# 概率论与数理统计

喵喵喵

2020 年 10 月 21 日

## 目录

# 1 概率论的基本概念

## 1.1 随机实验

## 1.2 样本空间、随机事件

将随机试验 $E$ 的所有可能结果组成的集合称为 $E$ 的 **样本空间**，记为 $S$，样本空间的元素，即 $E$ 的每个结果，称为 **样本点**

称试验 $E$ 的样本空间 $S$ 的子集为 $E$ **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**

由一个样本点组成的单点集称为 **基本事件**

样本空间 $S$ 包含所有的样本点，他是 $S$ 自身的子集，在每次试验中它总是出现，$S$ 称为 **必然事件**，空集 $\emptyset$ 称为 **不可能事件**

设试验 $E$ 的样本空间为 $S$，而 $A, B, A_k \subseteq S$

1. 若 $A \subset B$，则称事件 $B$ 包含事件 $A$，事件 $A$ 的发生必导致事件 $B$ 发生
   若 $A \subset B, B \subset A$ 即 $A = B$，则称事件 $A$ 与事件 $B$ **相等**
2. 事件 $A \cup B = \{x \mid x \in A \text{ or } x \in B\}$ 称为事件 $A$ 与事件 $B$ 的 **和事件**，当且仅当 $A, B$ 中至少由一个发生时，事件 $A \cup B$ 发生
3. 事件 $A \cap B = \{x \mid x \in A \text{ and } x \in B\}$ 称为事件 $A$ 与事件 $B$ 的 **积事件**，也记作 $AB$
   称 $\bigcap_{k=1}^{\infty} A_k$ 为可列个事件 $A_1, A_2, ...$ 的积事件
4. 事件 $A - B = \{x \mid x \in A \text{ and } x \notin B\}$ 称为事件 $A$ 与事件 $B$ 的 **差事件**
5. 若 $A \cap B = \emptyset$，则称事件 $A$ 与事件 $B$ **互不相容**或 **互斥**的
6. 若 $A \cup B = S, A \cap B = \emptyset$，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**

## 1.3 频率与概率

**Definition 1.1.** 在相同的条件下，进行了 $n$ 次试验，在这 $n$ 次试验中，事件 $A$ 发生次数 $n_A$ 称为事件 $A$ 发生的 **频数**，比值 $n_A/n$ 称为事件 $A$ 发生的 **频率**，并记成 $f_n(A)$

基本性质

1. $0 \leq f_n(A) \leq 1$
2. $f_n(S) = 1$
3. 若 $A_1, \dots, A_k$ 是两两互不相容的事件，则

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

**Definition 1.2.** 设 $E$ 是随机试验，$S$ 是它的样本空间，对于 $E$ 的每一事件 $A$ 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，有 $P(A) = 1$
3. **可列可加性**：设 $A_1, A_2, \dots$ 是两两互不相容事件，有

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

**Proposition 1.3.** $P(\emptyset) = 0$

证明. 令 $A_n = \emptyset$，则 $\bigcup_{n=1}^{\infty} A_n = \emptyset$，且 $A_i A_j = \emptyset$，由可列可加性

$$P(\emptyset) = P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_n) = \sum_{n=1}^{\infty} P(\emptyset)$$

由概率的非负性，$P(\emptyset) = 0$ □

**Proposition 1.4** (有限可加性)**.** *若 $A_1, \dots, A_n$ 是两两互不相容的事件，则有*

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

证明. 令 $A_{n+1} = A_{n+2} = \dots = \emptyset$，即有 $A_i A_j = \emptyset$

$$\begin{aligned} & P(A_1 \cup A_2 \cup \dots \cup A_n) \\ & \quad = P(\bigcup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k) \\ & \quad = \sum_{k=1}^{\infty} P(A_k) + 0 = P(A_1) + \dots + P(A_n) \end{aligned}$$

□

**Proposition 1.5.** 设 $A, B$ 是两个事件，若 $A \subset B$, 则有

$$P(B-A) = P(B) - P(A)$$
$$P(B) \geq P(A)$$

证明. $B = A \cup (B-A)$ □

**Proposition 1.6.** 对于任一事件 $A$

$$P(A) \leq 1$$

证明. $P(A) \leq P(S) = 1$ □

**Proposition 1.7** (逆事件的概率). 对于任一事件 $A$, 有

$$P(\overline{A}) = 1 - P(A)$$

证明. $P(S) = P(A \cup \overline{A})$ □

**Proposition 1.8** (加法公式). 对于任一两事件 $A, B$ 有

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

证明. $A \cup B = A \cup (B - AB)$

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

□

可推广到

$$P(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{i=1}^{n} P(A_i) - \sum_{1 \leq i < j \leq n} P(A_i A_j)$$
$$+ \sum_{1 \leq i < j < k \leq n} P(A_i A_j A_k) + \cdots + (-1)^{n-1} P(A_1 \ldots A_n)$$

## 1.4 等可能概型（古典概型）

**等可能概型**（古典概型）

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

设试验的样本空间为 $S = \{e_1, \dots, e_n\}$，由于在试验中每个基本事件发生的可能性相同，即

$$P(\{e_1\}) = \cdots = P(\{e_n\})$$

又由于基本事件两两互不相容，于是

$$\begin{aligned} 1 = P(S) &= P(\{e_1\} \cup \cdots \cup \{e_n\}) \\ &= P(\{e_1\}) + \cdots + P(\{e_n\}) \\ &= nP(\{e_i\}) \end{aligned}$$

于是

$$P(\{e_i\}) = \frac{1}{n}$$

若事件 $A$ 包含 $k$ 个基本 i 事件，即 $A = \{e_{i_1}\} \cup \cdots \cup \{e_{i_k}\}$，则有

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

**Example 1.1.** 设有 $N$ 件产品，其中有 $D$ 件次品，今从中任取 $n$ 件，文其中恰有 $k(k \leq D)$ 件次品的概率

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

**Example 1.2.** 袋中有 $a$ 只白球，$b$ 只红球，$k$ 个人依次在袋中取一只球，求第 $i$ 人取到白球（记为事件 $B$）的概率（$k \leq a + b$）

共有 $A_{a+b}^{k}$ 个基本事件，事件 $B$ 发生时，第 $i$ 人取的应是白球，有 $a$ 中取法，剩余 $k-1$ 只球有 $A_{a+b-1}^{k-1}$ 种取法，则

$$P(B) = \frac{a \cdot A_{a+b-1}^{k-1}}{A_{a+b}^{k}} = \frac{a}{a+b}$$

## 1.5 条件概率

**Definition 1.9.** 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**

条件概率 $P(\cdot|A)$ 符合

1. **非负性**：对于每一事件 $B$，有 $P(B|A) \geq 0$
2. **规范性**：对于必然事件 $S$，有 $P(S|A) = 1$
3. **可列可加性**：设 $B_1, B_2, \ldots$ 是两两互不相容的事件，则有

$$P(\bigcup_{i=1}^{\infty} B_i|A) = \sum_{i=1}^{\infty} P(B_i|A)$$

**Theorem 1.10** (乘法定理)**.** 设 $P(A) > 0$，则有

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

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

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

**Example 1.3.** 设袋中装有 $r$ 只红球，$t$ 只白球，每次自袋中任取一只球，观察其颜色再放回，并再放入 $a$ 只与所取出的那只球同色的球，若在袋中连续取球四次，试求第一、二次取到红球且第三、四次取到白球的概率

以 $A_i$ 表示事件“第 $i$ 次取到红球”，则

$$\begin{aligned} P(A_1 A_2 \overline{A_3}\,\overline{A_4}) &= P(\overline{A_4}|A_1 A_2 \overline{A_3})P(\overline{A_3}|A_1 A_2)P(A_2|A_1)P(A_1) \\ &= \frac{t+a}{r+t+3a} \cdot \frac{t}{r+t+2a} \cdot \frac{r+a}{r+t+a} \cdot \frac{r}{r+t} \end{aligned}$$

**Definition 1.11.** 设 $S$ 为试验 $E$ 的样本空间，$B_1, \ldots, B_n$ 为 $E$ 的一组事件，若

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

则称 $B_1, \dots, B_n$ 为样本空间 $S$ 的一个 **划分**

**Theorem 1.12.** 设试验 $E$ 的样本空间为 $S$，$A$ 为 $E$ 的事件，$B_1, \dots, B_n$ 为 $S$ 的一个划分，且 $P(B_i) > 0$，则

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n)$$

称为 **全概率公式**

证明.

$$A = AS = A(B_1 \cup \cdots \cup B_n) = AB_1 \cup \cdots \cup AB_n$$

□

**Theorem 1.13.** 设试验 $E$ 的样本空间 $S$，$A$ 为 $E$ 的事件，$B_1, \dots, B_n$ 为 $S$ 的一个划分，且 $P(A) > 0, P(B_i) > 0$，则

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

称为 **贝叶斯公式**

特别的，取 $n = 2$，则

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

**Example 1.4.** 患肺癌的概率约为 0.1%, 在人群中有 20% 是吸烟者，他们患肺癌的概率约为 0.4%，求不吸者患肺癌的概率

以 $C$ 记事件 “患肺癌”，以 $A$ 记事件 “吸烟”，则 $P(C) = 0.001, P(A) = 0.2, P(C|A) = 0.004$，由全概率公式

$$P(C) = P(C|A)P(A) + P(C|\overline{A})P(\overline{A})$$

因此

$$P(C|\overline{A}) = 0.00025$$

### 1.6 独立性

**Definition 1.14.** 设 $A, B$ 是两事件，如果满足

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

则称事件 $A, B$ **相互独立**，简称 $A, B$ **独立**

**Theorem 1.15.** 设 $A, B$ 是两事件，且 $P(A) > 0$，若 $A, B$ 相互独立，则 $P(B|A) = P(B)$

**Theorem 1.16.** 若事件 $A, B$ 相互独立，则 $A$ 与 $\overline{B}$，$\overline{A}$ 与 $B$，$\overline{A}$ 与 $B$ 也相互独立

**Definition 1.17.** 设 $A, B, C$ 是三个事件，满足

$$\begin{cases} P(AB) = P(A)P(B) \\ P(BC) = P(B)P(C) \\ P(AC) = P(A)P(C) \\ P(ABC) = P(A)P(B)P(C) \end{cases}$$

则称事件 $A, B, C$ **相互独立**

一般地，设 $A_1, \dots, A_n$，如果对于 q 其中任意 $2, 3, \dots, n$ 个事件的积事件的概率，都等于各事件概率之积，则称事件 $A_1, \dots, A_n$ **相互独立**

