

## 第一章 概率论的基本概念

1. 互斥(互不相容):  $AB = \emptyset$  对立(互逆):  $\bar{A} = B$

完备(完全)事件组:  $A_1, A_2, \dots, A_n$ , 满足  $A_i A_j = \emptyset, i \neq j$ , 且  $\bigcup_{i=1}^n A_i = \Omega$

2. 德摩根律:  $\overline{A \cup B} = A \cap B$   $\overline{A \cap B} = A \cup B$   $\overline{A - B} = \overline{A \bar{B}} = \bar{A} \cup B$

3.  $A \subset B, AB = A$

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

$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(AB) - P(BC) - P(AC) + P(ABC)$

$P(A - B) = P(A) - P(AB) = P(A\bar{B})$

5. 古典概率计算公式:  $P(A) = \frac{k}{n} = \frac{\text{事件 A 中的基本事件数}}{\text{样本空间 S 中的基本事件总数}}$

6. 不放回抽取, 连续取 n 次每次取 1 个  $\Leftrightarrow$  一次取 n 个

7. 条件概率:  $P(B|A) = \frac{P(AB)}{P(A)}$  乘法定理:  $P(AB) = P(B|A)P(A)$

缩减样本空间解法

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

8. 全概率公式: 离散:  $P(A) = \sum_{i=1}^n P(AB_i) = \sum_{i=1}^n P(A|B_i)P(B_i)$

连续:  $p_X(x) = \int_{-\infty}^{+\infty} p(x|y)p_Y(y) dy$   $p_Y(y)$  同理

9. 贝叶斯公式: 离散:  $P(B_i|A) = \frac{P(A|B_i)P(B_i)}{P(A)} = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^n P(A|B_j)P(B_j)}$

连续:  $p(y|x) = \frac{p(x|y)p_Y(y)}{p_X(x)} = \frac{p(x|y)p_Y(y)}{\int_{-\infty}^{+\infty} p(x|y)p_Y(y) dy}$   $p(x|y)$  同理

10. A、B 独立:  $P(AB) = P(A)P(B)$

## 第二章 随机变量及其分布

1. 离散型随机变量 (概率质量函数 PMF)

a) **0-1 分布**: 抛硬币, 二选一

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

b) **二项分布**:  $X \sim B(n, p)$  n 重伯努利, 出现 k 次“是”

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

$E(X) = np$   $D(X) = np(1-p)$

c) **几何分布**:  $X \sim Ge(p)$  n 重伯努利, 第 k 次首次出现“是”

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

无记忆性 为负二项分布的特例  $r=1$

d) **负二项分布(帕斯卡分布)**:  $X \sim Nb(r, p)$  几何分布的和

$P\{X = k\} = C_{k-1}^{r-1} p^r (1-p)^{k-r}, k = r, r+1, \dots$   $E(X) = \frac{r}{p}$   $D(X) = \frac{r(1-p)}{p^2}$

$X = \text{第 } k \text{ 次实验, 正好发生 } r \text{ 次“是”}$

e) **超几何分布**:  $X \sim h(n, N, M)$  不放回抽样的二项分布

$P\{X = k\} = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$   $E(X) = n \frac{M}{N}$   $D(X) = n \frac{M}{N} (1 - \frac{M}{N}) (1 - \frac{n-1}{N-1})$

N 件物品, 有 M 件次品, 抽 n 件 (不放回) 有 k 件次品概率

分子: k 件从不合格品中抽取, 剩下的在合格品中抽取

分母: 从 N 件中随便抽取 n 件

若 N 巨大, 放不放回区别不大, 近似为二项分布

f) **松柏分布**:  $X \sim P(\lambda)$  二项分布的极限,  $p = \lambda/n$

$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$   $E(X) = \lambda$   $D(X) = \lambda$

条件: 平稳性、独立性、普通性

意义: 单位时间内随机事件发生的次数; 例: 汽车站台的候客人数

2. 连续型随机变量 (概率密度函数 PDF)

a) **均匀分布**:  $X \sim U(a, b)$  古典派的几何概型

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

$$E(X) = \frac{a+b}{2}$$

$$D(X) = \frac{(b-a)^2}{12}$$

b) **正态分布**:  $X \sim N(\mu, \sigma^2)$  二项分布的另一种极限

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

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

重要结论: ①  $aX + b \sim N(a\mu + b, a^2\sigma^2)$

② **标准化**: 令  $Z = \frac{X-\mu}{\sigma}$ , 则  $Z \sim N(0, 1)$

③  $Z \sim N(0, 1)$ ,  $P(Z > z_\alpha) = \alpha$ ,  $0 < \alpha < 1$ ,  $z_\alpha$  称为上  $\alpha$  分位点

