

# 数理统计

Didnelpsun

## 目录

<b>1</b>	<b>统计量</b>	<b>1</b>
<b>2</b>	<b>三大分布</b>	<b>1</b>
2.1	$\chi^2$ 分布 . . . . .	1
2.2	$t$ 分布 . . . . .	2
2.3	$F$ 分布 . . . . .	2
2.4	函数分布 . . . . .	3
<b>3</b>	<b>参数估计</b>	<b>3</b>
3.1	矩估计 . . . . .	3
3.1.1	一阶矩 . . . . .	3
3.1.2	二阶矩 . . . . .	3
3.2	最大似然估计 . . . . .	4
3.2.1	离散型 . . . . .	4
3.2.2	连续型 . . . . .	4
<b>4</b>	<b>估计量评价标准</b>	<b>5</b>
4.1	无偏性 . . . . .	5
4.2	有效性 . . . . .	5
4.3	一致性 . . . . .	5
<b>5</b>	<b>置信区间</b>	<b>5</b>
5.1	方差已知 . . . . .	5
5.2	方差未知 . . . . .	5

6	假设检验	6
7	两类错误	7

# 1 统计量

利用期望和方差等数学特征之间的关系进行计算统计量, 往往以  $\sum_{i=1}^n X_i$  或类似的形式。

**例题:** 已知总体  $X$  的期望为  $EX = 0$ , 方差  $DX = \sigma^2$ 。从总体抽取容量为  $n$  的简单随机样本, 其均值和方差分别为  $\bar{X}, S^2$ 。记  $S_k^2 = \frac{n}{k}\bar{X}^2 + \frac{1}{k}S^2$  ( $k = 1, 2, 3, 4$ ), 则 ()。

$$A. E(S_1^2) = \sigma^2 \quad B. E(S_2^2) = \sigma^2$$

$$C. E(S_3^2) = \sigma^2 \quad D. E(S_4^2) = \sigma^2$$

解:  $E(S_k^2) = E\left(\frac{n}{k}\bar{X}^2 + \frac{1}{k}S^2\right) = \frac{n}{k}E\bar{X}^2 + \frac{1}{k}E(S^2) = \frac{n}{k}((E\bar{X})^2 + D\bar{X}) + \frac{1}{k}E(S^2) = \frac{n}{k}\left(0 + \frac{\sigma^2}{n}\right) + \frac{1}{k}\sigma^2 = \frac{2\sigma^2}{k}, \therefore k = 2$ 。

**例题:** 设  $X_i$  为来自总体  $E(\lambda)$  ( $\lambda > 0$ ) 的简单随机样本, 记统计量  $T = \frac{1}{n} \sum_{i=1}^n X_i^2$ , 求  $ET$ 。

$$\begin{aligned} \text{解: } ET &= E\frac{1}{n} \sum_{i=1}^n X_i^2 = \frac{1}{n} \sum_{i=1}^n EX_i^2 = \frac{1}{n} \sum_{i=1}^n (DX_i + E^2 X_i) = \frac{1}{n} \sum_{i=1}^n \left(\frac{1}{\lambda^2} + \frac{1}{\lambda^2}\right) \\ &= \frac{1}{n} \cdot \frac{2n}{\lambda^2} = \frac{2}{\lambda^2}。 \end{aligned}$$

**例题:** 设  $X_i$  为来自总体  $X$  的简单随机样本, 而  $X \sim B\left(1, \frac{1}{2}\right)$ 。记  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , 求  $P\left\{\bar{X} = \frac{k}{n}\right\}$ 。 ( $0 \leq k \leq n$ )

解:  $\because X \sim B\left(1, \frac{1}{2}\right), \therefore \sum_{i=1}^n X_i \sim B\left(n, \frac{1}{2}\right)$ 。

$$\begin{aligned} P\left\{\bar{X} = \frac{k}{n}\right\} &= P\left\{\frac{1}{n} \sum_{i=1}^n X_i = \frac{k}{n}\right\} = P\left\{\sum_{i=1}^n X_i = k\right\} = C_n^k \left(\frac{1}{2}\right)^k \left(\frac{1}{2}\right)^{n-k} \\ &= C_n^k \cdot \left(\frac{1}{2}\right)^n。 \end{aligned}$$

