

SCNUSoS  
《概率论与数理统计》讲义



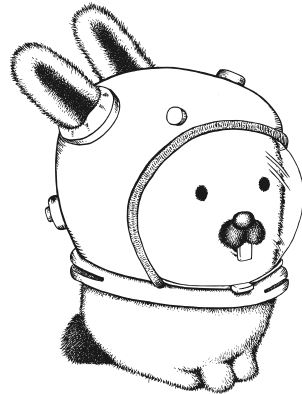
EVAN GAO

大二上



Made with X<sub>Y</sub>L<sup>A</sup>T<sub>E</sub>X  
*And Love*

evan@notch1p.xyz



Compiled on D:20231212



# CONTENTS

<b>CHAPTER 1</b>	<b>随机事件及其概率</b>	<b>PAGE 4</b>
1.1	随机事件	4
1.2	频率	4
1.3	概率 概率的性质 — 5	5
1.4	条件概率 乘法公式 — 6 • 全概率公式 — 7 • Bayes 公式 — 8	5
1.5	事件的独立性	9
1.6	Bernoulli 试验	9
<b>CHAPTER 2</b>	<b>随机变量及其分布</b>	<b>PAGE 10</b>
2.1	随机变量	10
2.2	离散型随机变量及其分布 常用离散分布 — 10	10
2.3	累积分布函数 随机变量的分布函数 — 11 • 离散型随机变量的分布函数 — 12	11
2.4	连续型随机变量及其概率密度 常用连续型分布 — 13	12
2.5	随机变量函数的分布 离散型随机变量函数的分布 — 15 • 连续型随机变量函数的分布 — 15	15
<b>CHAPTER 3</b>	<b>多维随机变量及其分布</b>	<b>PAGE 17</b>
3.1	二维随机变量及其分布 二维随机变量 — 17 • 二维随机变量的分布函数 — 17 • 二维离散型随机变量及其分布律 — 18 • 二维连续型随机变量及其概率密度 — 19 • 二维均匀分布 — 20	17
3.2	随机变量的独立性 离散型随机变量的独立性 — 21 • 连续型随机变量的独立性 — 22	21
<b>CHAPTER 4</b>	<b>随机变量的数字特征</b>	<b>PAGE 24</b>
4.1	数学期望 随机变量的数学期望 — 24 • 随机变量函数的数学期望 — 24 • 数学期望的性质 — 25	24
4.2	方差 方差的定义 — 26 • 方差的计算 — 26 • 方差的性质 — 27	26
4.3	大数定律与中心极限定理 Chebyshev 不等式 — 28 • 大数定律 — 29 • 中心极限定理 CLT — 30	28
<b>CHAPTER 5</b>	<b>数理统计的基础知识</b>	<b>PAGE 32</b>
5.1	基本概念 统计量 — 32 • 常用统计量 — 32	32

5.2	常用统计分布	33
	卡方分布 — 33 • t 分布 — 34 • 分位数 — 35 • F 分布 — 35	
5.3	抽样分布	36
	单正态总体的抽样分布 — 36	

## CHAPTER 6 参数估计 PAGE 37

6.1	点估计问题概述	37
	点估计的概念 — 37 • 评价估计的标准 — 37	
6.2	点估计的常用方法	38
	矩估计法 — 38 • 最大似然估计法 — 39	
6.3	置信区间	40
	置信区间的概念 — 40	
6.4	正态总体的置信区间	41
	正态总体均值的置信区间 — 41 • 正态总体方差的置信区间 — 42	

## CHAPTER 7 假设检验 PAGE 43

7.1	基本概念	43
7.2	单正态总体的假设检验	43
	总体均值的假设检验 — 43 • 总体方差的假设检验 — 44	

# Chapter 1

## 随机事件及其概率

### 1.1 随机事件

事件是样本空间的子集

#### Definition 1.1.1: 样本空间

考虑样本空间集合  $S$ , 我们有  $S \stackrel{\text{def}}{=} \{\text{所有样本点}\}$ .

由定义, 我们可以得到几种特殊的样本空间:

- $\emptyset$  事件: 不可能发生的事件.
- $S - \emptyset$  发生的事件.
- 基本事件  $\omega$ :  $|\omega| = 1$  i.e. 基本事件只含有一个样本点.

#### Theorem 1.1.1 De Morgan 律

由于  $\emptyset$  事件和  $S - \emptyset$  事件是互为对偶的, 我们可以得到

$$\begin{aligned}\overline{A \cup B} &= \bar{A} \cap \bar{B} \\ \overline{A \cap B} &= \bar{A} \cup \bar{B}\end{aligned}\tag{1.1}$$

我们定义和, 积事件为

$$\begin{aligned}\bigcup_{i=1}^n A_i \text{ happen} &\iff \exists i \in [1, n] \text{ s.t. } A_i \text{ happens} \\ \bigcap_{i=1}^n A_i \text{ happen} &\iff \forall i \in [1, n] \text{ s.t. } A_i \text{ happens}\end{aligned}\tag{1.2}$$

### 1.2 频率

#### Definition 1.2.1: 频率

考虑事件  $A$ , 其发生的频率是

$$f_n(A) \stackrel{\text{def}}{=} \frac{r_n(A)}{n} \in [0, 1]$$

其中频数  $r_n(A) \in [0, n]$ . 显然有  $f_n(S) = 1$ .

### Corollary 1.2.1 有限可加性

若  $A_i \cap A_j = \emptyset$  且  $i \neq j, i, j \in [1, k]$  i.e. 互斥事件, 则

$$f_n \left( \bigcap_{i=1}^k A_i \right) = \sum_{i=1}^k f_n(A_i)$$

## 1.3 概率

### Definition 1.3.1: 概率

(Kolmogorov 公理化定义) 设有随机试验  $E$  且与之对应的样本空间  $S$ , 考虑事件  $A$

for  $\forall A \in E$ , if

- ①  $0 \leq P(A) \leq 1$
- ②  $P(S) = 1$
- ③  $\Pr\{\bigcup_{i=1}^{\infty} A_i\} = \sum_{i=1}^{\infty} \Pr\{A_i\}$  i.e. 可列可加性

则称  $P$  为  $S$  上的**概率**.

### 1.3.1 概率的性质

由上面的定义我们能够得到概率的性质:

1.  $\Pr\{\emptyset\} = 0$
2.  $\Pr\{\bigcup_{i=1}^n A_i\} = \sum_{i=1}^n \Pr\{A_i\} \iff \forall i, j (i \neq j \rightarrow A_i A_j = \emptyset)$
3.  $\Pr\{\bar{A}\} + \Pr\{A\} = 1$
4.  $\Pr\{A - B\} = \Pr\{A\} - \Pr\{AB\} \Rightarrow (A - B) \cap B = \emptyset$
5. (单调性)  $B \subseteq A \Rightarrow \Pr\{B\} \leq \Pr\{A\}$
6. 若满足 5, 由 4 可得  $\Pr\{A - B\} = \Pr\{A\} - \Pr\{B\}$
7. (容斥原理)  $\Pr\{A \cup B\} = \Pr\{A\} + \Pr\{B\} - \Pr\{AB\}$

## 1.4 条件概率

### Definition 1.4.1: 条件概率

设  $A, B$  是两个事件, 且  $P(A) \neq 0$ , 则称

$$\Pr\{A|B\} \stackrel{\text{def}}{=} \frac{\Pr\{AB\}}{\Pr\{B\}} \quad (1.3)$$

为在事件  $A$  发生的条件下, 事件  $B$  的**条件概率**.

### Corollary 1.4.1 条件概率之性质

概率满足的性质条件概率都满足.

**Theorem 1.4.1**

设  $A_1, A_2, \dots, A_n$  是  $n$  个互斥事件, 则有

$$\Pr\left\{\bigcup_{i=1}^n A_i \mid A\right\} = \sum_{i=1}^n \Pr\{A_i \mid A\} \quad (1.4)$$

**Question 1: 证明**

$$\Pr\{\bar{B} \mid A\} = 1 - \Pr\{B \mid A\} \quad (1.5)$$

**Proof:**

$$\begin{aligned} \Pr\{\bar{B} \mid A\} &= \frac{\Pr\{\bar{B}A\}}{\Pr\{A\}} \\ &= \frac{\Pr\{A\} - \Pr\{BA\}}{\Pr\{A\}} \\ &= 1 - \frac{\Pr\{BA\}}{\Pr\{A\}} \\ &= 1 - \Pr\{B \mid A\} \end{aligned}$$

**1.4.1 乘法公式**

由条件概率的定义, 我们可以得到**乘法公式**:

**Theorem 1.4.2 乘法公式**

$$\begin{aligned} \Pr\{A_1 A_2 \cdots A_n\} &= \Pr\{A_1\} \Pr\{A_2 \mid A_1\} \Pr\{A_3 \mid A_1 A_2\} \cdots \Pr\{A_n \mid A_1 A_2 \cdots A_{n-1}\} \\ &= \prod_{i=1}^n \Pr\left\{A_i \mid \bigcup_{i=1}^{n-1} A_i\right\} \end{aligned} \quad (1.6)$$

**Example 1.4.1 (“彩票”)**

第一次买中的概率为  $\frac{1}{2}$ , 第二次买中而第一次未中的概率是  $\frac{7}{10}$ , 第三次买中而前两次未中的概率是  $\frac{9}{10}$ , 求三次都未中的概率.

(1.6) **Solution:** 以  $A_i (i = 1, 2, 3)$  表示事件“第  $i$  次买中”, 以  $B$  表示事件“三次都未中”, 由乘法公式

$$\begin{aligned} \because B &= \overline{A_1 A_2 A_3} \\ \therefore \Pr\{B\} &= \Pr\{\overline{A_1 A_2 A_3}\} \\ &= \Pr\{\bar{A}_1\} \Pr\{\bar{A}_2 \mid \bar{A}_1\} \Pr\{\bar{A}_3 \mid \bar{A}_1 \bar{A}_2\} \\ &= \left(1 - \frac{1}{2}\right) \left(1 - \frac{7}{10}\right) \left(1 - \frac{9}{10}\right) \\ &= \frac{3}{200} \end{aligned}$$

### Question 2: P<sub>28</sub>-19

袋中装有  $a$  个红球,  $b$  个白球, 每次自袋中有放回地任取一球, 并同时再放入  $m$  个与之相同的球, 连续如此进行  $2n$  次, 求前  $n$  次为红球, 后  $n-1$  次为白球, 第  $2n$  次为红球的概率.

**Solution:** 设事件  $A_i$  表示第  $i$  次取出的球为红球,  $B_i$  表示白球, 那么可知所求概率为

$$\Pr\{A_1 A_2 A_3 \cdots A_n B_{n+1} B_{n+2} \cdots B_{2n-1} A_{2n}\}$$

(1.6) 利用乘法公式展开上式, 立即有

$$\Pr\{A_1\} \Pr\{A_2 | A_1\} \cdots \Pr\{A_n | A_1 A_2 \cdots A_{n-1}\} \times \Pr\{B_{n+1} | A_1 A_2 \cdots A_n\} \Pr\{B_{n+2} | \cdots B_{n+1}\} \cdots \Pr\{A_{2n} | A_1 \cdots A_n B_{n+1} B_{n+2} \cdots B_{2n-1}\}$$

此式太长, 不妨将其分成三部分:

- ①  $\Pr\{A_1\} \Pr\{A_2 | A_1\} \cdots \Pr\{A_n | A_1 A_2 \cdots A_{n-1}\}$  (前  $n$  个红球)
- ②  $\Pr\{B_{n+1} | A_1 A_2 \cdots A_n\} \Pr\{B_{n+2} | \cdots B_{n+1}\} \cdots \Pr\{B_{2n-1} | \cdots B_{2n-2}\}$  (后  $n-1$  个白球)
- ③  $\Pr\{A_{2n} | A_1 \cdots A_n B_{n+1} B_{n+2} \cdots B_{2n-1}\}$  (最后的红球)

三式相乘, 有

$$\prod_{i=1}^n \frac{a + (i-1)m}{a + b + (i-1)m} \prod_{i=n+1}^{2n-1} \frac{b + (i-n-1)m}{a + b + (i-1)m} \times \frac{a + (n-1)m}{a + b + (2n-1)m}$$

即为所求.

## 1.4.2 全概率公式

### Lemma 1.4.1 完备事件组

设  $A_1, A_2, \dots, A_n$  是有限或可数个事件, 若其满足 1. 两两互斥, 2.  $\bigcup_{i=1}^n A_i = S$ , 则称  $A_1, A_2, \dots, A_n$  是一个**完备事件组**.

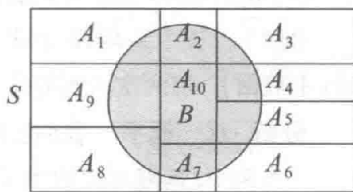


图 1-4-2

Figure 1.1: 全概公式示意

(1.7)指出在复杂情况下不易计算  $\Pr\{B\}$  时应该根据具体情况构造完备事件组  $\{A_i\}$ , 使得事件  $B$  发生的概率是各事件  $A_i$  发生的条件下造成事件  $B$  发生概率的和.

### Theorem 1.4.3 全概率公式

设  $A_1, A_2, \dots, A_n$  是一个完备事件组, 且  $\forall i (i \in [1, n] \rightarrow \Pr\{A_i\} > 0)$ , 则对任一事件  $B$ , 有

$$\Pr\{B\} = \sum_{i=1}^n \Pr\{A_i\} \Pr\{B | A_i\} \quad (1.7)$$



### Question 3: P<sub>14</sub> 10

从 1 到 9 的整数中有放回地依次随机抽取 3 次, 求取出的 3 个数之积能被 10 整除的概率.

