阶原点矩, 简称为  $n$  阶矩; 若

$$\mu_n = E[(X - E(X))^n], \quad (n > 1) \quad (4.24)$$

存在, 则称它为  $X$  的  $n$  阶中心矩。

 **笔记** 数学期望  $E(X)$  是  $X$  的一阶原点矩, 方差  $D(X)$  是  $X$  的二阶中心矩。

原点矩和中心矩满足下面的公式

$$\begin{aligned}
 \mu_n &= E\{[X - E(X)]^n\} \\
 &= E\left\{\sum_{k=0}^n C_n^k X^k [-E(X)]^{n-k}\right\} \\
 &= \sum_{k=0}^n C_n^k E(X^k) [-E(X)]^{n-k} \\
 &= \sum_{k=0}^n C_n^k \alpha_k (-\alpha_1)^{n-k}, \quad (\text{规定 } \alpha_0 = 1)
 \end{aligned} \tag{4.25}$$

所以, 若能求得  $X$  的前  $n$  阶原点矩, 则它的  $n$  阶中心矩可由上式确定。当  $n = 2$  时, 有

$$\mu_2 = \alpha_2 - \alpha_1^2 \tag{4.26}$$

此即式 4.14  $D(X) = E(X^2) - [E(X)]^2$  的等价表示。

#### 定义 4.6 (随机向量的混合原点矩和混合中心矩)

设  $(X, Y)$  为二维随机向量,  $k, l$  为非负整数。若

$$\alpha_{kl} = E(X^k Y^l) \tag{4.27}$$

存在, 则称它为  $(X, Y)$  的  $k + l$  阶混合原点矩; 若

$$\mu_{kl} = E\{[X - E(X)]^k [Y - E(Y)]^l\} \tag{4.28}$$

存在, 则称它为  $(X, Y)$  的  $k + l$  阶混合中心矩。

 **笔记**  $X$  与  $Y$  的协方差可表示为  $\text{Cov}(X, Y) = \mu_{11}$ , 相关系数可表示为  $\rho(X, Y) = \frac{\mu_{11}}{\sqrt{\mu_{20}}\sqrt{\mu_{02}}}$ 。

### 4.3.4 协方差矩阵和正态向量

设  $n$  维随机向量  $(X_1, X_2, \dots, X_n)$  的二阶混合中心矩

$$v_{ij} = \text{Cov}(X_i, X_j) \quad (i, j = 1, 2, \dots, n)$$

都存在, 则称矩阵

$$\mathbf{V} = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ v_{21} & v_{22} & \cdots & v_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{n1} & v_{n2} & \cdots & v_{nn} \end{pmatrix} \tag{4.29}$$

为  $n$  维随机向量  $(X_1, X_2, \dots, X_n)$  的协方差矩阵。利用协方差矩阵的定义可以推广正态分布的定义到  $n$  维随机向量的情形。设  $(X_1, X_2, \dots, X_n)$  为  $n$  维随机向量, 引入下面的行矩阵

$$\mathbf{X} = (X_1, X_2, \dots, X_n), \quad \mathbf{x} = (x_1, x_2, \dots, x_n)$$

规定  $E(X_1, X_2, \dots, X_n) = (E(X_1), E(X_2), \dots, E(X_n))$ , 则向量

$$\mathbf{a} = (\mu_1, \mu_2, \dots, \mu_n) = (E(X_1), E(X_2), \dots, E(X_n)) = E(\mathbf{X}) \tag{4.30}$$

若  $\mathbf{X}$  的概率密度为

$$f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{n/2} |\mathbf{V}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \mathbf{a})\mathbf{V}^{-1}(\mathbf{x} - \mathbf{a})^T\right\} \tag{4.31}$$

其中  $\mathbf{V}$  是  $\mathbf{X}$  的协方差矩阵, 则称  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  服从  $n$  维正态分布, 记做  $\mathbf{X} \sim N(\mathbf{a}, \mathbf{V})$ , 并称  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  为  $n$  维正态向量。接下来不加证明的介绍一些相关性质。

**性质** 正态向量具有以下重要性质:

1. 随机向量  $(X_1, X_2, \dots, X_n)$  服从  $n$  维正态分布的充要条件是  $X_1, X_2, \dots, X_n$  的任意线性组合  $\sum_{k=1}^n l_k X_k$  服从一维正态分布, 这里  $l_1, l_2, \dots, l_n$  是任意取定的一组不全为零的实数。
2. 设  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  为  $n$  维正态向量, 则由  $\mathbf{X}$  的  $m$  ( $m < n$ ) 个分量所构成的  $m$  维随机向量  $\tilde{\mathbf{X}} = (X_{i_1}, X_{i_2}, \dots, X_{i_m})$  ( $1 \leq i_1 < i_2 < \dots < i_m \leq n$ ) 是  $m$  维正态向量, 且其数学期望为

$$\tilde{\mathbf{a}} = (\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_m}) = (E(X_{i_1}), E(X_{i_2}), \dots, E(X_{i_m}))$$

协方差矩阵为

$$\tilde{\mathbf{V}} = \begin{pmatrix} \text{Cov}(X_{i_1}, X_{i_1}) & \text{Cov}(X_{i_1}, X_{i_2}) & \cdots & \text{Cov}(X_{i_1}, X_{i_m}) \\ \text{Cov}(X_{i_2}, X_{i_1}) & \text{Cov}(X_{i_2}, X_{i_2}) & \cdots & \text{Cov}(X_{i_2}, X_{i_m}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_{i_m}, X_{i_1}) & \text{Cov}(X_{i_m}, X_{i_2}) & \cdots & \text{Cov}(X_{i_m}, X_{i_m}) \end{pmatrix}$$

特别当  $m = 1$  时, 有  $X_j \sim N(\mu_j, D(X_j))$  ( $1 \leq j \leq n$ ), 即  $n$  维正态向量  $\mathbf{X}$  的每个分量  $X_j$  均为一维正态变量。

3. 若  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  服从  $n$  维正态分布  $N(\mathbf{a}, \mathbf{V})$ ,  $\mathbf{C}$  为  $m \times n$  矩阵, 其中元素  $C_{ij}$  ( $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ ) 均为常数且  $\text{rank}(\mathbf{C}) = m$ , 则  $m$  维随机向量

$$\mathbf{Y}^T = \mathbf{C}\mathbf{X}^T$$

服从  $m$  维正态分布  $N(\mathbf{C}\mathbf{a}^T, \mathbf{C}^T\mathbf{V}\mathbf{C})$ 。

4. 若  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  服从  $n$  维正态分布  $N(\mathbf{a}, \mathbf{V})$ ,  $l_1, l_2, \dots, l_n$  是一组不全为零的常数, 则  $Y = \sum_{i=1}^n l_i X_i$  服从一维正态分布

$$N\left(\sum_{j=1}^n l_j \mu_j, \sum_{j=1}^n \sum_{k=1}^n l_j l_k \text{Cov}(X_j, X_k)\right)$$

5. 设  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  是  $n$  维正态向量, 则  $X_1, X_2, \dots, X_n$  相互独立的充要条件是它们两两不相关。

## 4.4 条件期望与条件方差

类似于条件概率、条件分布函数等定义, 可以对期望和方差做“条件”定义。

### 定义 4.7 (条件期望与条件方差)

设  $X, Y$  为随机变量, 在  $Y = y$  条件下,  $X$  的条件期望定义为

$$E(X | Y = y) = \begin{cases} \sum x_i P\{X = x_i | Y = y\}, & X, Y \text{ 是离散型的} \\ \int_{-\infty}^{+\infty} x f_{X|Y}(x | y) dx, & X, Y \text{ 是连续型的} \end{cases} \quad (4.32)$$

$X$  的条件方差定义为

$$D(X | Y = y) = \begin{cases} \sum [x_i - E(X | Y = y)]^2 P\{X = x_i | Y = y\}, & X, Y \text{ 是离散型的} \\ \int_{-\infty}^{+\infty} [x - E(X | Y = y)]^2 f_{X|Y}(x | y) dx, & X, Y \text{ 是连续型的} \end{cases} \quad (4.33)$$

本小节介绍条件期望和条件方差主要是为了介绍下述两大重要定理。

**定理 4.4 (全期望公式)**

设随机变量  $X, Y$ , 则有全期望公式

$$\begin{aligned} E(X) &= E[E(X|Y)] \\ &= \begin{cases} \sum E(X|Y=y_j)P\{Y=y_j\}, & X, Y \text{ 是离散型的} \\ \int_{-\infty}^{+\infty} E(X|Y=y)f_Y(y)dy, & X, Y \text{ 是连续型的} \end{cases} \end{aligned} \quad (4.34)$$



**例题 4.5** 黄金周到某旅游风景区的游客数是随机变量  $N$  ( $N$  的可能值为正整数),  $X_i$  是第  $i$  位游客在进入景区后所花费的钱数, 假定  $X_1, X_2, \dots, X_N$  是独立同分布的随机变量列, 且与  $N$  相互独立, 求黄金周期间该旅游风景区的平均收入。

**解** 由全期望公式得

$$\begin{aligned} E\left(\sum_{i=1}^N X_i\right) &= E\left[E\left(\sum_{i=1}^N X_i \middle| N\right)\right] \\ &= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^N X_i \middle| N=n\right) P\{N=n\} \\ &= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^n X_i \middle| N=n\right) P\{N=n\} \\ &= \sum_{n=1}^{\infty} E\left(\sum_{i=1}^n X_i\right) P\{N=n\} \quad (\text{因 } \sum_{i=1}^n X_i \text{ 与 } N \text{ 独立}) \\ &= \sum_{n=1}^{\infty} nE(X_1) P\{N=n\} \quad (\text{因 } X_1, X_2, \dots \text{ 独立同分布}) \\ &= \left[\sum_{n=1}^{\infty} nP\{N=n\}\right] E(X_1) \\ &= E(N) E(X_1) \end{aligned}$$

这表明此类题目可以把求和上标随机变量  $N$  通过期望的形式挪出来, 消去求和项。

**定理 4.5 (全方差公式)**

设随机变量  $X, Y$ , 则有全方差公式

$$D(X) = E[D(X|Y)] + D[E(X|Y)] \quad (4.35)$$



**笔记** 全方差公式说明, 当有另一个随机变量  $Y$  对  $X$  起作用时,  $X$  的方差可以分解为两部分——一部分是  $Y$  对  $X$  的系统影响部分  $E(X|Y)$  的方差, 另一部分是条件方差  $D(X|Y)$  的平均。

## 第五章 大数定律与中心极限定理

### 内容提要

□ 随机变量的收敛性

□ Chebyshev 不等式

□ 大数定律

□ 中心极限定理

**注** 本章所描述的大数定律和中心极限定理等是概率论和统计学的基础理论，在理论研究和应用实践中都有着重要应用。因此本章涉及的定义、定理等较为抽象，如果读者不想局限于应试考试的要求，可参考下述网址链接 [https://www.tup.tsinghua.edu.cn/upload/books/yz/094807-01.pdf] 增进对本章的了解。(戳链接直达 ~)

### 5.1 随机变量的收敛性

高等数学中介绍了数列的收敛性概念，那么对随机变量序列的收敛性又当如何定义呢？由于随机变量序列  $\{X_n, n = 1, 2, \dots\}$  除了下标  $n$  从 1 变化至无穷外，每个  $X_n$  都含有样本点  $\omega$ ，这使得对  $X_n$  定义的收敛性与数列的收敛性有很大不同。在正式开始本章内容之前，先介绍随机变量序列的两种收敛性：依概率收敛和依分布收敛。

#### 5.1.1 依概率收敛

##### 定义 5.1 (依概率收敛)

设  $\{X_n, n = 1, 2, \dots\}$  是随机变量序列， $X$  也是一个随机变量，若对任意  $\varepsilon > 0$ ，有

$$\lim_{n \rightarrow \infty} P\{|X_n - X| \geq \varepsilon\} = 0 \quad (5.1)$$

则称随机变量序列  $\{X_n\}$  依概率收敛于  $X$ ，记作  $(p) \lim_{n \rightarrow \infty} X_n = X$  或者  $X_n \xrightarrow{P} X$ 。

依概率收敛表明，随机变量  $X_n$  对  $X$  的绝对偏差不小于任一给定正数的概率随着  $n$  增大而越来越接近于零。由概率的性质可知定义 5.1 中的式 5.1 等价于下式。

$$\lim_{n \rightarrow \infty} P\{|X_n - X| < \varepsilon\} = 1 \quad (5.2)$$

特别地，当随机变量  $X$  为单点分布，即  $P\{X = a\} = 1$  时，则称序列  $\{X_n\}$  依概率收敛于  $a$  (记  $X_n \xrightarrow{P} a$ )。

**性质** 依概率收敛与常数的随机变量序列有以下性质：

1.  $(p) \lim_{n \rightarrow \infty} (X_n \pm Y_n) = a \pm b$
2.  $(p) \lim_{n \rightarrow \infty} X_n Y_n = ab$
3.  $(p) \lim_{n \rightarrow \infty} \frac{X_n}{Y_n} = \frac{a}{b} (Y_n \neq 0, b \neq 0)$
4. 若  $(p) \lim_{n \rightarrow \infty} X_n = a, (p) \lim_{n \rightarrow \infty} Y_n = b$ ，又  $g(x, y)$  在点  $(a, b)$  处连续，则有

$$(p) \lim_{n \rightarrow \infty} g(X_n, Y_n) = g(a, b)$$

**例题 5.1** 设随机变量  $X_n$  服从 Cauchy 分布  $C\left(0, \frac{1}{n}\right)$ ，其概率密度如下，证明  $(p) \lim_{n \rightarrow \infty} X_n = 0$ 。

$$f_n(x) = \frac{n}{\pi(1+n^2x^2)}, \quad (-\infty < x < +\infty)$$

**证明** 由依概率收敛的定义， $\forall \varepsilon > 0$  有

$$P\{|X_n - 0| < \varepsilon\} = P\{|X_n| < \varepsilon\} = \int_{-\varepsilon}^{\varepsilon} \frac{n}{\pi(1+n^2x^2)} dx = \frac{2}{\pi} \arctan n\varepsilon$$

则  $\lim_{n \rightarrow \infty} P\{|X_n - 0| < \varepsilon\} = \lim_{n \rightarrow \infty} \frac{2}{\pi} \arctan n\varepsilon = 1$ ，因此  $X_n$  依概率收敛于 0。

### 5.1.2 依分布收敛

#### 定义 5.2 (依分布收敛)

设  $\{X_n, n = 1, 2, \dots\}$  为随机变量序列, 其对应的分布函数序列为  $\{F_n(x), n = 1, 2, \dots\}$ ,  $X$  是另一随机变量, 其分布函数为  $F(x)$ 。若对  $F(x)$  的每个连续点  $x$  有

$$\lim_{n \rightarrow \infty} F_n(x) = F(x) \quad (5.3)$$

则称随机变量序列  $\{X_n\}$  依分布收敛于  $X$ , 记作  $X_n \xrightarrow{L} X$ , 或称分布函数序列  $\{F_n(x), n = 1, 2, \dots\}$  弱收敛于  $F(x)$ , 记为  $F_n(x) \xrightarrow{w} F(x)$ 。

下面的定理刻画了这两种收敛的关系。

#### 定理 5.1 (依概率收敛和依分布收敛的关系)

对于随机变量序列  $\{X_n, n = 1, 2, \dots\}$  和随机变量  $X$

- 若  $X_n \xrightarrow{P} X$ , 则  $X_n \xrightarrow{L} X$
- 对于常数  $C$ ,  $X_n \xrightarrow{P} C$  的充分必要条件为  $X_n \xrightarrow{L} C$

### 5.1.3 其他类型的收敛

最后来介绍其他两种收敛形式。

#### 定义 5.3 ( $r$ 阶平均收敛与几乎处处收敛)

设  $\{X_n, n = 1, 2, \dots\}$  为随机变量序列,  $X$  为随机变量

1. 若对于  $r > 0$ ,  $X, X_1, X_2, \dots, X_n, \dots$  的  $r$  阶矩都存在, 并且有

$$\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$$

称该序列依  $r$  阶平均收敛于  $X$ , 记作  $X_n \xrightarrow{r} X$

2. 若等式

$$P\{\omega | \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\} = 1$$

恒成立, 称该序列几乎处处收敛于  $X$ , 记作  $X_n \xrightarrow{a.s.} X$

 **笔记** 若  $X_n \xrightarrow{r} X$ , 或者  $X_n \xrightarrow{a.s.} X$ , 则  $X_n \xrightarrow{P} X$  成立。

## 5.2 Chebyshev 不等式

下面介绍一重要定理, 其揭示了随机变量  $X$  的分布未知的情况下, 利用期望和方差估算随机变量  $X$  和均值  $E(X)$  的差在不同偏差  $\varepsilon$  内概率  $P\{|X - E(X)| < \varepsilon\}$  的下界。