④  $P(\mu - \sigma < X < \mu + \sigma) \approx 68.26\%$

$P(\mu - 3\sigma < X < \mu + 3\sigma) \approx 99.72\%$

c) **标准正态分布**:  $X \sim N(0, 1)$  用  $Z$  表示

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

d) **指数分布**:  $X \sim \text{Exp}(\lambda)$  泊松分布的间隔, 连续的几何分布

$$p(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases} \quad F(x) = \begin{cases} 1 - e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$E(X) = \frac{1}{\lambda} \quad D(X) = \frac{1}{\lambda^2}$$

无记忆性+泊松分布的三个条件

指数函数取对数→对数正态分布

3. 伯努利分布  $\xrightarrow{+ \text{伯努利分布}}$  二项分布  $\xrightarrow{+ \text{二项分布}}$  二项分布  $\xrightarrow{n \text{ 很大}}$  正态分布  $\xrightarrow{+ \text{正态分布}}$  正态分布

4. (累积)分布函数 CDF:  $F(x) = P(X \leq x)$

离散:  $F(x) = P(X \leq x) = \sum_{a \leq x} p(a)$

连续:  $F(x) = P(X \leq x) = \int_{-\infty}^x p(t) dt \quad \frac{d}{dx} F(x) = p(x)$

$F(x)$  为分布函数的充分必要条件: ①  $F(x)$  单调非减; ②  $F(x)$  右连续;

$$\textcircled{3} \lim_{x \rightarrow -\infty} F(x) = 0, \lim_{x \rightarrow +\infty} F(x) = 1;$$

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

$P(X < a) = 1 - P(X \geq a) = F(a - 0) \leftarrow "a-0"$  为  $a$  左极限, 离散时有意义

$$P(X = a) = P(X \leq a) - P(X < a) = F(a) - F(a - 0)$$

5. 中心极限定理: 正态分布是所有分布的最终归宿

6. 泊松过程:  $P(X = k, t) = \frac{(\lambda t)^k}{k!} e^{-\lambda}$  时间可变的泊松分布 ( $t=1$ )

7. 唯二无记忆性的分布: 几何分布、指数分布

8. 随机变量的函数分布: PDF 为  $p_X(x)$  的  $X$ , PDF 为  $p_Y(y)$  的  $Y = g(X)$

$h(y)$  是单调函数  $y = g(x)$  的反函数

$$\text{a) 公式法: } p_Y(y) = \begin{cases} p_X[h(y)] \cdot |h'(y)|, & a < y < b \\ 0, & \text{其他} \end{cases}$$

其中  $a, b$  为函数  $g(X)$  在  $X$  可能取值区间上的值域

b) 定义法: ① 写出  $p_X(x)$ ;

$$\textcircled{2} F_Y(y) = P\{Y \leq y\} = P\{g(X) \leq y\} = P\{X \leq h(y)\} = F_X(h(y));$$

$$\textcircled{3} p_Y(y) = F'_Y(y)$$

### 第三章 多为随机变量及其分布

#### 1. 离散:

联合概率质量函数 JPMF

边缘概率质量函数 MPMF (边缘分布):

$$p_{i.} = P\{X = x_i\} = \sum_{j=1}^{\infty} P\{X = x_i, Y = y_j\} = \sum_{j=1}^{\infty} p_{ij}, i \in N \quad Y \text{ 同理}$$

$$\text{条件概率质量函数: } P\{X = x_i | Y = y_i\} = \frac{P(X=x_i, Y=y_i)}{P(Y=y_i)} \quad Y \text{ 同理}$$

#### 2. 连续:

联合概率密度函数 JPDF

边缘概率密度函数 MPDF (边缘密度):  $p_X(x) = \int_{-\infty}^{+\infty} p(x, y) dy \quad Y \text{ 同理}$

$$\text{条件概率密度函数: } p_{X|Y}(x|y) = \frac{p(x, y)}{p_Y(y)} \quad X \text{ 为条件同理}$$

3. 条件概率  $\xrightarrow{\text{组成}}$  条件分布

4. 联合累积分布函数 JCDF:  $F(x, y) = P(\{X \leq x\} \text{ 且 } \{Y \leq y\}) = P(X \leq x, Y \leq y)$