显然这三个数中必有两数为 5 和 {2, 4, 6, 8}, 因此

法一: 分情况讨论

- ①  $A = \{\text{三个数里有两个 5 和一个偶数}\}$
- ②  $B = \{\text{三个数里有一个 5 和两个偶数}\}$
- ③  $C = \{\text{三个数里有一个 5 一个偶数和一个其他的奇数}\}$

$$\text{那么有 } \Pr\{A\} = \frac{\binom{4}{1} \cdot 3}{9^3} \quad \Pr\{B\} = \frac{\binom{4}{1} \cdot 3 + \binom{4}{2} A_3^3}{9^3} \quad \Pr\{C\} = \frac{\binom{4}{1} \binom{4}{1} A_3^3}{9^3} \text{ 相加得 } \frac{156}{729}.$$

法二: 对立事件

不妨设三个数中出现 5 的事件为  $A$ , 出现偶数的事件为  $B$ , 那么

$$\begin{aligned} \Pr\{AB\} &= 1 - \Pr\{\overline{AB}\} \\ &= 1 - \left( \Pr\{\overline{A}\} + \Pr\{\overline{B}\} - \Pr\{\overline{A}\overline{B}\} \right) \\ &= 1 - \left( \frac{8^3}{9^3} + \frac{5^3}{9^3} - \frac{4^3}{9^3} \right) \\ &= \frac{156}{729} \end{aligned}$$

法三: 另一种分情况讨论 (笔者的分法)

- ①  $A = \{\text{三个数中有两个相同的}\}$
- ②  $B = \{\text{三个数全不同}\}$ 
  - (a)  $B_1 = \{\text{三个数全不同且有两个偶数}\}$
  - (b)  $B_2 = \{\text{三个数全不同且有两个奇数}\}$

此项须要分成下述两项, 否则会重

$$\text{那么有 } \Pr\{A\} = \frac{\binom{4}{1} \binom{3}{1} + \binom{4}{1} \binom{3}{1}}{9^3} \quad \Pr\{B_1\} = \frac{\binom{4}{1} \binom{4}{1} \binom{3}{1}}{9^3} \quad \Pr\{B_2\} = \frac{\binom{4}{1} \binom{4}{1} A_3^3}{9^3} \text{ 相加得 } \frac{156}{729}.$$

#### Note

讨论各种情况的概率进而求得所求事件的概率的方法实际上是**全概率公式**的一种体现.

### 1.4.3 Bayes 公式

**全概率公式**是通过计算某一事件会发生的所有原因和情况的可能性大小来计算该事件发生的概率, 而**Bayes 公式**则与之相反, 考察一件已经发生的事情的各种原因或或情况的可能性大小.

#### Theorem 1.4.4 Bayes 公式

设  $A_1, A_2, \dots, A_n$  是一个完备事件组, 且  $\forall i (i \in [1, n] \rightarrow \Pr\{A_i\} > 0)$ , 则对任一可能发生的事件  $B$ , 有

$$\Pr\{A_i | B\} = \frac{\Pr\{A_i B\}}{\Pr\{B\}} = \frac{\Pr\{A_i\} \Pr\{B | A_i\}}{\sum_{j=1}^n \Pr\{A_j\} \Pr\{B | A_j\}} \quad (1.8)$$

特别地, 当  $n = 2$  时, Bayes 公式也可以写成

$$\Pr\{A | B\} = \frac{\Pr\{A\} \Pr\{B | A\}}{\Pr\{B\}} \quad (1.9)$$

$\Pr\{A_i\}$ ,  
 $\Pr\{A_i | B\}$  分别称为原因的**先验概率**和**后验概率**

## 1.5 事件的独立性

### Definition 1.5.1: 两事件独立性

独立与互斥是两种不同的概念

设  $A, B$  是两个事件, 若

$$\Pr\{AB\} = \Pr\{A\} \Pr\{B\}$$

则称  $A, B$  相互独立.

### Theorem 1.5.1

设  $A, B$  两事件相互独立且  $\Pr\{B\} > 0$ , 则有

$$\Pr\{A|B\} = \Pr\{A\}$$

### Theorem 1.5.2

若  $A, B$  两事件相互独立, 则它们对立事件和其本身 (不同事件间) 的组合也相互独立.

根据三个事件 (略) 独立性的定义, 可以类推到  $n$  个事件的独立性: 设  $A_1, A_2, \dots, A_n$  是  $n$  个事件, 若对于其中任意  $k (2 \leq k \leq n)$  个事件  $A_{i_1}, A_{i_2}, \dots, A_{i_k}$ , 有

$$\Pr\{A_{i_1} A_{i_2} \cdots A_{i_k}\} = \Pr\{A_{i_1}\} \Pr\{A_{i_2}\} \cdots \Pr\{A_{i_k}\} \quad (1.10)$$

则称  $A_1, A_2, \dots, A_n$  相互独立.

### Definition 1.5.2

设  $A_1, A_2, \dots, A_n (n > 2)$  是  $n$  个事件, 若其中任意两个事件相互独立, 则称这  $n$  个事件两两独立.

由此可以得到多个独立事件所具备的性质

### Corollary 1.5.1

设  $A_1, A_2, \dots, A_n (n > 2)$  相互独立, 则其中任意  $k (2 \leq k \leq n)$  个 (它们的对立) 事件也相互独立.

## 1.6 Bernoulli 试验

即两点分布.

### Theorem 1.6.1 Bernoulli 定理

一次试验中, 事件  $A$  发生的概率为  $p$ , 进行这样的试验  $n$  次, 事件  $A$  发生  $k$  次的概率为

$$b(k; n; p) = \binom{n}{k} p^k (1-p)^{n-k}$$

这相当于在实验结果序列中任取  $k$  次发生  $A$  事件

### Note

$n$  次 Bernoulli 试验的概率分布就是二项分布. 其概率密度函数为

$$f(k, n, p) = \Pr(k; n, p) = \Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$$

**Glossary:** 两点分布 = Bernoulli 试验; 二项分布 = 多次 Bernoulli 试验 = Bernoulli 概型.

**辨析:**  $\Pr\{A\}$  和  $\Pr(X)$  使用的括号不同. 大括号代表是事件, 即一系列样本点集合 (故而用大括号), 小括号代表是随机变量 (因为此时是概率密度函数). ☺

# Chapter 2

## 随机变量及其分布

### 2.1 随机变量

#### Definition 2.1.1: 随机变量

设随机试验的样本空间为  $S$ , 称定义在  $S$  上的实值单值函数  $X = X(\omega)$  为**随机变量**.  
亦即

$$X : \omega \mapsto x_i \in \mathbb{R} \text{ i.e. } S \rightarrow \mathbb{R}$$

实际上,  $\Pr(X = a)$  是  $A = \{\omega | X(\omega) = a\}$  的简记.

### 2.2 离散型随机变量及其分布

设  $X$  是一个随机变量, 如果  $X$  的所有可能取值为有限个或可列无限多个, 则称  $X$  为**离散型随机变量**.

#### Definition 2.2.1: 离散型随机变量分布律

设  $X$  是一个离散型随机变量, 如果

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

则称  $p_k$  为  $X$  的**分布律**或**概率分布**, 亦称**概率质量函数** (PMF). 记为  $f_X(x)$ .

就是高中所学之分布列.

#### 2.2.1 常用离散分布

##### 两点分布

#### Definition 2.2.2: 两点分布

设  $X$  的分布律为

$$\Pr(X = k) = \begin{cases} p, & k = x_1 \\ 1 - p, & k = x_2 \end{cases} \quad (2.1)$$

其中  $0 < p < 1$ , 则称  $X$  服从以  $p$  为参数的**两点分布**.

若  $X$  服从  $x_1 = 1, x_2 = 0$  处参数为  $p$  的两点分布, 则称其服从参数为  $p$  的**0-1分布**.

## 二项分布

### Definition 2.2.3: 二项分布

设  $X$  的分布律为

$$\Pr(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n \quad (2.2)$$

其中  $n$  为正整数,  $0 < p < 1$ , 则称  $X$  服从参数为  $n, p$  的**二项分布**, 记为  $X \sim B(n, p)$ .

当  $n = 1$  时, 二项分布退化为 0-1 分布.

二项概率总存在一个最大值  $M$  s.t.  $(n+1)p - 1 \leq M < (n+1)p$ .

$\sim$  记号就是**服从**的意思

## Poisson 分布

### Definition 2.2.4: Poisson 分布

设  $X$  的分布律为

$$\Pr(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots \quad (2.3)$$

其中  $\lambda > 0$ , 则称  $X$  服从参数为  $\lambda$  的**Poisson 分布**, 记为  $X \sim P(\lambda)$ .

因为  $e^x = 1 + \sum_{n=1}^{\infty} \frac{x^n}{n!}$ , 将  $x$  换为  $\lambda$  可得泊松分布的概率和为 1.

## 2.3 累积分布函数

### 2.3.1 随机变量的分布函数

#### Definition 2.3.1: 随机变量 CDF

设  $X$  是一个随机变量, 对任意实数  $x$ , 定义

$$F(x) \stackrel{\text{def}}{=} \Pr(X \leq x) \in [0, 1]$$

称  $F(x)$  为  $X$  的**累积分布函数**(CDF). 有时记作  $F_X(x)$  或  $X \sim F(x)$ .

分布函数  $F(x)$  的值就表示  $X$  落在区间  $(-\infty, x]$  的概率.

#### 性质

1. 单调非减性:  $x_1 < x_2 \Rightarrow F(x_1) \leq F(x_2)$
2.  $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0$ ,  $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = 1$ .
3. 右连续性:  $\lim_{x \rightarrow x_0^+} F(x) = F(x_0)$

具有上述性质的函数一定是某个随机变量的 CDF.

### 2.3.2 离散型随机变量的分布函数

在上述基础上,对于离散型随机变量  $X$  有 CDF:

$$F(x) = \Pr(X \leq x) = \sum_{x_i \leq x} \Pr(X = x_i)$$

$F(x)$  是一个**阶梯型函数**,在点  $x_i$  处的跃度为  $\Pr(X = x_i)$ . 可以看出阶梯型函数符合 CDF 之性质:总是右连续的;而作为一个分段函数来说,其各段区间总是(习惯上)左闭右开的.

#### Note

考题中由 CDF 反推分布列可以根据每一个左侧端点得到对应的随机变量  $x_i$ .

计算分布列时,有公式  $\Pr\{a \leq X < b\} = F(b) - F(a)$ .

随机变量  $X$  的 CDF 为阶梯型函数  $\Leftrightarrow X$  是离散型随机变量.

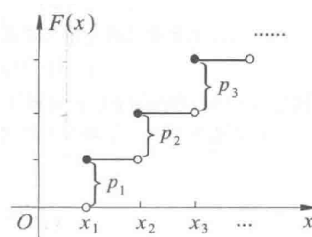


图 2-3-1

Figure 2.1: 阶梯型 CDF

### 2.4 连续型随机变量及其概率密度

#### Definition 2.4.1: PDF

设  $X$  是一个随机变量,如果存在非负可积函数  $f(x)$ ,使对任意实数  $x$  有

$$F(x) = \Pr\{X \leq x\} = \int_{-\infty}^x f(t) dt \quad (2.4)$$

则称  $X$  为**连续型随机变量**, $f(x)$  为  $X$  的**概率密度函数(PDF)**.

立即可以得到 PDF 的性质

1.  $f(x) \geq 0$ ;
2.  $\int_{-\infty}^{+\infty} f(x) dx = 1$ .
3.  $F'(x) = f(x)$ . 若  $f(x)$  在  $x$  处连续

#### Note

若某个随机变量的 CDF 可以写成 PDF 积分函数的形式,则称其为连续型随机变量.

**证明**  $X$  取任一特值的概率为 0:

$$\text{考虑极限 } \Pr\{X = a\} = \lim_{\Delta x \rightarrow 0^+} \Pr\{a - \Delta x \leq X \leq a\}$$

对 RHS 积分,有

$$\begin{aligned} \Pr\{X = a\} &= \lim_{\Delta x \rightarrow 0^+} \int_{a-\Delta x}^a f(x) dx \\ &= 0. \end{aligned}$$

☺

由上面的证明我们可以知道以下的性质

$$\Pr\{a < X < b\} = F(b) - F(a) = \int_a^b f(x) dx. \quad (2.5)$$

这个等式无论  $X$  两侧端点开闭都成立.

## 2.4.1 常用连续型分布

### 均匀分布

#### Definition 2.4.2: 均匀分布

设  $X$  的概率密度为

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \quad (2.6)$$

U for Uniform

其中  $a < b$ , 则称  $X$  服从参数为  $a, b$  的**均匀分布**, 记为  $X \sim U(a, b)$ .

设  $[a, b]$  之间的任意子区间为  $[c, d]$ , 由上式可得  $\Pr\{c < X < d\} = \int_c^d \frac{1}{b-a} dx = \frac{d-c}{b-a}$ . 因而我们可以得到

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a \leq x < b \\ 1, & x \geq b \end{cases} \quad (2.7)$$

### 指数分布

#### Definition 2.4.3: 指数分布

若随机变量  $X$  的概率密度为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}, \lambda > 0 \quad (2.8)$$

则称  $X$  服从参数为  $\lambda$  的**指数分布**, 记为  $X \sim E(\lambda)$ .

积分得

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad (2.9)$$

### 正态分布

#### Definition 2.4.4: 正态分布

若随机变量  $X$  的概率密度为

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (-\infty < x < +\infty) \quad (2.10)$$

其中  $\mu, \sigma (\sigma > 0)$  为常数, 则称  $X$  服从参数为  $\mu, \sigma^2$  的**正态分布**, 记为  $X \sim N(\mu, \sigma^2)$ .

#### Question 4: Poisson 积分

不使用 Wallis 公式, 求 Poisson 积分  $\int_{-\infty}^{+\infty} e^{-t^2} dt$ .