#### 定理 5.2 (Chebyshev 不等式)

若随机变量  $X$  存在数学期望  $E(X)$  和方差  $D(X)$ , 则对任意正数  $\varepsilon$ , 有 Chebyshev 不等式

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{D(X)}{\varepsilon^2} \quad (5.4)$$

或等价地有

$$P\{|X - E(X)| < \varepsilon\} \geq 1 - \frac{D(X)}{\varepsilon^2} \quad (5.5)$$

**证明** 在 Markov 不等式 4.15 中, 以  $X - E(X)$  取代  $X$  并令正实数  $r = 2$ , 则对任意正数  $\varepsilon$  有

$$P\{|X - E(X)| \geq \varepsilon\} \leq \frac{E\{[X - E(X)]^2\}}{\varepsilon^2} = \frac{D(X)}{\varepsilon^2}$$

下面用一道题目说明 Chebyshev 不等式的具体应用。

**例题 5.2(P119 第 1 题)** 设随机变量  $X$  的分布未知, 但已知  $E(X) = \mu, D(X) = \sigma^2$ , 试用 Chebyshev 不等式估计  $X$  落在区间  $(\mu - 3\sigma, \mu + 3\sigma)$  内的概率下界。

**解** 根据定理 5.2, 取  $\varepsilon = 3\sigma$  有

$$P\{|X - \mu| < 3\sigma\} \geq 1 - \frac{\sigma^2}{(3\sigma)^2} = 1 - \frac{1}{9} = \frac{8}{9}$$

这表明  $X$  落在区间  $(\mu - 3\sigma, \mu + 3\sigma)$  内的概率不低于  $\frac{8}{9}$ 。

 **笔记** Chebyshev 不等式需要在后续定理的证明中频繁用到, 考试中也经常考察, 请牢记。

## 5.3 大数定律

### 5.3.1 大数定律的概念

当我们大量重复某一相同的实验的时候, 其最后的实验结果可能会稳定在某一数值附近。就像抛硬币一样, 当我们不断地抛, 抛个上千次, 甚至上万次, 我们会发现正面或者反面向上的次数都会接近一半。除了抛硬币, 现实中还有许许多多这样的例子, 像掷骰子, 最著名的实验就是蒲丰投针实验。这些实验都像我们传达了一个共同的信息, 那就是大量重复实验最终的结果都会比较稳定。那稳定性到底是什么? 怎样去用数学语言把它表达出来? 这个满足的规律就是**大数定律**。

大数定律有若干个表现形式, 下面我们介绍几种常见形式。

### 5.3.2 Chebyshev 大数定律

#### 定理 5.3 (Chebyshev(切比雪夫) 大数定律)

设  $\{X_n, n = 1, 2, \dots\}$  是相互独立的随机变量序列, 且分别存在数学期望  $E(X_k)$  和方差  $D(X_k)$  ( $k = 1, 2, \dots$ )。若存在常数  $C$ , 使得  $D(X_k) \leq C$  ( $k = 1, 2, \dots$ ), 则对任意  $\varepsilon > 0$  有

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{k=1}^n X_k - \frac{1}{n} \sum_{k=1}^n E(X_k)\right| < \varepsilon\right\} = 1 \quad (5.6)$$

**证明** 因  $\{X_n, n = 1, 2, \dots\}$  相互独立, 亦两两不相关, 故

$$D\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n^2} \sum_{k=1}^n D(X_k) \leq \frac{C}{n}$$

令  $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$ , 则由 Chebyshev 不等式得

$$1 \geq P\{|Y_n - E(Y_n)| < \varepsilon\} \geq 1 - \frac{D(Y_n)}{\varepsilon^2} \geq 1 - \frac{C}{n\varepsilon^2}$$

令  $n \rightarrow \infty$ , 由数列极限的夹逼法则则知

$$\lim_{n \rightarrow \infty} P\{|Y_n - E(Y_n)| < \varepsilon\} = 1$$

注意到  $E(Y_n) = E\left(\frac{1}{n} \sum_{k=1}^n X_k\right) = \frac{1}{n} \sum_{k=1}^n E(X_k)$ , 把  $Y_n$  和  $E(Y_n)$  的表达式代入到上式中即证。

根据 Chebyshev 大数定律可以得到下面的更常用的形式。

**定理 5.4 (独立同分布大数定律)**

设  $X_n, n = 1, 2, \dots$  是相互独立的随机变量序列, 且存在相同的数学期望  $E(X_k) = \mu$  与方差  $D(X_k) = \sigma^2 (k = 1, 2, \dots)$ , 则对任意  $\varepsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1 \quad (5.7)$$

♡

定理 5.4 表明, 若随机变量序列  $\{X_n, n = 1, 2, \dots\}$  相互独立, 且具有相同的数学期望与方差  $E(X_k) = \mu, D(X_k) = \sigma^2 (k = 1, 2, \dots)$ , 则随机变量序列  $Y_n = \frac{1}{n} \sum_{k=1}^n X_k$  依概率收敛于  $\mu$ , 即  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu$

 **笔记** 人们在大量实践中发现: 大概率事件在一次试验中几乎肯定要发生, 而小概率事件在一次试验中几乎不可能发生。这一规律称为“实际推断原理”(详见第 8 章假设检验), 它对于我们在实际工作中处理概率接近于 1 或者 0 的事件具有一定的指导意义。

例如, 设  $X$  是一个概率分布未知的随机变量, 我们无法由定义算出  $E(X)$ , 为了估算  $E(X)$ , 对  $X$  做一次  $n$  重观测试验, 记  $X_k$  为第  $k$  个试验的结果 ( $k = 1, 2, \dots, n$ ), 由于这一连串独立重复观测只是设想要做而尚未实施, 所以  $X_k (k = 1, 2, \dots, n)$  是随机变量, 并且  $X_1, X_2, \dots, X_n$  是相互独立的, 每一个随机变量  $X_k (k = 1, 2, \dots, n)$  的概率分布都与  $X$  的概率分布一致。因此有

$$E(X_k) = E(X), \quad D(X_k) = D(X), \quad (\text{假设期望、方差都存在}; k = 1, 2, \dots, n)$$

根据定理 5.4, 对任意给定的一个比较小的正数  $\varepsilon$ , 当  $n$  充分大时, 事件  $\left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - E(X) \right| < \varepsilon \right\}$  为大概率事件, 因此, 对一次  $n$  重观测试验的结果:  $x_1, x_2, \dots, x_n$ , 由实际推断原理知, 几乎可以肯定有

$$\left| \frac{x_1 + x_2 + \dots + x_n}{n} - E(X) \right| < \varepsilon \quad (5.8)$$

故  $E(X) \approx \frac{1}{n} \sum_{k=1}^n x_k$ , 这种估算  $E(X)$  的方法称为数学期望的统计估计。

**5.3.3 Bernoulli 大数定律**

Bernoulli 大数定律是概率论历史上最早出现的大数定律, 是由瑞士数学家雅科布·伯努利在他的遗著《猜度术》中首次提出的, 现叙述如下。

**定理 5.5 (Bernoulli(伯努利) 大数定律)**

设  $n_A$  是  $n$  次独立重复试验中事件  $A$  发生的次数,  $p$  是事件  $A$  在每次试验发生的概率, 则对任意正数  $\varepsilon$  有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| < \varepsilon \right\} = 1, \quad \lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| \geq \varepsilon \right\} = 0 \quad (5.9)$$

即  $\frac{n_A}{n} \xrightarrow{P} p$  成立。

♡

**证明** 引入随机变量

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

显然有

$$n_A = X_1 + X_2 + \dots + X_n$$

因为  $X_k$  只依赖第  $k$  次试验的结果, 又假设各次试验是独立的, 所以  $X_1, X_2, \dots, X_n, \dots$  是相互独立的。另外, 因为  $X_k \sim B(1, p)$ , 所以  $E(X_k) = p, D(X_k) = p(1-p) (k = 1, 2, \dots)$ 。根据定理 5.4 有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - p \right| < \varepsilon \right\} = 1$$



亦即

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{n_A}{n} - p \right| < \varepsilon \right\} = 1$$

### 5.3.4 Khinchin 大数定律

Khinchin 大数定律是大数定律的著名表述之一，由苏联数学家亚历山大·雅科夫列维奇·辛钦提出。

#### 定理 5.6 (Khinchin(辛钦) 大数定律)

设  $\{X_n, n = 1, 2, \dots\}$  是独立同分布的随机变量序列，且  $E(X_k) = \mu (k = 1, 2, \dots)$  存在，则对任意  $\varepsilon > 0$ ，有

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{1}{n} \sum_{k=1}^n X_k - \mu \right| < \varepsilon \right\} = 1 \quad (5.10)$$

即  $\frac{1}{n} \sum_{k=1}^n X_k \xrightarrow{P} \mu$  成立。



**注** Khinchin 大数定律不要求随机变量的方差存在，只要求随机变量序列独立同分布且期望存在。Bernoulli 大数定律是 Khinchin 大数定律的特殊情况。关于定理 5.6 的证明可参考 [https://zhuanlan.zhihu.com/p/709397691]。

**例题 5.3** 设有独立同分布的随机变量序列  $\{X_1, X_2, \dots, X_n, \dots\}$ ，其分布函数为

$$F(x) = A + \frac{1}{\pi} \arctan \frac{x}{B}, \quad (B \neq 0)$$

判断 Khinchin 大数定律是否适用于该随机变量序列。

**解** 只需要判断  $E(X_n)$  是否存在即可，即广义积分  $\int_{-\infty}^{+\infty} \left| x \frac{dF}{dx} \right| dx$  是否收敛。不难计算

$$\begin{aligned} \int_{-\infty}^{+\infty} \left| x \frac{dF}{dx} \right| dx &= \int_{-\infty}^{+\infty} \frac{|B||x|}{\pi(B^2 + x^2)} dx \\ &= \frac{2|B|}{\pi} \int_0^{+\infty} \frac{x}{B^2 + x^2} dx \\ &= \frac{|B|}{\pi} \int_0^{+\infty} \frac{d(B^2 + x^2)}{B^2 + x^2} \\ &= \frac{|B|}{\pi} \lim_{a \rightarrow +\infty} \ln(B^2 + x^2) \Big|_0^a = +\infty \end{aligned}$$

因此不适用 Khinchin 大数定律。

## 5.4 中心极限定理

### 5.4.1 中心极限定律的概念

**中心极限定理 (Central Limit Theorem, CLT)** 是概率论中一个非常重要的结论，它指出在一定条件下，独立随机变量  $X_1, X_2, \dots, X_n$  的和  $Y = \sum_{i=1}^n X_i$  收敛于正态分布。

中心极限定理不要求随机变量本身是正态分布的，所以它表明了概率统计中一个非常重要的事实——在一定条件下，我们可以用正态分布去近似任何随机分布。用具体事例来表达，如果我们抛 500 次硬币，由于每次抛硬币正面朝上的概率为  $p = \frac{1}{2}$ ，则抛硬币正面朝上的次数  $X$  可以视作服从二项分布  $B(n, p)$ 。现在中心极限定理告诉我们，可以将  $X$  近似地视作一个服从正态分布  $N(\mu, \sigma^2)$  的随机变量。下图展现了另一非正态分布随机变量在  $N$  不断变大后和正态曲线的逼近。

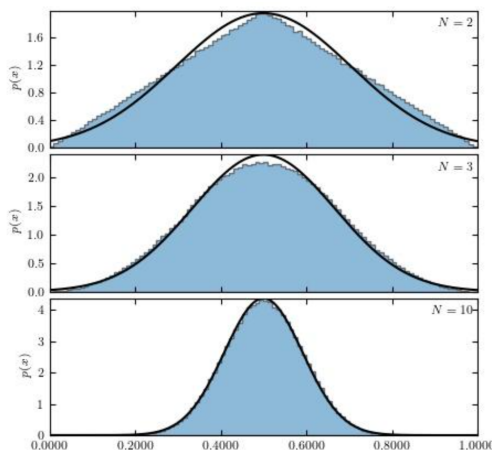


图 5.1: 中心极限定理的图示

### 5.4.2 Levy-Lindberg 中心极限定理

#### 定理 5.7 (Levy-Lindberg 中心极限定理)

设  $\{X_n, n = 1, 2, \dots\}$  是独立同分布的随机变量序列, 且有有限的数学期望和方差  $E(X_k) = \mu, D(X_k) = \sigma^2 \neq 0 (k = 1, 2, \dots)$ , 设随机变量

$$Y_n = \frac{\sum_{k=1}^n X_k - E\left(\sum_{k=1}^n X_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k\right)}} = \frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \quad (5.11)$$

则  $Y_n$  的分布函数  $F_n(x)$  对于任意实数  $x$ , 都有

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P\left\{\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \leq x\right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (5.12)$$

亦表明  $F_n(x) \xrightarrow{w} \Phi(x)$

**笔记** 由定理 5.7 可知, 当  $n$  很大时,  $Y_n$  近似地服从标准正态分布  $N(0, 1)$ 。因此, 当  $n$  很大时,  $\sum_{k=1}^n X_k = \sqrt{n}\sigma Y_n + n\mu$

近似地服从正态分布  $N(n\mu, n\sigma^2)$ 。若令  $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ , 则当  $n$  很大时,  $\bar{X}$  近似地服从正态分布  $N\left(\mu, \frac{\sigma^2}{n}\right)$ 。由此可见, 在独立同分布的场合, 无论  $X_1, X_2, \dots, X_n$  本身的分布函数为何, 它们的算术平均数  $\bar{X}$  在  $n$  很大时总是近似地服从正态分布。这一结论具有重要的理论及应用价值。

**例题 5.4** 一加法器同时收到 20 个噪声电压  $V_k (k = 1, 2, \dots, n)$ , 设它们是相互独立的随机变量, 且都在区间  $(0, 10)$  上服从均匀分布。记  $V = \sum_{k=1}^n V_k$ , 试估计  $P\{V > 105\}$  的值。

**解** 根据均匀分布的特点, 易知  $E(V_k) = 5, D(V_k) = \frac{100}{12} (k = 1, 2, \dots, 20)$ , 由定理 5.7 知

$$V = \sum_{k=1}^{20} V_k \text{ 近似服从 } N\left(20 \times 5, \frac{100}{12} \times 20\right)$$

于是有

$$\begin{aligned}
 P\{V > 105\} &= P\left\{\frac{V - 20 \times 5}{\sqrt{20} \cdot \sqrt{\frac{100}{12}}} > \frac{105 - 20 \times 5}{\sqrt{20} \cdot \sqrt{\frac{100}{12}}}\right\} \\
 &= P\left\{\frac{V - 20 \times 5}{\sqrt{20} \cdot \sqrt{\frac{100}{12}}} > 0.387\right\} \\
 &= 1 - P\left\{\frac{V - 20 \times 5}{\sqrt{20} \cdot \sqrt{\frac{100}{12}}} \leq 0.387\right\} \\
 &\approx 1 - \Phi(0.387) = 0.348
 \end{aligned}$$

即有  $P\{V > 105\} \approx 0.348$ 。

### 5.4.3 De Moivre-Laplace 中心极限定理

#### 定理 5.8 (De Moivre-Laplace 中心极限定理)

设随机变量  $\eta_n (n = 1, 2, \dots)$  服从参数为  $n, p (0 < p < 1)$  的二项分布, 则对于任意区间  $(a, b]$ , 恒有

$$\lim_{n \rightarrow \infty} P\left\{a < \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq b\right\} = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt \quad (5.13)$$

 **笔记** 定理 5.8 表明, 当  $n$  很大且  $0 < p < 1$  是一个定值时 (或者说  $np(1-p)$  不太小时), 二项变量  $\eta_n$  的分布近似正态分布  $N(np, np(1-p))$ 。结合第二章中的泊松近似可知:

- 当  $np$  大小适中时, 可采用泊松近似式 2.8
- 当  $np$  大小较大时, 可依据定理 5.8 用正态分布近似计算, 即

$$\begin{aligned}
 \sum_{k=n_1}^{n_2} C_n^k p^k (1-p)^{n-k} &= P\{n_1 \leq \eta_n \leq n_2\} \\
 &= P\left\{\frac{n_1 - np}{\sqrt{np(1-p)}} \leq \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq \frac{n_2 - np}{\sqrt{np(1-p)}}\right\} \\
 &\approx \Phi\left(\frac{n_2 - np}{\sqrt{np(1-p)}}\right) - \Phi\left(\frac{n_1 - np}{\sqrt{np(1-p)}}\right)
 \end{aligned} \quad (5.14)$$