边缘累积分布函数 MCDF:  $F_X(x) = F(x, +\infty) = \lim_{y \rightarrow +\infty} F(x, y) = P(X \leq x)$

$$F_Y(y) = F(+\infty, y)$$

条件累积分布函数: 连续:  $F_{X|Y}(x|y) = \int_{-\infty}^x \frac{p(u, y)}{p_Y(y)} du$  X 为条件同理

5. 相互独立: CDF:  $F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i)$

PMF:  $P(X = x_1, X = x_2, \dots, X = x_n) = \prod_{i=1}^n P(X = x_i)$

PDF:  $p(x_1, x_2, \dots, x_n) = \prod_{i=1}^n p_i(x_i)$

离散:  $p_{ij} = p_i \cdot p_j$  连续:  $p(x, y) = p_X(x)p_Y(y)$

6. 二维均匀分布:  $p(x, y) = \begin{cases} 1/S \\ 0, \text{其他} \end{cases}$

二维正态分布:  $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$   $X \sim N(\mu_1, \sigma_1^2)$   $Y \sim N(\mu_2, \sigma_2^2)$

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

$X, Y$  相互独立  $\Leftrightarrow \rho = 0$   $aX + bY \sim N(a\mu_1 + b\mu_2, a\sigma_1^2 + 2ab\sigma_1\sigma_2\rho + b\sigma_2^2)$

7.  $Z = (X, Y)$  的分布:

a)  $X, Y$  离散:  $P(X + Y = k) = \sum_{i=0}^k P(X = i)P(Y = k - i)$  卷积公式

b)  $X, Y$  连续:  $Z = Y + X$   $p_Z(z) = \int_{-\infty}^{+\infty} p(x, z - x) dx = \int_{-\infty}^{+\infty} p(z - y, y) dy$

$$\stackrel{X, Y \text{ 相互独立}}{=} \int_{-\infty}^{+\infty} p_X(x)p_Y(z - x) dx = \int_{-\infty}^{+\infty} p_X(z - y)p_Y(y) dy$$

$$Z = \frac{Y}{X} \quad p_{Y/X}(z) = \int_{-\infty}^{+\infty} |x| P(X, xz) dx \quad |x| = \left| \frac{dy}{dz} \right|$$

$$Z = YX \quad p_{YX}(z) = \int_{-\infty}^{+\infty} \frac{1}{|x|} P\left(x, \frac{z}{x}\right) dx$$

c)  $X$  离散,  $Y$  连续:  $F_Z(z) = P\{Z \leq z\} = \sum_i p_i P\{g(x_i, Y) \leq z | X = x_i\}$

8.  $X_i \sim F_i(x)$   $Y = \max(X_1, X_2, \dots, X_n)$   $F_Y(y) = \prod_{i=1}^n F_i(y)$

$$Z = \min(X_1, X_2, \dots, X_n) \quad F_Z(z) = 1 - \prod_{i=1}^n (1 - F_i(y))$$

## 第四章 随机变量的数字特征

1. 数学期望 (随机变量的一阶矩)

意义: ①对不确定性的计量; ②加权平均 (重心)

a) 离散:  $\mu = \mu_X = E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$  前提:  $\sum_{i=1}^{\infty} |x_i| p(x_i) < \infty$

b) 连续:  $E(X) = \int_{-\infty}^{\infty} xp(x) dx$

性质: ①  $E[g(X)] = \sum_{i=1}^{\infty} g(x_i) p(x_i) = \int_{-\infty}^{\infty} g(x) p(x) dx$ ;

$$E[g(X, Y)] = \sum_j \sum_i g(x_i, y_j) p(x_i, y_j) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) p(x, y) dx dy$$

②  $E(c) = c \rightarrow E(E(X)) = E(X)$  ③ 齐次性:  $E(aX) = aE(X)$

④ 可加性:  $E(X \pm Y) = E(X) \pm E(Y)$

⑤ 施瓦茨不等式:  $[E(XY)]^2 \leq E(X^2)E(Y^2)$

⑥  $X, Y$  独立:  $E(XY) = E(X)E(Y)$

2.  $A_i$  的示性变量 (函数):  $X_i = \begin{cases} 1, \text{事件 } A_i \text{ 发生} \\ 0, \text{事件 } A_i \text{ 没发生} \end{cases}$  用于求  $E(X)$