**Solution:** 转化为求累次积分

要求原式, 不妨先求其平方, 亦即

$$\left( \int_{-\infty}^{+\infty} e^{-t^2} dt \right)^2 = \int_{-\infty}^{+\infty} e^{-x^2} dx \int_{-\infty}^{+\infty} e^{-y^2} dy$$

化为累次积分有

$$= \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} dx dy.$$

$\mathbb{R}^2$  即  $xOy$  面, 化为极坐标有

$$\begin{aligned} \iint_{\mathbb{R}^2} e^{-\rho^2} \rho \, d\rho \, d\theta &= \int_0^{2\pi} d\theta \int_0^{+\infty} e^{-\rho^2} \rho \, d\rho \\ &= \pi \\ \therefore \int_{-\infty}^{+\infty} e^{-t^2} \, dt &= \sqrt{\pi}. \end{aligned} \quad (2.11)$$

通过计算 Poisson 积分, 我们可以得到正态分布的 PDF 重要性质:

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx = 1 \quad (2.12)$$

### 图形特征

$\sigma$  越小, 方差越大,  
曲线中峰越陡

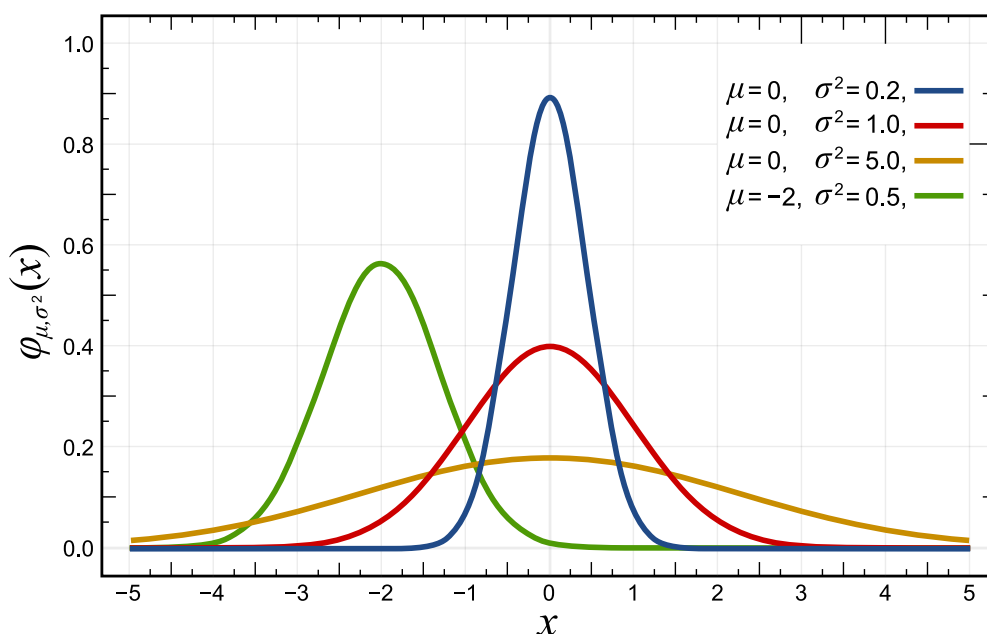


Figure 2.2: PDF

1.  $\mu$  确定曲线位置,  $\sigma$  确定曲线中峰的陡峭程度
2. 密度曲线关于  $x = \mu$  对称
3. 密度曲线在  $x = \mu$  处有最大值  $\frac{1}{\sqrt{2\pi}\sigma}$
4. 密度曲线在  $x = \mu \pm \sigma$  处有拐点且以  $x$  轴为渐近线

当  $\mu = 0, \sigma = 1$  时, 称为**标准正态分布**, 记为  $X \sim N(0, 1)$ , PDF 用  $\varphi(x)$  表示, CDF 用  $\Phi(x)$  表示:

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad \Phi(x) = \int_{-\infty}^x \varphi(t) \, dt \quad (2.13)$$

显然  $\varphi(x)$  为偶函数.

#### Theorem 2.4.1 重要定理

设  $X \sim N(\mu, \sigma^2)$ , 则  $Y = \frac{X-\mu}{\sigma} \sim N(0, 1)$ .

计算  $\Pr\{Y \leq x\}$  时, 可转化为计算  $\Pr\{X \leq \mu + \sigma x\}$ , 计算此时的  $\Phi(x)$  即得证.

这说明任意一般的正态分布都可以通过**线性变换**转化为标准正态分布. 因此  $X$  的分布函数也可以写成

$$F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right)$$

$$\Pr\{a < X < b\} = \Pr\left\{\frac{a-\mu}{\sigma} < Y < \frac{b-\mu}{\sigma}\right\} = \Phi\left(\frac{b-\mu}{\sigma}\right) - \Phi\left(\frac{a-\mu}{\sigma}\right)$$

## 2.5 随机变量函数的分布

前面提到计算  $\Pr\{Y \leq x\}$  时, 可转化为计算  $\Pr\{X \leq \mu + \sigma x\}$ , 其中的  $Y = \frac{X-\mu}{\sigma}$  是随机变量  $X$  的函数, 因此  $Y$  也是随机变量.

### Definition 2.5.1: 随机变量函数

若存在一个函数  $g(x)$  s.t. 随机变量  $X, Y$  满足

$$Y = g(X)$$

称  $Y$  为  $X$  的一个**随机变量函数**.

### 2.5.1 离散型随机变量函数的分布

根据上述定义, 显然  $X$  的函数  $Y$  也是一个随机变量.

#### 导出 $Y$ 的分布律

首先确定  $Y$  的所有取值, 通过  $Y$  的每一个值  $y_i (i = 1, 2, \dots)$  确定相应的 (注: 我们用  $C_i$  来表示  $X$  的取值为  $x_i$  时, 与对应  $Y$  匹配的  $X$  的取值的集合)

$$C_i = \{x_i | g(x_j) = y_i\},$$

$$\{Y = y_i\} = \{X \in C_i\},$$

那么有

$$\Pr\{Y = y_i\} = \Pr\{X \in C_i\} = \sum_{x_j \in C_i} \Pr\{X = x_j\} \quad (2.14)$$

为  $Y$  的分布律. 这表明:  $Y$  的分布律完全由  $X$  的分布律确定.

### 2.5.2 连续型随机变量函数的分布

分布律: 离散变量  
PDF: 连续变量

通过上述定义, 不难发现连续随机变量的函数不一定是连续型随机变量.

#### Note

考虑符号函数作为映射法则:

$$Y = \text{sign}(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1 & x < 0 \end{cases}$$

显然  $Y = 0, \pm 1$ , 是一个离散型随机变量. 因此, 在连续型的情况下,  $Y$  的连续性由其自身决定.

#### 导出 $Y$ 的 PDF

已知  $X$  的 CDF  $F_X(x)$  或 PDF  $f_X(x)$ , 则  $Y = g(x)$  的 CDF 可以通过以下公式求得:

$$F_Y(y) = \Pr\{Y \leq y\} = \Pr\{g(X) \leq y\} = \Pr\{X \in C_y\}, \text{ 其中 } C_y = \{x | g(x) \leq y\}. \quad (2.15)$$

之后通过积分  $\int_{C_y} f_X(x) dx$  求得 CDF.

#### Exercise 2.5.1 离散变量函数导出 PMF

给出随机变量  $X$  的分布律如下, 试求  $Y = X^2$  的分布律:

$X$	-2	-1	0	1	3
$p_i$	1/5	1/6	1/5	1/15	11/30



(2.14) **Solution:** 由题意得  $Y = 0, 1, 4, 9,$ , 计算得

$$\begin{aligned} \Pr\{Y = 0\} &= \Pr\{X = 0\} = \frac{1}{5} & \Pr\{Y = 1\} &= \Pr\{X = \pm 1\} = \frac{7}{30} \\ \Pr\{Y = 4\} &= \Pr\{X = -2\} = \frac{1}{5} & \Pr\{Y = 9\} &= \Pr\{X = 3\} = \frac{11}{30} \end{aligned}$$

$$\therefore \begin{array}{c|cccc} Y & 0 & 1 & 4 & 9 \\ \hline p_i & 1/5 & 7/30 & 1/5 & 11/30 \end{array}$$

### Exercise 2.5.2 连续变量函数导出 PDF

设  $X$  服从标准正态分布, 求  $Y = 2X^2 + 1$  的 PDF

**Solution:** 要求 PDF, 不妨先求  $Y$  的 CDF  $F_Y(y)$  i.e.  $\Pr\{Y \leq y\}$ , 代入有

$$\begin{aligned} F_Y(y) &= \Pr\{2X^2 + 1 \leq y\} = \Pr\left\{\sqrt{-\frac{y-1}{2}} \leq X \leq \sqrt{\frac{y-1}{2}}\right\} \\ &= \begin{cases} \int_{-\sqrt{\frac{y-1}{2}}}^{\sqrt{\frac{y-1}{2}}} f_X(x) dx & y > 1 \\ 0 & \text{otherwise} \end{cases} \\ &= \begin{cases} F_X\left(\sqrt{\frac{y-1}{2}}\right) - F_X\left(-\sqrt{\frac{y-1}{2}}\right) & y > 1 \\ 0 & \text{otherwise} \end{cases} \end{aligned}$$

$$\therefore f_Y(y) = F'_Y(y) = \begin{cases} \sqrt{2} \frac{f_X\left(\sqrt{\frac{y-1}{2}}\right)}{4\sqrt{y-1}} + \sqrt{2} \frac{f_X\left(-\sqrt{\frac{y-1}{2}}\right)}{4\sqrt{y-1}} & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\therefore \varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2.14) \quad \text{注意到 } \varphi(x) \text{ 为偶函数, 合并上述两项}$$

$$\implies f_Y(y) = \begin{cases} \frac{1}{2\sqrt{\pi(y-1)}} e^{-\frac{y-1}{4}} & y > 1 \\ 0 & \text{otherwise} \end{cases}$$

# Chapter 3

## 多维随机变量及其分布

### 3.1 二维随机变量及其分布

#### 3.1.1 二维随机变量

##### Definition 3.1.1: 二维随机变量

设随机试验的样本空间为  $S$ , 定义在  $S$  上的实值单值函数  $X = X(\omega)$   $Y = Y(\omega)$  为两个随机变量. 称向量  $(X, Y)$  为定义在  $S$  上的**二维随机变量**

$$\begin{aligned} X &: S \rightarrow \mathbb{R} \\ \text{i.e. } Y &: S \rightarrow \mathbb{R} \\ (X, Y) &: S \rightarrow \mathbb{R}^2 \text{ i.e. } \omega \mapsto (X(\omega), Y(\omega)) \end{aligned}$$

类似地, 我们可以得到  $n$  维随机变量的定义.

#### 3.1.2 二维随机变量的分布函数

##### Definition 3.1.2: 二维随机变量的分布函数

设  $(X, Y)$  是二维随机变量, 对于任意实数  $x, y$ , 定义二元函数

$$F(x, y) = \Pr\{X \leq x \cap Y \leq y\} \stackrel{\text{def}}{=} \Pr\{X \leq x, Y \leq y\} \quad (3.1)$$

称为二维随机变量  $(X, Y)$  的**分布函数**, 或**联合分布函数 (Joint CDF)**.

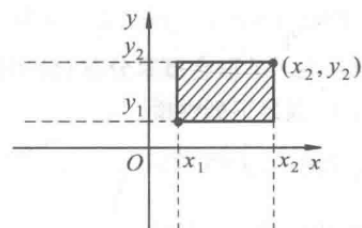


图 3-1-2

Figure 3.1: CDF

将这个向量视作平面内随机点的坐标, 分布函数就是随机点  $(X, Y)$  落入矩形域的概率. 通过容斥原理计算得到

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

## 联合 CDF 的性质

右侧几个等式可通过矩形域的面积来理解.

1.  $0 \leq F(x, y) \leq 1$

(a) 固定  $y, F(-\infty, y) = 0$ ;

(b) 固定  $x, F(x, -\infty) = 0$ ;

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

2. (单调性)  $F(x, y)$  关于  $x$  和  $y$  都是单调不减的.

(a) 固定  $y, F(x, y)$  关于  $x$  单调不减;

(b) 固定  $x, F(x, y)$  关于  $y$  单调不减.

3. (连续性)  $F(x, y)$  关于  $x$  和  $y$  都是右连续的.

若已知  $(X, Y)$  的分布函数  $F(X, Y)$ , 则可由之导出各个参数 (在固定另一个参数的情况下) 各自的分布函数:

$$F_X(x) = \Pr\{X \leq x\} = \Pr\{X \leq x\} = \lim_{y \rightarrow +\infty} F(x, y), \quad (3.3)$$

$$F_Y(y) = \Pr\{Y \leq y\} = \Pr\{Y \leq y\} = \lim_{x \rightarrow +\infty} F(x, y), \quad (3.4)$$

如此导出的 CDF 称为**边缘分布函数**.

### 3.1.3 二维离散型随机变量及其分布律

#### Definition 3.1.3: 联合概率分布律

设  $(X, Y)$  是二维离散型随机变量, 若存在非负函数  $p_{ij}$ , 使得

$$\Pr\{X = x_i, Y = y_j\} = p_{ij}, \quad i, j = 1, 2, \dots \quad (3.5)$$

则称  $(X, Y)$  为二维离散型随机变量, 并称  $p_{ij}$  为  $(X, Y)$  的**联合概率分布律**.

