

# 概率论与数理统计笔记

刘承杰

南京大学软件学院

2021 年 8 月 13 日

## 目录

<b>1</b>	<b>概率论的基本概念</b>	<b>1</b>
1.1	随机试验	1
1.2	样本空间、随机事件	1
1.3	频率与概率	2
1.4	等可能概型（古典概型）	3
1.5	条件概率	3
1.6	独立性	4
<b>2</b>	<b>随机变量及其分布</b>	<b>5</b>
2.1	随机变量	5
2.2	离散型随机变量及其分布律	5
2.3	随机变量的分布函数	5
2.4	连续型随机变量及其概率密度	6
2.5	随机变量的函数的分布	7
<b>3</b>	<b>多维随机变量及其分布</b>	<b>8</b>
3.1	二维随机变量	8
3.2	边缘分布	8

3.3	条件分布	9
3.4	相互独立的随机变量	9
3.5	两个随机变量的函数的分布	10
<b>4</b>	<b>随机变量的数字特征</b>	<b>12</b>
4.1	数学期望	12
4.2	方差	13
4.3	协方差及相关系数	14
4.4	矩、协方差矩阵	15
<b>5</b>	<b>大数定律及中心极限定理</b>	<b>16</b>
5.1	大数定律	16
5.2	中心极限定理	16
<b>6</b>	<b>样本及抽样分布</b>	<b>19</b>
6.1	随机样本	19
6.2	直方图和箱线图	19
6.3	抽样分布	19
<b>7</b>	<b>参数估计</b>	<b>24</b>
7.1	点估计	24
<b>8</b>	<b>假设检验</b>	<b>26</b>
<b>9</b>	<b>方差分析及回归分析</b>	<b>27</b>

# 1 概率论的基本概念

## 1.1 随机试验

**定义 1.1 (随机试验)** 在概率论中, 我们将具有以下三个特点的试验称为随机试验:

1. 可以在相同的条件下重复的进行;
2. 每次试验的可能结果不止一个, 并且能事先明确试验的所有可能结果;
3. 进行一次试验之前不能确定哪个结果会出现

## 1.2 样本空间、随机事件

**定义 1.2 (样本空间)** 随机事件 $E$ 的所有可能结果组成的集合, 记为 $S$

**定义 1.3 (样本点)** 样本空间的元素, 即 $E$ 的每个结果

**定义 1.4 (随机事件)** 试验 $E$ 的样本空间 $S$ 的子集为 $E$ 的随机事件, 简称事件. 在每次试验中, 当且仅当这一子集中的一个样本点出现时, 称这一事件发生

**定义 1.5 (基本事件)** 由一个样本点组成的单点集, 称为基本事件

**定义 1.6 (必然事件和不可能事件)** 样本空间 $S$ 包含所有的样本点, 它是 $S$ 自身的子集, 在每次试验中它总是发生的,  $S$ 称为 **必然事件**. 空集 $\emptyset$ 不包含任何样本点, 它也作为样本空间的子集, 它在每次试验中都不可能发生,  $\emptyset$ 称为**不可能事件**.

**定义 1.7 (事件间的关系与事件的运算)** 设试验 $E$ 的样本空间为 $S$ , 而 $A, B, A_k (k = 1, 2, \dots)$ 是 $S$ 的子集.

1. 若 $A \subset B$ , 则称事件 $B$ 包含事件 $A$ , 这指的是事件 $A$ 发生, 则事件 $B$ 必然发生. 若 $A \subset B$ 且 $B \subset A$ , 即 $A = B$ , 则称事件 $A$ 与事件 $B$ 相等。

2. 事件 $A \cup B = \{x | x \in A \text{ 或 } x \in B\}$ 称为事件 $A$ 与事件 $B$ 的**和事件**。当且仅当 $A, B$ 中至少有一个发生时, 事件 $A \cup B$ 发生。

类似地, 称 $\bigcup_{k=1}^n A_k$ 为 $n$ 个事件 $A_1, A_2, \dots, A_n$ 的**和事件**; 称 $\bigcup_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \dots$ 的**和事件**

3. 事件 $A \cap B = \{x | x \in A \text{ 且 } x \in B\}$ 称为事件 $A$ 与事件 $B$ 的**积事件**。当且仅当 $A, B$ 同时发生时, 事件 $A \cap B$ 发生。

类似地, 称 $\bigcap_{k=1}^n A_k$ 为 $n$ 个事件 $A_1, A_2, \dots, A_n$ 的**积事件**; 称 $\bigcap_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, \dots$ 的**积事件**

4. 事件 $A - B = \{x | x \in A \text{ 且 } x \notin B\}$ 称为事件 $A$ 与事件 $B$ 的**差事件**。当且仅当  $A$ 发生、 $B$ 不发生时事件 $A - B$ 发生。

5. 若 $A \cap B = \emptyset$ , 则称事件 $A$ 与 $B$ 是**互不相容的**, 或**互斥的**。这指的是事件 $A$ 与 $B$  不能同时发生。基本事件是两两互不相容的。

6. 若  $A \cup B = S$  且  $A \cap B = \emptyset$ , 则称事件  $A$  与事件  $B$  互为逆事件, 又称事件  $A$  与事件  $B$  互为对立事件。这指的是对每次试验而言, 事件  $A$ 、 $B$  中必有一个发生, 且仅有一个发生。 $A$  的对立事件记为  $\bar{A}$

**定理 1.1 (集合运算定律)** 设  $A$ 、 $B$ 、 $C$  为事件, 则有:

交换律:

$$A \cup B = B \cup A$$

$$A \cap B = B \cap A$$

结合律:

$$A \cup (B \cap C) = (A \cup B) \cap C$$

$$A \cap (B \cup C) = (A \cap B) \cup C$$

分配律:

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

德摩根律:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$

$$\overline{A \cap B} = \bar{A} \cup \bar{B}$$

### 1.3 频率与概率

**定义 1.8 (频率)** 在相同的条件下, 进行了  $n$  次试验, 在这  $n$  次试验中, 事件  $A$  发生的次数  $n_A$  称为事件  $A$  发生的频数。比值  $\frac{n_A}{n}$  称为事件  $A$  发生的频率, 并记为  $f_n(A)$ 。

**定义 1.9 (概率)** 设  $E$  是随机试验,  $S$  是它的样本空间, 对于  $E$  的每一个事件  $A$  赋予一个实数, 记为  $P(A)$ , 称为事件  $A$  的概率, 如果集合函数  $P(\cdot)$  满足下列条件:

1. 非负性: 对于每一个事件  $A$ , 有  $P(A) \geq 0$ ;
2. 规范性: 对于必然事件  $S$ , 有  $P(S) = 1$ ;
3. 可列可加性: 设  $A_1, A_2, \dots$  是两两不相容的事件, 即对于  $A_i A_j = \emptyset, i \neq j, i, j = 1, 2, \dots$ , 有