3. 方差 (二阶矩): 衡量集中程度

$$\text{Var}(X) = \sigma^2 = \sigma_X^2 = D(X) = E[(X - E(X))^2] = \sum_{i=1}^{\infty} (x_i - \mu)^2 p(x_i)$$

性质: ①  $D(X) = E(X^2) - E(X)^2$ ;

②  $D(aX + b) = a^2 D(X)$ ; b 的几何意义为平移量

③  $D(c) = 0 \rightarrow D(D(x)) = D(x)$

④  $D(X \pm Y) = D(X) + D(Y) \pm 2\text{Cov}(X, Y)$

⑤  $D(X) = 0 \Leftrightarrow$  存在常数 c 使得  $P(X = c) = 1$   $X = c$  与  $P(X = c) = 1$  不同

4. 标准差: 解决方差单位不一致  $\sigma = \sigma_X = \sqrt{\text{Var}(X)}$

5. 马尔可夫不等式:  $P(X \geq a) \leq \frac{E(X)}{a}$  切比雪夫不等式:  $P(|X - \mu| \geq k) \leq \frac{\sigma^2}{k^2}$

6. 协方差:  $\text{Cov}(X, Y) = E[(X - \mu_X)(Y - \mu_Y)] = E(XY) - E(X)E(Y) = \text{Cov}(Y, X)$

$\text{Cov}(X, Y) > 0$ , 正相关;  $< 0$ , 负相关;  $= 0$ , 不相关 不相关  $\Leftarrow$  独立

性质: ①  $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$

②  $\text{Cov}(X_1 + X_2, Y) = \text{Cov}(X_1, Y) + \text{Cov}(X_2, Y)$

7. 相关系数:  $\rho_{XY} = \frac{\text{Cov}(X,Y)}{\sigma_X \sigma_Y}$  因为标准差有单位  $\rho_{XY} \in [-1,1]$
- $\rho > 0$ , 正相关;  $\rho < 0$ , 负相关;
- $\rho = 1$ , 完全正相关;  $\rho = -1$ , 完全负相关;  $\rho = 0$ , (线性) 不相关;
- $|\rho_{XY}| = 1 \Leftrightarrow$  存在常数  $a(a \neq 0), b$  使得  $P(Y = aX + b) = 1$
8. 满足二维正态分布的  $X, Y$  独立  $\Leftrightarrow \rho = 0$ , 即不相关

## 第五章 大数定律和中心极限定理

1. 概率收敛: 记为  $Y_n \xrightarrow{P} a$   $\lim_{n \rightarrow \infty} P\{|Y_n - a| < \varepsilon\} = 1$
2. 弱大数定律 统计存在的基础
- a) 伯努利大数定律: 记为  $\frac{f_A}{n} \xrightarrow{P} p, n \rightarrow \infty$
- $$X_n \sim B(n, p) \quad \lim_{n \rightarrow \infty} P\left\{\left|\frac{f_A}{n} - p\right| < \varepsilon\right\} = 1$$
- b) 辛钦大数定律: 记为  $\bar{X} \xrightarrow{P} \mu, n \rightarrow \infty$
- $$X_n \text{ 具有相同的分布} \quad \lim_{n \rightarrow \infty} P\{|X - E(\bar{X})| < \varepsilon\} = 1$$
- c) 切比雪夫大数定律: 记为  $\bar{X} \xrightarrow{P} \mu, n \rightarrow \infty$
- $$\text{Var}(X_i) \leq c \quad \lim_{n \rightarrow \infty} P\{|X - E(\bar{X})| < \varepsilon\} = 1$$

	分布	独立性	方差
伯努利大数定律	伯努利分布	独立	无要求
辛钦大数定律	同分布	独立	无要求
切比雪夫大数定律	无要求	不相关	同上界

3. 强大数定律:  $P\left(\lim_{n \rightarrow \infty} |\bar{X} - \mu| < \varepsilon\right) = 1$  约束条件比弱大数更强——考研不考
4. 中心极限定理 解释了为什么生活中正态分布处处可见
- a) 棣莫弗-拉普拉斯定理 理解: 伯努利分布的和的极限是正态分布
- $$\lim_{n \rightarrow \infty} P\left(\frac{X - np}{\sqrt{np(1-p)}} \leq x\right) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt \quad \mu = np, \sigma = \sqrt{np(1-p)}$$
- b) 列维-林德伯格定理:
- $$\lim_{n \rightarrow \infty} P\left(\frac{\sum_{i=1}^n X_i - n\mu}{\sqrt{n}\sigma} \leq x\right) = \Phi(x) \quad E(\sum_i X_i) = n\mu, D(\sum_i X_i) = n\sigma^2$$