显然,  $p_{ij}$  满足性质: 1.  $p_{ij} \geq 0$ ; 2.  $\sum_i \sum_j p_{ij} = 1$ . 利用其分布律, 易见  $(X, Y)$  在  $D$  上的概率为

$$\Pr\{(X, Y) \in D\} = \sum_{(x_i, y_j) \in D} p_{ij} \quad (3.6)$$

而其联合 CDF 为

$$F(x, y) = \Pr\{X \leq x, Y \leq y\} = \sum_{x_i \leq x, y_j \leq y} p_{ij} \quad (3.7)$$

#### 边缘分布律

通过联合概率分布, 我们可以得到  $X, Y$  各自的概率分布:

即行和, 列和

$$p_{i \cdot} = \Pr\{X = x_i\} = \sum_j p_{ij} \quad (3.8)$$

$$p_{\cdot j} = \Pr\{Y = y_j\} = \sum_i p_{ij} \quad (3.9)$$

这称为  $(X, Y)$  的**边缘分布**.

#### Question 5: 例 1

设随机整数变量  $X = 1, 2, 3, 4, Y = 1 \cdots X$  等可能地取值, 试求  $(X, Y)$  的分布律.

**Solution:**

$$\begin{aligned} \text{由题意可得 } \Pr\{X = i, Y = j\} &= \Pr\{Y = j | X = i\} \Pr\{X = i\} \quad (1.6) \\ &= \frac{1}{i} \cdot \frac{1}{4}, i = 1, 2, 3, 4, j \leq i. \end{aligned}$$

$X, Y$	1	2	3	4
1	1/4	0	0	0
2	1/8	1/8	0	0
3	1/12	1/12	1/12	0
4	1/16	1/16	1/16	1/16

### 3.1.4 二维连续型随机变量及其概率密度

#### Definition 3.1.4: 二维连续型随机变量

设  $(X, Y)$  是二维随机变量,  $F(X, Y)$  为其分布函数, 若存在非负可积函数  $f(x, y)$  s.t.

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

则称  $(X, Y)$  为**二维连续型随机变量**, 称  $f(x, y)$  为其**概率密度**

显然,  $f(x, y)$  满足性质: ①  $f(x, y) \geq 0$ ; ②  $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) dx dy = F(+\infty, +\infty) = 1$ .

**重要性质** 设  $D$  是  $xOy$  平面上的区域,  $(X, Y)$  落入其中的概率为

$$\Pr\{(X, Y) \in D\} = \iint_D f(x, y) dx dy \quad (3.10)$$

若  $f(x, y)$  在  $(x, y)$  处连续, 则有

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y). \quad (3.11)$$

#### 边缘概率密度

我们知道此处的  $X$  和  $Y$  都是都是连续型随机变量, 可得它们的**边缘密度**为

$$f_X(x) = \int_{-\infty}^{+\infty} f(x, y) dy; \quad f_Y(y) = \int_{-\infty}^{+\infty} f(x, y) dx; \quad (3.12)$$

结合之前的性质, 我们可以知道对  $F(x, y)$  求一个变量的偏导可以得到另一个变量的边缘密度.

#### Exercise 3.1.1 习题 3-1-6 改

设随机变量  $(X, Y)$  的概率密度为

$$f(x, y) = \begin{cases} k(6 - x - y), & 0 < x < 2, 2 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

求① 常数  $k$ ; ②  $\Pr\{X + Y \leq 4\}$ ; ③  $X, Y$  的边缘密度.

**Solution:** (1) 因为

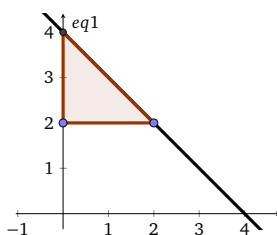
$$\iint_{\mathbb{R}^2} f(x, y) dx dy = \begin{cases} \iint_D k(6 - x - y) dx dy, & x \in D = \{(x, y) | x \in (0, 2) y \in (2, 4)\} \\ 0, & \text{otherwise} \end{cases} = 1$$

$$\text{i.e. } \iint_D k(6 - x - y) dx dy = \int_0^2 \int_2^4 k(6 - x - y) dx dy = 1, \text{ 计算得 } k = \frac{1}{8}.$$

通常都是分段函数, 且一段为0 (不然很难算)

(2) 由题意, 所有符合的  $(X, Y)$  至少落在直线  $x + y = 4$  的左侧. 由 (1) 知点只可能落入  $D$  中, 故令  $G = D \cap \{x + y \leq 4\}$ , 即下图棕色区域

w/ Geogebra



那么, 我们有

$$k \cdot \iint_G 6 - x - y \, dx \, dy = \Pr\{X + Y \leq 4\} = \frac{2}{3} \text{ 即为所求.}$$

定义 (3.1.4) (3) 由边缘密度的定义可以得到

$$f_X(x) = \begin{cases} \frac{1}{8} \cdot \int_2^{4-x} (6 - x - y) \, dy = \frac{1}{8}(6 - 4x + \frac{1}{2}x^2), & 0 < x < 2 \\ 0, & \text{otherwise} \end{cases}$$

相当于固定一个变量, 对另一个变量积分

同理可得

$$f_Y(y) = \begin{cases} \frac{1}{8} \cdot \int_0^{4-y} (6 - x - y) \, dx = (\dots \text{略}), & 2 < y < 4 \\ 0, & \text{otherwise} \end{cases}$$

### 3.1.5 二维均匀分布

#### Definition 3.1.5: 二维均匀分布

设  $G$  是平面上的有界区域, 其面积为  $A$ , 若二维随机变量  $(X, Y)$  具有概率密度函数

$$f(x, y) = \begin{cases} \frac{1}{A}, & (x, y) \in G \\ 0, & \text{otherwise} \end{cases} \quad (3.13)$$

则称  $(X, Y)$  在  $G$  上服从**二维均匀分布**.

若  $(X, Y)$  在  $G$  上服从均匀分布, 则其概率密度函数反映在几何上为定义在  $xOy$  平面内区域  $G$  上的空间的一块平面. 考虑向平面  $G$  上投掷一质点, 若点落在区域  $B \subseteq G$  内的概率与  $B$  的面积成正比且与  $B$  的位置无关, 则称  $G$  为均匀分布的区域, 坐标  $(X, Y)$  在  $G$  上服从均匀分布.

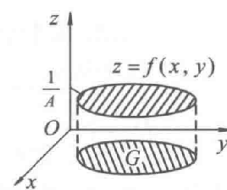


图 3-1-5

#### Corollary 3.1.1

均匀分布  $(X, Y)$  的两个边缘分布仍为均匀分布且分别为

$$\begin{aligned} f_X(x) &= \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases} \\ f_Y(y) &= \begin{cases} \frac{1}{d-c}, & c < y < d \\ 0, & \text{otherwise} \end{cases} \end{aligned} \quad (3.14)$$

对于**矩形域**  $G$  该公式成立, 其他形状则不一定.

Figure 3.2: 均匀分布

#### Question 6: 例 5

设  $(X, Y)$  服从单位圆域上的均匀分布, 求对应的两个边缘分布.

**Solution:** 由题意得, 其 PDF 为

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

(3.12) 只需考虑点在圆域内的情况, 易得

$$f_X(x) = \frac{1}{\pi} \int_{-\sqrt{1-x^2}}^{\sqrt{1-x^2}} dy = \frac{2}{\pi} \sqrt{1-x^2}.$$

$$\therefore f_X(x) = \begin{cases} \frac{2}{\pi} \sqrt{1-x^2}, & -1 \leq x \leq 1 \\ 0, & \text{otherwise} \end{cases}$$

因为  $x^2 + y^2 \leq 1$  是轮换式, 将  $x$  换成  $y$  即可得到  $f_Y(y)$ .

## 3.2 随机变量的独立性

### Definition 3.2.1: 独立性

设  $(X, Y)$  是二维随机变量, 若对于任意  $x, y$ ,

$$\Pr\{X \leq x, Y \leq y\} = \Pr\{X \leq x\} \Pr\{Y \leq y\}$$

i.e.  $F(x, y) = F_X(x)F_Y(y)$  充要条件

(3.15)

则称  $X$  和  $Y$  是**相互独立**的.

### Theorem 3.2.1

随机变量  $X, Y$  相互独立的充要条件是对  $X$  和  $Y$  所生成的任何事件相互独立, 即

$$\Pr\{X \in A, Y \in B\} = \Pr\{X \in A\} \Pr\{Y \in B\}$$
(3.16)

### Theorem 3.2.2

如果  $X, Y$  相互独立, 则对任意函数  $g_1(x), g_2(y)$  均有两者相互独立.

### 3.2.1 离散型随机变量的独立性

若  $X, Y$  是离散型随机变量, 可以由此判断其独立性:

**定义 2** 若对  $(X, Y)$  的所有可能取值  $(X_i, Y_j)$  有

$$\Pr\{X = X_i, Y = Y_j\} = \Pr\{X = X_i\} \Pr\{Y = Y_j\},$$

i.e.  $p_{ij} = p_{i \cdot} p_{\cdot j}$

(3.17)

则称  $X, Y$  相互独立.

#### Exercise 3.2.1 P<sub>77</sub>-11

设  $X, Y$  相互独立, 完成下面的联合分布:

$Y, X$	$x_1$	$x_2$	$x_3$	$p_{i \cdot}$
$y_1$	$a$	$1/9$	$c$	
$y_2$	$1/9$	$b$	$1/3$	
$p_{\cdot j}$				$1$

**Solution:** 首先求  $a, b, c$ :

$$\because p_{21} = \frac{1}{9} = p_2 \cdot p_{\cdot 1}$$

$$p_{23} = \frac{1}{3} = p_2 \cdot p_{\cdot 3}$$

联立上两式可得  $c = 3a$ . 又

$$\sum_i \sum_j p_{ij} = 1$$

得  $a + b + c = \frac{4}{9}$ . 又

$$p_{22} = b = p_2 \cdot p_{\cdot 2}$$

解得  $b = \frac{2}{9}$ .

注意计算准确 综上可以得到  $a = \frac{1}{18}, b = \frac{2}{9}, c = \frac{1}{6}$ . 即

$Y, X$	$x_1$	$x_2$	$x_3$	$p_{\cdot i}$
$y_1$	1/18	1/9	1/6	1/3
$y_2$	1/9	2/9	1/3	2/3
$p_{\cdot j}$	1/6	1/3	1/2	1

### 3.2.2 连续型随机变量的独立性

若  $X, Y$  是连续型随机变量, 可以由此判断其独立性:

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

严谨的说法是:  
几乎处处成立, 亦  
即不包括平面上面  
积为 0 的集合.

则称  $X, Y$  相互独立.

#### Proposition 3.2.1

(3.18)和(3.15)是等价命题.

**证明:** (充分性) 对(3.18)两边积分有

$$\begin{aligned} \int_{-\infty}^x \int_{-\infty}^y f(x, y) dx dy &= F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_X(x)f_Y(y) dx dy \\ &= \int_{-\infty}^x f_X(x) dx \int_{-\infty}^y f_Y(y) dy \quad (\text{分离变量}) \\ &= F_X(x)F_Y(y). \end{aligned}$$

(必要性) 对(3.15)两边求两次偏导有

$$\frac{\partial^2 F(x, y)}{\partial x \partial y} = f(x, y) = \frac{\partial F_X}{\partial x} \cdot \frac{\partial F_Y}{\partial y} = f_X(x)f_Y(y)$$

同时, 我们通过上述证明, 得到了重要的推论:

#### Corollary 3.2.1 快速判断独立性

若  $f(x, y)$  是可分离变量的, 且  $x, y$  的取值范围独立  $\iff X, Y$  独立.



Exercise 3.2.2 P<sub>78</sub>14

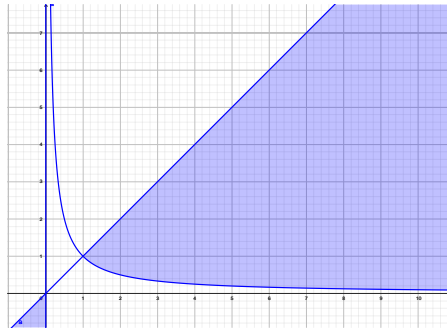
$$f(x, y) = \begin{cases} \frac{1}{2x^2y}, & 1 \leq x \leq +\infty, \frac{1}{x} \leq y \leq x, \\ 0, & \text{otherwise} \end{cases}$$

判断  $X, Y$  是否独立.

**Solution:** 要判断其独立性, 先求两个边缘密度:

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

对于  $Y$ , 在下图第一象限阴影区域内对  $x$  积分



此处的  $x$  的范围并不是  $[1, +\infty]$ . 应将  $D$  投影到  $y$  轴上来讨论  $x$  的范围 (想累次积分定义)

设积分区域 (阴影部分) 为  $D$ , 则将  $D$  投影到  $y$  轴上. (复习)

$$D = D_1 + D_2 = \{x \geq y, 1 \leq y\} + \{x \geq 1/y, 0 < y \leq 1\}$$

$$\begin{aligned} \text{i.e. } f_Y(y) &= \int_{D_1} f(x, y) dx + \int_{D_2} f(x, y) dx \\ &= \begin{cases} \int_{1/y}^{+\infty} \frac{1}{2x^2y} dx, & 0 < y \leq 1, \\ \int_y^{+\infty} \frac{1}{2x^2y} dx, & 1 \leq y, \\ 0, & \text{otherwise} \end{cases} = (\dots) \end{aligned}$$

要写出乘积. 若判断独立性先枚举特值找反例.

两者相乘, 显然不与  $f(x, y)$  相等.  $\therefore XY$  不独立.



# Chapter 4

## 随机变量的数字特征

### 4.1 数学期望

#### 4.1.1 随机变量的数学期望

##### Definition 4.1.1

设离散型随机变量  $X$  的概率分布为

$$\Pr\{X = x_i\} = p_i, \quad i = 1, 2, \dots$$

若级数  $\sum_{i=1}^{\infty} x_i p_i$  绝对收敛, 则定义  $X$  的**数学期望**为

$$\mathbb{E}[X] = \sum_i x_i p_i. \quad (4.1)$$

复习: 绝对收敛:  
 $\sum_{i=1}^{+\infty} |x_i p_i| < \infty$

$\mathbb{E}[X], E(X), EX$   
都是等价的写法.