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

当  $n \rightarrow \infty$  时频率  $f_n(A)$  在一定意义下接近于概率  $P(A)$

## 1.4 等可能概型（古典概型）

**定义 1.10 (等可能概型)** 具有以下两个特点的试验称为等可能概型：

1. 试验的样本空间只包含有限个元素；
2. 试验中每个基本事件发生的可能性相同。

**定理 1.2 (等可能概型中事件A的概率计算公式)** 若事件A包含K个基本事件，即  $A = e_{i_1} \cup e_{i_2} \cup \dots \cup e_{i_k}$ ，这里  $i_1, i_2, \dots, i_k$  是  $1, 2, \dots, n$  中某  $k$  个不同的数，则有：

$$P(A) = \sum_{j=1}^k P(\{e_{i_j}\}) = \frac{k}{n} = \frac{A \text{ 包含的基本事件数}}{S \text{ 中基本事件的总数}}$$

**定理 1.3 (超几何分布的概率公式)** 设共有  $N$  件产品，其中有  $D$  件次品，从中取  $n$  件，其中恰好有  $k (k \leq D)$  件次品的概率为：

$$p = \frac{\binom{D}{k} \binom{N-D}{n-k}}{\binom{N}{n}}$$

## 1.5 条件概率

**定义 1.11 (条件概率)** 设  $A, B$  是两个事件，且  $P(A) > 0$ ，称

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

为在事件  $A$  发生的条件下事件  $B$  发生的条件概率。

**定理 1.4 (乘法定理)** 设  $P(A) > 0$ ，则有

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

上式可以推广到多个事件的积事件的情况。例如，设  $A, B, C$  为事件，且  $P(AB) > 0$ ，则有

$$P(ABC) = P(C|AB)P(B|A)P(A)$$

一般地，设  $A_1, A_2, \dots, A_n$  为  $n$  个事件， $n \geq 2$ ，且  $P(A_1 A_2 \dots A_{n-1}) > 0$ ，则有

$$P(A_1 A_2 \dots A_n) = P(A_n | A_1 A_2 \dots A_{n-1}) P(A_{n-1} | A_1 A_2 \dots A_{n-2}) \dots P(A_2 | A_1) P(A_1)$$

**定义 1.12 (样本空间的划分)** 设  $S$  为试验  $E$  的样本空间， $B_1, B_2, \dots, B_n$  为  $E$  的一组事件，若

1.  $B_i B_j = \emptyset, i \neq j, i, j = 1, 2, \dots, n$
2.  $B_1 \cup B_2 \cup \dots \cup B_n = S$

则称  $B_1, B_2, \dots, B_n$  为样本空间的一个划分。

若  $B_1, B_2, \dots, B_n$  为样本空间的一个划分，那么对于每次试验，事件  $B_1, B_2, \dots, B_n$  中必有且只有一个发生。

**定理 1.5 (全概率公式)** 设试验 $E$ 的样本空间为 $S$ ,  $A$ 为 $E$ 的事件,  $B_1, B_2, \dots, B_n$ 为 $S$ 的一个划分, 且 $P(B_i) > 0, i = 1, 2, \dots, n$ , 则

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \dots + P(A|B_n)P(B_n)$$

**定理 1.6 (贝叶斯(Bayes)公式)** 设试验 $E$ 的样本空间为 $S$ ,  $A$ 为 $E$ 的事件,  $B_1, B_2, \dots, B_n$ 为 $S$ 的一个划分, 且 $P(A) > 0, P(B_i) > 0, i = 1, 2, \dots, n$ , 则

$$P(B_i|A) = \frac{P(B_iA)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}, i = 1, 2, \dots, n$$

## 1.6 独立性

**定义 1.13 (独立性)** 设 $A, B$ 是两事件, 如果满足等式

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

则称事件 $A, B$ 相互独立

同理, 对于 $A, B, C$ 三个事件, 如果满足等式

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

$$P(BC) = P(B)P(C)$$

$$P(AC) = P(A)P(C)$$

$$P(ABC) = P(A)P(B)P(C)$$

则称事件 $A, B, C$ 相互独立。

一般地, 设 $A_1, A_2, \dots, A_n$ 是 $n(n \geq 2)$ 个事件, 如果对于其中任意2个, 任意3个,  $\dots$ , 任意 $n$ 个事件的积事件的概率都等于各事件概率的积, 则称事件 $A_1, A_2, \dots, A_n$ 相互独立。

## 2 随机变量及其分布

### 2.1 随机变量

**定义 2.1 (随机变量)** 设随机试验的样本空间为  $S = \{e\}$ ,  $X = X(e)$  是定义在样本空间  $S$  上的实值单值函数, 称  $X = X(e)$  为随机变量

### 2.2 离散型随机变量及其分布律

**定义 2.2 (离散型随机变量)** 全部可能取到的值是有限个或可列无限多个的随机变量, 称为离散型随机变量。

**定义 2.3 ((0-1)分布)** 设随机变量  $X$  只能取 0, 1 两个值, 它的分布律是

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

**定义 2.4 (伯努利试验)** 设试验  $E$  只有两个可能结果:  $A$  和  $\bar{A}$ , 则称  $E$  为伯努利(Bernoulli)试验. 设  $P(A) = p$  ( $0 < p < 1$ ), 此时  $P(\bar{A}) = 1 - p$ , 将  $E$  独立重复进行  $n$  次, 则称这一串重复的独立试验为  $n$  重伯努利试验。

**定义 2.5 (二项分布)** 以  $X$  表示  $n$  重伯努利试验中事件  $A$  发生的次数, 事件  $A$  在指定的  $k$  ( $0 \leq k \leq n$ ) 次试验中发生的概率为

$$P\{X = k\} = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, 2, \dots, n$$

随机变量  $X$  服从参数为  $n, p$  的二项分布, 并记为  $X \sim b(n, p)$ 。

**定义 2.6 (泊松分布)** 设随机变量  $X$  所有可能取的值为  $0, 1, 2, \dots$ , 而取各个值的概率为

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

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

**定理 2.1 (泊松定理)** 设  $\lambda > 0$  是一个常数,  $n$  是任意正整数, 设  $np_n = \lambda$ , 则对于任一固定的非负整数  $k$ , 有

$$\lim_{n \rightarrow \infty} \binom{n}{k} p_n^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$$

### 2.3 随机变量的分布函数

**定义 2.7 (随机变量的分布函数)** 设  $X$  是一个随机变量,  $x$  是任意实数, 函数

$$F(X) = P\{X \leq x\}, -\infty < x < \infty$$

称为  $X$  的分布函数

## 定理 2.2 (分布函数性质)

1.  $F(x)$  是一个不减函数
2.  $0 \leq x \leq 1$ , 且  $F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = 0, F(\infty) = \lim_{x \rightarrow \infty} F(x) = 1$
3.  $F(x+0) = F(x)$ , 即  $F(x)$  是右连续的