**Example 1.5.** 要验收一批（100）件乐器，验收方案如下：自该批乐器中随机地取 3 件测试（相互独立），如果 3 件中至少有一件在测试中被认为音色不纯，则这批乐器被拒绝接收。设一件音色不纯的乐器经测试查出其音色不纯的概率为 0.95，而一件音色纯的乐器被误认为不纯的概率为 0.01，已知 100 中有 4 件音色不纯，试问这批乐器被接收的概率是多少

设以 $H_i$ 表示 3 件中恰有 $i$ 件不纯，$A$ 表示这批批乐器被接收，则

$$P(A|H_0) = 0.99^3, P(A|H_1) = 0.99^2 \times 0.05$$
$$P(A|H_2) = 0.99 \times 0.05^2, P(A|H_3) = 0.05^3$$

而

$$P(H_0)=\frac{\binom{96}{3}}{\binom{100}{3}},P(H_1)=\frac{\binom{4}{1}\binom{96}{2}}{\binom{100}{3}}$$
$$P(H_2)=\frac{\binom{4}{2}\binom{96}{1}}{\binom{100}{3}},P(H_3)=\frac{\binom{4}{3}}{\binom{100}{3}}$$

故

$$P(A)=\sum P(A|H_i)P(H_i)$$

# 2 随机变量及其分布

## 2.1 随机变量

**Definition 2.1.** 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

设离散型随机变量 $X$ 所有可能取的值为 $x_k(k=1,2,\dots)$，$X$ 取各个可能值的概率，即事件 $\{X=x_k\}$ 的概率，为

$$P\{X=x_k\}=p_k,k=1,2,... \tag{2.2.1}$$

由概率的定义,$p_k$ 满足如下两个条件

1. $p_k\geq 0,k=1,2,...$
2. $\sum_{k=1}^{\infty}p_k=1$

我们称 (2.2.1) 为离散型随机变量 $X$ 的 **分布律**，分布律也可以用表格表示

| $X$ | $x_1$ | $x_2$ | ... | $x_n$ | ... |
|---|---|---|---|---|---|
| $p_k$ | $p_1$ | $p_2$ | ... | $p_n$ | ... |

### 2.2.1 （0-1）分布

设随机变量 $X$ 只可能取 0 与 1 两个值，它的分布律是

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

则称 $X$ 服从以 $p$ 为参数的（0-1）分布或两点分布

### 2.2.2 伯努利试验、二项分布

设试验 $E$ 只有两个可能结果：$A$ 及 $\overline{A}$, 则称 $E$ 为 **伯努利试验**。设 $P(A)=p(0<p<1)$，此时 $P(\overline{A}=1-p)$。将 $E$ 独立重复地进行 $n$ 次，则称这一串重复的独立试验为 $n$ **重伯努利试验**

以 $X$ 表示 $n$ 重伯努利事件 $A$ 发生的次数，$X$ 是一个随机变量。记 $q=1-p$，即有

$$P\{X=k\}=\binom{n}{k}p^kq^{1-k}$$

注意到 $\binom{n}{k}p^kq^{1-k}$ 刚好是 $(p+q)^n$ 的展开式中出现 $p^k$ 的那一项，我们称随机变量 $X$ 服从参数 $n,p$ 的 **二项分布**，并记为 $X\sim b(n,p)$

### 2.2.3 泊松分布

设随机变量 $X$ 所有可能的值为 $0,1,2,...$，而各个值的概率为

$$P\{X=k\}=\frac{\lambda^ke^{-\lambda}}{k!}$$

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

易知 $P\{X=k\}\geq 0$，$k=0,1,2,...$，且有

$$\sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^ke^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$$

**Theorem 2.2** (泊松定理)**.** 设 $\lambda>0$ 是一个常数，$n$ 是任意正整数，设 $np_n=\lambda$，则对于任意固定的非负整数 $k$，有

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

证明. 设 $p_n = \frac{\lambda}{n}$，有

$$\begin{aligned}\binom{n}{k} p_n^k (1-p_k)^{n-k} &= \frac{n(n-1)\ldots(n-k+1)}{k!}(\frac{\lambda}{n})^k(1-\frac{\lambda}{n})^{n-k} \\ &= \frac{\lambda^k}{k!}[(1-\frac{1}{n})\ldots(1-\frac{k-1}{n})](1-\frac{\lambda}{n})^n(1-\frac{\lambda}{n})^{-k}\end{aligned}$$

当 $n \to \infty$ 时 $(1-\frac{\lambda}{n})^n \to e^{-\lambda}$，故有

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

□

也就是说以 $n, p$ 为参数的二项分布的概率值可以由参数为 $\lambda = np$ 的泊松分布的概率值近似

### 2.2.4 超几何分布

$$P\{X=k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}, k = l_1, \ldots, l_2$$

其中 $l_1 = \max\{0, n-N+M\}, l_2 = \min\{M, n\}$，则称随机变量 $X$ 服从参数 $n, N, M$ 的超几何分布

如果 $N$ 件产品中含有 $M$ 件次品，从中任意一次取出 $n$ 件，令 $X =$ 抽取的 $n$ 件产品中的次品件数，则 $X$ 服从参数 $n, N, M$ 的超几何分布

如果 $N$ 件产品中含有 $M$ 件次品，从中一件接一件有放回地取 $n$ 次，则 $X$ 服从 $B(n, \frac{M}{N})$

**Lemma 2.3.** $C_{m+n}^r = \sum_{t=0}^{r} C_m^{r-t} C_n^t$

证明.

$$(x+y)^{m+n} = (x+y)^m (x+y)^n \tag{2.2.2}$$

$$\sum_{r=0}^{m+n} C_{m+n}^r x^r y^{m+n-r} = \left[\sum_{s=0}^{m} C_m^s x^s y^{m-s}\right]\left[\sum_{t=0}^{n} C_n^t x^t y^{n-t}\right] \tag{2.2.3}$$

$$\sum_{r=0}^{m+n} C_{m+n}^r x^r y^{m+n-r} = \sum_{s=0}^{m}\sum_{t=0}^{n} C_m^s C_n^t x^{s+t} y^{m+n-s-t} \tag{2.2.4}$$

□

**Proposition 2.4.** $E(X) = \frac{nM}{N}$

x=k,s=M,N=N,n=n

证明.

$$\begin{aligned}
E(X) &= \sum_{k=0}^{n} kP\{X=k\} = \sum_{k=0}^{n} \frac{kC_M^k C_{N-M}^{n-k}}{C_N^n} \\
&= \sum_{k=0}^{n} C_{N-M}^{n-k} \frac{kM!}{k!(M-k)!} \frac{n!(N-n)!}{N!} \\
&= \sum_{k=1}^{n} C_{N-M}^{n-k} \frac{M(M-1)!}{(k-1)!(M-k)!} \frac{n(n-1)!(N-n)!}{N(N-1)!} \\
&= \frac{nM}{N} \sum_{k=1}^{n} \frac{C_{N-M}^{n-k} C_{M-1}^{k-1}}{C_{N-1}^{n-1}} \\
&= \frac{nM}{N} \frac{1}{C_{N-1}^{n-1}} \sum_{k=1}^{n} C_{N-M}^{n-k} C_{M-1}^{k-1} \\
&= \frac{nM}{N} \frac{1}{C_{N-1}^{n-1}} \sum_{k=0}^{n-1} C_{N-M}^{n-k-1} C_{M-1}^{k} \\
&= \frac{nM}{N} \frac{C_{N-1}^{n-1}}{C_{N-1}^{n-1}} \\
&= \frac{nM}{N}
\end{aligned}$$

□

## 2.3 随机变量的分布函数

**Definition 2.5.** 设 $X$ 是一个随机变量，$x$ 是任意实数，函数

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

称为 $X$ 的 **分布函数**

对于任意实数 $x_1, x_2(x_1 < x_2)$，有

$$\begin{aligned} P\{x_1 < X \le x_2\} &= P\{X \le x_2\} - P\{X \le x_1\} \\ &= F(x_2) - F(x_1) \end{aligned}$$

分布函数 $F(x)$ 具有以下的基本性质

1. $F(x)$ 是一个不减函数
2. $0 \le F(x) \le 1$，且

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

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

3. $F(x+0) = F(x)$

## 2.4 连续型随机变量及其概率密度

如果对于随机变量 $X$ 的分布函数 $F(x)$，存在非负函数 $f(x)$ 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**，其中函数 $f(x)$ 称为 $X$ 的 **概率密度函数**，简称 **概率密度**

概率密度 $f(x)$ 具有以下性质

1. $f(x) \ge 0$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. 对于任意实数 $x_1, x_2(x_1 \le x_2)$，

$$P\{x_1 < X \le x_2\} = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx$$

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

### 2.4.1 均匀分布

若连续型随机变量 $X$ 有概率密度

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

则称 $X$ 在区间 $(a,b)$ 上服从 **均匀分布**，记为 $X\sim U(a,b)$。分布函数

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

### 2.4.2 指数分布

若连续型随机变量 $X$ 的概率密度为

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

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

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

对于任意 $s,t>0$，有

$$P\{X>s+t|X>s\}=P\{X>t\}$$

事实上

$$\begin{aligned}P\{X>s+t|X>s\}&=\frac{P\{(X>s+t)\cap(X>s)\}}{P\{X>s\}}\\&=\frac{P\{X>s+t\}}{P\{X>s\}}=\frac{1-F(s+t)}{1-F(s)}\\&=e^{-t/\theta}=P\{X>t\}\end{aligned}$$

这个性质称为无记忆性

### 2.4.3 正态分布

若连续型随机变量 $X$ 的概率密度为

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

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

令 $(x-\mu)/\sigma=t$,记 $I=\int_{-\infty}^{\infty}e^{-t^2/2}dt$,则有 $I^2=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}e^{-(t^2+u^2)/2}dtdu$,利用极坐标得

$$I^2=\int_0^{2\pi}\int_0^{\infty}re^{-r^2/2}drd\theta=2\pi$$

$f(x)$ 有以下性质

1. 曲线关于 $x=\mu$ 对称
2. 当 $x=\mu$ 时取得最大值
$$f(\mu)=\frac{1}{\sqrt{2\pi}\sigma}$$

特别的，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，即

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

易知

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

**Lemma 2.6.** *若* $X\sim N(\mu,\sigma^2)$，*则* $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

于是，若 $X\sim N(\mu,\sigma^2)$，则

$$F(x)=P\{X\leq x\}=P\{\frac{X-\mu}{\sigma}\leq\frac{x-\mu}{\sigma}\}=\Phi(\frac{x-\mu}{\sigma})$$

## 2.5 随机变量的函数的分布

**Example 2.1.** 设随机变量 $X$ 具有概率密度 $f_X(x), -\infty < x < \infty$, 求 $Y = X^2$ 的概率密度

分别记 $X, Y$ 的分布函数为 $F_X(x), F_Y(y)$，当 $y > 0$ 时

$$\begin{aligned} F_Y(y) &= P\{Y \leq y\} = \{X^2 \leq y\} \\ &= P\{-\sqrt{y} \leq X \leq \sqrt{y}\} \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