## 第六章 样本及抽样分布

1. 统计量(样本数字特征): 不含未知参数

- a) 样本均值:  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$
- b) 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:  $S = \sqrt{S^2}$
- c) k(原点)阶矩:  $E(X^k)$
- k 阶中心矩:  $E\{[X - E(X)]^k\}$
- k+l 阶混合矩:  $E(X^k Y^l)$
- k+l 阶混合中心矩  $E\{[X - E(X)]^k [Y - E(Y)]^l\}$

性质: ①  $E(\bar{X}) = E(X) = \mu$ ; ②  $D(\bar{X}) = \frac{1}{n} D(X)$

③ 样本方差和方差的关系:  $E(S^2) = \sigma^2$

2. 抽样分布: 统计量的分布

a) 卡方分布: 记作  $\chi^2 \sim \chi^2(n)$   $\chi^2 = \sum_{i=1}^n X_i^2$   $n$  越大, 越接近正态分布

$$f(y) = \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} (y > 0) \quad E(X) = n \quad D(X) = 2n$$

$$\text{伽马函数: } \Gamma(x) = \int_0^{\infty} \frac{t^{x-1}}{e^t} dt$$

若  $n$  为正整数,  $\Gamma(n) = (n-1)!$

性质: ①可加性:  $\chi_1^2 + \chi_2^2 \sim \chi^2(n_1 + n_2)$ ;

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

b) t 分布: 记作  $T \sim t(n)$   $T = \frac{X}{\sqrt{Y/n}}$ ,  $\begin{cases} X \sim N(0,1) \\ Y \sim \chi^2(n) \end{cases}$   $n$  越大, 越接近正态分布

$$h(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}} \quad E(X) = 0 (n > 1) \quad D(X) = \frac{n}{n-2} (n > 2)$$

$$\text{性质: ①上}\alpha\text{分位点: } P\{t > t_{\alpha}(n)\} = \int_{t_{\alpha}(n)}^{\infty} h(t) dt = \alpha$$

②  $h(t)$  为偶函数;

③  $n$  充分大时,  $t(n)$  分布近似于  $N(0,1)$

c) F 分布: 记作  $F \sim F(n_1, n_2)$   $F = \frac{X/n_1}{Y/n_2}$ ,  $\begin{cases} X \sim \chi^2(n_1) \\ Y \sim \chi^2(n_2) \end{cases}$

$$p(x) = \frac{\Gamma(\frac{n_1+n_2}{2})\Gamma(\frac{n_1}{2})^{\frac{n_1}{2}}}{\Gamma(\frac{n_1}{2})\Gamma(\frac{n_2}{2})} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2}x\right)^{-\frac{n_1+n_2}{2}}$$

$$E(X) = \frac{n_2}{n_2-2} (n_2 > 2) \quad D(X) = \frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} (n_2 > 4)$$

$$\text{性质: ①上}\alpha\text{分位点: } P\{F > F_{\alpha}(n_1, n_2)\} = \int_{F_{\alpha}(n_1, n_2)}^{\infty} \psi(y) dy = \alpha$$

$$\text{② } F \sim F(n_1, n_2), \text{ 则 } \frac{1}{F} \sim F(n_2, n_1), \quad F_{1-\alpha}(n_1, n_2) = \frac{1}{F_{\alpha}(n_1, n_2)}$$

d) 一个正态总体的抽样分布: 总体  $X \sim N(\mu, \sigma^2)$

i. 样本均值的分布( $\sigma^2$ 已知):  $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$  或  $Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$

ii. 样本方差的分布( $\mu$ 未知): 量化  $S^2$  逼近  $\sigma^2$  的靠谱程度

$$\bar{X} \text{ 与 } S^2 \text{ 相互独立, 且 } \chi^2 = \frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2} \sim \chi^2(n-1)$$

iii. 样本均值的分布( $\sigma^2$ 未知):  $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t(n-1)$  量化  $\bar{X}$  逼近  $\mu$  的靠谱程度

iv. 样本方差的分布( $\mu$ 已知):  $\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$  ???

e) 两个正态总体的抽样分布: 总体  $X \sim N(\mu_1, \sigma_1^2)$  和 总体  $X \sim N(\mu_2, \sigma_2^2)$

i. 样本均值的差( $\sigma_1^2, \sigma_2^2$ 已知):