反之, 具备上述三条性质的函数必是某个随机变量的分布函数。

## 2.4 连续型随机变量及其概率密度

**定义 2.8 (连续型随机变量、概率密度)** 如果对于随机变量  $X$  的分布函数  $F(x)$ , 存在非负可积函数  $f(x)$ , 使对于任意实数  $x$  有

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

则称  $X$  为连续型随机变量,  $f(x)$  称为  $X$  的概率密度函数, 简称概率密度。

**定义 2.9 (均匀分布)** 若连续型随机变量  $X$  具有概率密度

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

则称  $X$  在区间  $(a, b)$  上服从均匀分布, 记为  $X \sim U(a, b)$

$X$  的分布函数为

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

**定义 2.10 (指数分布)** 若连续型随机变量  $X$  的概率密度为

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

其中  $\theta > 0$  为常数, 则称  $X$  服从参数为  $\theta$  的指数分布

$X$  的分布函数为

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

**定义 2.11 (正态分布)** 若连续型随机变量  $X$  的概率密度为

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

其中  $\mu, \sigma (\sigma > 0)$  为常数, 则称  $X$  服从参数为  $\mu, \sigma$  的正态分布或高斯分布, 记为  $X \sim N(\mu, \sigma^2)$

$X$  的分布函数为

$$F(x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$$



特别地, 当 $\mu = 0, \sigma = 1$ 时, 称随机变量 $X$ 服从标准正态分布, 其概率密度和分布函数分别用 $\varphi(x), \Phi(x)$ 表示, 有

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

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

对于一般的正态分布, 只需通过一个线性变换就能化为标准正态分布:

$$\text{若 } X \sim N(\mu, \sigma^2), \text{ 则 } Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

## 2.5 随机变量的函数的分布

**定理 2.3** 设随机变量 $X$ 具有概率密度 $f_x(x), -\infty < x < \infty$ , 又设函数 $g(x)$ 处处可导且恒有 $g'(x) > 0$  (或恒有 $g'(x) < 0$ , 则 $Y = g(X)$ 是连续型随机变量, 其概率密度为

$$f_Y(y) = \begin{cases} f_x[h(y)]|h'(y)|, & \alpha < y < \beta \\ 0, & \text{其他} \end{cases}$$

其中 $\alpha = \min\{g(-\infty), g(\infty)\}, \beta = \max\{g(-\infty), g(\infty)\}, h(y)$ 是 $g(x)$ 的反函数

### 3 多维随机变量及其分布

#### 3.1 二维随机变量

**定义 3.1 (联合分布函数)** 设 $(X, Y)$ 是二维随机变量, 对于任意实数 $x, y$ , 二元函数

$$F(x, y) = P\{(X \leq x) \cap (Y \leq y)\} \equiv P\{X \leq x, Y \leq y\}$$

称为二维随机变量 $(X, Y)$ 的分布函数, 或称为随机变量 $X$ 和 $Y$ 的联合分布函数。

如果将二维随机变量 $(X, Y)$ 看成是平面上随机点的坐标, 那么分布函数 $F(x, y)$ 在 $(x, y)$ 处的函数值就是随机点 $(X, Y)$ 落在以点 $(x, y)$ 为顶点而位于该点左下方的无穷矩形区域内的概率。

所以, 随机点 $(X, Y)$ 落在矩形区域 $\{(x, y) | x_1 < x \leq x_2, y_1 < y \leq y_2\}$ 的概率为

$$P\{x_1 < x \leq x_2, y_1 < y \leq y_2\} = F(x_2, y_2) - F(x_2, y_1) + F(x_1, y_1) - F(x_1, y_2)$$

**定义 3.2 (联合概率密度)** 对于二维随机变量 $(X, Y)$ 的分布函数 $F(x, y)$ , 如果存在非负可积函数 $f(x, y)$ 使对于任意 $x, y$ 有

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

则称 $(X, Y)$ 是连续型的二维随机变量, 函数 $f(x, y)$ 称为二维随机变量 $(X, Y)$ 的概率密度, 或称为随机变量 $X$ 和 $Y$ 的联合概率密度。

#### 3.2 边缘分布

**定义 3.3 (边缘分布函数)** 二维随机变量 $(X, Y)$ 作为一个整体, 具有分布函数 $F(x, y)$ , 而 $X$ 和 $Y$ 都是随机变量, 各自也有分布函数, 将他们分别记为  $F_X(x), F_Y(y)$ , 依次称为二维随机变量 $(X, Y)$ 关于 $X$ 和关于 $Y$ 的边缘分布函数, 且

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

**定义 3.4 (边缘分布律)** 记

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

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

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

**定义 3.5 (边缘概率密度)** 记

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

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

分别称 $f_X(x), f_Y(y)$ 为 $(X, Y)$ 关于 $X$ 和 $Y$ 的边缘概率密度。

### 3.3 条件分布

**定义 3.6 (条件分布律)** 设 $(X, Y)$ 是二维离散型随机变量, 对于固定的 $j$ , 若 $P\{Y = y_j\} > 0$ , 则称

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

为在 $Y = y_j$ 条件下随机变量 $X$ 的条件分布律

同样, 对于固定的 $i$ , 若 $P\{X = x_i\} > 0$ , 则称

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

为在 $X = x_i$ 条件下随机变量 $Y$ 的条件分布律

**定义 3.7 (条件概率密度)** 设二维随机变量 $(X, Y)$ 的概率密度为 $f(x, y)$ ,  $(X, Y)$ 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 $\frac{f(x, y)}{f_Y(y)}$ 为在 $Y = y$ 的条件下 $X$ 的条件概率密度, 记为

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)}$$

称 $\int_{-\infty}^x f_{X|Y}(x|y) dx = \int_{-\infty}^x \frac{f(x, y)}{f_Y(y)} dx$ 为在 $Y = y$ 的条件下 $X$ 的条件分布函数, 记为 $P\{X \leq x | Y = y\}$ 或 $F_{X|Y}(x|y)$ , 即

$$F_{X|Y}(x|y) = P\{X \leq x | Y = y\} = \int_{-\infty}^x \frac{f(x, y)}{f_Y(y)} dx$$

类似地, 可以定义 $f_{Y|X}(y|x) = \frac{f(x, y)}{f_X(x)}$ 和 $F_{Y|X}(y|x) = \int_{-\infty}^y \frac{f(x, y)}{f_X(x)} dy$