**例题 5.5** 某单位设置有一电话总机, 共有 200 架电话分机. 设每个电话分机有 5% 的时间要使用外线通话, 假定每个分机是否使用外线是相互独立的, 问总机要备多少条外线才能以 90% 的概率保证每个分机要使用外线时不必等候?

**解** 设总机需要设置  $k$  条外线, 再设同时使用外线通话的分机数为  $X$ , 不难看出  $X \sim B(200, 0.05)$ , 由 De Moivre-Laplace 中心极限定理知

$$P\{X \leq k\} = P\left\{\frac{X - np}{\sqrt{np(1-p)}} \leq \frac{k - np}{\sqrt{np(1-p)}}\right\} \approx \Phi\left(\frac{k - 10}{\sqrt{9.5}}\right) \geq 0.9$$

查表可知要有  $k \geq 13.96$ , 因此取  $k_{\min} = 14$ 。

**例题 5.6** 试分别用 Chebyshev 不等式和 De Moivre-Laplace 中心极限定理求一枚均匀硬币需要抛多少次, 才能使得正面出现的频率在 0.4 与 0.6 之间的概率至少为 0.9?

**解** 用  $X$  表示抛掷一枚均匀硬币  $n$  次正面出现的次数, 则  $X \sim B(n, 0.5)$ ,  $E(X) = \frac{n}{2}$ ,  $D(X) = \frac{n}{4}$ , 则正面出现的频

率为  $\frac{X}{n}$ , 且  $E\left(\frac{X}{n}\right) = \frac{1}{2}, D\left(\frac{X}{n}\right) = \frac{1}{4n}$ , 因此:

1. 由 Chebyshev 不等式 5.4

$$\begin{aligned} P\left\{0.4 \leq \frac{X}{n} \leq 0.6\right\} &= P\left\{\left|\frac{X}{n} - 0.5\right| \leq 0.1\right\} \\ &\geq 1 - \frac{1}{0.1^2} D\left(\frac{X}{n}\right) = 1 - \frac{25}{n} \end{aligned}$$

令  $1 - \frac{25}{n} \geq 0.9$ , 解得  $n \geq 250$ , 即至少要抛掷 250 次硬币。

2. 由 De Moivre-Laplace 中心极限定理和式 5.13 有

$$\begin{aligned} P\left\{0.4 \leq \frac{X}{n} \leq 0.6\right\} &= P\left\{\left|\frac{\frac{X}{n} - 0.5}{\sqrt{\frac{1}{4n}}}\right| \leq \frac{0.1}{\sqrt{\frac{1}{4n}}}\right\} \\ &\approx 2\Phi(0.2\sqrt{n}) - 1 \geq 0.9 \end{aligned}$$

由  $0.2\sqrt{n} \geq 1.65$ , 解得  $n \geq 68.0625$ 。因此至少要抛掷 69 次硬币。

比较这两个结果, 可以看出, 用 Chebyshev 不等式近似计算比用中心极限定理近似计算所得结果更“粗糙”一些, 这是因为 Chebyshev 不等式给出的是概率的下界, 而中心极限定理得到的是概率的近似值。

#### 5.4.4 其他形式的中心极限定理

前面介绍的两种中心极限定理的表述形式都基于“独立同分布”的随机变量序列, 下面的定理则表述了不同分布的情况, 我们一般称之为 Lyapunov 中心极限定理或 Lindberg 中心极限定理 (此种表述省去)。

##### 定理 5.9 (Lyapunov 中心极限定理)

设随机变量  $X_1, X_2, X_3, \dots, X_n, \dots$  相互独立, 他们具有数学期望和方差

$$E(X_k) = \mu_k, \quad D(X_k) = \sigma_k^2 > 0, \quad (k = 1, 2, 3, \dots)$$

记  $B_n^2 = \sum_{k=1}^n \sigma_k^2$ , 若存在正数  $\delta$ , 使得当  $n \rightarrow \infty$  时有

$$\frac{1}{B_n^{2+\delta}} \sum_{k=1}^n E(|X_k - \mu_k|^{2+\delta}) \rightarrow 0$$

则变量

$$Z_n = \frac{\sum_{k=1}^n X_k - E\left(\sum_{k=1}^n X_k\right)}{\sqrt{D\left(\sum_{k=1}^n X_k\right)}} = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n}$$

的分布函数  $F_n(x)$  对于任意  $x$ , 满足

$$F_n(x) \xrightarrow{w} \Phi(x)$$



# 概率论与数理统计 — 参数估计 精选习题 (20 道)

## 一、点估计 (§1)

### 题型 1: 矩估计

【1.3】 设总体  $X$  在  $[a, b]$  上服从均匀分布,  $(X_1, X_2, \dots, X_n)$  为其样本, 样本均值  $\bar{X}$ , 样本方差  $S^2$ , 则  $a, b$  的矩估计  $\hat{a} = \underline{\hspace{1cm}}$ ,  $\hat{b} = \underline{\hspace{1cm}}$ 。

解 由均匀分布的数字特征结论:

$$EX = \frac{a+b}{2}, \quad DX = \frac{(b-a)^2}{12}.$$

令  $EX = \bar{X}$ ,  $DX = S^2$ , 解得

$$\hat{a} = \bar{X} - \sqrt{3}S, \quad \hat{b} = \bar{X} + \sqrt{3}S,$$

即为  $a, b$  的矩估计。

点评 因为需要估计两个参数  $a, b$ , 所以应该构造两个方程: (1) 求出期望  $EX$  用  $\bar{X}$  代替; (2) 求出方差  $DX$  用  $S^2$  代替, 也可以求出  $E(X^2)$  用  $A_2 = \frac{1}{n} \sum_{i=1}^n X_i^2$  代替, 这样结果变成

$$\hat{a} = \bar{X} - \sqrt{3B_2}, \quad \hat{b} = \bar{X} + \sqrt{3B_2},$$

其中  $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  为二阶样本中心距。

### 题型 2: 最大似然估计

【1.6】 设某种元件的使用寿命  $X$  的概率密度为

$$f(x; \theta) = \begin{cases} 2e^{-2(x-\theta)}, & x > \theta \\ 0, & x \leq \theta \end{cases}$$

其中  $\theta > 0$  为未知参数, 又设  $x_1, x_2, \dots, x_n$  是  $X$  的一组样本观测值, 求参数  $\theta$  的最大似然估计值。

分析 多数情况下, 最大似然估计值可以由似然函数的驻点求得, 但是在有些情况下, 似然函数的驻点不存在, 此时, 可以通过参数的取值范围求最大似然估计。

解 由题意知, 似然函数为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \begin{cases} 2^n e^{-2 \sum_{i=1}^n (x_i - \theta)}, & x_i > \theta \ (i = 1, 2, \dots, n) \\ 0, & \text{其他} \end{cases}$$

当  $x_i > \theta$  时,  $L(\theta) > 0$ , 两边取对数

$$\ln L(\theta) = n \ln 2 - 2 \sum_{i=1}^n (x_i - \theta),$$

因为  $\frac{d \ln L(\theta)}{d\theta} = 2n > 0$ , 所以  $L(\theta)$  单调增加。

由于  $\theta$  要满足  $\theta < x_i \ (i = 1, 2, \dots, n)$ , 所以当  $\theta$  取  $x_1, x_2, \dots, x_n$  中的最小值时,  $L(\theta)$  取最大值。

故  $\theta$  的最大似然估计值为

$$\hat{\theta} = \min(x_1, x_2, \dots, x_n).$$

【1.8】 设总体  $X$  的概率密度为

$$f(x; \theta) = \begin{cases} \theta, & 0 < x < 1 \\ 1 - \theta, & 1 \leq x < 2 \\ 0, & \text{其他} \end{cases}$$

其中  $\theta$  是未知参数 ( $0 < \theta < 1$ )。  $X_1, X_2, \dots, X_n$  为来自总体  $X$  的简单随机样本, 记  $N$  为样本值  $x_1, x_2, \dots, x_n$  中小于 1 的个数。求

(1)  $\theta$  的矩估计;

(2)  $\theta$  的最大似然估计。

解 (1) 由于

$$\begin{aligned} EX &= \int_{-\infty}^{+\infty} x f(x; \theta) dx = \int_0^1 \theta x dx + \int_1^2 (1 - \theta) x dx \\ &= \frac{1}{2} \theta + \frac{3}{2} (1 - \theta) = \frac{3}{2} - \theta. \end{aligned}$$

令  $\frac{3}{2} - \theta = \overline{X}$ , 解得  $\theta = \frac{3}{2} - \overline{X}$ , 所以参数  $\theta$  的矩估计为  $\hat{\theta} = \frac{3}{2} - \overline{X}$ 。

(2) 似然函数为

$$L(\theta) = \prod_{i=1}^n f(x_i; \theta) = \theta^N (1 - \theta)^{n-N},$$

取对数, 得

$$\ln L(\theta) = N \ln \theta + (n - N) \ln(1 - \theta),$$

两边对  $\theta$  求导, 得

$$\frac{d \ln L(\theta)}{d\theta} = \frac{N}{\theta} - \frac{n - N}{1 - \theta}.$$

令  $\frac{d \ln L(\theta)}{d\theta} = 0$ , 得  $\theta = \frac{N}{n}$ 。

所以  $\theta$  的最大似然估计为  $\hat{\theta} = \frac{N}{n}$ 。

【1.11】 (1) 设  $X_1, X_2, \dots, X_n$  是来自概率密度为

$$f(x; \theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < 1 \\ 0, & \text{其他} \end{cases}$$

的总体样本,  $\theta$  未知, 求  $U = e^{-\frac{1}{\theta}}$  的最大似然估计值。

(2) 设  $X_1, X_2, \dots, X_n$  是来自正态总体  $N(\mu, 1)$  的样本。  $\mu$  未知, 求  $\theta = P\{X > 2\}$  的最大似然估计值。

解 (1) 先求  $\theta$  的最大似然估计。似然函数为

$$\begin{aligned} L(\theta) &= \prod_{i=1}^n \theta x_i^{\theta-1} = \theta^n \left( \prod_{i=1}^n x_i \right)^{\theta-1}, \\ \ln L(\theta) &= n \ln \theta + (\theta - 1) \ln \left( \prod_{i=1}^n x_i \right). \end{aligned}$$

令

$$\frac{d \ln L(\theta)}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^n \ln x_i = 0,$$

得  $\theta$  的最大似然估计值为

$$\hat{\theta} = \frac{-n}{\sum_{i=1}^n \ln x_i}.$$

$U = e^{-\frac{1}{\theta}}$  具有单调反函数, 故由最大似然估计的不变性知  $U$  的最大似然估计值为

$$\hat{U} = e^{-\frac{1}{\hat{\theta}}} = e^{-\frac{1}{\frac{n}{\sum_{i=1}^n \ln x_i}}} = \frac{1}{e^{\frac{n}{\sum_{i=1}^n \ln x_i}}}.$$

(2) 已知  $\mu$  的最大似然估计为  $\hat{\mu} = \bar{x}$ 。而  $\theta = P\{X > 2\} = 1 - P\{X \leq 2\} = 1 - \Phi(2 - \mu)$  具有单调反函数。由最大似然估计的不变性得  $\theta = P\{X > 2\}$  的最大似然估计值为

$$\hat{\theta} = 1 - \Phi(2 - \hat{\mu}) = 1 - \Phi(2 - \bar{x}).$$

**点评** 最大似然估计的不变性: 如果  $\hat{\theta}$  是  $\theta$  的最大似然估计, 则对  $\theta$  的任一函数  $g(\theta)$ , 其最大似然估计为  $g(\hat{\theta})$ 。

## 二、估计量的评选标准 (§2)

### 题型 1: 无偏性

**【2.4】** 已知总体  $X$  的概率密度为

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

其中未知参数  $\theta > 0$ 。设  $X_1, X_2, \dots, X_n$  为取自总体  $X$  的一个样本,

(1) 求  $\theta$  的最大似然估计量;

(2) 试问该估计量是否为无偏估计量? 说明理由。

**解** (1) 设  $x_1, x_2, \dots, x_n$  为相应于样本  $X_1, X_2, \dots, X_n$  的一个样本值, 似然函数为

$$L(\theta) = \begin{cases} \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^n x_i}, & x_1, x_2, \dots, x_n > 0 \\ 0, & \text{其他} \end{cases}$$

当  $x_1, x_2, \dots, x_n > 0$  时, 有

$$\ln L(\theta) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n x_i.$$

将上式对  $\theta$  求导数并令其等于零, 得

$$\frac{d \ln L(\theta)}{d\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n x_i = 0.$$

解得  $\theta$  的最大似然估计值为

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}.$$

因此,  $\theta$  的最大似然估计量为  $\hat{\theta} = \bar{X}$ 。

(2) 由于

$$E(\hat{\theta}) = E(\bar{X}) = E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i) = E(X),$$

而  $X$  服从指数分布,  $E(X) = \theta$ , 所以  $E(\hat{\theta}) = \theta$ , 故  $\hat{\theta} = \bar{X}$  为未知参数  $\theta$  的无偏估计量。

### 题型 2: 有效性

**【2.7】** 设总体  $X$  的样本是  $X_1, X_2, \dots, X_n$ , 试证明:

(1)  $\sum_{i=1}^n a_i X_i$  ( $a_i > 0, i = 1, 2, \dots, n, \sum_{i=1}^n a_i = 1$ ) 是  $E(X)$  的无偏估计量;

(2) 在  $E(X)$  的所有形如  $\sum_{i=1}^n a_i X_i$  的无偏估计量中,  $\bar{X}$  为最有效的估计。

**分析** 证明估计量的有效性时, 需要证明不等式成立, 因此采用 Cauchy-Schwarz 公式是很有效的方法。

**证** (1) 根据无偏性估计的定义有

$$E\left(\sum_{i=1}^n a_i X_i\right) = \sum_{i=1}^n a_i E(X_i) = E(X) \sum_{i=1}^n a_i = E(X),$$

故  $\sum_{i=1}^n a_i X_i$  是  $E(X)$  的无偏估计量。

(2) 由样本均值的性质可知

$$E(\bar{X}) = \frac{1}{n} E \sum_{i=1}^n X_i = EX,$$

因此  $\bar{X}$  也是  $E(X)$  的无偏估计量。

又由 Cauchy-Schwarz 不等式

$$\left(\sum_{i=1}^n x_i y_i\right)^2 \leq \left(\sum_{i=1}^n x_i^2\right) \left(\sum_{i=1}^n y_i^2\right),$$

令  $x_i = a_i, y_i = 1$ , 则

$$\left(\sum_{i=1}^n a_i\right)^2 = 1 \leq n \sum_{i=1}^n a_i^2,$$

故

$$\begin{aligned} D(\bar{X}) &= \frac{1}{n} D(X) = \frac{1}{n} D(X) \left(\sum_{i=1}^n a_i\right)^2 \\ &\leq D(X) \left(\sum_{i=1}^n a_i^2\right) = \sum_{i=1}^n D(a_i X_i) = D\left(\sum_{i=1}^n a_i X_i\right). \end{aligned}$$

证毕。

**点评** 本题也可以用导数知识求  $\sum_{i=1}^n a_i^2$  的最小值, 从而得出结论。另外本题的结论可以记住, 并当作定理应用。

### 题型 3: 相合性 (一致性)

**【2.9】** 设  $X_1, X_2, \dots, X_n$  是取自正态总体  $N(\mu, \sigma^2)$  的样本, 证明  $S^2$  是  $\sigma^2$  的一致估计。

**证法一** 由大数定律

$$\lim_{n \rightarrow \infty} P\left\{\left|\frac{1}{n} \sum_{i=1}^n X_i - \mu\right| < \varepsilon\right\} = 1,$$

所以  $\bar{X}$  是  $\mu$  的一致估计。

同理, 因  $X_1^2, X_2^2, \dots, X_n^2$  也独立同分布, 故  $\frac{1}{n} \sum_{i=1}^n X_i^2$  是  $E(X^2)$  的一致估计。

而

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2 = \frac{n}{n-1} \left(\frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2\right),$$



故当  $n \rightarrow \infty$  时,

$$\frac{n}{n-1} \rightarrow 1, \quad \frac{1}{n} \sum_{i=1}^n X_i^2 \xrightarrow{P} E(X^2), \quad \overline{X}^2 \xrightarrow{P} \mu^2,$$

即  $S^2 \xrightarrow{P} E(X^2) - \mu^2 = \sigma^2$ 。则  $S^2$  是  $\sigma^2$  的一致估计。

**证法二** 因为  $E(S^2) = \sigma^2$ ,  $D(S^2) = \frac{2\sigma^4}{n-1}$ , 故

$$\lim_{n \rightarrow \infty} E(S^2) = \sigma^2, \quad \lim_{n \rightarrow \infty} D(S^2) = \lim_{n \rightarrow \infty} \frac{2\sigma^4}{n-1} = 0,$$