上述定义限于离散型随机变量. 考虑连续型随机变量  $X$ , 其 PDF 为  $f(x)$ . 在数轴上取一段长度为  $\Delta x \rightarrow 0$  的区间  $[x, x + \Delta x]$ , 则  $X$  落入该区间的概率为

$$\Pr\{x \leq X \leq x + \Delta x\} = \int_x^{x+\Delta x} f(x) dx \approx f(x)\Delta x.$$

这称为离散化 在数轴上取无限多段这样的小区间并对概率求和, 我们有

$$\sum_i x_i f(x_i) \Delta x_i \approx \int_{\mathbb{R}} x f(x) dx.$$

并非所有随机变量  
都有数学期望

因此可以得到定义:

##### Definition 4.1.2

设  $X$  是连续型随机变量, PDF 为  $f(x)$ , 若上述积分绝对收敛, 定义  $X$  的**数学期望**为

$$\mathbb{E}[X] = \int_{\mathbb{R}} x f(x) dx. \quad (4.2)$$

#### 4.1.2 随机变量函数的数学期望

对于随机变量  $X$  的函数  $Y = g(X)$ , 我们可以通过定义求出  $X$  的分布而求出  $Y$  的分布, 进而由定义得  $Y$  的数学期望  $\mathbb{E}[g(X)]$ , 这做法较繁.

### Theorem 4.1.1

设  $X$  是一个随机变量,  $Y = g(X)$ , 且  $E(Y)$  存在, 于是

(1) 若  $X$  为离散型随机变量:

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \sum_i g(x_i)p_i, \quad (4.3)$$

(2) 若  $X$  为连续型随机变量:

$$\mathbb{E}[Y] = \mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x) dx. \quad (4.4)$$

推广至二维: 设  $(X, Y)$  是二维随机变量,  $Z = g(X, Y)$ , 若  $\mathbb{E}[Z]$  存在, 则

(1) 若为离散型:

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \sum_i \sum_j g(x_i, y_j)p_{ij}, \quad (4.5)$$

(2) 若为连续型:

$$\mathbb{E}[Z] = \mathbb{E}[g(X, Y)] = \iint_{\mathbb{R}^2} g(x, y)f(x, y) d\sigma. \quad (4.6)$$

**注意** 实际做题中, 积分区域  $D \subseteq \mathbb{R}$ , 根据所要求期望的随机变量来决定积分次序. E.g.  $\mathbb{E}[X]$ —先对  $y$  积分, 再对  $x$  积分;  $\mathbb{E}[Y]$ —先对  $x$  积分, 再对  $y$  积分;  $\mathbb{E}[XY]$ —两者谁先皆可, 但注意不满足 Fubini 定理的函数.

### 4.1.3 数学期望的性质

1.  $\mathbb{E}[c] = c$  ( $c$  为常数, 下同);
2.  $\mathbb{E}[cX] = c\mathbb{E}[X]$ ;
3.  $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$ ;
4. 若  $X, Y$  相互独立, 则  $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$ .

#### Exercise 4.1.1 P<sub>85</sub> 5

设随机变量  $X$  的分布律为

$X$	-2	0	2
$p_i$	0.4	0.3	0.3

求  $\mathbb{E}[X], \mathbb{E}[X^2], \mathbb{E}[3X^2 + 5]$

**Solution:** 由题意得,  $\mathbb{E}[X] = -0.2$ . 设  $Y = X^2$ , 由(4.5)得  $\mathbb{E}[X^2] = 2.8$ . 由性质 123 得  $\mathbb{E}[3X^2 + 5] = 3\mathbb{E}[X^2] + \mathbb{E}[5] = 13.4$ .

#### Exercise 4.1.2 P<sub>85</sub> 10

设  $(X, Y)$  的概率密度为

$$f(x, y) = \begin{cases} 12y^2, & 0 \leq y \leq x \leq 1, \\ 0, & \text{otherwise} \end{cases}$$

求  $\mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[XY], \mathbb{E}[X^2 + Y^2]$ .

**Solution:** 由题意得, 实际的积分区域为  $x$  轴,  $x = 1$  和  $y = x$  围成的三角形区域  $D$ .

(1) 求  $\mathbb{E}[X]$ , 由定义可得  $\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf_X(x) dx$ , 展开有

$$\mathbb{E}[X] = \iint_{\mathbb{R}^2} xf(x, y) dx dy$$

$D_1 = D_2$ , 分开写 注意到  $f(x, y)$  在  $D_1 = \{0 \leq x \leq 1, 0 \leq y \leq x\}$  上不为 0, 计算得  $\mathbb{E}[X] = \frac{4}{5}$ .  
是强调积分次序.

(2) 同理, 求  $\mathbb{E}[Y]$ ,  $D_2 = \{0 \leq y \leq 1, y \leq x \leq 1\}$ . 可得

$$\mathbb{E}[Y] = \underbrace{\int_0^1 \int_y^1}_{D_2} y f(x, y) dx dy = \frac{3}{5}.$$

这个次序计算简单 (3) 由于  $f_X(x)f_Y(y) \neq f(x, y)$ ,  $X, Y$  不具有独立性, 无法使用性质 4. 由于积分次序不影响结果, 因此先积  $y$  再积  $x$  可得

$$\mathbb{E}[XY] = \iint_{D_1} xy f(x, y) d\sigma = \frac{1}{2}.$$

(4) 由性质 3 得  $\mathbb{E}[X^2 + Y^2] = \mathbb{E}[X^2] + \mathbb{E}[Y^2]$ . 由 (1)(2) 可得

$$\mathbb{E}[X^2 + Y^2] = \iint_{D_1} x^2 f(x, y) d\sigma + \iint_{D_2} y^2 f(x, y) d\sigma = \frac{16}{15}.$$

## 4.2 方差

### 4.2.1 方差的定义

#### Definition 4.2.1: 方差

设  $X$  是一个随机变量, 若  $\mathbb{E}^2[X - \mu]$  存在, 则称其为  $X$  的**方差**, 记为  $\mathbb{V}[X]$ , 其中  $\mu = \mathbb{E}[X]$ . 方差亦记作  $\text{Var } X, D(X), DX$ .

方差的算术平方根  $\sqrt{\mathbb{V}[X]}$  称为  $X$  的**标准差**, 记为  $\sigma_X$ .

常见分布的期望和方差见书 P<sub>215</sub>

方差刻画了随机变量  $X$  的取值与其期望  $\mu$  的偏离程度, 方差越大, 随机变量的取值越分散. 而且, 若  $\mathbb{V}[X] = 0$ , 则随机变量  $X$  一定取常数值, 此时  $X$  不是随机变量.

### 4.2.2 方差的计算

若  $X$  是离散型随机变量, 其分布律为  $\Pr\{X = x_i\} = p_i, i = 1, 2, \dots$ , 则

$$\mathbb{V}[X] = \mathbb{E}^2[X - \mu] = \sum_i (x_i - \mu)^2 p_i. \quad (4.7)$$

若  $X$  是连续型随机变量, 其概率密度为  $f(x)$ , 则

$$\mathbb{V}[X] = \mathbb{E}^2[X - \mu] = \int_{\mathbb{R}} (x - \mu)^2 f(x) dx. \quad (4.8)$$

由数学期望的性质可得方差计算的重要公式:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X]. \quad (4.9)$$

证明:

$$\begin{aligned} \mathbb{V}[X] &= \mathbb{E}^2[X - \mu] = \mathbb{E}[X^2 - 2\mu X + \mu^2] \\ &= \mathbb{E}[X^2] - 2\mu \mathbb{E}[X] + \mu^2 \\ &= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 \\ &= \mathbb{E}[X^2] - \mathbb{E}^2[X]. \end{aligned}$$



### Question 7: Poisson 分布

设  $X \sim P(\lambda)$ , 求  $\mathbb{E}[X]$ ,  $\mathbb{V}[X]$ .

**Solution:** 随机变量  $X$  的分布律为

$$\Pr\{X = k\} = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots; \lambda > 0.$$

$$\text{由定义得 } \mathbb{E}[X] = \sum_{k=0}^{\infty} k \frac{\lambda^k e^{-\lambda}}{k!} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} = \lambda e^{-\lambda} \cdot e^{\lambda} = \lambda,$$

$$\begin{aligned} \mathbb{E}[X^2] &= \mathbb{E}[X(X-1) + X] = \mathbb{E}[X(X-1)] + \mathbb{E}[X] \\ &= \sum_{k=0}^{\infty} k(k-1) \frac{\lambda^k e^{-\lambda}}{k!} + \lambda = \lambda^2 e^{-\lambda} \underbrace{\sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!}}_{\text{Poisson 分布}=1} + \lambda \\ &= \lambda^2 e^{-\lambda} \cdot e^{\lambda} + \lambda = \lambda^2 + \lambda, \end{aligned}$$

$$\text{故方差 } \mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \lambda. \quad (4.10)$$

### Question 8: 均匀分布

设  $X \sim U(a, b)$ , 求  $\mathbb{E}[X]$ ,  $\mathbb{V}[X]$ .

**Solution:**  $X$  的概率密度为

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

$$\text{而 } \mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}^2[X] = \int_a^b x^2 \frac{1}{b-a} dx - \left(\frac{a+b}{2}\right)^2 = \frac{(b-a)^2}{12}. \quad (4.11)$$

#### 4.2.3 方差的性质

1.  $\mathbb{V}[c] = 0$ ;
2.  $\mathbb{V}[X] \geq 0$ ;
3.  $\mathbb{V}[cX] = c^2 \mathbb{V}[X]$ ;
4.  $\mathbb{V}[c + X] = \mathbb{V}[X]$ ;
5.  $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y] + 2 \text{Cov}(X, Y)$ ;
6. 若  $X, Y$  相互独立, 则  $\mathbb{V}[X + Y] = \mathbb{V}[X] + \mathbb{V}[Y]$ .

第六个性质可以推广到  $n$  维, 即若  $X_1, X_2, \dots, X_n$  相互独立, 则

$$\mathbb{V}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{V}[X_i].$$

#### Exercise 4.2.1 P<sub>90</sub>8

设  $X \sim N(1, 2)$ ,  $Y \sim E(3)$ , 且  $X, Y$  相互独立, 求  $\mathbb{V}[XY]$ .

**Solution:** (分析法) 要求  $\mathbb{V}[XY]$ , 即求  $\mathbb{E}[XY^2] - \mathbb{E}^2[XY]$ . 因为  $X, Y$  独立, 则

$$\mathbb{V}[XY] = \mathbb{E}[XY^2] - \mathbb{E}^2[X] \mathbb{E}^2[Y].$$

因为  $X, Y$  独立, 则  $X^2, Y^2$  也独立. 有

定理 (3.2.2) 
$$= \mathbb{E}[X^2] \mathbb{E}[Y^2] - \mathbb{E}^2[X] \mathbb{E}^2[Y].$$

$X^2, Y^2$  这俩玩意积不出来, 继续展开有

$$= (\mathbb{E}^2[X] + \mathbb{V}[X]) (\mathbb{E}^2[Y] + \mathbb{V}[Y]) - \mathbb{E}^2[X] \mathbb{E}^2[Y].$$

已知  $\mathbb{E}[X] = 1, \mathbb{E}[Y] = 3$ , 得  $= 27$ .

若设  $f(x) = \mathbb{E}^2[X - x], x \in \mathbb{R}$ , 求导可证明  $f(x)$  在  $x = \mu$  处取得最小值, 即  $\mathbb{V}[X]$  在  $\mu$  处取得最小值, 这说明随机变量的取值对其数学期望的偏离程度比其他任何值的偏离程度都要小.

### Theorem 4.2.1

若随机变量序列  $\{X_n\} \sim N$ , 且相互独立, 则其和常数的线性组合仍然服从正态分布.

$$\text{i.e. } \mathbf{A} \times \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \sim N, \text{ where } \mathbf{A} = (a_1 \ a_2 \ \cdots \ a_n). \quad (4.12)$$

### Example 4.2.1 (P<sub>91</sub>4)

若  $X_i \sim N(\mu_i, \sigma_i^2), i = 1, 2, \dots, n$ , 且相互独立, 求  $Y = \sum_{i=1}^n (a_i X_i + b_i)$  服从什么分布.

定理 (4.2.1) **Solution:** 由题意,  $Y$  是  $X$  的一个线性组合, 所以  $Y$  服从正态分布.

$$\mathbb{E}[Y] = \sum_{i=1}^n a_i \mathbb{E}[X_i] + \sum_{i=1}^n b_i,$$

$$\mathbb{V}[Y] = \sum_{i=1}^n a_i^2 \mathbb{V}[X_i].$$

$$\therefore Y \sim N \left( \sum_{i=1}^n (a_i \mu_i + b_i), \sum_{i=1}^n a_i^2 \sigma_i^2 \right)$$

## 4.3 大数定律与中心极限定理

### 4.3.1 Chebyshev 不等式

#### Theorem 4.3.1 Chebyshev 不等式

设随机变量  $X$  的期望为  $\mu$ , 方差为  $\sigma^2$ , 则对任意  $\varepsilon > 0$ , 有

$$\Pr\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}. \quad (4.13)$$

这个不等式从直观上表示事件  $X$  的取值都落在  $\mu$  的一个邻域  $(\mu - \varepsilon, \mu + \varepsilon)$  之外 (阴影部分) 的概率. 它同时指出随机变量  $X$  的方差越小, 事件  $\{|X - \mu| < \varepsilon\}$  发生的概率越大, 即  $X$  的取值基本上集中在  $\mu$  的邻域内. 因此称方差刻画了随机变量的离散程度.