### 3.4 相互独立的随机变量

**定义 3.8 (相互独立)** 若对于所有 $x, y$ , 满足下列条件之一, 则称随机变量 $X$ 和 $Y$ 是相互独立的:

1. 设 $F(x, y)$ 及 $F_X(x), F_Y(y)$ 分别是二维随机变量 $(X, Y)$ 的分布函数及边缘分布函数, 有

$$F(x, y) = F_X(x)F_Y(y)$$

2. 设 $f(x, y), f_X(x), f_Y(y)$ 分别为 $(X, Y)$ 的概率密度和边缘概率密度, 有

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

3. 当 $X, Y$ 是离散型随机变量是, 对于 $(X, Y)$ 的所有可能取值 $(x_i, y_i)$ , 有

$$P\{X = x_i, Y = y_i\} = P\{X = x_i\}P\{Y = y_i\}$$

**定理 3.1** 设 $(X_1, X_2, \dots, X_m)$ 和 $(Y_1, Y_2, \dots, Y_n)$ 相互独立, 则 $X_i (i = 1, 2, \dots, m)$ 和 $Y_j (j = 1, 2, \dots, n)$ 相互独立。又若 $h, g$ 是连续函数, 则 $h(X_1, X_2, \dots, X_m)$ 和 $g(Y_1, Y_2, \dots, Y_n)$ 相互独立。

### 3.5 两个随机变量的函数的分布

**定义 3.9 ( $Z = X + Y$ 分布)** 设 $(X, Y)$ 是二维连续型随机变量, 它具有概率密度 $f(x, y)$ , 则 $Z = X + Y$ 仍为连续型随机变量, 其概率密度为

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(z-y, y) dy, \quad (1)$$

或

$$f_{X+Y}(z) = \int_{-\infty}^{\infty} f(x, z-x) dx. \quad (2)$$

又若 $X$ 和 $Y$ 相互独立, 设 $(X, Y)$ 关于 $X, Y$ 的边缘密度分别为 $f_X(x), f_Y(y)$ , 则(1),(2)分别可化为

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

和

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

这两个公式称为 $f_X$ 和 $f_Y$ 的卷积公式, 记为 $f_X * f_Y$ , 即

$$f_X * f_Y = \int_{-\infty}^{\infty} f_X(z-y) f_Y(y) dy = \int_{-\infty}^{\infty} f_X(x) f_Y(z-x) dx$$

**定理 3.2** 有限个相互独立的正态随机变量的线性组合仍然服从正态分布

**定义 3.10 ( $Z = \frac{Y}{X}$ 分布、 $Z = XY$ 的分布)** 设 $(X, Y)$ 是二维连续型随机变量, 它具有概率密度 $f(x, y)$ , 则 $Z = \frac{Y}{X}, Z = XY$ 仍为连续型随机变量, 其概率密度分别为

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xz) dx; \quad (3)$$

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx. \quad (4)$$

又若 $X$ 和 $Y$ 相互独立, 设 $(X, Y)$ 关于 $X, Y$ 的边缘密度分别为 $f_X(x), f_Y(y)$ , 则(3)可化为

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) dx$$

(4)可化为

$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_X(x) f_Y(\frac{z}{x}) dx$$

**定义 3.11 ( $M = \max\{X, Y\}$ 分布及 $N = \min\{X, Y\}$ 分布)** 设 $X, Y$ 是两个相互独立的随机变量, 它们的分布函数分别为 $F_X(x)$ 和 $F_Y(y)$ , 可得 $M = \max\{X, Y\}$ 的分布函数为

$$F_{\max}(z) = F_X(z) F_Y(z)$$

类似地，可得到  $N = \min\{X, Y\}$  的分布函数为

$$F_{\min}(z) = 1 - [1 - F_X(z)][1 - F_Y(z)]$$

以上结果容易推广到  $n$  个相互独立的随机变量的情况。设  $X_1, X_2, \dots, X_n$  是  $n$  个相互独立的随机变量，它们的分布函数分别为  $F_{X_i}(x_i) (i = 1, 2, \dots, n)$ ，则  $M = \max\{X_1, X_2, \dots, X_n\}$  及  $N = \min\{X_1, X_2, \dots, X_n\}$  的分布函数分别为

$$F_{\max}(z) = F_{X_1}(z)F_{X_2}(z) \cdots F_{X_n}(z)$$

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

特别地，当  $X_1, X_2, \dots, X_n$  相互独立且具有相同分布函数  $F(x)$  时有

$$F_{\max}(z) = [F(z)]^n$$

$$F_{\min}(z) = 1 - [1 - F(z)]^n$$

## 4 随机变量的数字特征

### 4.1 数学期望

**定义 4.1 (数学期望)** 设离散型随机变量 $X$ 的分布律为 $P\{X = x_k\}, k = 1, 2, \dots$ . 若级数 $\sum_{k=1}^{\infty} x_k p_k$ 绝对收敛, 则称级数 $\sum_{k=1}^{\infty} x_k p_k$ 的和为随机变量 $X$ 的数学期望, 记为 $E(X)$ , 即

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

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, 则称积分 $\int_{-\infty}^{\infty} x f(x) dx$ 的值为随机变量 $X$ 的数学期望, 记为 $E(X)$ , 即

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

数学期望 $E(X)$ 完全由随机变量 $X$ 的概率分布所确定。

**定理 4.1** 设 $Y$ 是随机变量 $X$ 的函数:  $Y = g(X)$  ( $g$ 是连续函数)

(i) 如果 $X$ 是离散型随机变量, 它的分布律为 $P = \{X = x_k\} = p_k, k = 1, 2, \dots$ , 若 $\sum_{k=1}^{\infty} g(x_k) p_k$ 绝对收敛, 则有

$$E(Y) = E[g(x)] = \sum_{k=1}^{\infty} g(x_k) p_k$$

(ii) 如果 $X$ 是连续型随机变量, 它的概率密度为 $f(x)$ , 若 $\int_{-\infty}^{\infty} g(x) f(x) dx$ 绝对收敛, 则有

$$E(Y) = E[g(x)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

上述定理还可以推广到两个或两个以上随机变量的函数的情况。

例如设 $Z$ 是随机变量 $X, Y$ 的函数 $Z = g(X, Y)$  ( $g$ 是连续函数), 那么,  $Z$ 是一个一维随机变量。若二维随机变量 $(X, Y)$ 的概率密度为 $f(x, y)$ , 则有

$$E(Z) = E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy$$

这里设上式右边的积分绝对收敛。又若 $(X, Y)$ 为离散型随机变量, 其分布律为 $P = \{X = x_i, Y = y_j\} = p_{ij}, i, j = 1, 2, \dots$ , 则有

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

这里设上式右边的级数绝对收敛。

**定理 4.2** 数学期望的性质:

1° 设 $C$ 是常数, 则有 $E(C) = C$

2° 设 $X$ 是一个随机变量,  $C$ 是常数, 则有

$$E(CX) = CE(X)$$

3° 设 $X, Y$ 是两个随机变量, 则有

$$E(X + Y) = E(X) + E(Y)$$

这一性质可以推广到任意有限个随机变量之和的情况

4° 设 $X, Y$ 是两个相互独立的随机变量, 则有

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

这一性质可以推广到任意有限个相互独立的随机变量之积的情况

## 4.2 方差

**定义 4.2 (方差)** 设 $X$ 是一个随机变量, 若 $E\{[X - E(X)]^2\}$ 存在, 则称 $E\{[X - E(X)]^2\}$ 为 $X$ 的方差, 记为 $D(X)$ 或 $\text{Var}(X)$ , 即