由定理可知,  $S^2$  是  $\sigma^2$  的一致估计。

### 三、区间估计 (§3)

#### 题型 1: 正态总体均值 $\mu$ 的区间估计

**【3.4】** 从总体  $X_1 \sim N(\mu_1, 25)$  中取出一容量为  $n_1 = 10$  的样本, 其样本均值  $\bar{X}_1 = 19.8$ ; 从总体  $X_2 \sim N(\mu_2, 36)$  中取出容量为  $n_2 = 12$  的样本, 其样本均值  $\bar{X}_2 = 24.0$ , 已知两个样本之间相互独立, 求  $\mu_1 - \mu_2$  的 0.90 置信区间。

**解** 这是  $\sigma_1^2, \sigma_2^2$  都为已知时, 求均值差的区间估计问题。

由于  $1 - \alpha = 0.90$ , 故  $\frac{\alpha}{2} = 0.05$ ,  $u_{\frac{\alpha}{2}} = 1.645$ ,

又因为  $n_1 = 10, n_2 = 12, \sigma_1^2 = 25, \sigma_2^2 = 36$ , 所以

$$\begin{aligned} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= \sqrt{\frac{25}{10} + \frac{36}{12}} = \sqrt{5.5} = 2.345, \\ \bar{X}_1 - \bar{X}_2 - u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= 19.8 - 24.0 - 1.645 \times 2.345 \\ &= -4.2 - 3.858 = -8.06, \\ \bar{X}_1 - \bar{X}_2 + u_{\frac{\alpha}{2}} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} &= -4.2 + 3.858 = -0.34. \end{aligned}$$

因此, 所求的  $\mu_1 - \mu_2$  的 0.90 置信区间为  $[-8.06, -0.34]$ 。

#### 题型 2: 正态总体方差 $\sigma^2$ 的区间估计

**【3.9】** 两个正态总体  $N(\mu_1, \sigma_1^2)$ ,  $N(\mu_2, \sigma_2^2)$  的参数均未知, 分别从两个总体中抽取容量为 25 和 15 的两个独立样本, 测得样本方差分别为 6.38, 5.15, 求  $\frac{\sigma_1^2}{\sigma_2^2}$  的置信区间 ( $\alpha = 0.10$ )。

**解**  $n_1 = 25, S_1^2 = 6.38, n_2 = 15, S_2^2 = 5.15, \alpha = 0.10, \frac{\alpha}{2} = 0.05$ ,

查  $F$  分布表得

$$F_{0.05}(24, 14) = 2.35, \quad F_{0.05}(14, 24) = 2.13,$$

而

$$\frac{S_1^2}{S_2^2} = \frac{6.38}{5.15} \approx 1.24.$$

由置信区间公式得  $\frac{\sigma_1^2}{\sigma_2^2}$  的 90% 置信区间为

$$\left( \frac{S_1^2}{S_2^2} F_{0.95}(14, 24), \frac{S_1^2}{S_2^2} F_{0.05}(14, 24) \right) = (0.528, 2.641).$$

**【3.10】** 设  $X_1, X_2, \dots, X_n$  是来自分布  $N(\mu, \sigma^2)$  的样本,  $\mu$  已知,  $\sigma$  未知。

(1) 验证  $\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$ 。利用这一结果构造  $\sigma^2$  的置信水平为  $1 - \alpha$  的置信区间。

(2) 设  $\mu = 6.5$ , 且有样本值 7.5, 2.0, 12.1, 8.8, 9.4, 7.3, 1.9, 2.8, 7.0, 7.3。试求  $\sigma$  的置信水平为 0.95 的置信区间。

**解** (1) 因  $X_i \sim N(\mu, \sigma^2)$ , 故

$$\frac{X_i - \mu}{\sigma} \sim N(0, 1), \quad i = 1, 2, \dots, n.$$

由  $\frac{X_1 - \mu}{\sigma}, \frac{X_2 - \mu}{\sigma}, \dots, \frac{X_n - \mu}{\sigma}$  相互独立, 得

$$\sum_{i=1}^n \left( \frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n).$$

于是有

$$P \left\{ \chi_{1-\frac{\alpha}{2}}^2(n) < \sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} < \chi_{\frac{\alpha}{2}}^2(n) \right\} = 1 - \alpha,$$

即有

$$P \left\{ \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)} < \sigma^2 < \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right\} = 1 - \alpha.$$

得  $\sigma^2$  的置信水平为  $1 - \alpha$  的置信区间为

$$\left( \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\frac{\alpha}{2}}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\frac{\alpha}{2}}^2(n)} \right).$$

(2) 现在  $n = 10, \mu = 6.5, 1 - \alpha = 0.95, \alpha = 0.05$ , 由样本值经计算得  $\sum_{i=1}^{10} (X_i - \mu)^2 = 102.69$ , 查表知  $\chi_{0.025}^2(10) = 20.483, \chi_{0.975}^2(10) = 3.247$ 。

于是  $\sigma^2$  的置信水平为 0.95 的置信区间为 (5.013, 31.626)。 $\sigma$  的置信水平为 0.95 的置信区间为 (2.239, 5.624)。

### 题型 3: 单侧置信限

**【3.11】** 从一批电子元件中随机地抽取 10 只作寿命试验, 其寿命 (以小时计) 如下:

1498 1499 1501 1503 1500 1499 1499 1498 1500 1503

设寿命服从正态分布, 试求其平均寿命的 95% 置信下限。

**解** 本例中, 总体  $X \sim N(\mu, \sigma^2)$ , 且方差  $\sigma^2$  未知, 故应使用  $t$  分布。因

$$\frac{(\bar{X} - \mu) \cdot \sqrt{n}}{S} \sim t(n-1),$$

此时要求

$$P \left\{ \frac{(\bar{X} - \mu) \sqrt{n}}{S} < t_{\alpha}(n-1) \right\} = 1 - \alpha,$$

于是得  $\mu$  的置信度  $1 - \alpha$  单侧置信区间为

$$(\bar{X} - t_{\alpha}(n-1) \cdot \frac{S}{\sqrt{n}}, +\infty).$$

对于给定的数据, 具体计算如下:

$$\begin{aligned}\bar{x} &= \frac{1}{10}(1498 + 1499 + 1501 + 1503 + 1500 + 1499 + 1499 + 1498 + 1500 + 1503) = 1500, \\ s^2 &= \frac{1}{10-1}[(1498-1500)^2 + (1499-1500)^2 + (1501-1500)^2 \\ &\quad + (1503-1500)^2 + (1500-1500)^2 + (1499-1500)^2 + (1499-1500)^2 \\ &\quad + (1498-1500)^2 + (1500-1500)^2 + (1503-1500)^2] \\ &= \frac{10}{3},\end{aligned}$$

又

$$1 - \alpha = 0.95, \quad \alpha = 0.05, \quad t_{0.05}(10-1) = 1.8331,$$

故寿命均值的 95% 单侧置信区间为

$$(1500 - \frac{1}{\sqrt{10}} \times \sqrt{\frac{10}{3}} \times 1.8331, +\infty) \approx (1498.942, +\infty).$$

1498.942 就是所求的置信下限。

## 四、综合提高 (§4)

### 题型 1: 关于点估计

**【4.3】** 设  $X_1, X_2, \dots, X_n$  为总体的一个样本, 求下列各总体的密度函数或分布律中的未知参数的矩估计量和最大似然估计量。

$$(1) f(x) = \begin{cases} \theta c^\theta x^{-(\theta+1)}, & x > c, \\ 0, & \text{其他} \end{cases}, \text{ 其中 } c > 0 \text{ 为已知, } \theta > 1, \theta \text{ 为未知参数。}$$

$$(2) f(x) = \begin{cases} \sqrt{\theta} x^{\sqrt{\theta}-1}, & 0 \leq x \leq 1, \\ 0, & \text{其他} \end{cases}, \text{ 其中 } \theta > 0, \theta \text{ 为未知参数。}$$

$$(3) P\{X = x\} = C_m^x p^x (1-p)^{m-x}, \quad x = 0, 1, 2, \dots, m, \quad 0 < p < 1, \quad p \text{ 为未知参数。}$$

解 (1)

$$E(X) = \int_c^\infty x \theta c^\theta x^{-(\theta+1)} dx = \theta c^\theta \int_c^\infty x^{-\theta} dx = \theta c^\theta \left[ \frac{x^{-\theta+1}}{-\theta+1} \right]_c^{+\infty} = \frac{\theta c}{\theta-1}.$$

令  $\frac{\theta c}{\theta-1} = \bar{X}$ , 解得  $\hat{\theta} = \frac{\bar{X}}{\bar{X}-c}$ , 即为  $\theta$  的矩估计量。

似然函数为

$$L(\theta) = \prod_{i=1}^n \theta \cdot c^\theta x_i^{-(\theta+1)} = \theta^n c^{n\theta} \prod_{i=1}^n x_i^{-(\theta+1)},$$

$$\ln L(\theta) = n \ln \theta + n\theta \ln c - (\theta+1) \sum_{i=1}^n \ln x_i,$$

$$\text{令 } \frac{d}{d\theta} \ln L(\theta) = \frac{n}{\theta} + n \ln c - \sum_{i=1}^n \ln x_i = 0,$$

$$\text{解得 } \theta \text{ 的最大似然估计量 } \hat{\theta} = \frac{n}{\sum_{i=1}^n \ln X_i - n \ln c}.$$

(2)

$$EX = \int_0^1 x \sqrt{\theta} x^{\sqrt{\theta}-1} dx = \int_0^1 \sqrt{\theta} x^{\sqrt{\theta}} dx = \frac{\sqrt{\theta}}{\sqrt{\theta}+1} x^{\sqrt{\theta}+1} \Big|_0^1 = \frac{\sqrt{\theta}}{\sqrt{\theta}+1}.$$

$$\text{令 } \frac{\sqrt{\theta}}{\sqrt{\theta}+1} = \bar{X}, \text{ 解得 } \hat{\theta} = \left( \frac{\bar{X}}{\bar{X}-1} \right)^2, \text{ 即为 } \theta \text{ 的矩估计量。}$$

似然函数为:

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

$$\ln L(\theta) = \frac{n}{2} \ln \theta + (\sqrt{\theta} - 1) \sum_{i=1}^n \ln x_i,$$

$$\frac{d \ln L(\theta)}{d\theta} = \frac{n}{2\theta} + \frac{\sum_{i=1}^n \ln x_i}{2\sqrt{\theta}} = 0,$$

其最大似然估计值为  $\hat{\theta} = \frac{n^2}{(\sum_{i=1}^n \ln x_i)^2}$ ,  $\hat{\theta} = \frac{n^2}{(\sum_{i=1}^n \ln X_i)^2}$  即为  $\theta$  的最大似然估计量。

(3) 因为  $X \sim B(m, p)$ , 所以

$$E(X) = mp,$$

所以  $mp = \overline{X}$ ,  $\hat{p} = \frac{\overline{X}}{m}$  为  $p$  的矩估计。

似然函数为：

$$L(p) = \prod_{i=1}^n C_m^{x_i} p^{x_i} (1-p)^{m-x_i} = p^{\sum_{i=1}^n x_i} (1-p)^{nm - \sum_{i=1}^n x_i} \prod_{i=1}^n C_m^{x_i},$$

$$\ln L(p) = \sum_{i=1}^n x_i \ln p + (nm - \sum_{i=1}^n x_i) \ln(1-p) + \sum_{i=1}^n \ln C_m^{x_i},$$

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^n x_i}{p} + \frac{nm - \sum_{i=1}^n x_i}{1-p} (-1) = 0,$$

则  $\hat{p} = \frac{\overline{x}}{m}$  为  $p$  的最大似然估计值,  $\hat{p} = \frac{\overline{X}}{m}$  为  $p$  的最大似然估计量。

**【4.4】** 设总体  $X$  的概率分布为

| $X$ | 0          | 1                   | 2          | 3           |
|-----|------------|---------------------|------------|-------------|
| $P$ | $\theta^2$ | $2\theta(1-\theta)$ | $\theta^2$ | $1-2\theta$ |

其中  $0 < \theta < \frac{1}{2}$  是未知参数, 利用总体  $X$  的如下样本值

3, 1, 3, 0, 3, 1, 2, 3

求  $\theta$  的矩估计值和最大似然估计值。

**分析** 矩估计用基本求解方法即可。对于最大似然估计, 若似然函数出现多个驻点应该根据题意选择。

**解** 由离散型随机变量的期望公式

$$\begin{aligned} EX &= 0 \times \theta^2 + 1 \times 2\theta(1-\theta) + 2 \times \theta^2 + 3 \times (1-2\theta) \\ &= 2\theta - 2\theta^2 + 2\theta^2 + 3 - 6\theta = 3 - 4\theta, \end{aligned}$$

令  $EX = \overline{X}$ , 而由样本观测值得

$$\overline{X} = \frac{1}{8}(3+1+3+0+3+1+2+3) = \frac{1}{8} \times 16 = 2,$$

所以  $\theta$  的矩估计值为

$$\hat{\theta} = \frac{1}{4}(3 - \overline{X}) = \frac{1}{4}(3 - 2) = \frac{1}{4}.$$

根据题意, 似然函数为

$$L(\theta) = 4\theta^6(1-\theta)^2(1-2\theta)^4,$$

两边取对数可得

$$\ln L(\theta) = \ln 4 + 6 \ln \theta + 2 \ln(1 - \theta) + 4 \ln(1 - 2\theta),$$

$$\frac{d \ln L(\theta)}{d\theta} = \frac{6}{\theta} - \frac{2}{1 - \theta} - \frac{8}{1 - 2\theta} = \frac{24\theta^2 - 28\theta + 6}{\theta(1 - \theta)(1 - 2\theta)},$$

令  $\frac{d \ln L(\theta)}{d\theta} = 0$ , 得  $12\theta^2 - 14\theta + 3 = 0$ , 解之得  $\theta = \frac{7-\sqrt{13}}{12}$  或  $\frac{7+\sqrt{13}}{12}$ 。

因为已知  $0 < \theta < \frac{1}{2}$ , 故  $\theta = \frac{7-\sqrt{13}}{12}$ 。

因此  $\theta$  的最大似然估计值为  $\hat{\theta} = \frac{7-\sqrt{13}}{12}$ 。

**【4.9】** 设某种电子器件的寿命（以小时计） $T$  服从双参数的指数分布，其概率密度为

$$f(t) = \begin{cases} \frac{1}{\theta} e^{-\frac{t-c}{\theta}}, & t \geq c \\ 0, & \text{其他} \end{cases}$$

其中  $c, \theta$  ( $c, \theta > 0$ ) 为未知参数，自一批这种器件中随机地取  $n$  件进行寿命试验。设它们的失效时间依次为  $x_1 \leq x_2 \leq \cdots \leq x_n$ 。

(1) 求  $\theta$  与  $c$  的最大似然估计；

(2) 求  $\theta$  与  $c$  的矩估计。

**解** (1) 似然函数为

$$L(\theta, c) = \begin{cases} \frac{1}{\theta^n} e^{-\frac{1}{\theta} \sum_{i=1}^n (x_i - c)}, & x_i \geq c, i = 1, 2, \dots, n \\ 0, & \text{其他} \end{cases}$$

对数似然函数为：

$$\ln L(\theta, c) = -n \ln \theta - \frac{1}{\theta} \sum_{i=1}^n (x_i - c),$$

对数似然方程为：

$$\frac{\partial \ln L(\theta, c)}{\partial c} = \frac{n}{\theta} > 0, \quad \frac{\partial \ln L(\theta, c)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n (x_i - c) = 0,$$

故  $\ln L(\theta, c)$  关于  $c$  单调增加，故  $\hat{c} = x_1$ 。

由  $\frac{\partial \ln L(\theta, c)}{\partial \theta} = -\frac{n}{\theta} + \frac{1}{\theta^2} \sum_{i=1}^n (x_i - c) = 0$ , 得  $\theta$  的最大似然估计值为  $\hat{\theta} = \bar{x} - x_1$ 。

(2)

$$EX = \int_c^{+\infty} \frac{x}{\theta} e^{-\frac{x-c}{\theta}} dx = \theta + c,$$

$$E(X^2) = \int_c^{+\infty} \frac{x^2}{\theta} e^{-\frac{x-c}{\theta}} dx = \theta^2 + (\theta + c)^2,$$

令  $EX = \bar{x}$ ,  $E(X^2) = \frac{1}{n} \sum_{i=1}^n x_i^2$ , 那么  $\theta$  和  $c$  的矩估计为

$$\hat{\theta} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad \hat{c} = \bar{x} - \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}.$$