只证连续型的情况. 设  $X$  的概率密度为  $f(x)$ , 则

**证明:**

$$\Pr\{|X - \mu| \geq \varepsilon\} = \int_{|x-\mu| \geq \varepsilon} f(x) dx.$$

放缩, 取  $k = \frac{|x - \mu|^2}{\varepsilon^2}$ . (因为  $X$  都落在邻域外, 则  $k > 1$ .)

$$\leq \int_{|x-\mu| \geq \varepsilon} kf(x) dx$$

再次放缩, 变大区间

$$\leq \frac{1}{\varepsilon^2} \underbrace{\int_{\mathbb{R}} (x - \mu)^2 f(x) dx}_{\mathbb{V}[X]} = \frac{\sigma^2}{\varepsilon^2}$$

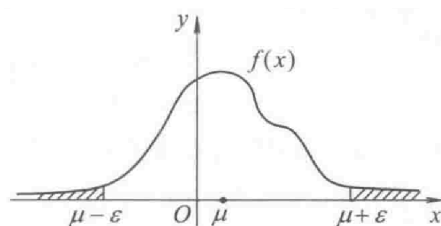


图 4-4-1

Figure 4.1:  $f(x)$  示意

### 4.3.2 大数定律

#### Lemma 4.3.1 随机变量序列相互独立

若对于任意  $n > 1$ ,  $X_1, X_2, \dots, X_n$  都相互独立, 则称  $X_1, X_2, \dots, X_n, \dots$  是相互独立的.

#### Theorem 4.3.2 大数定律

设随机变量  $X_1, X_2, \dots, X_n, \dots$  相互独立, 且具有相同的期望和方差

$$\mathbb{E}[X_i] = \mu, \mathbb{V}[X_i] = \sigma^2, i = 1, 2, \dots$$

记  $Y_n = \frac{1}{n} \sum_{i=1}^n X_i$ , 则对任意  $\varepsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} \Pr\{|Y_n - \mu| < \varepsilon\} = 1 \quad (4.14)$$

**证明:** 由  $Y_n$  的式子可以知道

$$\mathbb{E}[Y_n] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}[X_i] = \mu, \quad \mathbb{V}[Y_n] = \frac{1}{n^2} \sum_{i=1}^n \mathbb{V}[X_i] = \frac{\sigma^2}{n},$$

由 Chebyshev 不等式可得

$$\begin{aligned} \Pr\{|Y_n - \mu| < \varepsilon\} &\geq 1 - \Pr\{|Y_n - \mu| \geq \varepsilon\} \\ &\geq 1 - \frac{\mathbb{V}[Y_n]}{\varepsilon^2} \\ &\geq 1 - \frac{\sigma^2}{n\varepsilon^2}. \end{aligned} \quad (*)$$

夹逼法, (\*) 式两边取极限, 且因为概率不可能大于 1,

$$\begin{aligned} 1 &\geq \lim_{n \rightarrow \infty} \Pr\{|Y_n - \mu| < \varepsilon\} \geq \lim_{n \rightarrow \infty} \left(1 - \frac{\sigma^2}{n\varepsilon^2}\right) = 1. \\ \therefore \lim_{n \rightarrow \infty} \Pr\{|Y_n - \mu| < \varepsilon\} &= 1. \end{aligned}$$

这表明对任意  $\varepsilon > 0$ , 事件  $|Y_n - \mu| < \varepsilon$  发生的概率很大 ( $Y_n$  的取值一定落在  $\mu$  的邻域内), 当  $n$  很大时,  $Y_n$  集中于  $\mu$ . 像这样表现的收敛性成为随机变量序列  $Y_1, Y_2, \dots, Y_n, \dots$  依概率收敛于  $\mu$ , 记为

$$Y_n \xrightarrow{P} \mu. \quad (4.15)$$

这还表明, 随机变量  $X_1, X_2, \dots, X_n$  的算术平均值序列  $Y_n$  依概率收敛于  $\mu$ .

#### Corollary 4.3.1 Bernoulli 定理

设  $n_A$  是  $n$  重 Bernoulli 试验中事件  $A$  发生的次数,  $p$  是事件  $A$  在每次试验中发生的概率 ( $n_A \sim b(n, p)$ ), 则对任意的  $\varepsilon > 0$ , 有

$$\lim_{n \rightarrow \infty} \Pr \left\{ \left| \frac{n_A}{n} - p \right| < \varepsilon \right\} = 1. \quad (4.16)$$

证明同上.

$\frac{n_A}{n}$  是事件  $A$  发生的频率, 这个定理表明当  $n \rightarrow \infty$  时, 这个频率依概率收敛于事件  $A$  发生的概率  $p$ . 这说明频率的本质是一个随机变量.

### 4.3.3 中心极限定理 CLT

#### Theorem 4.3.3 中心极限定理

设随机变量  $X_1, X_2, \dots, X_n, \dots$  相互独立, 同分布, 且具有相同的期望  $\mu$  和方差  $\sigma^2$ , 记  $Y_n = \sum_{i=1}^n X_i$ , 则对任意  $x \in \mathbb{R}$ , 有

$$\lim_{n \rightarrow \infty} \Pr \left\{ \underbrace{\frac{Y_n - n\mu}{\sigma\sqrt{n}} \leq x}_{CDF} \right\} = \Phi(x). \quad (4.17)$$

其中  $\Phi(x)$  是标准正态分布的分布函数,  $\int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ . 这个定理实际上由 De Moivre-Laplace (二项分布) 定理和 Lindeberg-Levy (iid) 定理组成.

所以无论  $X$  服从什么分布, 只要  $n$  足够大, 都可以用正态分布来近似.

这个定理表明: 当  $n$  很大时, 随机变量  $Y_n$  的分布近似于正态分布  $N(n\mu, n\sigma^2)$ .

$$\text{i.e. } Y_n = \sum_{i=1}^n X_i \overset{\text{近似}}{\sim} N(n\mu, n\sigma^2). \xrightarrow{\text{定理(2.4.1)}} \frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{\text{近似}}{\sim} N(0, 1). \quad (4.18)$$

#### Note

复习定理 (2.4.1), 本节和下一章只要涉及到转化为标准正态分布都会经常用到. 可联系期望和方差的性质助记.

#### Exercise 4.3.1 P<sub>108</sub>26

设各零件的重量都是随机变量, 互相独立, 同分布, 期望为 0.5, 均方差为 0.1, 求 5000 个这样的零件的总重量超过 2510 的概率是多少.

**Solution:** 设第  $i$  个零件重量是  $X_i$ . 则  $X = \sum_{i=1}^{5000} X_i$ .  $\because X_i$  独立同分布, 由中心极限定理,

$$\begin{aligned} X &\sim N(0.5 \cdot 5000, 0.1 \cdot \sqrt{5000}), \\ \Pr\{X > 2510\} &= \Pr\left\{ \frac{X - 2500}{\sqrt{5000} \cdot 0.1} > \frac{2510 - 2500}{\sqrt{5000} \cdot 0.1} \right\}, \\ &= \Pr\{X > \sqrt{2}\} \approx 1 - \Phi(1.414). \end{aligned}$$

**Exercise 4.3.2** P<sub>108</sub>27

有 10000 盏功率相同的灯, 每一盏灯开着的概率是 0.7. 假设各盏灯的开关是相互独立的, 求开着的灯的盏数在 6800 到 7200 之间的概率.

**Solution:**

根据题意

$$X \sim B(10000, 0.7). \mu = 7000, \sigma^2 = 2100.$$

由中心极限定理

$$\begin{aligned} \frac{X - 7000}{\sqrt{2100}} &\sim N(0, 1). \\ \therefore \Pr\{6800 < X < 7200\} &= \Pr\left\{\frac{6800 - 7000}{\sqrt{2100}} < \frac{X - 7000}{\sqrt{2100}} < \frac{7200 - 7000}{\sqrt{2100}}\right\} \\ &\approx \Phi\left(\frac{20}{\sqrt{21}}\right) - \Phi\left(-\frac{20}{\sqrt{21}}\right) \\ &= 2\Phi\left(\frac{20}{\sqrt{21}}\right) - 1. \end{aligned}$$

**Note**

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

(4.19)



# Chapter 5

## 数理统计的基础知识

### 5.1 基本概念

#### 5.1.1 统计量

##### Definition 5.1.1: 统计量

设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的样本, 称此样本的任一不含总体分布未知参数的函数为该样本的统计量.

##### Example 5.1.1

设  $X \sim N(5, \sigma^2)$ ,  $\sigma$  未知.  $X_1, X_2, \dots, X_n$  为总体  $X$  的一个样本, 令

$$S_n = X_1 + X_2 + \dots + X_n, \quad \bar{X} = \frac{S_n}{n},$$

则  $S_n$  和  $\bar{X}$  都是样本的统计量. 但  $U = \frac{n(\bar{X} - 5)}{\sigma}$  不是统计量, 因为它含有未知参数  $\sigma$ .

#### 5.1.2 常用统计量

样本均值

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (5.1)$$

样本方差 (有偏, 无偏)

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \quad S^2 = \frac{1}{\underbrace{n-1}} \sum_{i=1}^n (X_i - \bar{X})^2 \quad (5.2)$$

样本标准差

$$S = \sqrt{S^2} \quad (5.3)$$

样本  $k$  阶原点矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k \in \mathbb{Z}_+ \quad (5.4)$$

当  $k = 1$  时,  $A_1 = \bar{X}$ .

样本  $k$  阶中心矩

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k \in \mathbb{Z}_+ \quad (5.5)$$

当  $k = 2$  时,  $\frac{n}{n-1}B_2 = S^2$ .

## 5.2 常用统计分布

### 5.2.1 卡方分布

#### Definition 5.2.1: $\chi^2$ 分布

设  $X_1, X_2, \dots, X_n$  是来自  $N(0, 1)$  的样本, 则称随机变量

$$\chi^2 = \underbrace{X_1^2 + X_2^2 + \dots + X_n^2}_{n \text{项}} \quad (5.6)$$

构造性定义

服从自由度为  $n$  的  $\chi^2$  分布, 记为  $\chi^2 \sim \chi^2(n)$ .

$\chi^2$  分布的 PDF 为

$$f(x) = \begin{cases} \frac{1}{2^{n/2} \Gamma\left(\frac{n}{2}\right)} x^{n/2-1} e^{-x/2}, & x > 0, \\ 0, & x \leq 0. \end{cases} \quad (5.7)$$

其中  $\Gamma$  函数定义为

$$\Gamma(\alpha) = \int_0^{+\infty} t^{\alpha-1} e^{-t} dt, \quad x > 0. \quad (5.8)$$

#### Question 9: 例 P<sub>123</sub>2

设  $X_1, \dots, X_6$  是来自总体  $N(0, 1)$  的样本, 又设

$$Y = (X_1 + X_2 + X_3)^2 + (X_4 + X_5 + X_6)^2,$$

试求常数  $C$ , s.t.  $CY \sim \chi^2$ .

**Solution:** 因为  $X_1 + X_2 + X_3 \sim N(0, 3)$ ,  $X_4 + X_5 + X_6 \sim N(0, 3)$ , 所以

$$\frac{X_1 + X_2 + X_3}{\sqrt{3}} \sim N(0, 1), \quad \frac{X_4 + X_5 + X_6}{\sqrt{3}} \sim N(0, 1),$$

定理 (3.2.2) 因为这两项相互独立,

$$\left( \frac{X_1 + X_2 + X_3}{\sqrt{3}} \right)^2 + \left( \frac{X_4 + X_5 + X_6}{\sqrt{3}} \right)^2 \sim \chi^2(2).$$

提公共常数  $C = \frac{1}{3}$  即得. 从而  $\frac{1}{3}Y \sim \chi^2(2)$ .

## 5.2.2 t 分布

### Definition 5.2.2: t 分布

设  $X \sim N(0, 1)$ ,  $Y \sim \chi^2(n)$ , 且  $X$  与  $Y$  相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{Y/n}} \quad (5.9)$$

服从自由度为  $n$  的 **t 分布**, 记为  $t \sim t(n)$ .

$t$  分布的 PDF 如下. 注意到这是一个偶函数.

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \quad x \in \mathbb{R}. \quad (5.10)$$

### Question 10: 例 P<sub>125</sub> 3

设随机变量  $X \sim N(2, 1)$ , 随机变量  $Y_1, Y_2, Y_3, Y_4$  均服从  $N(0, 4)$ , 且  $X, Y_i (i = 1, 2, 3, 4)$  都相互独立, 令

$$T = \frac{4(X-2)}{\sqrt{\sum_{i=1}^4 Y_i^2}},$$

试求  $T$  的分布. 并确定  $t_0$ , 使得  $\Pr\{|T| > t_0\} = 0.01$ .

**Solution:** 由于

$$X - 2 \sim N(0, 1), \quad Y_i/2 \sim N(0, 1), \quad i = 1, 2, 3, 4,$$

根据定义 (5.2.2),

$$T = \frac{4(X-2)}{\sqrt{\sum_{i=1}^4 Y_i^2}} = \frac{X-2}{\sqrt{\sum_{i=1}^4 \left(\frac{Y_i}{4}\right)^2}} = \frac{X-2}{\sqrt{\sum_{i=1}^4 \left(\frac{Y_i}{2}\right)^2 / 4}} \sim t(4)$$

由  $\Pr\{|T| > t_0\} = 0.01$ , 对  $n = 4$ ,  $\alpha = 0.01$ , 查表得  $t_0 = t_{\alpha/2}(4) = t_{0.005}(4) = 4.604$ .