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

在应用上还引入量 $\sqrt{D(X)}$ , 记为 $\sigma(X)$ , 称为标准差或均方差。

对于离散型随机变量, 有

$$D(X) = \sum_{k=1}^{\infty} [x_k - E(X)]^2 p_k$$

其中 $P = \{X = x_k\} = p_k, k = 1, 2, \dots$ 是 $X$ 的分布律

对于连续型随机变量, 有

$$D(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx$$

其中 $f(x)$ 是 $X$ 的概率密度

随机变量 $X$ 的方差可按下列公式计算

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

**定理 4.3** 方差的性质:

1° 设 $C$ 是常数, 则 $D(C) = 0$

2° 设 $X$ 是随机变量,  $C$ 是常数, 则有

$$D(CX) = C^2 D(X) \quad D(X + C) = D(X)$$

3° 设 $X, Y$ 是两个随机变量, 则有

$$D(X + Y) = D(X) + D(Y) + 2E\{[X - E(X)][Y - E(Y)]\}$$

特别地, 若 $X, Y$ 相互独立, 则有

$$D(X + Y) = D(X) + D(Y)$$

这一性质可以推广到任意有限多个相互独立的随机变量之和的情况

4°  $D(X) = 0$ 的充要条件是 $X$ 以概率1取常数 $E(X)$ , 即

$$P\{X = E(X)\} = 1$$

**定理 4.4 (切比雪夫(Chebyshev)不等式)** 设随机变量 $X$ 具有数学期望 $E(X) = \mu$ , 方差 $D(X) = \sigma^2$ , 则对于任意整数 $\varepsilon$ , 有不等式

$$P\{|X - \mu| \geq \varepsilon\} \leq \frac{\sigma^2}{\varepsilon^2}$$

也可写为

$$P\{|X - \mu| < \varepsilon\} \geq 1 - \frac{\sigma^2}{\varepsilon^2}$$

### 4.3 协方差及相关系数

**定义 4.3 (协方差、相关系数)** 量 $E\{[X - E(X)][Y - E(Y)]\}$ 称为随机变量 $X$ 与 $Y$ 的协方差, 记为 $\text{Cov}(X, Y)$ , 即

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

而

$$\rho_{XY} = \frac{\text{Cov}(X, Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

称为随机变量的 $X$ 和 $Y$ 的相关系数。

协方差的计算公式:

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

**定理 4.5** 协方差的性质:

1.  $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$ ,  $a, b$ 是常数;
2.  $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

**定理 4.6**  $\rho_{XY}$ 的性质:



1.  $|\rho_{XY}| \leq 1$
2.  $|\rho_{XY}| = 1$ 的充要条件是, 存在常数 $a, b$ , 使得 $P\{Y = a + bX\} = 1$

$\rho_{XY}$ 是一个可以用来表征 $X, Y$ 之间线性关系紧密程度的量, 当 $|\rho_{XY}|$ 较大时,  $X, Y$ 线性相关程度较好, 当 $|\rho_{XY}|$ 较小时,  $X, Y$ 线性相关程度较差, 当 $|\rho_{XY}| = 0$ 时, 称 $X$ 和 $Y$ 不相关。

## 4.4 矩、协方差矩阵

**定义 4.4** 设 $X$ 和 $Y$ 是随机变量。

若

$$E(X^k) \quad k = 1, 2, \dots$$

存在, 则称它为 $X$ 的 $k$ 阶原点矩, 简称 $k$ 阶矩。

若

$$E\{[X - E(X)]^k\} \quad k = 2, 3, \dots$$

存在, 则称它为 $X$ 的 $k$ 阶中心矩。

若

$$E\{[X - E(X)]^k[Y - E(Y)]^l\} \quad k, l = 1, 2, \dots$$

存在, 则称它为 $X$ 和 $Y$ 的 $k + l$ 阶混合中心矩。

**定义 4.5 (协方差矩阵)** 先以二维随机变量为例。二维随机变量 $(X_1, X_2)$ 有四个二阶中心矩 (设它们都存在), 分别记为

$$c_{11} = E\{[X_1 - E(X_1)]^2\}$$

$$c_{12} = E\{[X_1 - E(X_1)][X_2 - E(X_2)]\}$$

$$c_{21} = E\{[X_2 - E(X_2)][X_1 - E(X_1)]\}$$

$$c_{22} = E\{[X_2 - E(X_2)]^2\}$$

将它们排成矩阵的形式

$$\begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix}$$

这个矩阵称为随机变量 $(X_1, X_2)$ 的协方差矩阵。

设 $n$ 维随机变量 $(X_1, X_2, \dots, X_n)$ 的二阶混合中心矩 $c_{ij} = \text{Cov}(X_i, X_j) = E\{[X_i - E(X_i)][X_j - E(X_j)]\}, i, j = 1, 2, \dots, n$ 都存在, 则称矩阵

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} \\ c_{21} & c_{22} & \cdots & c_{2n} \\ \vdots & \vdots & & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} \end{pmatrix}$$

为 $n$ 维随机变量 $(X_1, X_2, \dots, X_n)$ 的协方差矩阵, 该矩阵是一个对称矩阵。

## 5 大数定律及中心极限定理

### 5.1 大数定律

**定理 5.1 (弱大数定律(辛钦大数定律))** 设 $X_1, X_2, \dots$ 是相互独立的, 服从同一分布的随机变量序列, 且具有数学期望 $E(X_k) = \mu (k = 1, 2, \dots)$ 。作前 $n$ 个变量的算数平均 $\frac{1}{n} \sum_{k=1}^n X_k$ , 则对于任意 $\varepsilon > 0$ , 有

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

**定义 5.1** 设 $Y_1, Y_2, \dots, Y_n, \dots$ 是一个随机变量序列,  $a$ 是一个常数。若对于任意正数 $\varepsilon$ , 有

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

则称序列 $Y_1, Y_2, \dots, Y_n, \dots$ 依概率收敛于 $a$ , 记为

$$Y_n \xrightarrow{P} a$$

**定理 5.2 (弱大数定律(辛钦大数定律))** 设随机变量 $X_1, X_2, \dots$ 相互独立, 服从同一分布, 且具有数学期望 $E(X_k) = \mu (k = 1, 2, \dots)$ , 则序列 $\bar{X} = \frac{1}{n} \sum_{k=1}^n X_k$ 依概率收敛于 $\mu$ , 即 $\bar{X} \xrightarrow{P} a$