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right) \text{ 或 } Z = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

ii. 样本均值的差( $\sigma_1^2 = \sigma_2^2$ 未知):

$$T = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{S_{\omega}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \quad S_{\omega}^2 = \frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1 + n_2 - 2}$$

iii. 样本方差的比例( $\mu_1, \mu_2$ 未知):  $F = \frac{S_1^2/\sigma_1^2}{S_2^2/\sigma_2^2} = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$

## 第七章 参数估计

分布函数已知，部分参数未知

### 1. 点估计：部分参数 $\theta$ ，估计量 $\hat{\theta}$

种类：①一致估计量： $\lim_{n \rightarrow \infty} P(|\hat{\theta} - \theta| < \varepsilon) = 1$        $\bar{\theta} \xrightarrow{P} \theta$       大样本容量

②无偏估计量： $E(\hat{\theta}) = \theta$       小样本容量

③更有效估计量： $D(\hat{\theta}_1) < D(\hat{\theta}_2)$ ， $\hat{\theta}_1$ 更有效

计算方法：

a) 矩估计法：

理论基础：样本  $k$  阶矩  $A_k = \frac{1}{n} \sum_{i=1}^n X_i^k$ ，随机变量  $k$  阶矩  $E(X^k) = \mu_k$

$$\lim_{n \rightarrow \infty} P(|A_k - \mu_k| < \varepsilon) = 1, \text{ 即 } A_k \xrightarrow{P} \mu_k, n \rightarrow \infty$$

步骤：列出一阶矩到  $k$  阶矩的方程。（考研最多两个方程）

b) 最大似然估计法：可能性最大的就是事实

i. 似然函数：离散： $L(\theta) = L(x_1, x_2, \dots, x_n; \theta) = \prod_{i=1}^n p(x_i; \theta), \theta \in \theta$

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

ii. 最大似然函数： $L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max_{\theta \in \theta} L(x_1, x_2, \dots, x_n; \theta)$

最大似然估计值： $\hat{\theta}(x_1, x_2, \dots, x_n)$ ；最大似然估计量： $\hat{\theta}(X_1, X_2, \dots, X_n)$

iii. 对数似然方程： $\frac{d}{d\theta} \ln L(\theta) = 0$       解法之一

### 2. 区间估计：置信水平 $1 - \alpha$

a) 单个总体  $N(\mu, \sigma^2)$

i.  $\mu$  的置信区间

$$\textcircled{1} \sigma^2 \text{ 已知: } \left( \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right) \quad \textcircled{2} \sigma^2 \text{ 未知: } \left( \bar{X} \pm \frac{s}{\sqrt{n}} t_{\alpha/2}(n-1) \right)$$

ii.  $\sigma^2$  的置信区间       $\sigma$  的置信区间同理

$$\textcircled{1} \mu \text{ 已知: } \left( \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\alpha/2}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\alpha/2}^2(n)} \right) \quad \textcircled{2} \mu \text{ 未知: } \left( \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \right)$$

b) 两个总体  $N(\mu_1, \sigma_1^2)$  和  $N(\mu_2, \sigma_2^2)$

i.  $\mu_1 - \mu_2$  的置信区间

$$\textcircled{1} \sigma_1^2, \sigma_2^2 \text{ 已知: } (\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$$

$$\textcircled{2} \sigma_1^2 = \sigma_2^2 = \text{未知: } (\bar{X} - \bar{Y} \pm t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$$

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

ii.  $\sigma_1^2 / \sigma_2^2$  的置信区间       $\mu_1, \mu_2$  未知

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

## 第八章 假设检验

1. 第一类错误： $H_0$  是对的，但我们拒绝了它（弃真）

第二类错误： $H_0$  错，接受（纳伪）

2. 显著性检测：只控制第一类错误

步骤：①提出  $H_0$ ；      ②给出显著性水平  $\alpha$ ；

③确定检验统计量及拒绝域形式；      ④求出拒绝域  $W$

3.