**【4.12】** 设总体  $X$  服从  $[0, \theta]$  上的均匀分布， $\theta$  未知 ( $\theta > 0$ )， $X_1, X_2, X_3$  是取自  $X$  的一个样本。

(1) 试证  $\hat{\theta}_1 = \frac{4}{3} \max_{1 \leq i \leq 3} X_i$ ,  $\hat{\theta}_2 = 4 \min_{1 \leq i \leq 3} X_i$  都是  $\theta$  的无偏估计；

(2) 上述两个估计中哪个更有效？

**证** (1) 设  $F(x)$  是  $X$  的分布函数，则

$$F(x) = \begin{cases} 1, & x > \theta \\ \frac{x}{\theta}, & 0 \leq x \leq \theta \\ 0, & x < 0 \end{cases}$$

$$Y = \max_{1 \leq i \leq 3} X_i, \quad Z = \min_{1 \leq i \leq 3} X_i,$$

$$F_Y(x) = [F(x)]^3, \quad f_Y(x, \theta) = \begin{cases} 3 \left(\frac{x}{\theta}\right)^2 \cdot \frac{1}{\theta}, & 0 \leq x \leq \theta \\ 0, & \text{其他} \end{cases}$$

故

$$EY = \frac{3}{\theta^3} \int_0^\theta x^3 dx = \frac{3}{4}\theta, \quad \text{则 } E(\hat{\theta}_1) = E\left(\frac{4}{3} \max X_i\right) = \theta,$$

同理

$$EZ = \frac{3}{\theta^3} \int_0^\theta x(\theta - x)^2 dx = \frac{1}{4}\theta, \quad \text{则 } E(\hat{\theta}_2) = E(4 \min X_i) = \theta.$$

(2)

$$DY = EY^2 - (EY)^2 = \frac{3}{\theta} \int_0^\theta x^2 \left(\frac{x}{\theta}\right)^2 dx - \left(\frac{3}{4}\theta\right)^2 = \frac{3}{80}\theta^2,$$

故

$$D(\hat{\theta}_1) = D\left(\frac{4}{3} \max X_i\right) = \frac{16}{9}DY = \frac{1}{15}\theta^2,$$

同理

$$DZ = \frac{3}{80}\theta^2, \quad \text{故 } D(\hat{\theta}_2) = D(4 \min X_i) = 16DZ = \frac{3}{5}\theta^2 > D(\hat{\theta}_1),$$

故  $\hat{\theta}_1$  更有效。

**【4.18】** 设总体  $X \sim U(\theta, 2\theta)$ , 其中  $\theta > 0$  是未知参数, 又  $X_1, X_2, \dots, X_n$  为取自该总体的样本, 证明  $\hat{\theta} = \frac{2}{3}\bar{X}$  是  $\theta$  的无偏估计和相合估计。

证

$$E(\hat{\theta}) = \frac{2}{3}E(\bar{X}) = \frac{2}{3}E(X) = \frac{2}{3} \cdot \frac{\theta + 2\theta}{2} = \theta,$$

故  $\hat{\theta}$  是  $\theta$  的无偏估计。

$$D(\hat{\theta}) = \frac{4}{9}D(\bar{X}) = \frac{4}{9} \cdot \frac{D(X)}{n} = \frac{4}{9n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{27n}.$$

当  $n \rightarrow \infty$  时,  $D(\hat{\theta}) \rightarrow 0$ , 故  $\hat{\theta}$  是  $\theta$  的相合估计。

## 题型 2: 关于区间估计

**【4.22】** 假如 0.50, 1.25, 0.80, 2.00 是来自总体  $X$  的简单随机样本值, 已知  $Y = \ln X$  服从正态分布  $N(\mu, 1)$ 。

(1) 求  $X$  的数学期望  $EX$  (记  $EX$  为  $b$ );

(2) 求  $\mu$  的置信度为 0.95 的置信区间;

(3) 利用上述结果求  $b$  的置信度为 0.95 的置信区间。

**解** (1) 由题意知  $Y$  的概率密度为

$$f(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2}}.$$

又由  $Y = \ln X$ , 得  $X = e^Y$ , 故

$$b = EX = Ee^Y = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^y e^{-\frac{(y-\mu)^2}{2}} dy = e^{\mu+\frac{1}{2}} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[y-(\mu+1)]^2} dy = e^{\mu+\frac{1}{2}}.$$

(2)  $\mu$  的置信区间公式为  $\left(\bar{Y} - \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}, \bar{Y} + \frac{\sigma}{\sqrt{n}} u_{\frac{\alpha}{2}}\right)$ 。

由  $1 - \alpha = 0.95$ , 查表得  $u_{\frac{\alpha}{2}} = 1.96$ 。

代入  $\sigma = 1$ ,  $n = 4$ ,  $\bar{y} = \frac{1}{4}(\ln 0.5 + \ln 1.25 + \ln 0.8 + \ln 2) = 0$ ,

得  $\left(-\frac{1}{2} \times 1.96, \frac{1}{2} \times 1.96\right)$ 。

故  $\mu$  的置信度为 0.95 的置信区间为  $(-0.98, 0.98)$ 。

(3) 由 (1) 可知,  $b = EX = e^{\mu+\frac{1}{2}}$ , 又由 (2) 知  $\mu$  的置信区间为  $(-0.98, 0.98)$ ,

因为  $e^x$  为严格增函数, 所以  $b$  的置信区间为  $(e^{-0.98+\frac{1}{2}}, e^{0.98+\frac{1}{2}})$ ,

即为  $(e^{-0.48}, e^{1.48})$ 。

**【4.27】** 随机地从  $A$  批导线中抽取 4 根, 又从  $B$  批导线中抽取 5 根, 测得电阻 (单位:  $\Omega$ ) 为

$A$  批导线: 0.143 0.142 0.143 0.137

$B$  批导线: 0.140 0.142 0.136 0.138 0.140

设测定数据分别来自分布  $N(\mu_1, \sigma^2), N(\mu_2, \sigma^2)$ , 且两样本相互独立, 又  $\mu_1, \mu_2, \sigma^2$  均为未知, 试求  $\mu_1 - \mu_2$  的置信度为 0.95 的置信区间。

**解**  $\mu_1 - \mu_2$  的置信区间为

$$\left[ \bar{X} - \bar{Y} - t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \bar{X} - \bar{Y} + t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \right],$$

在这里

$$\bar{X} = \frac{1}{4}(0.143 + 0.142 + 0.143 + 0.137) = 0.1413,$$

$$\bar{Y} = \frac{1}{5}(0.140 + 0.142 + 0.136 + 0.138 + 0.140) = 0.1392,$$

$$n_1 = 4, n_2 = 5, n_1 + n_2 - 2 = 7; \quad 1 - \alpha = 0.95, \alpha = 0.05, \frac{\alpha}{2} = 0.025.$$

查表得  $t_{\frac{\alpha}{2}}(7) = 2.3646$ ,

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = 6.509 \times 10^{-6}, \quad s_w = \sqrt{6.509 \times 10^{-6}} = 2.551 \times 10^{-3}.$$

将这些值代入上区间得  $(-0.002, 0.006)$ 。

### 题型 3: 与置信度、样本容量及区间长度有关

**【4.35】** 设样本  $(X_1, X_2, \dots, X_n)$  为总体  $X \sim N(\mu, \sigma^2)$  的样本, 其中  $\mu, \sigma^2$  为未知参数, 设随机变量  $L$  是关于  $\mu$  的置信度为  $1 - \alpha$  的置信区间的长度, 求  $E(L^2)$ 。

**解** 取  $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ , 则  $T \sim t(n-1)$ 。

由  $P\{|T| < t_{\frac{\alpha}{2}}(n-1)\} = 1 - \alpha$  可得  $\mu$  的  $1 - \alpha$  置信区间为

$$\left(\overline{X}-t_{\frac{\alpha}{2}}(n-1) \cdot \frac{S}{\sqrt{n}}, \quad \overline{X}+t_{\frac{\alpha}{2}}(n-1) \cdot \frac{S}{\sqrt{n}}\right),$$

易见区间长度  $L=2t_{\frac{\alpha}{2}}(n-1) \cdot \frac{S}{\sqrt{n}}$ 。

故

$$E(L^2)=E\left[4t_{\frac{\alpha}{2}}^2(n-1) \cdot \frac{S^2}{n}\right]=\frac{4}{n}t_{\frac{\alpha}{2}}^2(n-1) \cdot \sigma^2.$$

**【4.36】** 假定到某地旅游的一个游客的消费额  $X$  服从正态分布  $N(\mu, \sigma^2)$ ，且  $\sigma=500$ ， $\mu$  未知。要对平均消费额  $\mu$  进行估计，使这个估计的绝对误差小于 50 元，且置信度不小于 0.95，问至少需要随机调查多少个游客？

**解** 设  $n$  为需要调查的游客人数，要使

$$P\{|\overline{X}-\mu|<50\} \geqslant 0.95, \quad \text{即} \quad P\left\{\frac{|\overline{X}-\mu|}{\frac{\sigma}{\sqrt{n}}}<\frac{50}{\frac{\sigma}{\sqrt{n}}}\right\} \geqslant 0.95,$$

因为  $\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=U \sim N(0,1)$ ，由  $P\{|U|<u_{\frac{\alpha}{2}}\}=1-\alpha=0.95$ ，其中  $\alpha=0.05$ ，得

$$\frac{50}{\frac{\sigma}{\sqrt{n}}} \geqslant u_{\frac{0.05}{2}} \Rightarrow \sqrt{n} \geqslant \frac{1.96\sigma}{50} \Rightarrow n \geqslant \left(\frac{1.96\sigma}{50}\right)^2 = \left(\frac{1.96 \times 500}{50}\right)^2 = 384.16.$$

即至少需要随机调查 **385** 人，就有不小于 0.95 的把握，使得用调查所得的  $\overline{x}$  去估计平均消费额的真值  $\mu$ ，其绝对误差小于 50 元。

## 题目分布总结

| 板块      | 题型              | 题号                        | 数量        |
|---------|-----------------|---------------------------|-----------|
| §1 点估计  | 矩估计             | 1.3                       | 1         |
|         | 最大似然估计          | 1.6, 1.8, 1.11            | 3         |
| §2 评选标准 | 无偏性             | 2.4                       | 1         |
|         | 有效性             | 2.7                       | 1         |
|         | 相合性             | 2.9                       | 1         |
| §3 区间估计 | $\mu$ 区间估计      | 3.4                       | 1         |
|         | $\sigma^2$ 区间估计 | 3.9, 3.10                 | 2         |
|         | 单侧置信限           | 3.11                      | 1         |
| §4 综合提高 | 点估计             | 4.3, 4.4, 4.9, 4.12, 4.18 | 5         |
|         | 区间估计            | 4.22, 4.27                | 2         |
|         | 容量/长度           | 4.35, 4.36                | 2         |
| 合计      |                 |                           | <b>20</b> |

来源：《概率论与数理统计习题精选精解》第七章

## 假设检验

总体来说，假设检验这个模块的知识点不如前期的概率论和后续的随机过程，在数学上的难点较少，所以复习起来也相对比较容易。



# 假设检验的基本步骤

- 1. 假设 $H_0$ 和 $H_1$ 。
- 2. 根据第六章的内容选取检验统计量。
- 3. 写出拒绝域。
- 4. 查表，计算，下结论。

## $H_0$ 的小妙招

- 1. 等号肯定在  $H_0$  里面，所以对于一类认为是是否改变的题目，无脑把  $\mu = a$  放在原假设里即可。
- 2. 尽量控制第一类错误。单侧假设检验中，如果不论假设  $\mu \geq$  和  $\leq$ ，最后得出结论的判断是一致的時候，我们优先选择通过拒绝原假设得出结论的那种解法，因为  $\alpha$  可控且已知。

这里给出一道例题：

假设彩色显像管的寿命服从正态分布，并且已知标准差  $\sigma = 40$  小时，随机抽取100只，算得其平均寿命为10000小时，试求：

(2) 若显像管的平均寿命不小于10100小时被认定为合格，试在显著性水平  $\alpha = 0.005$  下检验这批显像管是否合格？

已知： $\Phi(2.576) = 0.995$ ,  $\Phi(2.33) = 0.99$

### 方法一

用  $H_0 : \mu \leq 10100$

$$\begin{cases} H_0 : \mu \leq 10100 & \text{(不合格)} \\ H_1 : \mu > 10100 & \text{(合格)} \end{cases}$$

单侧检验， $\alpha = 0.005$ ，右侧分位数  $z_\alpha = 2.576$ ，拒绝域： $Z > 2.576$

计算Z统计量：

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10000 - 10100}{40/\sqrt{100}} = \frac{-100}{4} = -25$$

判定：

$Z = -25 < 2.576$ ，不在拒绝域内，**不能拒绝原假设  $H_0$** ，即在  $\alpha = 0.005$  下认为这批显像管不合格。

### 方法二

用  $H_0 : \mu \geq 10100$

$$\begin{cases} H_0 : \mu \geq 10100 \\ H_1 : \mu < 10100 \end{cases}$$

左侧检验，拒绝域： $Z < -2.576$

统计量仍为  $Z = -25 < -2.576$ ，落在拒绝域，拒绝  $H_0$ ，结论同样是不合格。

这种情况下，由于第一类错误的概率  $\alpha$  是小的并且已知、可控，所以我们**优先选择第二种方法**，通过**拒接原假设进行假设检验**，我们犯第一类错误的概率不超过  $\alpha$ 。

# 总结

最关键的还是根据语义设定  $H_0$ 。比如，如果问你某个统计量是否有变动，你就假设没变动；问你因为题干给的一组新数据，是否抬高了原有的平均值，你就原假设平均值没变高（单侧假设检验）；如果问你与过去是否有显著差别，你就假设没显著差别……总之，默认以前得出来的规律是普遍规律，更可信，放在原假设；题干给的新的检验数据，只是一组特别的数据，相对不可信一些，放在备择假设。

# 其它类型的假设检验

给出一些典型，自己根据课本的例题梳理即可。

- 1. 非正态总体的样本假设检验：课本讲解了二项分布（本质是中心极限定理）和泊松分布（第六章课后习题25题的公式）
- 2. 二项数据检验和成对数据检验的区分（两组数据是否有关联？）；
- 3. 分布假设检验（皮尔逊定理， $\chi^2$ 拟合检验法以及两种情形）

# 来点猎奇的题目

## 第一题

设  $X_1, X_2, \dots, X_{16}$  是来自总体  $N(\mu, 4)$  的简单随机样本，考虑假设检验问题：

$$H_0 : \mu \leq 10, \quad H_1 : \mu > 10$$

若拒绝域为  $W = \{\overline{X} > 11\}$ ，求当  $\mu = 11.5$  时，该检验犯第二类错误的概率。

## 基础概念

第二类错误（取伪错误）： $H_1$  实际成立，但我们接受了  $H_0$ ，概率记为  $\beta$ 。  
本题真实均值  $\mu = 11.5$ （满足  $H_1 : \mu > 10$ ），接受  $H_0$  等价于  $\overline{X} \leq 11$ ，因此：

$$\beta = P(\overline{X} \leq 11 \mid \mu = 11.5)$$

## 步骤1：确定样本均值分布

总体  $X \sim N(\mu, \sigma^2)$ ,  $\sigma^2 = 4 \Rightarrow \sigma = 2$ ，样本量  $n = 16$

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) = N\left(\mu, \frac{4}{16}\right) = N(\mu, 0.25)$$

标准化：

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), \quad \frac{\sigma}{\sqrt{n}} = \frac{2}{\sqrt{16}} = \frac{2}{4} = 0.5$$

## 步骤2：代入 $\mu = 11.5$ 计算概率

$$\begin{aligned} \beta &= P(\overline{X} \leq 11 \mid \mu = 11.5) \\ &= P\left(\frac{\overline{X} - 11.5}{0.5} \leq \frac{11 - 11.5}{0.5}\right) \\ &= P\left(Z \leq \frac{-0.5}{0.5}\right) \\ &= P(Z \leq -1) \\ &= \Phi(-1) \end{aligned}$$

步骤3：化简结果

由标准正态分布对称性  $\Phi(-x) = 1 - \Phi(x)$ ：

$$\beta = \Phi(-1) = 1 - \Phi(1) \approx 1 - 0.8413 = 0.1587$$

结论

$\mu = 11.5$  时，犯第二类错误的概率为  $\Phi(-1) \approx 0.1587$

第二题

购入一批气球，出厂气球存在破损概率。商家承诺每个气球的破损率  $p < 1\%$ ，取显著性水平  $\alpha = 0.05$ ，求至少需要抽检多少只气球才能检验商家承诺是否真实。

建立假设

破损率为  $p$ ，商家宣称  $p < 0.01$ ，根据上面的原则，为了确保我们承人商家承诺是否真实，最好的办法还是通过**拒绝“气球的破损率>1%”自然得出商家的承诺真实**。所以将“商家承诺成立”设为备择假设，原假设包含等号：

$$\begin{cases} H_0 : p \geq 0.01 & (\text{商家说谎，破损率达标不上限}) \\ H_1 : p < 0.01 & (\text{商家承诺为真，破损率低于1\%}) \end{cases}$$