**定理 5.3 (伯努利大数定律)** 设 $f_A$ 是 $n$ 次独立重复试验中事件 $A$ 发生的次数,  $p$ 是事件 $A$ 在每次试验中发生的概率, 则对于任意正数 $\varepsilon > 0$ , 有

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

或

$$\lim_{n \rightarrow \infty} P \left\{ \left| \frac{f_A}{n} - p \right| \geq \varepsilon \right\} = 0$$

### 5.2 中心极限定理

**定理 5.4 (独立同分布的中心极限定理)** 设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 服从同一分布, 且具有数学期望和方差:  $E(X_k) = \mu, D(X_k) = \sigma^2 > 0 (k = 1, 2, \dots)$ , 则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

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

的分布函数 $F_n(x)$ 对于任意 $x$ 满足

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

**评论** 该定理表明, 均值为 $\mu$ , 方差为 $\sigma^2 > 0$ 的独立同分布的随机变量 $X_1, X_2, \dots, X_n$ 之和 $\sum_{k=1}^n X_k$ 的标准化变量, 当 $n$ 充分大时, 有

$$\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}\sigma} \sim N(0, 1)$$

也可改写为

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad \text{或} \quad \bar{X} \sim N(\mu, \sigma^2/n)$$

**定理 5.5 (李雅普诺夫(Lyapunov)定理)** 设随机变量 $X_1, X_2, \dots, X_n, \dots$ 相互独立, 它们具有数学期望和方差 $E(X_k) = \mu, D(X_k) = \sigma_k^2 > 0, k = 1, 2, \dots$ , 记 $B_n^2 = \sum_{k=1}^n \sigma_k^2$ 。若存在整数 $\delta$ , 使得当 $n \rightarrow \infty$ 时,

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

则随机变量之和 $\sum_{k=1}^n X_k$ 的标准化变量

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

的分布函数 $F_n(x)$ 对任意 $x$ , 满足

$$\lim_{n \rightarrow \infty} F_n(x) = \lim_{n \rightarrow \infty} P \left\{ \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

**评论** 该定理表明, 在定理的条件下, 随机变量

$$Z_n = \frac{\sum_{k=1}^n X_k - \sum_{k=1}^n \mu_k}{B_n}$$

当 $n$ 很大时, 近似地服从正态分布 $N(0, 1)$ 。由此, 当 $n$ 很大时,  $\sum_{k=1}^n X_k = B_n Z_n + \sum_{k=1}^n \mu_k$  近似地

服从正态分布 $N(\sum_{k=1}^n \mu_k, B_n^2)$ 。这也就是说, 无论各个随机变量 $X_k (k = 1, 2, \dots)$ 服从什么分布,

只要满足定理的条件, 那么它们的和 $\sum_{k=1}^n X_k$ 当 $n$ 很大时, 就近似地服从正态分布。

**定理 5.6 (棣莫弗-拉普拉斯(De Moivre-Laplace)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 $x$ , 有

$$\lim_{n \rightarrow \infty} P \left\{ \frac{\eta_n - np}{\sqrt{np(1-p)}} \leq x \right\} = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt = \Phi(x)$$

**评论** 该定理表明，正态分布是二项分布的极限分布。当 $n$ 充分大时，可以用该定理计算二项分布的概率。

## 6 样本及抽样分布

### 6.1 随机样本

**定义 6.1** 设 $X$ 是具有分布函数 $F$ 的随机变量, 若 $X_1, X_2, \dots, X_n$ 是具有同一分布函数 $F$ 的、相互独立的随机变量, 则称 $X_1, X_2, \dots, X_n$ 为从分布函数 $F$  (或总体 $F$ 、或总体 $X$ ) 得到的容量为 $n$ 的简单随机样本, 简称样本, 它们的观察值 $x_1, x_2, \dots, x_n$ 称为 样本值, 又称为 $X$ 的 $n$ 个独立的观察值。

由定义得: 若 $X_1, X_2, \dots, X_n$ 为 $F$ 的一个样本, 则 $X_1, X_2, \dots, X_n$ 相互独立, 且它们的分布函数都是 $F$ , 所以 $(X_1, X_2, \dots, X_n)$ 的分布函数为

$$F^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F(x_i)$$

又若 $X$ 具有概率密度 $f$ , 则 $(X_1, X_2, \dots, X_n)$ 的概率密度为

$$f^*(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$$

### 6.2 直方图和箱线图

直方图和箱线图的定义及画法略。

### 6.3 抽样分布

#### 定义 6.2

**定义 6.3** 设 $X_1, X_2, \dots, X_n$ 是来自总体 $X$ 的一个样本,  $g(X_1, X_2, \dots, X_n)$ 是 $X_1, X_2, \dots, X_n$ 的函数, 若 $g$ 中不含未知参数, 则称 $g(X_1, X_2, \dots, X_n)$ 是一统计量。

因为 $X_1, X_2, \dots, X_n$ 都是随机变量, 而统计量 $g(X_1, X_2, \dots, X_n)$ 是随机变量的函数, 因此统计量是一个随机变量. 设 $x_1, x_2, \dots, x_n$ 是相应于样本 $X_1, X_2, \dots, X_n$ 的样本值, 则称 $g(x_1, x_2, \dots, x_n)$ 是 $g(X_1, \dots, X_n)$ 的观察值。

下面列出几个常用的统计量, 设 $X_1, X_2, \dots, X_n$ 是来自总体 $X$ 的一个样本,  $x_1, x_2, \dots, x_n$ 是这一样本的观察值, 定义

样本平均值

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

样本方差

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

样本标准差

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

样本 $k$ 阶（原点）矩

$$A_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots$$

样本 $k$ 阶中心矩

$$B_k = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^k, \quad k = 2, 3, \dots$$

它们的观察值分别为

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

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

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

$$a_k = \frac{1}{n} \sum_{i=1}^n x_i^k, \quad k = 1, 2, \dots$$

$$b_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k, \quad k = 2, 3, \dots$$

**定义 6.4 (经验分布函数)** 与总体分布函数 $F(x)$ 相应的统计量——经验分布函数的做法为：设 $X_1, X_2, \dots, X_n$ 是总体 $F$ 的一个样本，用 $S(x), -\infty < x < \infty$ 表示 $X_1, X_2, \dots, X_n$ 中不大于 $x$ 的随机变量的个数.定义经验分布函数 $F_n(x)$ 为

$$F_n(x) = \frac{1}{n} S(x), \quad -\infty < x < \infty$$

**定义 6.5 ( $\chi^2$ 分布)** 设 $X_1, X_2, \dots, X_n$ 是来自总体 $N(0, 1)$ 的样本，则称统计量

$$\chi^2 = X_1^2 + X_2^2 + \dots + X_n^2$$

服从自由度为 $n$ 的 $\chi^2$ 分布，记为 $\chi^2 \sim \chi^2(n)$ .此处自由度是指上式右端包含的独立变量的个数。

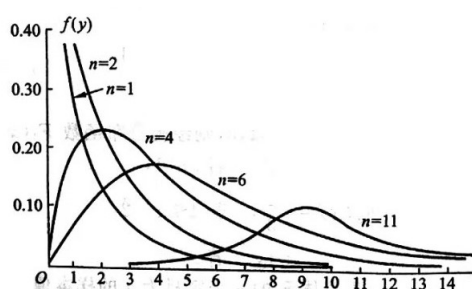


图 1:  $\chi^2(n)$ 分布的概率密度图

$\chi^2(n)$ 分布的概率密度为

$$f(y) = \begin{cases} \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}, & y > 0 \\ 0, & \text{其他} \end{cases}$$