三大抽样分布

	统计量		概率密度函数	$E(X)$	$D(X)$
卡方分布	$\chi^2 = \sum_{i=1}^n X_i^2$	$\chi^2 \sim \chi^2(n)$	$\frac{1}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} (x > 0)$	$n$	$2n$
t 分布	$T = \frac{X}{\sqrt{Y/n}}, \begin{cases} X \sim N(0,1) \\ Y \sim \chi^2(n) \end{cases}$	$T \sim t(n)$	$\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}$	$0 (n > 1)$	$\frac{n}{n-2} (n > 2)$
F 分布	$F = \frac{X/n_1}{Y/n_2}, \begin{cases} X \sim \chi^2(n_1) \\ Y \sim \chi^2(n_2) \end{cases}$	$F \sim F(n_1, n_2)$	$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right) \left(\frac{n_1}{n}\right)^{n_1/2}}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} x^{\frac{n_1}{2}-1} \left(1 + \frac{n_1}{n_2} x\right)^{-\frac{n_1+n_2}{2}}$	$\frac{n_2}{n_2-2} (n_2 > 2)$	$\frac{2n_2^2(n_1+n_2-2)}{n_1(n_2-2)^2(n_2-4)} (n_2 > 4)$

常用随机变量分布

		$P\{X = k\}/p(x)$	$F(x)$	$E(X)$	$D(X)$	
0-1 分布		$p^k(1-p)^{1-k}$		$p$	$p(1-p)$	抛硬币，二选一
二项分布	$X \sim B(n, p)$	$\binom{n}{k} p^k (1-p)^{n-k}$		$np$	$np(1-p)$	n 重伯努利，出现 k 次“是”
松柏分布	$X \sim P(\lambda)$	$\frac{\lambda^k}{k!} e^{-\lambda}$		$\lambda$	$\lambda$	二项分布的极限， $p = \frac{\lambda}{n}$
几何分布	$X \sim Ge(p)$	$p(1-p)^{k-1}$		$\frac{1}{p}$	$\frac{1-p}{p^2}$	n 重伯努利，第 k 次首次出现“是”
负二项分布	$X \sim Nb(r, p)$	$C_{k-1}^{r-1} p^r (1-p)^{k-r}$		$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	几何分布的和
超几何分布	$X \sim h(n, N, M)$	$\frac{C_M^k C_{N-M}^{n-k}}{C_N^n}$		$n \frac{M}{N}$	$n \frac{M}{N} (1 - \frac{M}{N}) (1 - \frac{n-1}{N-1})$	不放回抽样的二项分布
均匀分布	$X \sim U(a, b)$	$\begin{cases} \frac{1}{b-a}, a < x < b \\ 0, \text{其他} \end{cases}$	$\begin{cases} 0, x < a \\ \frac{x-a}{b-a}, a \leq x \leq b \\ 1, x > b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	古典派的几何概型
正态分布	$X \sim N(\mu, \sigma^2)$	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt$	$\mu$	$\sigma^2$	二项分布的另一种极限
标准正态分布	$X \sim N(0,1)$	$\frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$	0	1	
指数分布	$X \sim Exp(\lambda)$	$\begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases}$	$\begin{cases} 1 - e^{-\lambda x}, x \geq 0 \\ 0, x < 0 \end{cases}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	泊松分布的间隔，连续的几何分布
二维均匀分布		$p(x, y) = \begin{cases} 1/S \\ 0, \text{其他} \end{cases}$				
二维正态分布	$(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; \rho)$	$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]}$				



正态总体均值、方差的置信区间

	待估参数	其他参数	枢轴量的分布	置信区间
一个正态总体	$\mu$	$\sigma^2$ 已知	$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\left( \bar{X} \pm \frac{\sigma}{\sqrt{n}} z_{\alpha/2} \right)$
		$\sigma^2$ 未知	$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$	$\left( \bar{X} \pm \frac{S}{\sqrt{n}} t_{\alpha/2}(n-1) \right)$
	$\sigma^2$	$\mu$ 已知	$\chi^2 = \frac{1}{\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$	$\left( \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{\alpha/2}^2(n)}, \frac{\sum_{i=1}^n (X_i - \mu)^2}{\chi_{1-\alpha/2}^2(n)} \right)$
		$\mu$ 未知	$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\left( \frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \right)$
两个正态总体	$\mu_1 - \mu_2$	$\sigma_1^2, \sigma_2^2$ 已知	$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$(\bar{X} - \bar{Y} \pm z_{\alpha/2} \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}})$
		$\sigma_1^2 = \sigma_2^2 =$ 未知	$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$ $S_\omega^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$	$(\bar{X} - \bar{Y} \pm t_{\frac{\alpha}{2}}(n_1 + n_2 - 2) S_\omega \sqrt{\frac{1}{n_1} + \frac{1}{n_2}})$
	$\sigma_1^2/\sigma_2^2$	$\mu_1, \mu_2$ 未知	$F = \frac{S_1^2/S_2^2}{\sigma_1^2/\sigma_2^2} \sim F(n_1 - 1, n_2 - 1)$	$\left( \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{\alpha/2}(n_1 - 1, n_2 - 1)}, \frac{S_1^2}{S_2^2} \cdot \frac{1}{F_{1-\alpha/2}(n_1 - 1, n_2 - 1)} \right)$