左侧单侧二项分布检验，显著性水平  $\alpha = 0.05$ 。

设抽检  $n$  只，破损气球数量  $X \sim B(n, p)$ 。

拒绝域： $X \leq c$ （破损数极少，有充分证据认为破损率低于1%，拒绝  $H_0$ ）。

控制第一类错误： $p = 0.01$  时（ $H_0$  边界）

$$P(X \leq c \mid p = 0.01) \leq \alpha = 0.05$$

**临界简化：最小可区分情况  $c = 0$ （抽检中0只破损才判定商家合格）**

若  $c = 0$ ，则：

$$P(X = 0 \mid p = 0.01) = (1 - 0.01)^n \leq 0.05$$

解不等式：

$$0.99^n \leq 0.05$$

两边取自然对数（ $\ln x$  单调递增，不等号方向不变）：

$$n \cdot \ln 0.99 \leq \ln 0.05$$

$\ln 0.99 < 0$ ，两边除以负数，不等号反转：

$$\begin{aligned} n &\geq \frac{\ln 0.05}{\ln 0.99} \\ \ln 0.05 &\approx -2.9957, \quad \ln 0.99 \approx -0.01005 \\ n &\geq \frac{-2.9957}{-0.01005} \approx 298.07 \end{aligned}$$

$n$  为整数，向上取整  $n = 299$ 。

得出结论

在显著性水平  $\alpha = 0.05$ ，仅当抽检全部无破损时采信商家承诺的前提下，**至少需要检查299只气球**。

# 第十一章 随机过程的基本知识

**随机过程定义** 设  $\Omega$  为某试验的样本空间,  $T$  为一给定的集合, 若对每个  $t \in T$ ,  $X(t, \omega)$  是定义在  $\Omega$  上的随机变量, 则称  $\{X(t, \omega), t \in T\}$  为随机过程, 其中,  $T$  称为参数集。

随机过程可以认为是随机变量的每一个取值在时间轴上的延展后组成的一个“随机变量”, 这个理解也暗示了随机过程中最容易误解的点:

$\omega$  一旦选定, 在不同时刻的  $\omega$  应当是一样的。

想象你在追一部连载科幻剧 (整个随机过程):

- $\Omega$  是所有可能编出的剧本集合。
- $\omega_1$  代表“剧本 A”(比如主角拯救了世界),  $\omega_2$  代表“剧本 B”(主角毁灭了世界)。

当你按下了播放键 (即上帝掷了一次骰子, 选定了  $\omega_1$ ), 你就只能在同一个宇宙 ( $\omega_1$ ) 里, 随着时间  $t$  从第 1 集看到第 10 集。你绝对不能在第 2 秒时用  $\omega_1$  的剧情, 到了第 5 秒突然跳进  $\omega_2$  的宇宙。如果不同时刻的  $\omega$  可以不同, 那就不叫“一个过程的演变”, 而是每一秒都在疯狂穿越平行宇宙, 时间前后的相关性就崩塌了, 后面自相关函数、自协方差函数以及互相关函数, 互协方差函数的定义也就没有意义了。

从本质上讲, 随机过程  $X(\cdot, \cdot)$  是关于  $t \in T, \omega \in \Omega$  的二元映射, 即  $X(\cdot, \cdot): T \times \Omega \rightarrow \mathbb{R}$ 。

一方面, 当固定  $t \in T$  时,  $X(t, \cdot)$  是一个随机变量, 因此, 当  $t$  取遍  $T$  中所有元时,  $\{X(t, \cdot), t \in T\}$  代表了一族随机变量; 另一方面, 当进行一次随机试验, 得到样本点  $\omega \in \Omega$  时, 相应地得到了一个  $t$  的实函数  $X(\cdot, \omega)$ , 称之为随机过程的样本函数 (样本曲线, 路径, 轨道), 因而, 与试验的所有可能结果相对应,  $\{X(\cdot, \omega), \omega \in \Omega\}$  代表一族样本函数。此外, 与  $t \in T, \omega \in \Omega$  对应的  $X(t, \omega)$  是一个实数, 称为随机过程的一个状态, 集合

$$S \stackrel{\text{def}}{=} \{X(t, \omega) \mid t \in T, \omega \in \Omega\}$$

称为随机过程的状态空间 (值域, 相空间)。

随机过程  $\{X(t, \omega), t \in T\}$  通常简记为  $\{X(t), t \in T\}$ 。为了简明起见, 我们也用记号  $X_T$  表示随机过程  $\{X(t), t \in T\}$ , 即

$$X_T \stackrel{\text{def}}{=} \{X(t), t \in T\}$$

单个随机变量可由分布函数  $F(x) = P\{X \leq x\}$  完整刻画; 随机过程  $X_T$  是一族存在关联的随机变量, 需用有限维分布函数族描述其全部概率特性:

1. 一维分布函数: 对任意  $t \in T$ ,

$$F_t(x) \stackrel{\text{def}}{=} P\{X(t) \leq x\}$$

2. 二维分布函数: 对任意  $t_1, t_2 \in T$ ,

$$F_{t_1, t_2}(x_1, x_2) \stackrel{\text{def}}{=} P\{X(t_1) \leq x_1, X(t_2) \leq x_2\}$$

仅依靠一、二维分布无法完整描述随机过程。一般地, 任取有限个参数  $t_1, t_2, \dots, t_n \in T$ , 定义  $n$  维分布函数:

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) \stackrel{\text{def}}{=} P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$$

随机过程所有一、二、...、 $n$  维分布函数全体

$$\{F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) \mid t_1, t_2, \dots, t_n \in T, n = 1, 2, \dots\}$$

称为  $X_T$  的有限维分布函数族。

有限维分布函数族的**两条性质**

1. 对称性  
对下标  $1, 2, \dots, n$  的任意排列  $(j_1, j_2, \dots, j_n)$ :

$$F_{t_1, t_2, \dots, t_n}(x_1, x_2, \dots, x_n) = F_{t_{j_1}, t_{j_2}, \dots, t_{j_n}}(x_{j_1}, x_{j_2}, \dots, x_{j_n})$$

- 含义: 交换参数与对应变量顺序, 分布函数不变。
2. 相容性 (可减性)  
若  $m < n$ , 将高维分布多余变量取  $+\infty$ , 即可得到低维分布:

$$F_{t_1, t_2, \dots, t_m}(x_1, x_2, \dots, x_m) = F_{t_1, t_2, \dots, t_m, t_{m+1}, \dots, t_n}(x_1, x_2, \dots, x_m, \infty, \dots, \infty)$$

结论：有限维分布函数族**完全刻画**随机过程全部概率性质。

如何理解这个完全刻画呢？由于随机变无限维随机变量分布函数在数学上难以给出，所以我们便可以采用有限维进行代替。只要我们让维数不断趋近无穷，那么就可以解决无限维的问题。比如“这个随机过程在整个时间段  $[0, 1]$  内，最大值从来没有超过 5 的概率是多少？”

例1

已知随机过程  $X_T$  仅有两条样本轨道：

$$X(t, \omega_1) = k \cos t, \quad X(t, \omega_2) = k \sin t, \quad -\infty < t < +\infty, \quad k > 0$$

概率： $P(\omega_1) = \frac{1}{3}, P(\omega_2) = \frac{2}{3}$ 。

求：一维分布函数  $F_0(x)$ 、二维分布函数  $F_{0, \pi/2}(x_1, x_2)$ 。

## 随机变量的数字特征

有时获得有限的分布函数也是困难的，所以我们退而求其次，选择关注局部的数字特征。

设随机过程  $X_T = \{X(t), t \in T\}$

1. 均值函数

$$m(t) \stackrel{\text{def}}{=} E[X(t)] = \int_{-\infty}^{+\infty} x \, dF_t(x)$$

含义： $t$  时刻状态的统计平均，样本曲线的摆动中心。

2. 方差函数 & 标准差函数

方差：

$$D(t) \stackrel{\text{def}}{=} D[X(t)] = E\left\{[X(t) - m(t)]^2\right\}$$

简记  $V(t), \sigma^2(t)$ ；

标准差： $\sigma(t) = \sqrt{D(t)}$ ，刻画样本在  $t$  时刻偏离均值的程度。

3. 自相关函数

$$R_X(t_1, t_2) \stackrel{\text{def}}{=} E[X(t_1)X(t_2)], \quad t_1, t_2 \in T$$

简写  $R(t_1, t_2)$ 。

注意，自相关函数可简称为相关函数

4. 自协方差函数

$$C_X(t_1, t_2) \stackrel{\text{def}}{=} \text{Cov}(X(t_1), X(t_2))$$

简写  $C(t_1, t_2)$ ，绝对值越大，两时刻随机变量线性相关性越强； $C = 0$  则不相关。

**核心关系式**

$$\begin{aligned} C(t_1, t_2) &= R(t_1, t_2) - m(t_1)m(t_2) \\ D(t) &= C(t, t) = R(t, t) - m^2(t) \end{aligned}$$

**均值函数、自相关函数是随机过程最基础数字特征。**

5. 两个过程的互特征 ( $X_T, Y_T$  同  $\Omega$ 、同  $T$ )

- 互相关函数：

$$R_{XY}(t_1, t_2) \stackrel{\text{def}}{=} E[X(t_1)Y(t_2)]$$

- 互协方差函数：

$$\begin{aligned} C_{XY}(t_1, t_2) &\stackrel{\text{def}}{=} \text{Cov}(X(t_1), Y(t_2)) \\ &= E\{[X(t_1) - m_X(t_1)][Y(t_2) - m_Y(t_2)]\} \end{aligned}$$

$m_X(t), m_Y(t)$  分别为两过程均值函数。

注意这里xy必须要求是定义在同一个样本空间以及同一个参数集上的。这也就意味着刚才我们说到的 $\omega$ 这里也是一样的。无论自字头，or互字头。

## 例2

随机过程  $X_T = \{A \cos(\lambda t + \Phi), -\infty < t < +\infty\}$ ,  $\lambda > 0$  为常数;  $A, \Phi$  相互独立,  $A \sim U(0, 1)$ ,  $\Phi \sim U(0, 2\pi)$ , 求均值函数  $m(t)$ 、自相关函数  $R(t_1, t_2)$ 。

### 1. 均值函数 $m(t)$

$$\begin{aligned} m(t) &= E[A \cos(\lambda t + \Phi)] \\ &= E(A) \cdot E[\cos(\lambda t + \Phi)] \quad (A, \Phi \text{ 独立}) \\ &= \left( \int_0^1 a \, da \right) \cdot \left( \int_0^{2\pi} \cos(\lambda t + \varphi) \cdot \frac{1}{2\pi} d\varphi \right) \\ &= 0 \end{aligned}$$

### 2. 自相关函数 $R(t_1, t_2)$

$$\begin{aligned} R(t_1, t_2) &= E[A^2 \cos(\lambda t_1 + \Phi) \cos(\lambda t_2 + \Phi)] \\ &= E(A^2) \cdot E[\cos(\lambda t_1 + \Phi) \cos(\lambda t_2 + \Phi)] \\ &= \int_0^1 a^2 da \cdot \int_0^{2\pi} \frac{1}{2\pi} \cdot \frac{1}{2} [\cos(\lambda(t_2 - t_1)) + \cos(\lambda(t_1 + t_2) + 2\varphi)] d\varphi \\ &= \frac{\cos[\lambda(t_2 - t_1)]}{6} \end{aligned}$$

## 例3

两随机过程：

$$\begin{cases} X_T = \{A \cos(\lambda t + \Theta), -\infty < t < +\infty\} \\ Y_T = \{B \cos(\lambda t + \Theta - \varphi), -\infty < t < +\infty\} \end{cases}$$

$A, B, \lambda, \varphi$  为常数,  $\Theta \sim U(0, 2\pi)$ , 求互相关函数  $R_{XY}(t_1, t_2)$ 。

$$\begin{aligned} R_{XY}(t_1, t_2) &= E[A \cos(\lambda t_1 + \Theta) \cdot B \cos(\lambda t_2 + \Theta - \varphi)] \\ &= AB \int_0^{2\pi} \cos(\lambda t_1 + \theta) \cos(\lambda t_2 + \theta - \varphi) \cdot \frac{1}{2\pi} d\theta \\ &= \frac{AB}{2\pi} \int_0^{2\pi} \frac{1}{2} [\cos(\lambda(t_2 - t_1) - \varphi) + \cos(\lambda(t_1 + t_2) - \varphi + 2\theta)] d\theta \\ &= \frac{AB}{2} \cos[\lambda(t_1 - t_2) - \varphi] \end{aligned}$$

正好复习一下积化和差

## 两随机过程：不相关、相互独立、和过程数字特征

设  $X_T = \{X(t), t \in T\}$ ,  $Y_T = \{Y(t), t \in T\}$  同样本空间、同参数集  $T$

### 1. 不相关定义

若  $\forall t_1, t_2 \in T$ , 互协方差恒为0:

$$C_{XY}(t_1, t_2) \equiv 0$$

称  $X_T, Y_T$  **不相关**。

2. 相互独立定义

$\forall n, m \in \mathbb{N}^*$ , 任意  $t_1, \dots, t_n \in T, t'_1, \dots, t'_m \in T$ ,

随机向量  $(X(t_1), \dots, X(t_n))$  与  $(Y(t'_1), \dots, Y(t'_m))$  相互独立, 则两过程**相互独立**。

3. 关系

独立 (二阶矩存在)  $\implies$  不相关;

不相关  $\nRightarrow$  独立。

4. 和过程  $Z_T = \{X(t) + Y(t), t \in T\}$

- 均值函数

$$m_Z(t) = m_X(t) + m_Y(t)$$

- 自相关函数

$$R_Z(t_1, t_2) = R_X(t_1, t_2) + R_{XY}(t_1, t_2) + R_{YX}(t_1, t_2) + R_Y(t_1, t_2)$$

- 特殊简化条件

若  $m_X(t) \equiv 0, m_Y(t) \equiv 0$ , 且  $X_T, Y_T$  不相关:

$$R_Z(t_1, t_2) = R_X(t_1, t_2) + R_Y(t_1, t_2)$$

这里是由于  $C(t_1, t_2) = R(t_1, t_2) - m(t_1)m(t_2)$

## 随机过程的基本类型

### 随机过程按参数集、状态空间分类

依据参数集  $T \subset \mathbb{R}$ 、状态空间  $S$  是否可数, 分为4类:

- 可数参数、可数状态
- 可数参数、不可数状态
- 不可数参数、可数状态
- 不可数参数、不可数状态

- $T$  可数 (可列): **离散参数随机过程 / 随机序列 (时间序列)**
- $T$  为  $\mathbb{R}$  上区间: **连续参数随机过程**

### 按过程的性质特点分类

1. 二阶矩过程

**定义:** 设  $X_T = \{X(t), t \in T\}$ , 若  $\forall t \in T$ , 二阶原点矩  $E[X^2(t)]$  存在, 则称  $X_T$  为二阶矩过程。

**均值函数** 由柯西 - 施瓦茨不等式:

$$|E[X(t)]| = |E[X(t) \cdot 1]| \leq \sqrt{E[X^2(t)] \cdot E[1^2]} = \sqrt{E[X^2(t)]}$$

已知  $E[X^2(t)] < +\infty$ , 因此  $|E[X(t)]|$  有界,  $E[X(t)]$  必然存在。

**自相关函数**  $R(t_1, t_2) = E[X(t_1)X(t_2)]$  存在再次用柯西 - 施瓦茨不等式:

$$|E[X(t_1)X(t_2)]| \leq \sqrt{E[X^2(t_1)] \cdot E[X^2(t_2)]}$$

$E[X^2(t_1)], E[X^2(t_2)]$  均有限, 故  $E[X(t_1)X(t_2)]$  有界, 自相关函数存在。

- 二阶矩过程中有一些比较重要的子类: 它们是正态过程、正交增量过程和 (宽) 平稳过程。

这本书中讨论的“平稳过程”几乎都是指宽平稳过程。

证明这三种随机过程, 核心是验证它们各自的**数学定义**或**等价条件**。以下是简明、严谨的证明思路:

### 1. 证明正态过程 (高斯过程)

**核心目标:** 证明该过程在任意有限个时间点上的状态, 构成一个多维联合正态分布。

**常见证明方法:**

- 线性组合法 (最常用):** 根据多维正态分布的性质, 只需证明对任意  $n$  个时间点  $t_1, t_2, \dots, t_n$  以及任意实常数  $a_1, a_2, \dots, a_n$ , 其线性组合  $\sum_{i=1}^n a_i X(t_i)$  均服从一维正态分布。
- 特征函数法:** 计算  $(X(t_1), X(t_2), \dots, X(t_n))$  的联合特征函数, 验证其是否具有多维正态分布特征函数的标准形式。