### Exercise 5.2.1 P<sub>127</sub> 3

设  $X, Y$  相互独立且都服从  $N(0, 3^2)$ .  $X_1, \dots, X_9$  和  $Y_1, \dots, Y_9$  是分别取自总体  $X, Y$  的简单随机样本, 试证明

$$T = \frac{\sum_{i=1}^9 X_i}{\sqrt{\sum_{i=1}^9 Y_i^2}} \sim t(9)$$

**证明:** 由题意:  $X_i/3 \sim N(0, 1)$ ,  $Y_i/3 \sim N(0, 1) \Rightarrow \frac{\sum_{i=1}^9 X_i}{3} \sim N(0, 1)$  (①)

$$\text{构造 } \sum_{i=1}^9 Y_i^2 \sim \chi^2(9) \quad (②)$$

定义 (5.2.2) 由①②构造  $T \sim t(9)$ , 其中  $T$  为

$$T = \frac{\left(\sum_{i=1}^9 X_i/3\right)/3}{\sqrt{\sum_{i=1}^9 (Y_i/3)^2/9}} \quad \text{化简有} = \frac{\sum_{i=1}^9 X_i}{\sqrt{\sum_{i=1}^9 Y_i^2}}$$



### 5.2.3 分位数

#### Definition 5.2.3: 分位数

设  $X$  是一个连续型随机变量,  $0 < \alpha < 1$ , 若实数  $x_\alpha$  满足

$$\Pr\{X < x_\alpha\} = \alpha,$$

则称  $x_\alpha$  为  $X$  的水平为  $\alpha$  的**上侧分位数**.

若实数  $x_{\alpha/2}$  满足

$$\Pr\{|X| > x_{\alpha/2}\} = \alpha,$$

则称  $x_{\alpha/2}$  为  $X$  的水平为  $\alpha$  的**双侧分位数**. 一般不直接求解分位数, 对常用的统计分布一般查表来得到分位数的值.

**注** 记  $u_\alpha$  为标准正态分布的水平为  $\alpha$  的上侧分位数,  $\chi_\alpha^2(n)$  为自由度为  $n$  的  $\chi^2$  分布的水平为  $\alpha$  的上侧分位数,  $t_\alpha(n)$  为自由度为  $n$  的  $t$  分布的水平为  $\alpha$  的上侧分位数. 有结论 (利用 PDF 是偶函数的对称性)

$$u_\alpha = -u_{1-\alpha}, \quad t_\alpha(n) = -t_{1-\alpha}(n) \quad (5.11)$$

#### Example 5.2.1 (P<sub>121</sub> 例 1)

设  $\alpha = 0.05$ , 求标准正态分布的水平 0.05 的上侧分位数和双侧分位数.

**Solution:** 由于  $\Phi(u_{0.05}) = 1 - 0.05 = 0.95$ , 查表有  $u_{0.05} = 1.645$ . 双侧分位数  $u_{0.025}$  满足  $\Phi(u_{0.025}) = 1 - 0.025 = 0.975$ , 查表有  $u_{0.025} = 1.96$ .

### 5.2.4 F 分布

#### Definition 5.2.4: F 分布

设  $X \sim \chi^2(m)$ ,  $Y \sim \chi^2(n)$ , 且  $X$  与  $Y$  相互独立, 则称随机变量

$$F = \frac{X/m}{Y/n} \quad (5.12)$$

服从自由度为  $(m, n)$  的**F 分布**, 记为  $F \sim F(m, n)$ .

若  $X \sim t(n)$ , 则  $X^2 \sim F(1, n)$ ; 若  $F \sim F(m, n)$ , 则  $\frac{1}{F} \sim F(n, m)$ .

记  $F$  分布的分位数为  $F_\alpha(m, n)$ , 其中  $a, b$  分别为分位数的上侧概率和下侧概率.

#### Question 11: 证明

$$F_\alpha(m, n) = \frac{1}{F_{1-\alpha}(n, m)}. \quad (5.13)$$

**证明:** 设  $F \sim F(m, n)$ , 移项有

$$\text{定义 (5.2.3)} \quad \Pr\{F < F_\alpha(m, n)\} = \alpha = \Pr\left\{\frac{1}{F} \geq \frac{1}{F_\alpha(m, n)}\right\}.$$

定义 (2.3.1) 将  $1/F$  整体看作一个随机变量,

$$1 - \alpha = \Pr\left\{\frac{1}{F} < \frac{1}{F_\alpha(m, n)}\right\} \quad (*)$$

根据  $F$  分布的性质,  $1/F \sim F(n, m)$ , 其水平为  $1 - \alpha$  的分位数  $F_{1-\alpha}(n, m)$  满足

$$1 - \alpha = \Pr\left\{\frac{1}{F} < F_{1-\alpha}(n, m)\right\}$$

与(\*)式比较,

$$\therefore F_{1-\alpha}(n, m) = \frac{1}{F_{\alpha}(m, n)} \implies F_{\alpha}(m, n) = \frac{1}{F_{1-\alpha}(n, m)}.$$



## 5.3 抽样分布

### 5.3.1 单正态总体的抽样分布

#### Theorem 5.3.1

设总体  $X \sim N(\mu, \sigma)$ ,  $X_1, \dots, X_n$  是取自  $X$  的一个样本,  $\bar{X}$  和  $S^2$  分别是样本均值和样本方差, 则

$$(1) \bar{X} \sim N(\mu, \sigma^2/n); \quad (5.14)$$

$$(2^*) U = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1); \quad (5.15)$$

$$(3^*) \chi^2 = \frac{n-1}{\sigma^2} S^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \bar{X})^2 \sim \chi^2(n-1); \quad (5.16)$$

$$(4) \bar{X} \text{ 与 } S^2 \text{ 相互独立.} \quad (5.17)$$

$$(5) \chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n); \quad (5.18)$$

$$(6^*) T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1) \quad (5.19)$$

标\*号的是后面常用的定理.  
一些常用的记号:

总体  $\mu$  总体均值

•  $\sigma$  总体标准差

样本  $\bar{X}$  样本均值

•  $S$  样本标准差

# Chapter 6

## 参数估计

### 6.1 点估计问题概述

#### 6.1.1 点估计的概念

$\hat{\theta}$  是随机变量,也是样本的函数  
设  $X_1, \dots, X_n$  是取自总体  $X$  的一个样本,  $x_1, \dots, x_n$  是相应的一个样本值,  $\theta$  是总体分布中的未知参数, 为估计未知参数  $\theta$ , 需要构造一个适当的统计量

$$\hat{\theta}(X_1, \dots, X_n) \quad (6.1)$$

并用其观察值

$$\hat{\theta}(x_1, \dots, x_n) \quad (6.2)$$

来估计  $\theta$  的值. 式 (6.1) 和式 (6.2) 分别称为 **点估计量** 和 **点估计值**.

#### 6.1.2 评价估计的标准

无偏性

##### Definition 6.1.1

设  $\hat{\theta}$  是  $\theta$  的一个估计量, 若

$$\mathbb{E}[\hat{\theta}] = \theta, \quad (6.3)$$

则称  $\hat{\theta}$  是  $\theta$  的一个 **无偏估计量**, 否则称为 **有偏估计量**.

将  $\mathbb{E}[\hat{\theta}] - \theta$  称为用  $\hat{\theta}$  估计  $\theta$  的系统偏差. 对一般总体而言, 有

##### Theorem 6.1.1 Slutsky's theorem

设  $X_1, \dots, X_n$  是来自总体  $X$  的样本, 总体  $X$  的均值为  $\mu$ , 方差为  $\sigma^2$ , 则

1. 样本均值  $\bar{X}$  是  $\mu$  的无偏估计量;
2. 样本方差  $S^2$  是  $\sigma^2$  的无偏估计量.
3.  $B_2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$  是  $\sigma^2$  的渐进无偏 ( $n \rightarrow \infty$ ) 估计量.

##### Theorem 6.1.2

若  $\sum_i a_i = 1$ , 则估计量

$$(a_1 \quad \dots \quad a_n) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$$

是  $\mu$  的无偏估计量.

### Exercise 6.1.1 P<sub>140</sub>3

设  $\hat{\theta}$  是参数  $\theta$  的无偏估计. 且  $\mathbb{V}[\hat{\theta}] > 0$ , 试证  $\hat{\theta}^2 = (\hat{\theta})^2$  不是  $\theta^2$  的无偏估计.

**Proof:**

$$(4.9) \quad \begin{aligned} & \because \mathbb{V}[\hat{\theta}] > 0, \\ & \therefore \mathbb{E}[\hat{\theta}^2] - \mathbb{E}^2[\hat{\theta}] > 0. \end{aligned}$$

因为  $\mathbb{E}[\hat{\theta}] = \theta$ , 移项有

$$\therefore \mathbb{E}[\hat{\theta}^2] > \theta^2.$$



### 有效性

在没有系统偏差的情况下, 比较随机误差.

#### Definition 6.1.2

设  $\hat{\theta}_1$  和  $\hat{\theta}_2$  是  $\theta$  的两个无偏估计量, 有

$$\mathbb{V}[\hat{\theta}_1] \leq \mathbb{V}[\hat{\theta}_2], \quad (6.4)$$

则称  $\hat{\theta}_1$  比  $\hat{\theta}_2$  **有效**.

### 相合性

#### Definition 6.1.3

设  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$  为未知参数  $\theta$  的估计量. 若  $\hat{\theta} \xrightarrow{P} \theta$ , i.e.  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} \Pr\{|\hat{\theta}_n - \theta| < \varepsilon\} = 1, \quad (6.5)$$

则称  $\hat{\theta}_n$  是  $\theta$  的一个**相合估计量**.

## 6.2 点估计的常用方法

### 6.2.1 矩估计法

由大数定律我们知道, 当总体的  $k$  阶矩存在时, 样本的  $k$  阶矩依概率收敛于总体的  $k$  阶矩. 譬如说可用样本均值  $\bar{X}$  作为总体均值  $\mathbb{E}[X]$  的估计量. 一般地, 记

1. 总体  $k$  阶原点矩  $\mu_k = \mathbb{E}[X^k]$
2. 样本  $k$  阶原点矩  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$
3. 总体  $k$  阶中心矩  $\nu_k = \mathbb{E}^k[X - \mathbb{E}[X]]$
4. 样本  $k$  阶中心矩  $B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k$

#### Example 6.2.1 (P<sub>141</sub> 例 3)

设总体  $X$  的概率分布为

$X$	1	2	3
$p_i$	$\theta^2$	$2\theta(1-\theta)$	$(1-\theta)^2$

其中  $0 < \theta < 1$ . 现抽得一个样本  $x_1 = 1, x_2 = 2, x_3 = 1$ , 求  $\theta$  的矩估计量.

**Solution:**

$$\mu_1 = \mathbb{E}[X] = \theta^2 + 2 \times 2\theta(1-\theta) + 3(1-\theta)^2 = 3 - 2\theta,$$

$$A_1 = \frac{1}{3}(1 + 2 + 1) = \frac{4}{3}$$

体现矩估计思想: 样本矩等于总体矩  
 令之相等, 得  $3 - 2\theta = \frac{4}{3}$ , 即  $\theta$  矩估计  $\hat{\theta} = \frac{5}{6}$ .

## 6.2.2 最大似然估计法

### Definition 6.2.1

对任意给定样本值  $x_1, \dots, x_n$ ,

$$\exists \hat{\theta} = \hat{\theta}(\mathbf{x}), \text{ where } \mathbf{x} = (x_1, \dots, x_n) \text{ s.t. } \mathcal{L}(\hat{\theta}) = \underset{\theta}{\operatorname{argmax}} \mathcal{L}(\theta; \mathbf{x}),$$

则称  $\hat{\theta} = \hat{\theta}(\mathbf{x})$  为  $\theta$  的**最大似然估计值**,  $\hat{\theta}(X_1, X_2, \dots, X_n)$  为  $\theta$  的**最大似然估计量**. 统称为**最大似然估计 (MLE)**.

其中  $\mathcal{L}(\theta)$  称为样本的**似然函数**. 若为离散型, 设总体  $X$  的概率分布为  $\Pr\{X = x\} = p(x; \theta)$ , 如果  $X_1, X_2, \dots, X_n$  是来自  $X$  的样本, 观察值为  $\mathbf{x}$ , 记  $\mathcal{L}(\theta)$ :

$$\mathcal{L}(\theta) = \underset{\text{参数}}{\mathcal{L}}(\underset{\text{变量}}{\mathbf{x}}; \hat{\theta}) = \prod_{i=1}^n p(x_i; \theta) \quad (6.6)$$

若为连续型, 设总体  $X$  的 PDF 为  $f(x; \theta)$ ,  $X_1, X_2, \dots, X_n$  是来自  $X$  的样本 ( $n$  维连续型随机变量, 联合 PDF 为  $f(\mathbf{x})$ ), 观察值为  $\mathbf{x}$ . 因为  $(X_1, \dots, X_n)$  相互独立, 根据(3.15),

$$f(x_1, \dots, x_n) = f_{x_1}(x_1)f_{x_2}(x_2) \cdots f_{x_n}(x_n),$$

因此记  $\mathcal{L}(\theta)$ :

$$\mathcal{L}(\theta) = \mathcal{L}(\mathbf{x}; \theta) = \prod_{i=1}^n f(x_i; \theta) \quad (6.7)$$

为了方便计算, 经常在两侧取对数, 称为**对数似然函数**:

$$\ell(\mathbf{x}; \theta) = \ln \mathcal{L}(\mathbf{x}; \theta) \quad (6.8)$$

对似然函数而言, 驻点即为极大值点. 因此, 只要求出一个  $\theta$  s.t.  $\mathcal{L}(\theta)$  取最大值即可, 记这个  $\theta$  为  $\hat{\theta}$ , 对可微函数  $\mathcal{L}(\theta)$ , 令  $\mathcal{L}'(\theta) = 0$  或  $\ell'(\theta) = 0$ , 求出驻点.

### Example 6.2.2 (P<sub>143</sub> 例 4)

设  $X \sim b(1, p)$ ,  $X_1, X_2, \dots, X_n$  是来自  $X$  的样本, 求  $p$  的 MLE.