因此

$$f_Y(y) = \begin{cases} \frac{1}{2\sqrt{y}}[f_X(\sqrt{y} + f_X(-\sqrt{y}))] & y > 0 \\ 0 & y \leq 0 \end{cases}$$

**Theorem 2.7.** 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

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

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

**Proposition 2.8.** 设随机变量 $X \sim N(\mu, \sigma^2)$，试证明 $X$ 的线性函数 $Y = aX + b(a \neq 0)$ 也服从正态分布

证明. $X$ 的概率密度为

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

由 $Y = aX + b$

$$x = \frac{y-b}{a}, h'(y) = \frac{1}{a}$$

因此

$$
\begin{aligned}
f_Y(y) &= \frac{1}{|a|} f_X(\frac{y-b}{a}), -\infty < y < \infty \\
&= \frac{1}{|a|\sigma\sqrt{2\pi}} e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2(a\sigma)^2}}
\end{aligned}
$$

即有 $Y = aX + B \sim N(a\mu + b, (a\sigma)^2)$ □

# 3 多维随机变量及其分布

## 3.1 二维随机变量

设 $E$ 是一个随机试验,它的样本空间是 $S = \{e\}$,设 $X = X(e), Y = Y(e)$ 是定义在 $S$ 上的随机变量，它们构成的一个向量 $(X, Y)$ 叫做 **二维随机向量** 或 **二维随机变量**

**Definition 3.1.** 设 $(X, Y)$ 是二维随机变量，对于任意实数 $x, y$，二元函数

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

称为二维随机变量的 **分布函数**, 或称为随机变量 $X, Y$ 的 **联合分布函数**

分布函数 $F(x, y)$ 具有以下性质

1. $F(x, y)$ 是变量 $x, y$ 的不减函数
2. $0 \leq F(x, y) \leq 1$，且

$$
\begin{aligned}
&\forall y, F(-\infty, y) = 0 \\
&\forall x, F(x, -\infty) = 0 \\
&F(-\infty, -\infty) = 0, F(\infty, \infty) = 1
\end{aligned}
$$

3. $F(x+0, y) = F(x, y), F(x, y+0) = F(x, y)$，即 $F(x, y)$ 关于 $x$ 又连续，关于 $y$ 也右连续
4. 对于任意 $(x_1, y_1), (x_2, y_2), x_1 < x_2, y_1 < y_2$，下列不等式成立

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

如果二维随机变量 $(X,Y)$ 全部可能取到的值是有限对或可列无限多对，则称 $(X,Y)$ 是 **离散型的随机变量**

设二维离散型随机变量 $(X,Y)$ 所有可能取的值为 $(x_i,y_j), i,j=1,2,...$，记 $P\{X=x_i,Y=y_j\}=p_{ij}$，称为二维离散型随机变量 $(X,Y)$ 的 **分布律**，或随机变量 $X,Y$ 的 **联合分布律**

| $Y,X$ | $x_1$ | ... | $x_i$ | ... |
|---|---|---|---|---|
| $y_1$ | $p_{11}$ | ... | $p_{i1}$ | ... |
| $\vdots$ | | | | |
| $y_j$ | $p_{1j}$ | | $p_{ij}$ | |

对于二维随机变量 $(X,Y)$ 的分布函数 $F(x,y)$，如果存在非负的函数 $f(x,y)$ 使对于任意 $x,y$ 有

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

则称 $(X,Y)$ 是 **连续型的二维随机变量**，函数 $f(x,y)$ 称为二维随机变量的 **概率密度**，或称为随机变量 $X,Y$ 的 **联合概率密度**

概率密度 $f(x,y)$ 具有以下性质

1. $f(x,y)\geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f(x,y)dxdy=1$
3. 设 $G$ 是 $xOy$ 平面的区域，点 $(X,Y)$ 落在 $G$ 内的概率为
$$P\{(X,Y)\in G\}=\iint_G f(x,y)dxdy$$
4. 若 $f(x,y)$ 在点 $(x,y)$ 连续，则有
$$\frac{\partial^2F(x,y)}{\partial x\partial y}=f(x,y)$$

## 3.2 边缘分布

$F_X(x),F_Y(y)$ 称为二维随机变量 $(X,Y)$ 关于 $X,Y$ 的 **边缘分布函数**

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

记

$$p_{i\cdot}\sum_{j=1}^{\infty}p_{ij}=P\{X=x_i\}$$
$$p_{\cdot j}\sum_{i=1}^{\infty}p_{ij}=P\{X=y_j\}$$

分别称为 $p_{i\cdot}$ 和 $p_{\cdot j}$ 为 $(X,Y)$ 关于 $X,Y$ 的 **边缘分布律**

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

分别为 $X,Y$ 的 **边缘概率密度**

**Example 3.1.** 设二维随机变量 $(X,Y)$ 的概率密度

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

其中 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 都是常数，且 $\sigma_1,\sigma_2>0,-1<\rho<1$，我们称 $(X,Y)$ 服从参数为 $\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$ 的 **二维正态分布**，记为 $(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，试求二维正态分布随机变量的边缘概率密度

令 $t=\frac{1}{\sqrt{1-\rho^2}}(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1})$，则有

$$f_X(x)=\frac{1}{2\pi\sigma^1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{\infty}e^{-\frac{t^2}{2}}dt$$

即

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

## 3.3 条件分布

**Definition 3.2.** 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

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

为在 $Y=y_j$ 条件下随机变量 $X$ 的 **条件分布律**

**Definition 3.3.** 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 $F_{X|Y}(x|y)=\int_{-\infty}^{x}f_{X|Y}(x|y)dx=\int_{-\infty}^{x}\frac{f(x,y)}{f_Y(y)}dx$ 为在 $Y=y$ 下 $X$ 的条件分布函数

## 3.4 相互独立的随机变量

**Definition 3.4.** 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，若对于所有 $x,y$ 有

$$P\{X\le x,Y\le y\}=P\{X\le x\}P\{Y\le y\}$$
$$F(x,y)=F_X(x)F_Y(y)$$
$$f(x,y)=f_X(x)f_Y(y)$$

则称 $X,Y$ 是 **相互独立的**

下面考查二维正态随机变量 $(X,Y)$，它的概率密度为

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

如果 $\rho = 0$ 则对于所有 $x, y, f(x,y) = f_X(x) f_Y(y)$。如果 $X, Y$ 相互独立，令 $x = \mu_1, y = \mu_2$，则 $\rho = 0$

对于二维正态随机变量 $(X, Y)$，$X, Y$ 相互独立的充要条件是参数 $\rho = 0$

**Theorem 3.5.** 设 $(X_1, \ldots, X_m)$ 和 $(Y_1, \ldots, Y_n)$ 相互独立，则 $X_i$ 和 $Y_j$ 相互独立。又若 $h, g$ 是连续函数，则 $h(X_1, \ldots, X_m)$ 和 $g(Y_1, \ldots, Y_n)$ 相互独立

## 3.5 两个随机变量的函数的分布

### 3.5.1 $Z = X + Y$ 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

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

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

又设 $X, Y$ 相互独立，则

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

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

称为 $f_X, f_Y$ 的 **卷积公式**，记为 $f_X * f_Y$，即

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

证明.

$$F_Z(z) = P\{Z \leq z\} = \iint_{x+y \leq z} f(x, y) dx dy$$

则

$$F_z(z) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{z-y} f(x, y) dx \right] dy$$

$$F_z(z) = \int_{-\infty}^{\infty} \left[ \int_{-\infty}^{z} f(u-y, y) du \right] dy$$

$$F_z(z) = \int_{-\infty}^{z} \left[ \int_{-\infty}^{\infty} f(u-y, y) dy \right] du$$

□

有限个相互独立的正态随机变量的线性组合仍服从正态分布

**Example 3.2.** 设随机变量 $X, Y$ 相互独立，且分别服从参数为 $\alpha, \theta; \beta, \theta$ 的 $\Gamma$ 分布（分别记成 $X \sim \Gamma(a, \theta), Y \sim \Gamma(\beta, \theta)$），$X, Y$ 的概率密度分别为

$$f_X(x) = \begin{cases} \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta} & x > 0 \\ 0 \end{cases}, \alpha > 0, \theta > 0$$

$$f_Y(y) = \begin{cases} \frac{1}{\theta^\beta \Gamma(\beta)} y^{\beta-1} e^{-y/\theta} & y > 0 \\ 0 \end{cases}, \beta > 0, \theta > 0$$

试证明 $X + Y \sim \Gamma(\alpha + \beta, \theta)$

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

仅当 $0 < x < z$ 时被积函数不等于零，当 $z < 0$ 时 $f_Z(z) = 0$，当 $z > 0$ 有

$$\begin{aligned} f_Z(z) &= \int_0^x \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\theta} \frac{1}{\theta^\beta \Gamma(\beta)} (z-x)^{\beta-1} e^{-(z-x)/\theta} ds \\ &= \frac{e^{-z/\theta}}{\theta^{\alpha+\beta} \Gamma(\alpha)\Gamma(\beta)} \int_0^z x^{\alpha-1} (z-x)^{\beta-1} dx (\text{let } x = zt) \\ &= \frac{z^{\alpha+\beta-1} e^{-z/\theta}}{\theta^{\alpha+\beta} \Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1} (1-t)^{\beta-1} dt \\ &= A z^{\alpha+\beta-1} e^{-z/\theta} \end{aligned}$$

其中 $A = \frac{1}{\theta^{\alpha+\beta}\Gamma(\alpha)\Gamma(\beta)} \int_0^1 t^{\alpha-1}(1-t)^{\beta-1} dt$ 由概率密度的性质得

$$\begin{aligned} 1 = \int_{-\infty}^{\infty} f_Z(z) dz &= \int_0^{\infty} A z^{\alpha+\beta-1} e^{-z/\theta} dz \\ &= A\theta^{\alpha+\beta} \int_0^{\infty} (z/\theta)^{\alpha+\beta-1} e^{-z/\theta} d(z/\theta) \\ &= A\theta^{\alpha+\beta} \Gamma(\alpha + \beta) \end{aligned}$$

### 3.5.2 $Z = \frac{Y}{X}, Z = XY$ 的分布

$$f_{Y/X}(z) = \int_{-\infty}^{\infty} |x| f(x, xz) dx$$
$$f_{XY}(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f(x, \frac{z}{x}) dx$$