- **线性变换性质：**如果已知  $X(t)$  是由另一个已知的正态过程（如维纳过程 / 布朗运动）通过线性操作（如加法、常数乘法、极限、微积分）衍生而来的，可以直接利用“正态过程的线性变换依然是正态过程”这一性质来证明。

## 2. 证明正交增量过程

**核心目标：**证明在任意互不重叠的时间区间上，过程的增量彼此正交。对于均值为 0 的过程，正交等价于不相关（协方差为 0）。

**常见证明方法：**

任取四个时间点  $t_1 < t_2 \leq t_3 < t_4$ ，考察两个不重叠区间的增量  $\Delta X_1 = X(t_2) - X(t_1)$  和  $\Delta X_2 = X(t_4) - X(t_3)$ 。你需要计算这两个增量乘积的数学期望，并证明其等于 0：

$$E[(X(t_2) - X(t_1))(X(t_4) - X(t_3))] = 0$$

(注：如果该过程的数学期望不恒为 0，则需要证明这两个增量的协方差为 0，即它们不相关。)

## 3. 证明泊松过程

**核心目标：**证明一个计数过程  $N(t)$  满足泊松过程的三大公理。

**常见证明方法：**

必须同时验证以下三个条件：

1. **初始值：**验证  $N(0) = 0$ （起点事件发生次数为 0）。
2. **独立增量：**证明对于任意时间序列  $0 \leq t_1 < t_2 < \dots < t_n$ ，增量  $N(t_2) - N(t_1), N(t_3) - N(t_2), \dots, N(t_n) - N(t_{n-1})$  是相互独立的随机变量。
3. **增量分布（二选一验证）：**

- **定义法：**证明在任意长度为  $\tau$  的时间区间内，增量  $N(t + \tau) - N(t)$  服从参数为  $\lambda\tau$  的泊松分布：

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

- **微小区间法（极限定理）：**证明该过程具有平稳增量，且满足：

$$P(N(t + \Delta t) - N(t) = 1) = \lambda\Delta t + o(\Delta t)$$

$$P(N(t + \Delta t) - N(t) \geq 2) = o(\Delta t)$$

1. 设在时间区间  $(0, t]$  来到某商店门口的顾客数  $N(t)$  是强度为  $\lambda$  的泊松过程，每个来到商店门口的顾客进入商店的概率为  $p$ ，不进入商店就离去的概率是  $1 - p$ ，各个顾客进入商店与否相互独立，令  $X(t)$  为  $(0, t]$  内进入商店的顾客数，证明  $\{X(t), t \geq 0\}$  是强度为  $\lambda p$  的泊松过程。

**解析与证明：**

要证明  $\{X(t), t \geq 0\}$  是强度为  $\lambda p$  的泊松过程，我们需要验证泊松过程的定义条件。由于原过程  $N(t)$  的平稳独立增量性以及顾客选择的独立性，过程  $X(t)$  显然满足  $X(0) = 0$  以及独立增量性。

核心在于证明其**增量分布**服从参数为  $\lambda p\tau$  的泊松分布。

任取一时间区间  $(t, t + \tau]$ ，设该区间内来到商店的总顾客数为  $N = N(t + \tau) - N(t)$ ，实际进入商店的顾客数为  $X = X(t + \tau) - X(t)$ 。已知  $N(t)$  是强度为  $\lambda$  的泊松过程，故：

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

在已知总来到人数为  $n$  的条件下，每个顾客独立以概率  $p$  进入商店，因此进入商店的人数  $X$  服从二项分布  $B(n, p)$ ：

$$P(X = k \mid N = n) = \binom{n}{k} p^k (1 - p)^{n-k}, \quad k \leq n$$

(若  $n < k$ ，该概率为 0)

利用全概率公式，求  $X$  的边缘分布：



$$\begin{aligned}
 P(X=k) &= \sum_{n=k}^{\infty} P(X=k | N=n)P(N=n) \\
 &= \sum_{n=k}^{\infty} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \frac{(\lambda\tau)^n e^{-\lambda\tau}}{n!}
 \end{aligned}$$

消去  $n!$  并重排项：

$$= \frac{(\lambda\tau p)^k e^{-\lambda\tau}}{k!} \sum_{n=k}^{\infty} \frac{(\lambda\tau(1-p))^{n-k}}{(n-k)!}$$

令  $m = n - k$ ，则求和部分变为指数函数的泰勒展开：

$$P(X=k) = \frac{(\lambda p \tau)^k e^{-\lambda \tau}}{k!} \sum_{m=0}^{\infty} \frac{(\lambda \tau (1-p))^m}{m!} = \frac{(\lambda p \tau)^k e^{-\lambda \tau}}{k!} e^{\lambda \tau (1-p)}$$

合并指数项：

$$P(X=k) = \frac{(\lambda p \tau)^k}{k!} e^{-\lambda \tau + \lambda \tau - \lambda p \tau} = \frac{(\lambda p \tau)^k e^{-\lambda p \tau}}{k!}$$

**结论：**  $X(t)$  的增量服从参数为  $\lambda p \tau$  的泊松分布。结合初始条件和独立增量性， $\{X(t), t \geq 0\}$  是强度为  $\lambda p$  的泊松过程。得证。

2. 设  $\{N(t), t \geq 0\}$  是强度为  $\lambda$  的泊松过程， $S_n (n \geq 1)$  表示事件  $A$  第  $n$  次出现的发生时刻，试求  $(S_1, S_2)$  的联合概率密度。

**解析与计算：**

在泊松过程中，事件发生的**时间间隔**相互独立，且均服从参数为  $\lambda$  的指数分布。

令  $T_1 = S_1$  表示第一次事件发生的时间。

令  $T_2 = S_2 - S_1$  表示第一次与第二次事件发生的时间间隔。

已知  $T_1$  和  $T_2$  独立，且  $T_1 \sim \text{Exp}(\lambda), T_2 \sim \text{Exp}(\lambda)$ 。它们的边缘概率密度函数分别为：

$$f_{T_1}(t_1) = \lambda e^{-\lambda t_1} \quad (t_1 > 0)$$

$$f_{T_2}(t_2) = \lambda e^{-\lambda t_2} \quad (t_2 > 0)$$

由于  $T_1, T_2$  独立，它们的联合概率密度为：

$$f_{T_1, T_2}(t_1, t_2) = f_{T_1}(t_1) f_{T_2}(t_2) = \lambda^2 e^{-\lambda(t_1+t_2)} \quad (t_1 > 0, t_2 > 0)$$

现在进行变量代换。已知：

$$S_1 = T_1$$

$$S_2 = T_1 + T_2$$

反解得到：

$$T_1 = S_1$$

$$T_2 = S_2 - S_1$$

计算雅可比行列式 (Jacobian)：

$$J = \det \begin{pmatrix} \frac{\partial T_1}{\partial S_1} & \frac{\partial T_1}{\partial S_2} \\ \frac{\partial T_2}{\partial S_1} & \frac{\partial T_2}{\partial S_2} \end{pmatrix} = \det \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = 1$$

根据多维随机变量变换定理， $(S_1, S_2)$  的联合概率密度为：

$$\begin{aligned}
 f_{S_1, S_2}(s_1, s_2) &= f_{T_1, T_2}(s_1, s_2 - s_1) \cdot |J| \\
 &= \lambda^2 e^{-\lambda(s_1 + s_2 - s_1)} \cdot 1 \\
 &= \lambda^2 e^{-\lambda s_2}
 \end{aligned}$$

**确定定义域：**

由  $t_1 > 0$  得  $s_1 > 0$ ；由  $t_2 > 0$  得  $s_2 - s_1 > 0$ ，即  $s_2 > s_1$ 。

综合可得定义域为  $0 < s_1 < s_2$ 。

结论：

$(S_1, S_2)$  的联合概率密度函数为：

$$f_{S_1, S_2}(s_1, s_2) = \begin{cases} \lambda^2 e^{-\lambda s_2}, & 0 < s_1 < s_2 \\ 0, & \text{其他} \end{cases}$$

3. 设保险公司在  $t_i$  时发生理赔的金额为  $X_i$ ，各  $X_i$  相互独立同分布，其均值为  $\mu$ ，方差为  $\sigma^2$ 。若在  $[0, t]$  内发生理赔的次数  $N(t) (t \geq 0)$  是强度为  $\lambda$  的泊松过程，并且与  $\{X_i, i = 1, 2, \dots\}$  独立，试求保险公司在  $[0, t]$  内将要支付的总理赔额  $Y(t)$  的均值与方差。

解析与解答：

这是典型的**复合泊松过程**。总理赔额可以表示为随机和：

$$Y(t) = \sum_{i=1}^{N(t)} X_i$$

1. 求均值  $E[Y(t)]$ ：

利用全期望公式（或 Wald 等式），在给定  $N(t) = n$  的条件下：

$$E[Y(t) | N(t) = n] = E\left[\sum_{i=1}^n X_i\right] = nE[X_i] = n\mu$$

因此，条件期望  $E[Y(t) | N(t)] = N(t)\mu$ 。

再求无条件期望：

$$E[Y(t)] = E[E[Y(t) | N(t)]] = E[N(t)\mu] = \mu E[N(t)]$$

已知  $N(t)$  服从参数为  $\lambda t$  的泊松分布，故  $E[N(t)] = \lambda t$ 。

$$E[Y(t)] = \lambda t \mu$$

2. 求方差  $Var(Y(t))$ ：

利用全方差公式： $Var(Y) = E[Var(Y | N)] + Var(E[Y | N])$

• 计算  $E[Var(Y | N)]$ ：

已知  $X_i$  独立，故  $Var(Y(t) | N(t) = n) = Var(\sum_{i=1}^n X_i) = n\sigma^2$ 。

因此  $Var(Y(t) | N(t)) = N(t)\sigma^2$ 。

取期望： $E[Var(Y(t) | N(t))] = E[N(t)\sigma^2] = \sigma^2 \lambda t$ 。

• 计算  $Var(E[Y | N])$ ：

上文已得  $E[Y(t) | N(t)] = N(t)\mu$ 。

取方差： $Var(E[Y(t) | N(t)]) = Var(N(t)\mu) = \mu^2 Var(N(t))$ 。

由于泊松分布方差等于均值， $Var(N(t)) = \lambda t$ ，故该项为  $\mu^2 \lambda t$ 。

相加得到总方差：

$$Var(Y(t)) = \sigma^2 \lambda t + \mu^2 \lambda t = \lambda t (\sigma^2 + \mu^2)$$

(注： $\sigma^2 + \mu^2 = E[X^2]$ ，即二阶原点矩)

4. 设  $\{N(t), t \geq 0\}$  是强度为  $\lambda$  的泊松过程，随机变量  $T$  服从参数为  $\mu$  的指数分布，且与  $\{N(t), t \geq 0\}$  相互独立，试证明  $N(T)$  服从几何分布：

$$P\{N(T) = k\} = \frac{\mu}{\lambda + \mu} \left( \frac{\lambda}{\lambda + \mu} \right)^k, \quad k = 0, 1, 2, \dots$$

解析与证明：

利用全概率公式，对连续型随机变量  $T$  的概率密度进行积分。

已知  $T$  的概率密度函数为  $f_T(t) = \mu e^{-\mu t} (t > 0)$ 。

在给定  $T = t$  的条件下， $N(t)$  服从参数为  $\lambda t$  的泊松分布：

$$P\{N(T) = k | T = t\} = \frac{(\lambda t)^k e^{-\lambda t}}{k!}$$

由于  $T$  与  $N(t)$  独立，无条件概率为：

$$P\{N(T) = k\} = \int_0^\infty P\{N(T) = k | T = t\} f_T(t) dt$$

$$= \int_0^{\infty} \frac{(\lambda t)^k e^{-\lambda t}}{k!} \cdot \mu e^{-\mu t} dt$$

提取常数并合并指数项：

$$= \frac{\lambda^k \mu}{k!} \int_0^{\infty} t^k e^{-(\lambda+\mu)t} dt$$

利用伽马函数的积分公式  $\int_0^{\infty} x^n e^{-ax} dx = \frac{n!}{a^{n+1}}$ ，令  $x = t, n = k, a = \lambda + \mu$ ：

$$\int_0^{\infty} t^k e^{-(\lambda+\mu)t} dt = \frac{k!}{(\lambda + \mu)^{k+1}}$$

代入原式：

$$P\{N(T) = k\} = \frac{\lambda^k \mu}{k!} \cdot \frac{k!}{(\lambda + \mu)^{k+1}}$$

约去  $k!$  并整理：

$$P\{N(T) = k\} = \frac{\mu \lambda^k}{(\lambda + \mu)(\lambda + \mu)^k} = \frac{\mu}{\lambda + \mu} \left( \frac{\lambda}{\lambda + \mu} \right)^k$$

得证。

5. 设随机过程  $X(t) = X \cos(\omega t) + Y \sin(\omega t)$ ，其中  $\omega$  为常数， $X$  和  $Y$  是相互独立的正态随机变量，且  $E(X) = E(Y) = 0$ ， $Var(X) = Var(Y) = \sigma^2$ 。求  $X(t)$  的均值函数与自相关函数，并判断其是否为宽平稳过程。

**解析**

均值函数：

$$m_X(t) = E[X \cos(\omega t) + Y \sin(\omega t)] = E[X] \cos(\omega t) + E[Y] \sin(\omega t) = 0$$

自相关函数：

$$R_X(s, t) = E[X(s)X(t)] = E[(X \cos(\omega s) + Y \sin(\omega s))(X \cos(\omega t) + Y \sin(\omega t))]$$

由于  $X, Y$  独立且均值为 0，交叉项  $E[XY] = 0$ ，故：

$$\begin{aligned} R_X(s, t) &= E[X^2] \cos(\omega s) \cos(\omega t) + E[Y^2] \sin(\omega s) \sin(\omega t) \\ &= \sigma^2 [\cos(\omega s) \cos(\omega t) + \sin(\omega s) \sin(\omega t)] = \sigma^2 \cos(\omega(s - t)) \end{aligned}$$

判定：二阶矩  $E[X^2(t)] = \sigma^2 < \infty$ ；均值函数恒为 0；自相关函数仅依赖于时间差  $\tau = s - t$ 。因此， $X(t)$  是宽平稳过程。

6. 设随机过程  $X(t) = A \sin(\omega t + \Theta)$ ，其中  $A$  与  $\omega$  为实常数，随机相位  $\Theta$  服从  $(0, 2\pi)$  上的均匀分布。求  $X(t)$  的均值函数与自相关函数，并判断是否平稳。

**解析**

因为  $\Theta \sim U(0, 2\pi)$ ，其概率密度为  $f(\theta) = \frac{1}{2\pi}$ 。

均值函数：

$$m_X(t) = E[A \sin(\omega t + \Theta)] = \int_0^{2\pi} A \sin(\omega t + \theta) \frac{1}{2\pi} d\theta = 0$$

自相关函数：

$$R_X(s, t) = E[A^2 \sin(\omega s + \Theta) \sin(\omega t + \Theta)]$$

利用积化和差公式  $\sin \alpha \sin \beta = \frac{1}{2} [\cos(\alpha - \beta) - \cos(\alpha + \beta)]$ ：

$$R_X(s, t) = \frac{A^2}{2} E[\cos(\omega(s - t)) - \cos(\omega(s + t) + 2\Theta)]$$

其中  $E[\cos(\omega(s + t) + 2\Theta)] = \int_0^{2\pi} \cos(\omega(s + t) + 2\theta) \frac{1}{2\pi} d\theta = 0$ 。

$$R_X(s, t) = \frac{A^2}{2} \cos(\omega(s - t))$$

判定：均值为常数 0，自相关函数仅与时间差  $\tau = s - t$  有关，故  $X(t)$  为宽平稳过程。

7. 设随机过程  $Y(t) = At + B$ ，其中  $A$  与  $B$  是相互独立的随机变量，且  $E(A) = E(B) = 0$ ， $Var(A) = 1$ ， $Var(B) = 2$ 。求  $Y(t)$  的协方差函数，并判断其是否为宽平稳过程。

**解析**

均值函数：

$$m_Y(t) = E[At + B] = tE[A] + E[B] = 0$$

自相关函数（因均值为 0，与协方差函数相等）：

$$R_Y(s, t) = E[(As + B)(At + B)] = E[A^2 st + ABs + ABt + B^2]$$

由于  $A, B$  独立且均值为 0， $E[AB] = E[A]E[B] = 0$ 。

又  $E[A^2] = Var(A) + E[A]^2 = 1$ ,  $E[B^2] = Var(B) + E[B]^2 = 2$ 。  
故  $R_Y(s, t) = stE[A^2] + E[B^2] = st + 2$