$f(y)$ 的图形如图1所示。

**定理 6.1**  $\chi^2$ 分布的性质:

$\chi^2$ 分布的可加性 设 $\chi_1^2 \sim \chi^2(n_1)$ ,  $\chi_2^2 \sim \chi^2(n_2)$ , 并且 $\chi_1^2, \chi_2^2$ 相互独立, 则有

$$\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$$

$\chi^2$ 分布的数学期望和方差 若 $\chi^2 \sim \chi^2(n)$ , 则有

$$E(\chi^2) = n, \quad D(\chi^2) = 2n$$

$\chi^2$ 分布的上分位点 对于给定的正数 $\alpha, 0 < \alpha < 1$ , 满足条件

$$P\{\chi^2 > \chi_\alpha^2(n)\} = \int_{\chi_\alpha^2(n)}^{\infty} f(y) dy = \alpha$$

的点 $\chi_\alpha^2(n)$ 就是 $\chi^2(n)$ 分布的上 $\alpha$ 分位点。

**定义 6.6 (t分布)** 设 $X \sim N(0, 1)$ ,  $Y \sim \chi^2(n)$ , 且 $X, Y$ 相互独立, 则称随机变量

$$t = \frac{X}{\sqrt{Y/n}}$$

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

$t$ 分布又称学生氏(Student)分布. $t(n)$ 分布的概率密度函数为

$$h(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{\pi n} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2}, \quad -\infty < t < \infty$$

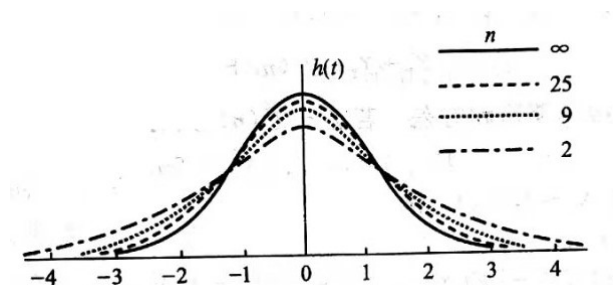


图 2:  $t$ 分布的概率密度图

**定义 6.7 (t分布的上分位点)** 对于给定的 $\alpha, 0 < \alpha < 1$ , 满足条件

$$P\{t > t_\alpha(n)\} = \int_{t_\alpha(n)}^{\infty} h(t) dt = \alpha$$

的点 $t_{\alpha}(n)$ 就是 $t(n)$ 分布的上 $\alpha$ 分位点。由定义及 $h(t)$ 图形的对称性可知

$$t_{1-\alpha}(n) = -t_{\alpha}(n)$$

**定义 6.8 (F分布)** 设 $U \sim \chi^2(n_1), V \sim \chi^2(n_2)$ , 且 $U, V$ 相互独立, 则称随机变量

$$F = \frac{U/n_1}{V/n_2}$$

服从自由度为 $(n_1, n_2)$ 的 $F$ 分布, 记为 $F \sim F(n_1, n_2)$ .

$F(n_1, n_2)$ 分布的概率密度为

$$\psi(y) = \begin{cases} \frac{\Gamma[(n_1+n_2)/2](n_1/n_2)^{n_1/2} y^{(n_1/2)-1}}{\Gamma(n_1/2)\Gamma(n_2/2)[1+(n_1y/n_2)]^{(n_1+n_2)/2}}, & y > 0, \\ 0, & \text{其他} \end{cases}$$

$\psi(y)$ 的图形如下图所示

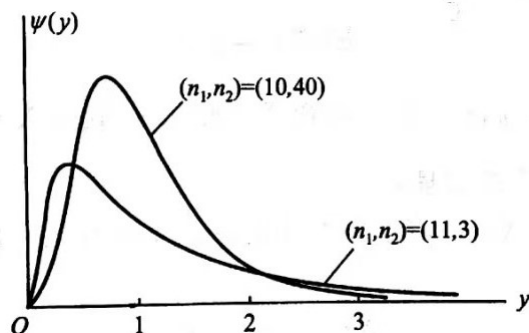


图 3:  $F$ 分布的概率密度图

由定义可知, 若 $F \sim F(n_1, n_2)$ , 则

$$\frac{1}{F} \sim F(n_2, n_1)$$

**定义 6.9 (F分布的上分位点)** 对于给定的 $\alpha, 0 < \alpha < 1$ , 满足条件

$$P\{F > F_{\alpha}(n_1, n_2)\} = \int_{F_{\alpha}(n_1, n_2)}^{\infty} \psi(y) dy = \alpha$$

的点 $F_{\alpha}(n_1, n_2)$ 就是 $F(n_1, n_2)$ 分布的上 $\alpha$ 分位点。

$F$ 分布的上 $\alpha$ 分布有如下的重要性质:

$$F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_2, n_1)}$$

**定理 6.2** 设 $X_1, X_2, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本,  $\bar{X}$ 是样本均值, 则有

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

**定理 6.3** 设 $X_1, X_2, \dots, X_n$ 是来自总体 $N(\mu, \sigma^2)$ 的样本,  $\bar{X}, S^2$ 分别是样本均值和样本方差, 则有



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

2°  $\bar{X}$  与  $S^2$  相互独立

**定理 6.4** 设  $X_1, X_2, \dots, X_n$  是来自总体  $N(\mu, \sigma^2)$  的样本,  $\bar{X}, S^2$  分别是样本均值和样本方差, 则有

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

**定理 6.5** 设  $X_1, X_2, \dots, X_{n_1}$  与  $Y_1, Y_2, \dots, Y_{n_2}$  分别是来自正态总体  $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$  的样本, 且这

两个样本相互独立. 设  $\bar{X} = \frac{1}{n_1} \sum_{i=1}^{n_1} X_i, \bar{Y} = \frac{1}{n_2} \sum_{i=1}^{n_2} Y_i$  分别是这两个样本的样本均值;  $S_1^2 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (X_i - \bar{X})^2$

,  $S_2^2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (Y_i - \bar{Y})^2$  分别是这两个样本的样本方差, 则有

$$1^\circ \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1-1, n_2-1);$$

2° 当  $\sigma_1^2 = \sigma_2^2 = \sigma^2$  时,

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_w \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2),$$