### 3.5.3 $M = \mathbf{max}\{X, Y\}, N = \mathbf{min}\{X, Y\}$ 的分布

对于 $n$ 个相互独立的随机变量

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

# 4 随机变量的数字特征

## 4.1 数学期望

**Definition 4.1.** 设离散型随机变量 $X$ 的分布律

$$P\{X = x_k\} = p_k, k = 1, 2, ...$$

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 $E(X)$

若连续型随机变量 $X$ 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 $X$ 设 **数学期望**，记为 $E(X)$

**Example 4.1.** 设 $X \sim \pi(\lambda)$，求 $E(X)$

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

**Theorem 4.2.** 设 $Y$ 是随机变量 $X$ 的函数：$Y = g(X)$（$g$ 是连续函数）

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

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

2. 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

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

数学期望的几个重要性质

1. 设 $C$ 是常数，则 $E(C) = C$
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

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

4. 设 $X, Y$ 是相互独立的随机变量，则有

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

## 4.2　方差

**Definition 4.3.** 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 $Var(X)$

$\sqrt{D(X)}$，记为 $\sigma(X)$，称为 **标准差**或 **均方差**

**Theorem 4.4.** $D(X) = E(X^2) - [E(X)]^2$

**Example 4.2.** 设随机变量 $X$ 具有数学期望 $E(X) = \mu$, 方差 $D(X) = \sigma^2 \neq 0$，记

$$X^* = \frac{X - \mu}{\sigma}$$

则

$$E(E^*) = 0$$
$$D(X^*) = 1$$

即 $X^* = \frac{X-\mu}{\sigma}$ 的数学期望为 $0$，方差为 $1$，$X^*$ 称为 $X$ 的 **标准化变量**

**Example 4.3.** 设随机变量 $X \sim \pi(\lambda)$，求 $D(X)$

随机变量 $X$ 的分布律为

$$P\{X = k\} = \frac{\lambda^k e^{-\lambda}}{k!}$$

因为 $E(X) = \lambda$，而

$$\begin{aligned} E(X^2) &= E[X(X-1) + X] = E[X(X-1)] + E(X) = \\ &= \sum_{k=0}^{\infty} k(k-1)\frac{\lambda^k e^{-\lambda}}{k!} + \lambda = \lambda^2 e^{-\lambda} \sum_{k=2}^{\infty} \frac{\lambda^{k-2}}{(k-2)!} + \lambda \\ &= \lambda^2 + \lambda \end{aligned}$$

所以方差

$$D(X) = \lambda$$

**Example 4.4.** 设随机变量 $X \sim U(a, b)$，求 $D(X)$

$E(X) = \frac{a+b}{2}$，方差为

$$\begin{aligned} D(X) &= E(X^2) - [E(X)]^2 \\ &= \int_a^b x^2 \frac{1}{b-a} dx - (\frac{a+b}{2})^2 = \frac{(b-a)^2}{12} \end{aligned}$$

**Example 4.5.** 设随机变量 $X$ 服从指数分布，其概率密度为

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

其中 $\theta > 0$，求 $E(X), D(X)$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} xf(x) = \theta \\ E(X^2) &= 2\theta^2 \\ D(X) &= \theta^2 \end{aligned}$$

方差的几个性质

1. 设 $C$ 是常数，则 $D(C) = 0$
2. 设 $X$ 是随机变量，$C$ 是常数，则有

$$D(CX) = C^2D(X), D(X+C) = D(X)$$

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

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

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

**Example 4.6.** 设随机变量 $X \sim b(n,p)$，求 $E(X), D(X)$

引入随机变量

$$X_k = \begin{cases} 1 & A \text{ happens at k}th \\ 0 & \end{cases}$$

易知 $X = X_1 + \cdots + X_n$。因为 $E(X_k) = p, D(X_k) = p(1-p)$，故

$$E(X) = E(\sum_{k=1}^{n} X_k) = np$$

$$D(X) = D(\sum_{k=1}^{n} X_k) = np(1-p)$$

**Example 4.7.** 设随机变量 $X \sim N(\mu, \sigma^2)$，求 $E(X), D(X)$

先求标准正态变量

$$Z = \frac{X - \mu}{\sigma}$$

的数学期望和方差，$Z$ 的概率密度为

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

于是

$$E(Z) = 0$$
$$D(Z) = E(Z^2) = 1$$

因 $X = \mu + \sigma Z$，即得

$$E(X) = E(\mu + \sigma Z) = \mu$$
$$D(X) = \sigma^2$$

推广得

$$C_1 X_1 + \cdots + C_n X_n \sim N(\sum_{i=1}^{n} C_i \mu_i, \sum_{i=1}^{n} C_i^2 \sigma_i^2)$$

**Definition 4.5** (切比雪夫不等式)**.** 设随机变量 $X$ 具有数学期望 $E(X) = \mu$，方差 $D(X) = \sigma^2$, 则对于任意正数 $\epsilon$ ，不等式

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

## 4.3 协方差及相关系数

**Definition 4.6.** $E\{[E - E(X)][Y - E(Y)]\}$ 称为随机变量 $X$ 与 $Y$ 的 **协方差**，记为 $Cov(X, Y)$，

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

称为随机变量 $X$ 与 $Y$ 的 **相关系数**

有

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

协方差有以下性质

1. $Cov(aX,bY)=abCov(X,Y)$
2. $Cov(X_1+X_2,Y)=Cov(X_1,Y)+Cov(X_2,Y)$

考虑以 $X$ 的线性函数 $a+bX$ 来近似表示 $Y$，我们以均方误差

$$\begin{aligned} e &= E[(Y-(a+bX))^2] \\ &= E(Y^2)+b^2E(X^2)+a^2-2bE(XY)+2abE(X)-2aE(Y) \end{aligned}$$

来衡量以 $a+bX$ 近似表达 $Y$ 的好坏程度。将 $e$ 分别关于 $a,b$ 求偏到，得

$$\begin{cases} \frac{\partial e}{\partial a}=2a+2bE(X)-2E(Y)=0 \\ \frac{\partial e}{\partial b}=2bE(X^2)-2E(XY)+2aE(X)=0 \end{cases}$$

解得

$$b_0=\frac{Cov(X,Y)}{D(X)}$$
$$a_0=E(Y)-b_0E(X)$$

代入得

$$\min_{a,b} E\{[Y-(a+bX)]^2\}=(1-\rho_{xy}^2)D(Y)$$

**Theorem 4.7.** *1.* $|\rho_{xy}|\le 1$

2. $|\rho_{xy}|=1$ 的充要条件是存在 $a,b$ 使得

$$P\{Y=a+bX\}=1$$

当 $\rho_{xy}=0$ 时，称 $X$ 和 $Y$ 不相关

**Example 4.8.** 设 $(X,Y)$ 服从二维正态分布，求 $X,Y$ 的相关系数

$\rho_{xy}=\rho$

### 4.4 矩、协方差矩阵

**Definition 4.8.** 设 $X,Y$ 是随机变量，若

$$E(X^k), k=1,2,...$$

存在，称它为 $X$ 的 $k$ **阶原点矩**，简称 $k$ **阶矩**

若

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

存在，称它为 $X$ 的 $k$ **阶中心矩**

若

$$E(X^kY^l), k,l=1,2,...$$

存在，称它为 $X,Y$ 的 $k+l$ **阶混合矩**

若

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

存在，称它为 $X,Y$ 的 $k+l$ **阶混合中心矩**

## 5 大数定律及中心极限定理

### 5.1 大数定律

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

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

对于独立同分布且具有均值 $\mu$ 的随机变量 $X_1,...,X_n$，当 $n$ 很大时它们的算术平均 $\frac{1}{n}\sum_{k=1}^n X_k$ 很可能接近 $\mu$

设 $Y_1,...,Y_n,...$ 是一个随机变量序列，$a$ 是一个常数，若对于任意正数 $\epsilon$ ，有

$$\lim_{n\to\infty} P\{|Y_n-a|<\epsilon\}=1$$

则称序列 $Y_1, \ldots, Y_n, \ldots$ **依概率收敛于** $a$，记为

$$Y_n \xrightarrow{P} a$$

设 $X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，又设函数 $g(x, y)$ 在点 $(a, b)$ 连续，则

$$g(X_n, Y_n) \xrightarrow{P} g(a, b)$$

这样上述定理可叙述为

**Theorem 5.2** (弱大数定理（辛钦大数定理）). 设随机变量 $X_1, \ldots, X_n, \ldots$ 相互独立，服从同一分布且具有数学期望 $E(X_k) = \mu$，则序列 $\overline{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ 一概率收敛于 $\mu$ ，即 $\overline{X} \xrightarrow{P} \mu$

**Theorem 5.3** (伯努利大数定理). 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 $\epsilon > 0$，有

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

或

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

## 5.2 中心极限定理

**Theorem 5.4** (独立同分布的中心极限定理). 设随机变量 $X_1, \ldots, X_n, \ldots$ 相互独立，服从同一分布，且具有数学期望和方差：$E(X_k) = \mu, D(X_k) = \sigma^2 > 0$，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

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

因此当 $n$ 充分大时，有

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

或

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

**Theorem 5.5** (李雅普诺夫定理)**.** 设随机变量 $X_1, \dots, X_n, \dots$ 相互独立，它们具有数学期望和方差

$$E(X_k) = \mu_k, D(X_k) = \sigma_k^2 > 0$$

记

$$B_n^2 = \sum_{k=1}^{n} \sigma_k^2$$

若存在正数 $\delta$ 使得当 $n \to \infty$ 时

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

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 $F_n(x)$ 对于任意 $x$ 满足

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

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，近似服从正态分布 $N(0,1)$

**Theorem 5.6** (棣莫弗-拉普拉斯定理)**.** 设随机变量 $\eta_n$ 服从参数为 $n, p$ 的二项分布，则对于任意 $x$ 有

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

# 6 样本及抽样分布

## 6.1 随机样本

我们将试验的全部可能的观察值称为 **总体**, 每一个可能观察值称为 **个体**, 总体中所包含的个体的个数称为总体的 **容量**, 容量为有限的称为 **有限总体**, 容量为无限的称为 **无限总体**

**Definition 6.1.** 设 $X$ 是具有分布函数 $F$ 的随机变量, 若 $X_1, \dots, X_n$ 是具有同一分布函数 $F$ 的、相互独立的随机变量, 则称 $X_1, \dots, X_n$ 为从分布函数 $F$ 得到的 **容 量为** $n$ **的简单随机样本**, 简称为 **样本**, 它们的观察值 $x_1, \dots, x_n$ 称为 **样本值**, 又称为 $X$ 的 $n$ 个 **独立的观察值**

## 6.2 抽样分布

**Definition 6.2.** 设 $X_1, \dots, X_n$ 是来自总体 $X$ 的一个样本 $g(X_1, \dots, X_n)$ 是 $X_1, \dots, X_n$ 的函数, 若 $g$ 中不含未知参数, 则称 $g(X_1, \dots, X_n)$ 是一 **统计量**