判定：虽然均值为常数 0，但自相关函数  $st + 2$  依赖于时间点  $s$  和  $t$  的具体取值，而不仅仅是时间差  $s - t$ 。因此， $Y(t)$  不是宽平稳过程。

8. 设相互独立的正态随机变量  $X_1, X_2 \sim N(0, \sigma^2)$ ，构造随机过程  $Z(t) = X_1 \cos t + X_2 \sin t + X_1 X_2$ 。求  $Z(t)$  的均值函数与自相关函数。

### 解析

均值函数：

$$m_Z(t) = E[X_1 \cos t + X_2 \sin t + X_1 X_2] = \cos t E[X_1] + \sin t E[X_2] + E[X_1] E[X_2] = 0$$

自相关函数：

$$R_Z(s, t) = E[Z(s)Z(t)] = E[(X_1 \cos s + X_2 \sin s + X_1 X_2)(X_1 \cos t + X_2 \sin t + X_1 X_2)]$$

展开并利用奇数阶中心矩为 0（即  $E[X_1^3] = 0, E[X_1^2 X_2] = 0$  等）：

$$R_Z(s, t) = E[X_1^2 \cos s \cos t + X_2^2 \sin s \sin t + X_1^2 X_2^2]$$

代入  $E[X_1^2] = E[X_2^2] = \sigma^2$ ，且  $E[X_1^2 X_2^2] = E[X_1^2] E[X_2^2] = \sigma^4$ ：

$$R_Z(s, t) = \sigma^2(\cos s \cos t + \sin s \sin t) + \sigma^4 = \sigma^2 \cos(s - t) + \sigma^4$$

9. 某十字路口的车辆到达服从泊松过程，总到达率为  $\lambda = 10$  辆/分钟。已知其中 60% 为小轿车，40% 为卡车，且车辆类型相互独立。求：

(1) 在任意 1 分钟内，恰好到达 3 辆小轿车和 2 辆卡车的概率。

(2) 从现在起，直到第一辆卡车到达的平均等待时间是多少秒？

### 解析

根据泊松过程的分流性质，小轿车和卡车的到达分别构成独立的泊松过程。

小轿车到达率  $\lambda_1 = 10 \times 0.6 = 6$  辆/分钟；卡车到达率  $\lambda_2 = 10 \times 0.4 = 4$  辆/分钟。

(1) 设 1 分钟内到达小轿车数量为  $X$ ，卡车数量为  $Y$ ，则  $X \sim P(6), Y \sim P(4)$ 。

$$P(X = 3, Y = 2) = P(X = 3)P(Y = 2) = \frac{6^3 e^{-6}}{3!} \times \frac{4^2 e^{-4}}{2!} = \frac{216}{6} e^{-6} \times \frac{16}{2} e^{-4} = 288 e^{-10}$$

(2) 第一辆卡车到达的时间  $T$  服从参数为  $\lambda_2 = 4$ （次/分钟）的指数分布。

平均等待时间  $E[T] = \frac{1}{\lambda_2} = \frac{1}{4}$  分钟 = 15 秒。

10. 某服务器接收请求的过程是强度为  $\lambda = 2$  次/秒的泊松过程。假定从  $t = 0$  时刻开始观察，求：

(1) 第 3 个请求到达的预期时间。

(2) 第 3 个请求到达时间超过 2 秒的概率。

### 解析

(1) 在泊松过程中，第  $n$  个事件到达的时刻  $S_n$  服从伽马分布（或厄兰分布），其期望为  $E[S_n] = \frac{n}{\lambda}$ 。

预期时间  $E[S_3] = \frac{3}{2} = 1.5$  秒。

(2) “第 3 个请求到达时间超过 2 秒”等价于“在  $[0, 2]$  秒内到达的请求总数少于 3 个”。

令  $N(2)$  为 2 秒内到达的请求数， $N(2) \sim P(\lambda t) = P(4)$ 。

$$P(S_3 > 2) = P(N(2) < 3) = P(N(2) = 0) + P(N(2) = 1) + P(N(2) = 2)$$

$$= \frac{4^0 e^{-4}}{0!} + \frac{4^1 e^{-4}}{1!} + \frac{4^2 e^{-4}}{2!} = e^{-4}(1 + 4 + 8) = 13e^{-4}$$

11. 一条装配线上的产品出现缺陷服从强度为  $\lambda = 5$  个/小时的泊松过程。已知每个缺陷是“严重缺陷”的概率为 0.1，是“轻微缺陷”的概率为 0.9。假设刚才发现了一个严重缺陷，求在下一个严重缺陷出现之前，恰好连续发现 4 个轻微缺陷的概率。

### 解析

这里利用泊松过程事件序列的性质。总事件流可以看作一系列独立的伯努利试验。

每一个到来的缺陷，属于“轻微”的概率为  $p = 0.9$ ，属于“严重”的概率为  $1 - p = 0.1$ 。

“两个严重缺陷之间恰好有 4 个轻微缺陷”意味着：接下来连续 4 个都是轻微缺陷，第 5 个是严重缺陷。此次数服从几何分布。

概率  $P = p^4(1 - p) = 0.9^4 \times 0.1 = 0.06561$ 。

12. 保险公司收到理赔请求的数量  $\{N(t), t \geq 0\}$  是强度为  $\lambda = 5$  次/天的泊松过程。每次理赔的金额  $X_i$  相互独立，且服从  $[10, 50]$ （单位：千元）的均匀分布。理赔金额与请求数量相互独立。求在  $t = 10$  天内，保险公司需要支付的总理赔额  $Y(t)$  的均值与方差。

### 解析

这是一个复合泊松过程， $Y(t) = \sum_{i=1}^{N(t)} X_i$ 。

10 天内请求次数  $N(10) \sim P(10 \times 5) = P(50)$ ，故  $E[N(10)] = 50, Var(N(10)) = 50$ 。

理赔金额  $X_i \sim U(10, 50)$ ，则  $E[X_i] = \frac{10+50}{2} = 30$ ；方差  $Var(X_i) = \frac{(50-10)^2}{12} = \frac{1600}{12} = \frac{400}{3}$ 。

二阶原点矩  $E[X_i^2] = Var(X_i) + (E[X_i])^2 = \frac{400}{3} + 900 = \frac{3100}{3}$ 。

总金额均值：

$$E[Y(10)] = E[N(10)] \cdot E[X_i] = 50 \times 30 = 1500 \text{ (千元)}。$$

总金额方差（根据复合泊松过程方差公式）：

$$Var(Y(10)) = \lambda t \cdot E[X_i^2] = 50 \times \frac{3100}{3} = \frac{155000}{3}。$$

13. 设  $\{W(t), t \geq 0\}$  是标准布朗运动（维纳过程）。求增量  $W(5) - W(2)$  的概率分布，并计算概率  $P(W(5) - W(2) < 0)$ 。

**解析**

标准布朗运动具有平稳独立增量，且  $W(0) = 0, W(t) \sim N(0, t)$ 。

对于任意  $0 \leq s < t$ ，增量  $W(t) - W(s) \sim N(0, t - s)$ 。

因此， $W(5) - W(2) \sim N(0, 5 - 2) = N(0, 3)$ 。

由于该正态分布均值为 0，关于纵轴对称，因此：

$$P(W(5) - W(2) < 0) = \Phi(0) = 0.5。$$

14. 设  $\{W(t), t \geq 0\}$  是标准布朗运动。计算协方差  $Cov(W(2), W(5))$ ，以及随机变量  $X = W(5) - 2W(2)$  的方差。

**解析**

布朗运动的协方差函数公式为： $Cov(W(s), W(t)) = \min(s, t)$ 。

$$\text{故 } Cov(W(2), W(5)) = \min(2, 5) = 2。$$

求  $X$  的方差展开：

$$Var(W(5) - 2W(2)) = Var(W(5)) + (-2)^2 Var(W(2)) + 2(1)(-2)Cov(W(5), W(2))$$

$$\text{代入 } Var(W(5)) = 5, Var(W(2)) = 2:$$

$$= 5 + 4 \times 2 - 4 \times 2 = 5。$$

15. 设  $\{W(t), t \geq 0\}$  是标准布朗运动。已知在  $t = 1$  时刻  $W(1) = 2$ ，求在此条件下  $W(4)$  的条件期望  $E[W(4) | W(1) = 2]$  与条件方差  $Var(W(4) | W(1) = 2)$ 。

**解析**

利用布朗运动的独立增量性，将  $W(4)$  拆分为  $W(1)$  与增量的和：

$$W(4) = W(1) + (W(4) - W(1))$$

其中  $W(4) - W(1)$  与  $W(1)$  是相互独立的。

条件期望：

$$E[W(4) | W(1) = 2] = E[W(1) + (W(4) - W(1)) | W(1) = 2]$$

$$= 2 + E[W(4) - W(1)] = 2 + 0 = 2$$

条件方差：

因增量与  $W(1)$  独立，条件方差等于无条件方差：

$$Var(W(4) | W(1) = 2) = Var(W(1) + (W(4) - W(1)) | W(1) = 2) = Var(W(4) - W(1))$$

由于  $W(4) - W(1) \sim N(0, 3)$ ，故条件方差为 3。

16. 设  $\{W(t), t \geq 0\}$  为标准布朗运动。定义新的随机过程  $B(t) = W(t) - tW(1)$ ，其中  $0 \leq t \leq 1$ （这被称为布朗桥）。求该过程的均值函数  $m_B(t)$ ，以及当  $0 \leq s \leq t \leq 1$  时的协方差函数  $C_B(s, t)$ 。

**解析**

均值函数：

$$m_B(t) = E[W(t) - tW(1)] = E[W(t)] - tE[W(1)] = 0 - 0 = 0$$

协方差函数：

$$C_B(s, t) = E[B(s)B(t)] = E[(W(s) - sW(1))(W(t) - tW(1))]$$

展开乘积：

$$= E[W(s)W(t) - tW(s)W(1) - sW(t)W(1) + stW^2(1)]$$

利用  $E[W(u)W(v)] = \min(u, v)$ ，且当前  $s \leq t \leq 1$ ：

$$E[W(s)W(t)] = \min(s, t) = s$$

$$E[W(s)W(1)] = \min(s, 1) = s$$

$$E[W(t)W(1)] = \min(t, 1) = t$$

$$E[W^2(1)] = 1$$

代入上式得：

$$C_B(s, t) = s - t(s) - s(t) + st(1) = s - ts - st + st = s(1 - t)$$

# 概率论第12章：平稳过程小习题集

建议在对傅立叶变换有系统了解后再食用这一章节

## 一、基础概念与基本知识

- 1. 什么是严平稳过程？它的有限维分布在时间平移下满足什么性质？
- 2. 严平稳过程的一维分布、二维分布分别有什么特点？
- 3. 若严平稳过程存在二阶矩，为什么它一定是宽平稳过程？
- 4. 什么是宽平稳过程？验证宽平稳过程需要检查哪两个条件？

5. 严平稳和宽平稳有什么区别？二者是否互相蕴含？
6. 对高斯过程而言，严平稳与宽平稳之间有什么特殊关系？（建议对有关高斯之类的特殊性质着重复习）
7. 什么是总能量？什么是平均功率？什么是功率谱密度？它和谱密度的关系？
8. 平稳过程自相关函数有哪些基本性质？
9. 什么叫两个平稳过程  $X(t)$ 、 $Y(t)$  平稳相关？互相关函数  $R_{XY}(\tau)$  如何定义？
10. 平稳过程的谱密度  $S_X(\omega)$  与自相关函数  $R_X(\tau)$  之间的傅里叶变换关系是什么？
11. 谱密度有哪些基本性质？
12. 什么是  $\delta$  函数？
13. 什么是互谱密度？互谱密度与互相关函数之间有什么关系？
14. 什么是数学期望的各态历经性？什么是相关函数的各态历经性？什么是各态历经性？
15. 均方连续平稳过程，期望和相关函数具有各态历经性的充要条件？

## 二、习题集

### 1. 随机符号翻转过程的平稳性

令

$$Y(t) = (-1)^{\mu_t} X, \quad t \in [0, +\infty),$$

其中：

1.  $X$  为随机变量，且

$$P\{X = -1\} = P\{X = 1\} = \frac{1}{2};$$

2.  $\{\mu_t, t \geq 0\}$  是强度为  $\lambda$  的泊松过程；
3.  $X$  与  $\{\mu_t, t \geq 0\}$  相互独立。

证明  $Y(t)$  是平稳过程。

### 2. 随机频率与随机相位过程

设  $\xi$  和  $\Theta$  为相互独立的随机变量， $\Theta \sim U[0, 2\pi]$ ， $\xi$  的概率密度为

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

令

$$X(t) = \cos(\xi t + \Theta), \quad -\infty < t < +\infty.$$

1. 证明  $\{X(t), -\infty < t < +\infty\}$  是平稳随机过程；
2. 求其自相关函数  $R_X(\tau)$  及谱密度  $S_X(\omega)$ 。

### 3. 随机相位余弦过程的功率谱密度

设

$$X(t) = A \cos(\omega_0 t + \Phi),$$

其中  $A, \omega_0$  是常数， $\Phi \sim U(0, 2\pi)$ 。求  $X(t)$  的功率谱密度。

### 4. 由自相关函数求谱密度

已知平稳过程  $\{X(t), -\infty < t < +\infty\}$  的自相关函数  $R_X(\tau)$ ，求其谱密度  $S_X(\omega)$ 。

$$R_X(\tau) = \sigma^2 e^{-a|\tau|} \left( \cos b\tau - \frac{a}{b} \sin(b|\tau|) \right), \quad a > 0, b > 0.$$

求对应的  $S_X(\omega)$ 。

### 5. 自相关函数与谱密度的一组特殊变换对

根据课件图片中的表格，写出下列自相关函数对应的谱密度。

1.  $R_X(\tau) = 1$ ;
2.  $R_X(\tau) = \delta(\tau)$ 。

## 6. 宽平稳但不是严平稳的反例

设  $\{X_n, n \in \mathbb{Z}\}$  相互独立, 且

$$X_{2k} \sim N(0, 1), \quad P\{X_{2k+1} = 1\} = P\{X_{2k+1} = -1\} = \frac{1}{2}.$$

证明  $\{X_n\}$  是宽平稳序列, 但不是严平稳序列。

## 7. 严平稳但不是宽平稳的反例

设  $Y$  服从标准柯西分布, 令

$$X(t) = Y, \quad -\infty < t < +\infty.$$

证明  $X(t)$  是严平稳过程, 但不是宽平稳过程。

## 8. 随机平移制造平稳过程

设  $f(t)$  是周期为  $T$  的确定性函数,  $\Theta \sim U(0, T)$ , 定义

$$X(t) = f(t + \Theta).$$

1. 证明  $X(t)$  是严平稳过程;
2. 若实值函数

$$f(t) = \sum_{n=-\infty}^{+\infty} c_n e^{i2\pi nt/T},$$

求  $R_X(\tau)$  和谱密度  $S_X(\omega)$  的形式。

## 9. 初始条件决定是否平稳

设  $\{N(t), t \geq 0\}$  是强度为  $\lambda$  的泊松过程。定义

$$X(t) = X(0)(-1)^{N(t)}.$$

分别讨论下面两种情况是否为宽平稳过程, 并求均值和自相关函数:

1.  $X(0) = 1$ ;
2.  $P\{X(0) = 1\} = P\{X(0) = -1\} = \frac{1}{2}$ , 且  $X(0)$  与  $N(t)$  独立。

## 10. 随机幅度余弦过程的各态历经性

设

$$X(t) = A \cos(\omega_0 t + \Phi),$$

其中  $\Phi \sim U(0, 2\pi)$ ,  $A$  与  $\Phi$  独立, 且  $E[A^2] < +\infty$ ,  $\omega_0 \neq 0$ 。

1. 证明  $X(t)$  是宽平稳过程;
2. 求  $R_X(\tau)$ ;
3. 判断它是否具有均值各态历经性;
4. 判断它是否具有相关函数各态历经性。

## 11. 白噪声滤波与随机正弦信号叠加

设  $\{\varepsilon_n, n \in \mathbb{Z}\}$  为零均值白噪声, 方差为  $\sigma^2$ 。令

$$Y_n = \sum_{k=0}^{\infty} a^k \varepsilon_{n-k}, \quad |a| < 1.$$

又设

$$Z_n = B \cos(\omega_0 n + \Phi),$$

其中  $B, \omega_0$  为常数,  $\Phi \sim U(0, 2\pi)$ , 且  $Z_n$  与  $\{\varepsilon_n\}$  独立。令

$$X_n = Y_n + Z_n.$$

1. 证明  $X_n$  是宽平稳序列;
2. 求  $R_X(m)$ ;
3. 求谱密度  $S_X(\omega)$ 。