其中

$$S_w^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}, \quad S_w = \sqrt{S_w^2}$$

## 7 参数估计

### 7.1 点估计

**定义 7.1** 点估计问题的一般提法如下：设总体 $X$ 的分布函数 $F(x; \theta)$ 的形式为已知， $\theta$ 是待估参数。 $X_1, X_2, \dots, X_n$ 是 $X$ 的一个样本， $x_1, x_2, \dots, x_n$ 是相应的一个样本值。点估计问题就是要构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ ，用它的观察值 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 作为未知参数 $\theta$ 的近似值。称 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 $\theta$ 的估计量，称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 $\theta$ 的估计值。在不致混淆的情况下统称估计量和估计值为估计。

**定义 7.2 (矩估计法)** 设 $X$ 为连续型随机变量，其概率密度为 $f(x; \theta_1, \theta_2, \dots, \theta_k)$ ，或 $X$ 为离散型随机变量，其分布律为 $P\{X = x\} = p(x; \theta_1, \theta_2, \dots, \theta_k)$ ，其中 $\theta_1, \theta_2, \dots, \theta_k$ 为待估参数， $X_1, X_2, \dots, X_n$ 是来自 $X$ 的样本。假设总体 $X$ 的前 $k$ 阶矩

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x; \theta_1, \theta_2, \dots, \theta_k) dx \quad X \text{ 为连续型}$$

或

$$\mu_l = E(X^l) = \sum_{x \in R_X} x^l p(x; \theta_1, \theta_2, \dots, \theta_k) \quad X \text{ 为离散型}$$

( $l = 1, 2, \dots, k$ , 其中 $R_X$ 是 $X$ 可能取值的范围) 存在。一般来说，它们是 $\theta_1, \theta_2, \dots, \theta_k$ 的函数。基于样本矩

$$A_l = \frac{1}{n} \sum_{i=1}^n X_i^l$$

依概率收敛于相应的总体矩 $\mu_l$  ( $l = 1, 2, \dots, k$ )，样本矩的连续函数依概率收敛于相应的总体矩的连续函数，我们就用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应的总体矩的连续函数的估计量。这种估计方法称为矩估计法。

矩估计法的具体做法如下：设

$$\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \dots, \theta_k), \\ \mu_2 = \mu_2(\theta_1, \theta_2, \dots, \theta_k), \\ \vdots \\ \mu_k = \mu_k(\theta_1, \theta_2, \dots, \theta_k), \end{cases}$$

这是一个包含 $k$ 个未知参数 $\theta_1, \theta_2, \dots, \theta_k$ 的联立方程组。一般来说，可以从中解出 $\theta_1, \theta_2, \dots, \theta_k$ ，得到

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \dots, \mu_k), \\ \theta_2 = \theta_2(\mu_1, \mu_2, \dots, \mu_k), \\ \vdots \\ \theta_k = \theta_k(\mu_1, \mu_2, \dots, \mu_k), \end{cases}$$

以 $A_i$ 分别代替上式中的 $\mu_i, i = 1, 2, \dots, k$ ，就以

$$\hat{\theta}_i = \theta_i(A_1, A_2, \dots, A_k), i = 1, 2, \dots, k$$

分别作为 $\theta_i, i = 1, 2, \dots, k$ 的估计量, 这种估计量称为矩估计量, 矩估计量的观察值称为矩估计值.

**定义 7.3 (最大似然估计法)** 若总体 $X$ 属离散型, 其分布律 $P\{X = x\} = p(x; \theta), \theta \in \Theta$ 的形式为已知,  $\theta$ 为待估参数,  $\Theta$ 是 $\theta$ 可能取值的范围. 设 $X_1, X_2, \dots, X_n$ 是来自 $X$ 的样本, 则 $X_1, X_2, \dots, X_n$ 的联合分布律为

$$\prod_{i=1}^n p(x_i; \theta)$$

又设 $x_1, x_2, \dots, x_n$ 是相应于样本 $X_1, X_2, \dots, X_n$ 的一个样本值. 易知样本 $X_1, X_2, \dots, X_n$ 取到观察值 $x_1, x_2, \dots, x_n$ 的概率, 亦即事件 $\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率为

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \theta \in \Theta$$

这一概率随 $\theta$ 的取值而变化, 它是 $\theta$ 的函数,  $L(\theta)$ 称为样本的似然函数. (注意, 这里 $x_1, x_2, \dots, x_n$ 是已知的样本值, 它们都是常数).

由费希尔引进的最大似然估计法, 就是固定样本观察值 $x_1, x_2, \dots, x_n$ , 在 $\theta$ 取值的可能范围 $\Theta$ 内挑选使似然函数 $L(x_1, x_2, \dots, x_n; \theta)$ 达到最大的参数值 $\hat{\theta}$ , 作为参数 $\theta$ 的估计值. 即取 $\hat{\theta}$ 使

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$$

这样得到的 $\hat{\theta}$ 与样本值 $x_1, x_2, \dots, x_n$ 有关, 常记为 $\hat{\theta}(x_1, x_2, \dots, x_n)$ , 称为参数 $\theta$ 的最大似然估计值, 而相应的统计量 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 称为参数 $\theta$ 的最大似然估计量.

若总体 $X$ 属连续型, 其概率密度 $f(x; \theta), \theta \in \Theta$ 的形式已知,  $\theta$ 为待估参数,  $\Theta$ 是 $\theta$ 可能取值的范围. 设 $X_1, X_2, \dots, X_n$ 是来自 $X$ 的样本, 则 $X_1, X_2, \dots, X_n$ 的联合密度为

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

设 $x_1, x_2, \dots, x_n$ 是相应于样本 $X_1, X_2, \dots, X_n$ 的一个样本值, 则随机点 $(X_1, X_2, \dots, X_n)$ 落在点 $(x_1, x_2, \dots, x_n)$ 的邻域 (边长分别为  $dx_1, dx_2, \dots, dx_n$  的 $n$ 维立方体) 内的概率近似地为

$$\prod_{i=1}^n f(x_i; \theta) dx_i \tag{1}$$

其值随 $\theta$ 的取值而变化. 与离散型的情况一样, 取 $\theta$ 的估计值 $\hat{\theta}$ 使概率(1)取到最大值, 但因 $\prod_{i=1}^n dx_i$ 不随 $\theta$ 而变, 故只考虑函数

$$L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$$

的最大值. 这里 $L(\theta)$ 称为样本的似然函数. 若

$$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$$

则称 $\hat{\theta}(x_1, x_2, \dots, x_n)$ 为 $\theta$ 的最大似然估计值, 称 $\hat{\theta}(X_1, X_2, \dots, X_n)$ 为 $\theta$ 的最大似然估计量.

## 8 假设检验

## 9 方差分析及回归分析