$g(x_1, \dots, x_n)$ 是 $g(X_1, \dots, X_n)$ 的观察值
定义
**样本平均值**

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

**样本方差**

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

**样本标准差**

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

**样本 $k$ 阶（原点）矩**

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

**样本 $k$ 阶中心矩**

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

若总体 $X$ 的 $k$ 阶矩 $E(X^k) = \mu_k$ 存在，则当 $n \to \infty$ 时，$A_k \xrightarrow{P} \mu_k$，这时因为 $X_1, ..., X_n$ 独立且与 $X$ 同分布，所以 $X_1^k, ..., X_n^k$ 独立且与 $X_k$ 同分布，故有

$$E(X_1^k) = \cdots = E(X_n^k) = \mu_k$$

从而由辛钦大数定理

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k$$

进而由依概率收敛的序列的性质知道

$$g(A_1, ..., A_k) \xrightarrow{P} g(\mu_1, ..., \mu_k)$$

### 6.2.1 $\chi^2$ 分布

设 $X_1, ..., X_n$ 是来自总体 $N(0,1)$ 的样本，则称统计量

$$\chi^2 = X_1^2 + \cdots + X_n^2$$

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

$\chi^2(n)$ 分布的概率密度为

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

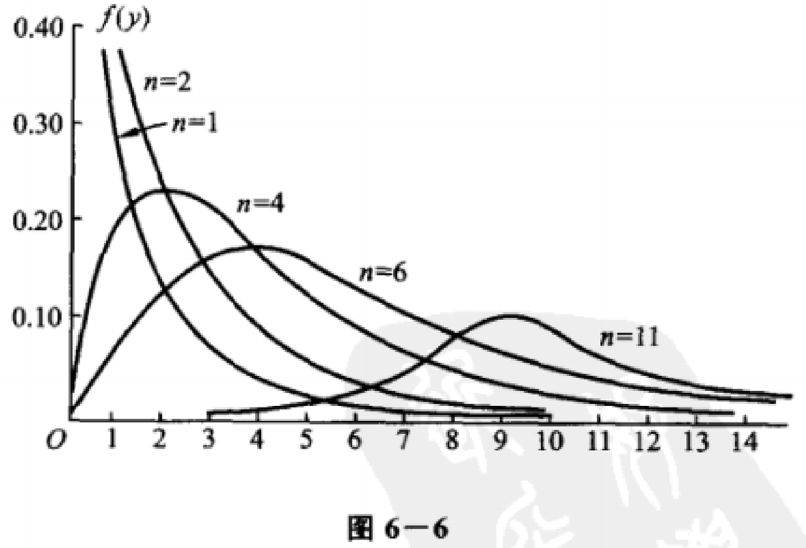


图 6—6

由 2.1 及 3.2 知 $\chi^2(1)$ 服从 $\Gamma(0.5,2)$ 分布。现 $X_i \sim N(0,1)$ 由定义 $X_i^2 \sim \chi^2(1)$，即 $X_i^2 \sim \Gamma(0.5,2)$，因此

$$\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \Gamma(\frac{n}{2},2)$$

$\chi^2$ **分布的可加性**设 $\chi_1^2 \sim \chi^2(n_1), \chi_2^2 \sim \chi^2(n_2)$ 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

$$E(\chi^2) = n, D(\chi_2) = 2n$$

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

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

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

### 6.2.2 $t$ 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 **学生式（Student）分布**，$t(n)$ 分布的概率密度函数为

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

$$\lim_{n\to\infty} h(t) = \frac{1}{\sqrt{2\pi}}e^{-t^2/2}$$

### 6.2.3 $F$ 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 $(n_1, n_2)$ 的 $F$ **分布**，记为 $F \sim F(n_1, n_2)$

由定义得

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

$F$ 分布的上 $\alpha$ 分位点有如下性质

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

## 6.3 正态总体的样本均值与样本方差的分布

设总体 $X$ 的均值为 $\mu$，方差为 $\sigma^2$，$X_1, \dots, X_n$ 是来自 $X$ 的一个样本，$\overline{X}, S^2$ 分别是样本均值和样本方差，则

$$E(\overline{X}) = \mu, \quad D(\overline{X}) = \sigma^2/n$$

而

$$E(S^2) = E\left[\frac{1}{n-1}(\sum_{i=1}^{n} X_i^2 - n\overline{X}^2)\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} E(X_i^2) - nE(\overline{X}^2)\right]$$
$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(\sigma^2+\mu^2) - n(\sigma^2/n+\mu^2)\right] = \sigma^2$$

即

$$E(S^2) = \sigma^2$$

进而，设 $X \sim N(\mu, \sigma^2)$，则 $\overline{X}$ 也服从正态分布，则

**Theorem 6.3.** 设 $X_1, \dots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的样本，$\overline{X}$ 是样本均值，则有

1. $\overline{X} \sim N(\mu, \sigma^2/n)$
2. $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$
3. $\overline{X}$ 与 $S^2$ 相互独立
4. $\frac{\overline{X}}{S/\sqrt{n}} \sim t(n-1)$

证明. 2. 设 $Z_i = \frac{X_i-\mu}{\sigma}$，则 $Z_1, \dots, Z_n$ 相互独立，而

$$\overline{Z} = \frac{1}{n}\sum_{i=1}^{n} Z_i = \frac{\overline{X}-\mu}{\sigma}$$
$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2} = \sum_{i=1}^{n}\left[\frac{(X_i-\mu)-(\overline{X}-\mu)}{\sigma}\right]^2$$
$$= \sum_{i=1}^{n}(Z_i-\overline{Z})^2 = \sum_{i=1}^{n} Z_i^2 - n\overline{Z}^2$$

取一 $n$ 阶正交矩阵 $\mathbf{A} = (a_{ij})$，其中第一行的元素均为 $1/\sqrt{n}$，作正交变换

$$\mathbf{Y} = \mathbf{A}\,\mathbf{Z}$$

其中

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{Z} = \begin{bmatrix} Z_1 \\ \vdots \\ Z_n \end{bmatrix}$$

由于 $Y_i=\sum_{j=1}^{n}a_{ij}Z_j$，故 $Y_1,\dots,Y_n$ 仍为正交变量，由 $Z_i\sim N(0,1)$ 知

$$E(Y_i)=E(\sum_{j=1}^{n}a_{ij}Z_j)=0$$

又由 $Cov(Z_i,Z_j)=\delta_{ij}$，知

$$\begin{aligned}Cov(Y_i,Y_k)&=Cov(\sum_{j=1}^{n}a_{ij}Z_j,\sum_{l=1}^{1}a_{kl}Z_l)\\&=\sum_{j=1}^{n}\sum_{l=1}^{n}a_{ija_{kl}}Cov(Z_j,Z_l)=\sum_{j=1}^{n}a_{ij}a_{kj}=\delta_{jk}\end{aligned}$$

故 $Y_1,dots,Y_n$ 两两不相关，又由 $n$ 维随机变量 $(Y_1,\dots,Y_n)$ 是由 $n$ 维正态随机变量 $(X_1,\dots,X_n)$ 经由线性变换得到的，因此 $(Y_1,\dots,Y_n)$ 也是 $n$ 维正态随机变量，因此 $Y_1,\dots,Y_n$ 相互独立，且有 $Y_i\sim N(0,1)$，而

$$Y_1=\sum_{j=1}^{n}a_{1j}Z_j=\sum_{j=1}^{n}\frac{1}{\sqrt{n}}Z_j=\sqrt{n}\overline{Z}$$

$$\sum_{i=1}^{n}Y_i^2=\mathbf{Y}^T\mathbf{Y}=(\mathbf{A}\,\mathbf{Z})^T(\mathbf{A}\,\mathbf{Z})=\mathbf{Z}(\mathbf{A}^T\,\mathbf{A})\,\mathbf{Z}=\mathbf{Z}^T\,\mathbf{E}\,\mathbf{Z}=\mathbf{Z}^T\,\mathbf{Z}=\sum_{i=1}^{n}Z_i^2$$

于是

$$\frac{(n-1)S^2}{\sigma^2}=\sum_{i=1}^{n}Z_i^2-n\overline{Z}^2=\sum_{i=1}^{n}Y_i^2-y_1^2=\sum_{i=2}^{n}Y_i^2$$

3. $\overline{X}=\sigma\overline{Z}+\mu=\frac{\sigma Y_1}{\sqrt{n}}+\mu$ 仅依赖于 $Y_1$
4. 由 1，2，3

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1),\quad \frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$$

且两者独立，由 $t$ 分布的定义知

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\Bigg/\sqrt{\frac{(n-1)S^2}{\sigma^2(n-1)}}\sim t(n-1)$$

□

**Theorem 6.4.** 设 $X_1,\dots,X_{n_1},Y_1,\dots,Y_{n_2}$ 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 $\overline{X}=\frac{1}{n_1}\sum_{i=1}^{n_1}X_i,\overline{Y}=\frac{1}{n_2}\sum_{i=1}^{n_2}Y_i$ 为别是两样本的均值,$S_1^2,S_2^2$ 分别是两样本的样本方差,则有

1. $\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2}\sim F(n_1-1,n_2-1)$
2. 当 $\sigma_1^2=\sigma_2^2=\sigma^2$ 时

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

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

证明. 1.

$$\frac{(n_1-1)S_1^2}{\sigma_1^2}\sim\chi^2(n_1-1),\frac{(n_2-1)S_2^2}{\sigma_2^2}\sim\chi^2(n_2-1)$$

由假设 $S_1,S_2$ 相互独立,因此

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

即

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

2. 易知 $\overline{X}-\overline{Y}\sim N\left(\mu_1-\mu_2,\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}\right)$,即有

$$U=\frac{(\overline{X}-\overline{Y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}\sim N(0,1)$$

又因为

$$\frac{(n_1-1)S_1^2}{\sigma^2}\sim\chi^2(n_1-1),\frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_2-1)$$

且它们相互独立,故由 $\chi^2$ 分布的可加性知

$$V=\frac{(n_1-1)S_1^2}{\sigma^2}+\frac{(n_2-1)S_2^2}{\sigma^2}\sim\chi^2(n_1+n_2-2)$$