正态总体参数的假设检验

检验参数	其他参数	$H_0$	$H_1$	检验统计量的分布	拒绝域
$\mu$	$\sigma^2$ 已知	$\mu = \mu_0$ $\mu \leq \mu_0$ $\mu \geq \mu_0$	$\mu \neq \mu_0$ $\mu > \mu_0$ $\mu < \mu_0$	$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1)$	$ U  \geq u_{\alpha/2}$ $U \geq u_{\alpha}$ $U \leq u_{\alpha}$
	$\sigma^2$ 未知			$T = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$	$ T  \geq t_{\alpha/2}(n-1)$ $T \geq t_{\alpha}(n-1)$ $T \leq t_{\alpha}(n-1)$
$\sigma^2$	$\mu$ 已知	$\sigma^2 = \sigma_0^2$ $\sigma^2 \leq \sigma_0^2$ $\sigma^2 \geq \sigma_0^2$	$\sigma^2 \neq \sigma_0^2$ $\sigma^2 > \sigma_0^2$ $\sigma^2 < \sigma_0^2$	$\chi^2 = \frac{1}{\sigma_0^2} \sum_{i=1}^n (X_i - \mu)^2 \sim \chi^2(n)$	$\chi^2 \leq \chi_{1-\alpha/2}^2(n)$ or $\chi^2 \geq \chi_{\alpha/2}^2(n)$ $\chi^2 \geq \chi_{\alpha}^2(n)$ $\chi^2 \leq \chi_{1-\alpha}^2(n)$
	$\mu$ 未知			$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$	$\chi^2 \leq \chi_{1-\alpha/2}^2(n-1)$ or $\chi^2 \geq \chi_{\alpha/2}^2(n-1)$ $\chi^2 \geq \chi_{\alpha}^2(n-1)$ $\chi^2 \leq \chi_{1-\alpha}^2(n-1)$
$\mu_1 - \mu_2$	$\sigma_1^2, \sigma_2^2$ 已知	$\mu_1 - \mu_2 = \mu_0$ $\mu_1 - \mu_2 \leq \mu_0$ $\mu_1 - \mu_2 \geq \mu_0$	$\mu_1 - \mu_2 \neq \mu_0$ $\mu_1 - \mu_2 > \mu_0$ $\mu_1 - \mu_2 < \mu_0$	$Z = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$	$ U  \geq u_{\alpha/2}$ $U \geq u_{\alpha}$ $U \leq u_{\alpha}$
	$\sigma_1^2 = \sigma_2^2 =$ 未知			$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{S_{\omega} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2)$	$ T  \geq t_{\alpha/2}(n_1 + n_2 - 2)$ $T \geq t_{\alpha}(n_1 + n_2 - 2)$ $T \leq t_{\alpha}(n_1 + n_2 - 2)$
$\sigma_1^2/\sigma_2^2$	$\mu_1, \mu_2$ 已知	$\sigma_1^2 = \sigma_2^2$ $\sigma_1^2 \leq \sigma_2^2$ $\sigma_1^2 \geq \sigma_2^2$	$\sigma_1^2 \neq \sigma_2^2$ $\sigma_1^2 > \sigma_2^2$ $\sigma_1^2 < \sigma_2^2$	$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2), \begin{cases} X \sim \chi^2(n_1) \\ Y \sim \chi^2(n_2) \end{cases}$	$F \leq F_{1-\alpha/2}(n_1, n_2)$ or $F \geq F_{\alpha/2}(n_1, n_2)$ $F \leq F_{\alpha}(n_1, n_2)$ $F \geq F_{1-\alpha}(n_1, n_2)$
	$\mu_1, \mu_2$ 未知			$F = S_1^2/S_2^2 \sim F(n_1 - 1, n_2 - 1)$	$F \leq F_{1-\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$ or $F \geq F_{\frac{\alpha}{2}}(n_1 - 1, n_2 - 1)$ $F \leq F_{\alpha}(n_1 - 1, n_2 - 1)$ $F \geq F_{1-\alpha}(n_1 - 1, n_2 - 1)$