**Solution:** 设  $x_1, \dots, x_n$  是相应于样本  $X_1, \dots, X_n$  的观察值,  $X$  的分布律为

$$\Pr\{X = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1.$$

于是似然函数为

$$\mathcal{L}(p) = p^{\sum_{i=1}^n x_i} \times (1-p)^{\sum_{i=1}^n 1-x_i},$$



取对数

$$\ln \mathcal{L}(p) = \left( \sum_{i=1}^n x_i \right) \ln p + \left( n - \sum_{i=1}^n x_i \right) \ln(1-p),$$

求导

$$\frac{d \ln \mathcal{L}(p)}{dp} = \frac{\sum_{i=1}^n x_i}{p} - \frac{n - \sum_{i=1}^n x_i}{1-p} = 0$$

最大似然估计值 解出  $\hat{p}(x_1, \dots, x_n)$

$$\hat{p} = \frac{\sum_{i=1}^n x_i}{n} = \bar{x},$$

最大似然估计量 可得  $\hat{p}(X_1, \dots, X_n)$

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n} = \bar{X},$$

### Example 6.2.3 (P<sub>144</sub> 例 5)

设  $X \sim e(\lambda)$ ,  $X_1, X_2, \dots, X_n$  是来自  $X$  的样本, 求  $\lambda$  的 MLE.

**Solution:** 似然函数为

$$\begin{aligned} \mathcal{L}(x_1, \dots, x_n; \lambda) &= \prod_{i=1}^n f_{x_i}(x_i) = \prod_{i=1}^n \lambda^n e^{-\lambda x_i} \\ &= \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, \end{aligned}$$

取对数

$$\ln \mathcal{L}(x_1, \dots, x_n; \lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i,$$

求导

$$\frac{d \ln \mathcal{L}(x_1, \dots, x_n; \lambda)}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0,$$

解得  $\lambda$  的最大似然估计值  $\hat{\lambda}(x_1, \dots, x_n) = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$

## 6.3 置信区间

### 6.3.1 置信区间的概念

设  $\theta$  是总体  $X$  的未知参数,  $\hat{\theta}_1(X_1, \dots, X_n)$  和  $\hat{\theta}_2(X_1, \dots, X_n)$  是  $\theta$  的两个估计量, 如果对任意给定的  $\alpha > 0$ , 有

$$\Pr\{\hat{\theta}_1 < \theta < \hat{\theta}_2\} = 1 - \alpha, \quad (6.9)$$

则称随机区间

$$\hat{\theta}_1 < \theta < \hat{\theta}_2$$

要求置信区间尽可能短

是  $\theta$  的一个置信度为  $1 - \alpha$  的置信区间,  $\hat{\theta}_1$  和  $\hat{\theta}_2$  分别称为  $\theta$  的置信下限和置信上限. 在保证置信度的条件下, 尽可能提高估计精度.

## 6.4 正态总体的置信区间

### 6.4.1 正态总体均值的置信区间

设  $X_1, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本,  $\mu$  未知而  $\sigma$  已知, 对给定的置信水平  $1 - \alpha$ , 由标准正态分布 PDF 的对称性可得  $\mu$  的置信区间为

$$\left( \bar{X} - u_{\alpha/2} \times \frac{\sigma}{\sqrt{n}}, \bar{X} + u_{\alpha/2} \times \frac{\sigma}{\sqrt{n}} \right). \quad (6.10)$$

其中  $u_{\alpha/2}$  是标准正态分布的上侧  $\alpha/2$  分位数.

#### Example 6.4.1 (P<sub>151</sub> 例 1)

设  $X_1, \dots, X_{100}$  是来自正态总体的样本,  $\bar{x} = 80, \sigma = 12$ , 求  $\mu$  的置信度为 0.95 的置信区间.

(6.10) **Solution:** 由题意  $1 - \alpha = 0.95 \Rightarrow \alpha = 0.05$ , 由标准正态分布表可得  $u_{\alpha/2} = u_{0.025} = 1.96$ , 代入有区间 (77.6, 82.4).

定理 (6.1.1)-2 设  $X_1, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本,  $\mu, \sigma$  都未知, 对给定的置信水平  $1 - \alpha$ , 用  $\sigma^2$  的无偏估计  $S^2$  代替, 构造枢轴变量

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

定理 (5.3.1)-6 因为  $T \sim t(n-1)$ , 由

$$\Pr \left\{ -t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1) \right\} = 1 - \alpha,$$

得  $\mu$  的置信区间为

$$\left( \bar{X} - t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}} \right). \quad (6.11)$$

#### Example 6.4.2 (P<sub>151</sub> 例 2)

设  $X_1, \dots, X_{25}$  是来自正态总体的样本,  $\bar{x} = 80$ , 样本标准差  $s = 12$ , 求  $\mu$  的置信度为 0.95 的置信区间.

**Solution:** 根据题意,  $t_{\alpha/2}(n-1) = t_{0.025}(24) = 2.0639$ . 将  $\sigma$  用  $s$  代替, 代入有区间 (75.05, 84.95).

#### Exercise 6.4.1 P<sub>161</sub> 18

某公司生产电池, 现从其产品中抽取 50 只电池做寿命试验, 这些电池的平均寿命  $\bar{x} = 2.266$ ,  $s = 1.935$ . 求电池平均寿命的置信度为 95% 的置信区间.

**Solution:** 设电池寿命为  $X$ . 由题意构造枢轴变量

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}},$$

$$-t_{\alpha/2}(n-1) < \frac{\bar{X} - \mu}{S/\sqrt{n}} < t_{\alpha/2}(n-1).$$

可得  $\mu$  的置信区间为

$$\left( \bar{X} - t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}}, \bar{X} + t_{\alpha/2}(n-1) \times \frac{S}{\sqrt{n}} \right),$$

将  $t_{0.025}(49) = 1.96, \bar{x} = 2.266, s = 1.935, n = 49$  代入可得 (1.730, 2.802).

### 6.4.2 正态总体方差的置信区间

定理 (5.3.1)-3 设  $X_1, \dots, X_n$  是来自正态总体  $N(\mu, \sigma^2)$  的样本,  $\mu, \sigma$  都未知, 对给定的置信水平  $1 - \alpha$ , 构造枢轴变量

$$\frac{n-1}{\sigma^2} S^2 \sim \chi^2(n-1)$$

因为卡方分布的 PDF 只在正半轴上有值 (不具有对称性), 所以两侧为  $\chi_{1-\alpha/2}^2(n-1), \chi_{\alpha/2}^2(n-1)$ , 且

$$\Pr \left\{ \chi_{1-\alpha/2}^2(n-1) < \frac{n-1}{\sigma^2} S^2 < \chi_{\alpha/2}^2(n-1) \right\} = 1 - \alpha$$

求解  $\sigma^2$ , 得方差为  $\sigma^2$  的  $1 - \alpha$  置信区间为

$$\left( \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \right). \quad (6.12)$$

对上式开根号即得标准差为  $\sigma$  的  $1 - \alpha$  置信区间.

# Chapter 7

## 假设检验

假设检验就是反证法. 为了验证某个假设是否正确, 先使其成立, 根据抽取的样本计算出一个统计量 (方差, 均值), 若该统计量的值落在一个不合理的范围 (拒绝域) 内, 则认为不成立, 反之成立.

### 7.1 基本概念

上述的不合理可以表示为小概率事件. 小概率事件发生的概率越小, 说明假设成立的可能性越小, 因此拒绝假设的理由越充分. 但是, 不能因为拒绝域内的事件发生了就拒绝假设 (如果拒绝域十分小, 仍然可以接受假设), 因为这样会导致**第一类错误**. 犯这个错误的概率就是小概率事件发生的概率, 即

考虑法庭判决: 判错无辜的人, 这就是第一类错误.

$$\Pr\{\text{reject } H_0 | H_0 \text{ is true}\} = \alpha$$

为了避免第一类错误, 需要选择一个合适的拒绝域, 使得第一类错误的概率不超过预先规定的值  $\alpha$ , 这个值称为**显著性水平** (就是上侧分位数). 一般取  $\alpha = 0.05$  或  $\alpha = 0.01$ . 我们把要检验的假设  $H_0$  称为**原假设**, 因此有**对立假设**  $H_1$  存在. 例如:

$$H_0 : \mu = \mu_0; \quad H_1 : \mu \neq \mu_0.$$

这称为双侧假设检验.

经常需要构造**检验统计量** (一般记为  $U, T, \chi$ ), 当检验统计量的值落在某个区域  $W$  内时, 拒绝原假设, 称  $W$  为**拒绝域**. 一般来说, 置信区间外的值都是拒绝域内的值. 显著性水平正好和置信水平相反.

### 7.2 单正态总体的假设检验

#### 7.2.1 总体均值的假设检验

若是检验总体均值, 最后就代入样本均值; 若是检验总体方差, 最后就代入样本方差.

方差已知

与前面求置信区间类似, 设总体  $X \sim N(\mu, \sigma^2)$ . 总体方差  $\sigma^2$  已知.  $X_1, \dots, X_n$  是取自总体  $X$  的样本, 给定显著性水平  $\alpha$ , 检验假设  $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$ . 以  $H_0$  成立为前提, 构造检验统计量

定理 (5.3.1)-2, 就是构造标准正态分布

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1) \quad (7.1)$$

记观察值为  $u$ . 这称为  $u$  检验法.

根据拒绝域的定义, 可以得到

$$W = (-\infty, -u_{\alpha/2}) \cup (u_{\alpha/2}, +\infty). \quad (7.2)$$

根据样本计算样本均值  $\bar{x}$ , 将  $\mu_0, \sigma$  代入, 计算  $u$ , 若  $u \in W$ , 则拒绝  $H_0$ , 否则接受  $H_0$ .

Example 7.2.1 (P<sub>165</sub> 例 1)

设总体  $X \sim N(500, 2^2)$ . 取其中的 9 个样本如下:

505    499    502    506    498    498    497    510    503

总体标准差不变. 检验假设  $H_0: \mu = 500, H_1: \mu \neq 500$ . ( $\alpha = 0.05$ )

**Solution:** 以  $H_0$  成立为前提, 构造检验统计量

$$U = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} = \frac{\bar{X} - 500}{2/3} \sim N(0, 1)$$

将样本均值  $\bar{x} = 502$  代入, 计算观察值得  $u = 3$ .

$$\therefore u \in W = (-\infty, -u_{\alpha/2}) \cup (u_{\alpha/2}, +\infty) = (-1.96, -\infty) \cup (1.96, +\infty)$$

因此不接受  $H_0$ .

### 方差未知

定理 (5.3.1)-6 题设同上, 但方差未知. 同理可以构造检验统计量  $T$ , 因为  $\bar{X}$  为  $\mu$  的无偏估计量,  $S^2$  是  $\sigma^2$  的无偏估计量, 我们用样本方差 (计算无偏方差) 来代替总体方差, 以  $H_0$  成立为前提,

$$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1) \quad (7.3)$$

记观察值为  $t$ . 这称为  $t$  检验法.

根据拒绝域的定义, 可以得到 ( $t$  分布的 PDF 是偶函数)

$$W = (-\infty, -t_{\alpha/2}(n-1)) \cup (t_{\alpha/2}(n-1), +\infty). \quad (7.4)$$

将样本均值  $\bar{x}$ , 样本方差  $s^2$  代入, 计算  $t$ , 若  $t \in W$ , 则拒绝  $H_0$ , 否则接受  $H_0$ .

## 7.2.2 总体方差的假设检验

定理 (5.3.1)-3 设  $X \sim N(\mu, \sigma^2)$ .  $X_1, \dots, X_n$  是取自  $X$  的样本, 给定显著性水平  $\alpha$ , 检验假设  $H_0: \sigma^2 = \sigma_0^2, H_1: \sigma^2 \neq \sigma_0^2$ . 以  $H_0$  成立为前提, 构造检验统计量  $\chi^2$

$$\chi^2 = \frac{n-1}{\sigma_0^2} S^2 \sim \chi^2(n-1), \quad (7.5)$$

由  $\chi^2$  分布的 PDF 图象可以知道, 两侧分别为  $\chi_{1-\alpha/2}^2(n-1), \chi_{\alpha/2}^2(n-1)$ , 记观察值为  $\chi^2$ , 这称为  $\chi^2$  检验法. 因此

$$W = [0, \chi_{\alpha/2}^2(n-1)) \cup (\chi_{1-\alpha/2}^2(n-1), +\infty). \quad (7.6)$$

将样本方差  $s^2$  代入, 计算  $\chi^2$ , 若  $\chi^2 \in W$ , 则拒绝  $H_0$ , 否则接受  $H_0$ .

### Exercise 7.2.1 P<sub>171</sub> 5

某种导线的电阻服从正态分布  $N(\mu, 0.005^2)$ . 从一批导线中抽取 9 根, 测定电阻, 得  $s = 0.008$ . 对  $\alpha = 0.05$ , 能否认为这批导线电阻的标准差仍为 0.005?

**Solution:** 设导线的电阻为  $X$ . 由题意  $X \sim N(\mu, 0.005^2)$ . 设  $H_0: \sigma = 0.005, H_1: \sigma \neq 0.005$ . 在  $H_0$  成立的前提下, 构造

$$\chi^2 = \frac{n-1}{0.005^2} s^2 \sim \chi^2(8)$$

因为  $\alpha = 0.05, \alpha/2 = 0.025, \chi_{0.025}^2(8) = 17.535, \chi_{0.975}^2(8) = 2.180$ , 因此

$$W = [0, 2.180) \cup (17.54, +\infty)$$

计算观察值,

$$\chi^2 = \frac{8}{0.005^2} \times 0.008^2 = 20.48 \in W$$

因而不接受假说  $H_0$ . 不能认为这批导线电阻的标准差仍为 0.005.