$U,V$ 独立

□

# 7 参数估计

## 7.1 点估计

### 7.1.1 矩估计法

设 $X$ 为连续型随机变量，其概率密度函数为 $f(x;\theta_1,\dots,\theta_k)$，或 $X$ 为离散型随机变量，其分布律为 $P\{X=x\}=p(x;\theta_1,\dots,\theta_k)$，其中 $\theta_1,\dots,\theta_k$ 为待估计参数，$X_1,\dots,X_n$ 是来自 $X$ 的样本，假设总体 $X$ 的前 $k$ 阶矩

$$\mu_l = E(X^l) = \int_{-\infty}^{\infty} x^l f(x;\theta_1,\dots,\theta_k)dx \quad \text{or}$$
$$\mu_l = E(X^l) = \sum_{x\in R_X} x^l p(x;\theta_1,\dots,\theta_k)$$

其中 $R_X$ 是 $X$ 的取值范围，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 $\mu_l$, 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**。具体做法如下，设

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

这是一个包含 $k$ 个未知参数 $\theta_1,\dots,\theta_k$ 的联立方程组，可解出

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

以 $A_i$ 分别代替上式中的 $\mu_i$，就以

$$\hat{\theta}_i = \theta_i(A_1, \dots, A_k)$$

分别作为 $\theta_i$ 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**

**Example 7.1.** 设总体 $X$ 在 $[a, b]$ 上服从均匀分布，$a, b$ 未知，$X_1, \dots, X_n$ 是来自 $X$ 的样本，试求 $a, b$ 的矩估计量

$$\begin{aligned}
&\mu_1 = E(X) = (a+b)/2 \\
&\mu_2 = E(X^2) = D(X) + [E(X)]^2 = (b-a)^2/12 + (a+b)^2/4
\end{aligned}$$

解得

$$a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}, b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$$

e 分别以 $A_1, A_2$ 代替 $\mu_1, \mu_2$

**Example 7.2.** 设总体 $X$ 的均值 $\mu$ 及方差 $\sigma^2$ 都存在，且有 $\sigma^2 > 0$，但 $\mu, \sigma^2$ 均未知，又设 $X_1, \dots, X_n$ 是来自 $X$ 的样本，试求 $\mu, \sigma^2$ 的矩估计量

$$\begin{cases}
\mu_1 = E(X) = \mu \\
\mu_2 = E(X^2) = \sigma^2 + \mu^2
\end{cases}$$

解得

$$\begin{cases}
\mu = \mu_1 \\
\sigma^2 = \mu_2 - \mu_1^2
\end{cases}$$

因此

$$\begin{aligned}
&\hat{\mu} = A_1 = \overline{X} \\
&\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2
\end{aligned}$$

### 7.1.2 最大似然估计法

若总体 $X$ 属于离散型，其分布律 $P\{X=x\}=p(x;\theta),\theta\in\Theta$ 的形式已知，$\theta$ 为待沽参数，$\Theta$ 是 $\theta$ 可能取值的范围，设 $X_1,\dots,X_n$ 是来自 $X$ 的样本，则 $X_1,\dots,X_n$ 的联合分布

$$\prod_{i=1}^{n} p(x_1;\theta)$$

又设 $x_1,\dots,x_n$ 是相应于 $X_1,\dots,X_n$ 的样本值，则事件 $\{X_1=x_1,\dots,X_n=x_n\}$ 发生的概率为

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

$L(\theta)$ 称为样本的 **似然函数**。取 $\hat{\theta}$ 使

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

这样得到的 $\hat{\theta}$ 与样本值 $x_1,\dots,x_n$ 有关，常记为 $\hat{\theta}(x_1,\dots,x_n)$, 称为参数 $\theta$ 的 **最大似然估计值**，而相应的统计量 $\hat{\theta}(X_1,\dots,X_n)$ 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,设 $X_1,\dots,X_n$ 是来自 $X$ 的样本，它们的联合密度为

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

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

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

因因子 $\prod_{i=1}^{n} dx_i$ 不随 $\theta$ 改变，故只需考虑函数

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

的最大值，这里 $L(\theta)$ 称为样本的 **似然函数**，若

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

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

很多时候 $p(x;\theta), f(x;\theta)$ 关于 $\theta$ 可微，因此 $\theta$ 的最大似然估计可以从方程

$$\frac{d}{d\theta}\ln L(\theta) = 0$$

求得，这个方程称为 **对数似然方程**

**Example 7.3.** 设 $X \sim (1,p)$，$X_1,\dots,X_n$ 是来自 $X$ 的一个样本，求参数 $p$ 的最大 r 似然估计

设 $x_1,\dots,x_n$ 是相应于样本 $X_1,\dots,X_n$ 的一个样本值，$X$ 的分布律为

$$P\{X=x\} = p^2(1-p)^{1-x}, x=0,1$$

故

$$\begin{aligned}
L(p) &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^n x_i}(1-p)^{n-\sum_{i=1}^n x_i} \\
\ln L(p) &= (\sum_{i=1}^{n} x_i)\ln p + (n-\sum_{i=1}^{n} x_i)\ln(1-p) \\
\frac{d}{dp}\ln L(p) &= \frac{\sum_{i=1}^n x_i}{p} - \frac{n-\sum_{i=1}^n x_i}{1-p} = 0 \\
\hat{p} &= \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}
\end{aligned}$$

**Example 7.4.** 设 $X \sim N(\mu,\sigma^2)$，$\mu$,$\sigma$ 未知，$x_1,\dots,x_n$ 是来自 $X$ 的一个样本值，求 $\mu,\sigma^2$ 的最大似然估计函数

$X$ 的概率密度函数

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}(x-\mu^2)\right]$$

似然函数为

$$\begin{aligned} L(\mu, \sigma^2) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x-\mu^2)\right] \\ &= (2\pi)^{-n/2}(\sigma^2)^{-n/2} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right] \\ \ln L &= -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2 \end{aligned}$$

令

$$\begin{cases} \frac{\partial}{\partial\mu}\ln L = \frac{1}{\sigma^2}(\sum_{i=1}^{n} x_i - n\mu) = 0 \\ \frac{\partial}{\partial\sigma^2}\ln L = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases}$$

解得

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

**Example 7.5.** 设总体 $X$ 在 $[a,b]$ 上服从均匀分布，$a,b$ 未知，$x_1,\dots,x_n$ 是一个样本值，求 $a,b$ 的最大似然估计函数记 $x_{(1)} = \min\{x_1,\dots,x_n\}, x_{(n)} = \max\{x_1,\dots,x_n\}$，$X$ 的概率密度是

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

似然函数为

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n} & a \le x_1,\dots,x_n \le b \\ 0 \end{cases}$$

由于 $a \le x_1,\dots,x_n \le b$ 等价于 $a \le x_{(1)}, x_{(n)} \le b$ 似然函数可写成

$$L(a,b) = \begin{cases} \frac{1}{(b-a)^n} & a \le x_{(1)}, b \ge x_{(n)} \\ 0 \end{cases}$$

于是对于满足条件 $a \le x_{(1)}, b \ge x_{(n)}$ 的任意 $a,b$ 有

$$L(a,b) = \frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)} - x_{(1)})^n}$$

即 $L(a,b)$ 在 $a=x_{(1)}, b=x_{(n)}$ 时取得最大值，故 $a,b$ 的最大似然估计值为

$$\hat{a}=x_{(1)}, \hat{b}=x_{(n)}$$

此外，最大似然估计具有下述性质：设 $\theta$ 的函数 $u=u(\theta), \theta\in\Theta$ 具有单值反函数 $\theta=\theta(u), u\in\mathcal{U}$，又假设 $\hat{\theta}$ 是 $X$ 的概率分布中参数 $\theta$ 的最大似然估计，则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计，这一性质称为最大似然估计的**不变性**

## 7.2 基于截尾样本的最大似然估计

将随机抽取的 $n$ 个产品在 $t=0$ 时投入试验直到每个产品都失效，记录每一个产品的失效时间，这样的道德样本叫完全样本。截尾寿命试验有两种：定时截尾寿命试验，假设将随机抽取的 $n$ 个产品在 $t=0$ 投入试验，试验进行到事先规定的截尾时间 $t_0$ 停止，如试验截止时共有 $m$ 个产品失效，它们的失效时间分别为

$$0\le t_1\le ...\le t_m\le t_0$$

此时 $m$ 是一个随机变量，所得的样本 $t_1,...,t_m$ 称为 **定时截尾样本**。另一种是定数截尾样本，$t_m$ 是第 $m$ 个产品的失效时间，所得到的样本 $t_1 ... t_m$ 称为 **定数截尾样本**

设产品的寿命分布为指数分布

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

设有 $n$ 个产品投入定数截尾试验。一个产品在 $(t_i,t_i+dt_i]$ 失效的概率近似地为 $f(t_i)dt_i=\frac{1}{\theta}e^{-t_i/\theta}dt_i$，其余 $n-m$ 个产品寿命超过 $t_m$ 的概率为 $(\int_{t_m}^{\infty}\frac{1}{\theta}e^{-t/\theta}dt)^{n-m}=(e^{-t_m/\theta})^{n-m}$，故上述观察结果出现的概率近似地为

$$\begin{aligned}&\binom{n}{m}(\frac{1}{\theta}e^{-t_1/\theta}t_1)...(\frac{1}{\theta}e^{-t_m/\theta}t_m)(e^{-t_m/\theta})^{n-m}\\&=\binom{n}{m}\frac{1}{\theta^m}e^{-\frac{1}{\theta}[t_1+\cdots+t_m+(n-m)t_m]}dt_1 ... dt_m\end{aligned}$$

其中 $dt_1, \dots, dt_m$ 为常数，因忽略常数不影响 $\theta$ 的最大似然估计，故可取似然函数为

$$L(\theta) = \frac{1}{\theta^m} e^{-\frac{1}{\theta}[t_1+\cdots+t_m+(n-m)t_m]}$$

则

$$\frac{d}{d\theta} \ln L(\theta) = -\frac{m}{\theta} + \frac{1}{\theta^2}[t_1 + \cdots + t_m + (n-m)t_m] = 0$$

于是

$$\hat{\theta} = \frac{s(t_m)}{m}$$

其中 $s(t_m) = t_1 + \cdots + t_m + (n-m)t_m$ 称为总试验时间

对于截尾样本

$$0 \leq t_1 \leq \dots \leq t_m \leq t_0$$

可得

$$L(\theta) = \frac{1}{\theta^m} e^{-\frac{1}{\theta}[t_1+\cdots+t_m+(n-m)t_0]}$$

$$\hat{\theta} = \frac{s(t_0)}{m}$$

其中 $s(t_0) = t_1 + \cdots + t_m + (n-m)t_0$

## 7.3 评估量的评选标准

### 7.3.1 无偏性

设 $X_1, \dots, X_n$ 是总体 $X$ 的一个样本，$\theta \in \Theta$ 是包含在总体 $X$ 的分布中的待沽参数，这里 $\Theta$ 是 $\theta$ 的取值范围

**Definition 7.1.** 若估计量 $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对于任意 $\theta \in \Theta$ 有

$$E(\hat{\theta}) = \theta$$

则称 $\hat{\theta}$ 是 $\theta$ 的 **无偏估计量**

**Example 7.6.** 设总体 $X$ 的 $k$ 阶矩 $\mu_k = E(X^k)$ 存在，又设 $X_1, \dots, X_n$ 是 $X$ 的一个样本，试证明不论总体服从什么分布，$k$ 阶样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 $k$ 阶总体矩 $\mu_k$ 的无偏估计

$X_1, \ldots, X_n$ 与 $X$ 同分布，故有

$$E(X_i^k) = E(X^k) = \mu_k$$

**Example 7.7.** 设总体 $X$ 服从指数分布，其概率密度为

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

其中 $\theta > 0$ 未知，又设 $X_1, \ldots, X_n$ 是来自 $X$ 的样本，证明 $nZ = n(\min\{X_1, \ldots, X_n\})$ 是 $\theta$ 的无偏估计

$Z = \min\{X_1, \ldots, X_n\}$ 具有概率密度

$$f_{\min}(x;\theta) = \begin{cases} \frac{n}{\theta} e^{-nx/\theta} & x > 0 \\ 0 \end{cases}$$

故

$$E(Z) = \frac{\theta}{n}$$

### 7.3.2 有效性

**Definition 7.2.** 设 $\hat{\theta}_1 = \hat{\theta}_1(X_1, \ldots, X_n), \hat{\theta}_2 = \hat{\theta}_2(X_1, \ldots, X_n)$ 都是 $\theta$ 的无偏估计量，若对于任意 $\theta \in \Theta$

$$D(\hat{\theta}_1) \leq D(\hat{\theta}_2)$$

且至少对于某一个 $\theta \in \Theta$ 上式不等号成立，则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ **有效**

### 7.3.3 相合性

**Definition 7.3.** 设 $\hat{\theta}(X_1, \ldots, X_n)$ 为参数 $\theta$ 的估计量，若对于任意 $\theta \in \Theta$，当 $n \to \infty$ 时 $\hat{\theta}(X_1, \ldots, X_n)$ 依概率收敛于 $\theta$，则称 $\hat{\theta}$ 为 $\theta$ 的 **相合估计量**

## 7.4 区间估计

**Definition 7.4.** 设总体 $X$ 的分布函数 $F(x;\theta)$ 含有一个未知数 $\theta \in \Theta$，对于给定值 $\alpha(0<\alpha<1)$ 若由来自 $X$ 的样本 $X_1,\dots,X_n$ 确定的两个统计量 $\underline{\theta}<\bar{\theta}$ 对于任意 $\theta\in\Theta$ 满足

$$P\{\underline{\theta}(X_1,\dots,X_n)<\theta<\bar{\theta}(X_1,\dots,X_n)\}\geq 1-\alpha$$

则称随机区间 $(\underline{\theta},\bar{\theta})$ 是 $\theta$ 的置信水平为 $1-\alpha$ 的 **置信区间**，$\underline{\theta},\bar{\theta}$ 分别是双侧置信区间的 **置信下限**和 **置信下限**，$1-\alpha$ 为 **置信水平**

**Example 7.8.** 设总体 $X\sim N(\mu,\sigma^2)$，$\sigma^2$ 已知，$\mu$ 未知，设 $X_1,\dots,X_n$ 是来自 $X$ 的样本，求 $\mu$ 的置信水平为 $1-\alpha$ 的置信区间

$\overline{X}$ 是 $\mu$ 的无偏估计，且有

$$\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$$

因此按正态分布

$$P\{\left|\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\right|<z_{\alpha/2}\}=1-\alpha$$

即

$$P\{\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2}<\mu<\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\}=1-\alpha$$

因此得到了 $\mu$ 的一个置信水平为 $1-\alpha$ 的置信区间

$$\left(\overline{X}-\frac{\sigma}{\sqrt{n}}z_{\alpha/2},\overline{X}+\frac{\sigma}{\sqrt{n}}z_{\alpha/2}\right)$$

寻求未知参数 $\theta$ 的置信区间的做法

1. 寻求一个样本 $X_1,\dots,X_n$ 和 $\theta$ 的函数 $W=W(X_1,\dots,X_n;\theta)$ 使得 $W$ 的分布不依赖于 $\theta$ 以及其他未知参数，称具有这种性质的函数 $W$ 为 **枢轴量**
2. 对于给定的置信水平 $1-\alpha$ 定出两个常数 $a,b$，使得

$$P\{a<W(X_1,\dots,X_n;\theta)<b\}=1-\alpha$$

## 7.5 正态总体均值与方差的区间估计

### 7.5.1 单个总体 $N(\mu, \sigma^2)$ 的情况

设已给定置信水平为 $1-\alpha$，并设 $X_1, \dots, X_n$ 为总体 $N(\mu, \sigma^2)$ 的样本，$\overline{X}, S^2$ 分别是样本均值和样本方差

1. 均值 $\mu$ 的置信区间

   (a) $\sigma^2$ 为已知，得到

   $$\left(\overline{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2}\right)$$

   (b) $\sigma^2$ 未知，考虑 $S^2$ 是 $\sigma^2$ 的无偏估计，由定理 6.3

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

   因此得 $\mu$ 的一个置信水平为 $1-\alpha$ 的置信区间

   $$\left(\overline{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1)\right)$$

2. 方差 $\sigma^2$ 的置信区间 $\mu$ 未知，$\sigma^2$ 的无偏估计为 $S^2$

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

   有

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

   因此得到方差 $\sigma^2$ 的一个置信水平为 $1-\alpha$ 的置信区间

   $$\left(\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)}, \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}\right)$$

### 7.5.2 两个总体 $N(\mu_1, \sigma_1^2), N(\mu_2, \sigma_2^2)$ 的情况

设给定置信水平 $1-\alpha$，并设 $X_1, \dots, X_{n_1}$ 是来自第一个总体的样本；$Y_1, \dots, Y_{n_2}$ 是来自第二个总体的样本，这两个样本相互独立，且设 $\overline{X}, \overline{Y}, S_1^2, S_2^2$ 分别是第一、二个总体的样本均值、方差

1. 两个总体均值差 $\mu_1 - \mu_2$ 的置信区间

(a) $\sigma_1^2, \sigma_2^2$ 均为已知，则

$$\overline{X} - \overline{Y} \sim N(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2})$$

因此得到 $\mu_1 - \mu_2$ 的一个置信水平为 $1 - \alpha$ 的置信区间

$$\left(\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$$

(b) $\sigma_1^2 = \sigma_2^2 = \sigma^2$，但 $\sigma^2$ 未知，此时由定理 6.4

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

可得 $\mu_1 - \mu_2$ 的一个置信水平为 $1 - \alpha$ 的置信区间为

$$\left(\overline{X} - \overline{Y} \pm t_{\alpha/2}(n_1 + n_2 - 2)S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\right)$$

此处

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

2. 两个总体方差比 $\sigma_1^2/\sigma_2^2$ 的置信区间 $\mu_1, \mu_2$ 均未知，由定理 6.4

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

有

$$P\left\{F_{1-\alpha/2}(n_1 - 1, n_2 - 1) < \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} < F_{\alpha/2}(n_1 - 1, n_2 - 1)\right\} = 1 - \alpha$$

因此得 $\sigma_1^2/\sigma_2^2$ 的一个置信水平为 $1 - \alpha$ 的置信区间

$$\left(\frac{S_1^2}{S_2^2}\frac{1}{F_{\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2}\frac{1}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)}\right)$$

## 7.6 $(0-1)$ 分布参数的区间估计

设有一容量 $n>50$ 的大样本，它来自 $(0-1)$ 分布的总体 $X$,$X$ 的分布律为

$$f(x;p)=p^x(1-p)^{1-x},x=0,1$$

其中 $p$ 为未知参数，现在求 $p$ 的置信水平为 $1-\alpha$ 的置信区间

已知 $(0-1)$ 分布的均值和方差分别为

$$\mu=p,\sigma^2=p(1-p)$$

设 $X_1,\dots,X_n$ 是一个样本，因样本容量 $n$ 较大，由中心极限定理，知

$$\frac{n\overline{X}-np}{\sqrt{np(1-p)}}\sim N(0,1)$$

于是有

$$P\left\{-z_{\alpha/2}<\frac{n\overline{X}-np}{\sqrt{np(1-p)}}<z_{\alpha/2}\right\}\approx 1-\alpha$$

而不等式

$$-z_{\alpha/2}<\frac{n\overline{X}-np}{\sqrt{np(1-p)}}<z_{\alpha/2}$$

等价于

$$(n+z_{\alpha/2}^2)p^2-(2n\overline{X}+z_{\alpha/2}^2)p+n\overline{X}^2<0$$

记

$$p_1=\frac{1}{2a}(-b-\sqrt{b^2-4ac})$$
$$p_2=\frac{1}{2a}(-b+\sqrt{b^2-4ac})$$

此处 $a=n+z_{\alpha/2}^2,b=-(2n\overline{X}+z_{\alpha/2}^2),c=n\overline{X}^2$, 于是得置信区间

$$(p_1,p_2)$$

### 7.7 单侧置信区间

对于给定值 $\alpha(0 < \alpha < 1)$ 若由样本 $X_1, \dots, X_n$ 确定的统计量 $\underline{\theta} = \underline{\theta}(X_1, \dots, X_n)$ 对于任意 $\theta \in \Theta$ 满足

$$P\{\theta > \underline{\theta}\} \geq 1 - \alpha$$

称随机区间 $(\underline{\theta}, \infty)$ 是 $\theta$ 的置信水平为 $1-\alpha$ 的 **单侧置信区间**，$\underline{\theta}$ 称为 $\theta$ 的置信为凭为 $1-\alpha$ 的单侧置信上限

若对于正态总体 $X$, 若均值 $\mu$ ，方差 $\sigma^2$ 均未知，设 $X_1, \dots, X_n$ 是一个样本，由

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

有

$$P\{\frac{\overline{X} - \mu}{S/\sqrt{n}} < t_\alpha(n-1)\} = 1 - \alpha$$

因此得到 $\mu$ 的一个置信水平为 $1-\alpha$ 的单侧置信区间

$$\left(\overline{X} - \frac{S}{\sqrt{n}} t_\alpha(n-1), \infty\right)$$

# 8 假设检验

### 8.1 假设检验

**Example 8.1.** 某车间用一台包装机包装葡萄糖，袋装糖的净重是一个随机变量，它服从正态分布，当机器正常时，其均值为 0.5 kg，标准差为 0.015kg，某日开工后为检验包装机是否正常，随机地抽取它所包装的糖 9 袋，称得净重为（kg）

0.497 0.506 0.518 0.524 0.498 0.511 0.520 0.515 0.512

问机器是否正常

以 $\mu$,$\sigma$ 分别表示这一天袋装糖的净重总体 $X$ 的均值和标准差。由于标准差比较稳定，设 $\sigma = 0.015$，于是 $X \sim N(\mu, 0.015^2)$，这里 $\mu$ 未知，问题

是根据样本值来判断 $\mu$ 是否等于 0.5。为此我们提出两个相互对立的假设

$$H_0 : \mu = \mu_0 = 0.5$$
$$H_1 : \mu \neq \mu_0$$

$\overline{X}$ 是 $\mu$ 的无偏估计，$\overline{X}$ 的观察值 $\bar{x}$ 的大小在一定程度上反映 $\mu$ 的大小。如果假设 $H_0$ 为真，则观察值 $\bar{x}$ 与 $\mu_0$ 的偏差 $|\bar{x}-\mu_0|$ 一般不应太大

$P_{\mu_0}\{\cdot\}$ 表示参数 $\mu$ 取 $\mu_0$ 时事件的概率，$P_{\mu\in H_0}\{\cdot\}$ 表示 $\mu$ 取 $H_0$ 规定的值时事件 $\{\cdot\}$ 的概率。我们希望

$$P\{\text{当 } H_0 \text{ 为真拒绝 } H_0\} \leq \alpha$$

由于只允许犯这类错误的概率最大为 $\alpha$，因此

$$P_{\mu_0}\left\{|\overline{X}-\mu_0|\sigma/\sqrt{n} \geq k\right\} = \alpha$$

由于当 $H_0$ 为真时，$Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$，因此

$$k = a_{\alpha/2}$$

因此若 $Z$ 的观察值满足

$$|z| = \frac{\bar{x}-\mu_0}{\sigma/\sqrt{n}} \geq k = z_{\alpha/2}$$

则拒绝 $H_0$，反之接受

数 $\alpha$ 称为 **显著水平**，统计量 $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 称为 **检验统计量**

在显著水平 $\alpha$ 下，检验假设

$$H_0 : \mu = \mu_0, \quad H_1 : \mu \neq \mu_0$$

$H_0$ 称为 **原假设**或 **零假设**，$H_1$ 称为 **备择假设**

当检验统计量取某个区域 $C$ 中的值时，我们拒绝原假设 $H_0$，则称区域 $C$ 为 **拒绝域**，拒绝域的边界点称为 **临界点**

$H_0$ 为真时拒绝称为第 I 类错误，$H_0$ 不真时接收 $H_0$ 称为第 II 类错误

只对犯第 I 类错误的概率加以控制而不考虑第 II 类错误的概率的检验称为 **显著性检验**。

$$H_0: \mu \le \mu_0, \quad H_1: \mu > \mu_0$$

称为 **右边检验**

$$H_0: \mu \ge \mu_0, \quad H_1: \mu < \mu_0$$

称为 **左边检验**。左边检验和右边检验统称为 **单边检验**

设总体 $X \sim N(\mu, \sigma^2)$，$\mu$ 未知，$\sigma$ 已知，$X_1, ..., X_n$ 是来自 $X$ 的样本，给定显著性水平 $\alpha$ ，求检验问题

$$H_0: \mu \le \mu_0, \quad H_1: \mu > \mu_0$$

的拒绝域

因 $H_0$ 中的全部 $\mu$ 都比 $H_1$ 中的 $\mu$ 要小，当 $H_1$ 为真时，观察值 $\bar{x}$ 往往偏大，因此拒绝域的形式为

$$\bar{x} \ge k$$

则

$$\begin{aligned} P_{\mu \in H_0} \overline{X} \ge k &= P_{\mu \le \mu_0} \left\{ \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \ge \frac{k - \mu_0}{\sigma/\sqrt{n}} \right\} \\ &\le P_{\mu \le \mu_0} \left\{ \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ge \frac{k - \mu_0}{\sigma/\sqrt{n}} \right\} \end{aligned}$$

要控制 $P\{$当 $H_0$ 为真拒绝 $H_0\} \le \alpha$ 只需令

$$P_{\mu \le \mu_0} \left\{ \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \ge \frac{k - \mu_0}{\sigma/\sqrt{n}} \right\} = \alpha$$

因此 $\frac{k-\mu_0}{\sigma/\sqrt{n}} = z_\alpha, k = \mu_0 + \frac{\sigma}{sqrtn z_\alpha}$ ，即得拒绝域为

$$\bar{x} \ge \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha$$

同理得左边检验问题

$$H_0: \mu \ge \mu_0, \quad H_1: \mu < \mu_0$$

的拒绝域为

$$z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \le -z_\alpha$$

## 8.2 正态总体均值的假设检验

### 8.2.1 单个总体 $N(\mu,\sigma^2)$ 均值 $\mu$ 检验

1. $\sigma^2$ 已知，关于 $\mu$ 检验（$Z$ 检验）利用统计量 $Z=\frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$ 来确定拒绝域，称为 $Z$ **检验法**
2. $\sigma^2$ 未知，关于 $\mu$ 的检验（$t$ 检验）设总体 $X\sim N(\mu,\sigma^2)$，其中 $\mu,\sigma^2$ 未知，求检验问题

$$H_0:\mu=\mu_0,\quad H:\mu\neq\mu_0$$

的拒绝域（显著性水平 $\alpha$ ）

设 $X_1,\dots,X_n$ 是来自总体 $X$ 的样本，用

$$t=\frac{\overline{X}-\mu_0}{S/\sqrt{n}}$$

作为检验统计量，当观察值 $|t|=\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|$ 过分大时就拒绝 $H_0$，拒绝域的形式为

$$|t|=\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|\geq k$$

当 $H_0$ 为真时，$\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\sim t(n-1)$，故由

$$P_{\mu_0}\left\{\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|\geq k\right\}=\alpha$$

得 $k=t_{\alpha/2}(n-1)$，即得拒绝域为

$$|t|=\left|\frac{\overline{X}-\mu_0}{S/\sqrt{n}}\right|\geq t_{\alpha/2}(n-1)$$

$t$ **检验法**

### 8.2.2 两个正态总体均值差的检验 ($t$ 检验)

求检验问题

$$H_0:\mu_1-\mu_2=\delta,\quad H_1:\mu_1-\mu_2\neq\delta$$

的拒绝域，取显著水平为 $\alpha$ 用下述 $t$ 统计量作为检验统计量

$$t = \frac{(\overline{X} - \overline{Y}) - \delta}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$$

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

$$P_{\mu_1-\mu_2=\delta}\left\{\left|\frac{(\bar{x} - \bar{y}) - \delta}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)\right| \geq k\right\} = \alpha$$

得 $k = t_{\alpha/2}(n_1 + n_2 - 2)$，于是拒绝域为

$$|t| = \left|\frac{(\bar{x} - \bar{y}) - \delta}{S_w\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)\right| \geq t_{\alpha/2}(n_1 + n_2 - 2)$$

## 8.3 正态总体方差的假设检验

### 8.3.1 单个总体的情况

设总体 $X \sim N(\mu, \sigma^2)$，$\mu, \sigma^2$ 均未知，$X_1, \dots, X_n$ 是来自 $X$ 的样本，要求检验假设（显著性水平为 $\alpha$ ）

$$H_0 : \sigma^2 = \sigma_0^2, \quad H_1 : \sigma^2 \neq \sigma_0^2$$

$\sigma_0^2$ 为已知常数

当 $H_0$ 为真时

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

取

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

则拒绝域有以下形式

$$\frac{(n-1)s^2}{\sigma_0^2} \leq k_1 \quad \text{or} \quad \frac{(n-1)s^2}{\sigma_0^2} \geq k_2$$

则

$$P_{\sigma_0^2}\left\{\left(\frac{(n-1)s^2}{\sigma_0^2}\le k_1\right)\cup\left(\frac{(n-1)s^2}{\sigma_0^2}\ge k_2\right)\right\}=\alpha$$

一般取

$$P_{\sigma_0^2}\left\{\frac{(n-1)s^2}{\sigma_0^2}\le k_1\right\}=\frac{\alpha}{2},P_{\sigma_0^2}\left\{\frac{(n-1)s^2}{\sigma_0^2}\ge k_2\right\}=\frac{\alpha}{2},$$

故 $k_1=\chi^2_{1-\alpha/2}(n-1),k_2=\chi^2_{\alpha/2}(n-1)$

下面求单边检验问题（显著性水平 $\alpha$ ）

$$H_0:\sigma^2\le\sigma_0^2,\quad H_1:\sigma^2>\sigma_0^2$$

的拒绝域，拒绝域的形式为

$$s^2\ge k$$

因为

$$\begin{aligned}P_{\sigma^2\le\sigma_0^2}\{S^2\ge k\}&=P_{\sigma^2\le\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma_0^2}\ge\frac{(n-1)k}{\sigma_0^2}\right\}\\&\le P_{\sigma^2\le\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma^2}\ge\frac{(n-1)k}{\sigma_0^2}\right\}\end{aligned}$$

只需令

$$P_{\sigma^2\le\sigma_0^2}\left\{\frac{(n-1)S^2}{\sigma^2}\ge\frac{(n-1)k}{\sigma_0^2}\right\}=\alpha$$

因此拒绝域

$$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}\ge\chi^2_\alpha(n-1)$$

左边检验问题

$$H_0:\sigma^2\ge\sigma_0^2,\quad H_1:\sigma^2<\sigma_0^2$$

的拒绝域为

$$\chi^2=\frac{(n-1)s^2}{\sigma_0^2}\le\chi^2_{1-\alpha}(n-1)$$

以上检验法称为 $\chi^2$ **检验法**

### 8.3.2 两个总体的情况

$\mu_1, \mu_2, \sigma_1^2, sigma_2^2$ 均未知，检验假设（显著性水平为 $\alpha$ ）

$$H_0 : \sigma_1^2 \le \sigma_2^2, \quad H_1 : \sigma_1^2 > \sigma_2^2$$

当 $H_1$ 为真时，观察值 $\frac{S_1^2}{S_2^2}$ 有偏大的趋势，故拒绝域有形式

$$\frac{s_1^2}{s_2^2} \ge k$$

因为

$$P_{\sigma_1^2 \le \sigma_2^2}\left\{\frac{S_1^2}{S_2^2} \ge k\right\} \le P_{\sigma_1^2 \le \sigma_2^2}\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \ge k\right\}$$

只需令

$$P_{\sigma_1^2 \le \sigma_2^2}\left\{\frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \ge k\right\} = \alpha$$

因此拒绝域为

$$F = \frac{s_1^2}{s_2^2} \ge F_\alpha(n_1 - 1, n_2 - 1)$$

**$F$ 检验法**