# 2 三大分布

## 2.1 $\chi^2$ 分布

**例题:** 设  $X_1, X_2, X_3, X_4$  是来自正态总体  $N(0, 4)$  的简单随机样本, 记  $X = a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$ 。求  $X$  服从  $\chi^2$  分布下的参数与自由度。

解: 若  $X_1, X_2, X_3, X_4$  同一个正态分布, 所以  $EX_1 = EX_2 = EX_3 = EX_4 = 0$ ,  $DX_1 = DX_2 = DX_3 = DX_4 = 4$ 。

$$E(X_1 - 2X_2) = EX_1 - 2EX_2 = 0, \quad D(X_1 - 2X_2) = DX_1 - 4DX_2 = 20。$$

$\therefore X_1 - 2X_2 \sim N(0, 20)$ , 同理  $3X_3 - 4X_4 \sim N(0, 100)$ 。

对其标准化:  $\frac{X_1 - 2X_2 - 0}{\sqrt{20}} \sim N(0, 1)$ ,  $\frac{3X_3 - 4X_4 - 0}{\sqrt{100}} \sim N(0, 1)$ 。

若要让  $X$  满足  $\chi^2$  分布, 则要将  $a(X_1 - 2X_2)^2 + b(3X_3 - 4X_4)^2$  两项标准化。

$\therefore \frac{(X_1 - 2X_2)^2}{20} + \frac{(3X_3 - 4X_4)^2}{100} \sim \chi^2(2)$ , 所以  $a = \frac{1}{20}$ ,  $b = \frac{1}{100}$ 。

## 2.2 $t$ 分布

**例题:** 设  $X_1, X_2, \dots, X_8$  是来自正态总体  $N(0, 3^2)$  的简单随机样本, 则统计量  $Y = \frac{X_1 + X_2 + X_3 + X_4}{\sqrt{X_5^2 + X_6^2 + X_7^2 + X_8^2}}$  服从什么分布?

解:  $\because X_1, \dots, X_8 \sim N(0, 9)$ ,  $\therefore X_1 + X_2 + X_3 + X_4 \sim N(0, 36)$ 。

$\therefore \frac{X_1 + X_2 + X_3 + X_4 - 0}{6} \sim N(0, 1)$ 。

$$\begin{aligned} \frac{X_5^2 + X_6^2 + X_7^2 + X_8^2}{9} &= \left(\frac{X_5 - 0}{3}\right)^2 + \left(\frac{X_6 - 0}{3}\right)^2 + \left(\frac{X_7 - 0}{3}\right)^2 + \left(\frac{X_8 - 0}{3}\right)^2 \\ &\sim \chi^2(4) \\ \therefore \frac{\frac{X_1 + X_2 + X_3 + X_4 - 0}{6}}{\sqrt{\frac{X_5^2 + X_6^2 + X_7^2 + X_8^2}{9}}/4} &= \frac{X_1 + X_2 + X_3 + X_4}{\sqrt{X_5^2 + X_6^2 + X_7^2 + X_8^2}} \sim t(4)。 \end{aligned}$$

## 2.3 $F$ 分布

**例题:** 设  $X_1, X_2, \dots, X_{15}$  是来自正态总体  $N(0, 3^2)$  的简单随机样本, 则统计量  $Y = \frac{X_1^2 + X_2^2 + \dots + X_{10}^2}{2X_{11}^2 + X_{12}^2 + \dots + X_{15}^2}$  服从什么分布?

解:  $\because \frac{X_i - 0}{3} \sim N(0, 1)$ ,  $\left(\frac{X_i - 0}{3}\right)^2 = \frac{x_i^2}{9} \sim \chi^2(1)$ 。

$\therefore \frac{X_1^2 + X_2^2 + \dots + X_{10}^2}{9} \sim \chi^2(10)$ ,  $\frac{X_{11}^2 + X_{12}^2 + \dots + X_{15}^2}{9} \sim \chi^2(5)$ 。

$$\begin{aligned} \frac{\frac{X_1^2 + X_2^2 + \dots + X_{10}^2}{9}/10}{\frac{X_{11}^2 + X_{12}^2 + \dots + X_{15}^2}{9}/5} &= \frac{X_1^2 + X_2^2 + \dots + X_{10}^2}{2X_{11}^2 + X_{12}^2 + \dots + X_{15}^2} = Y \sim F(10, 5)。 \end{aligned}$$

**例题:** 已知  $(X, Y)$  的概率分布函数为  $f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2 - 2y + 1)}$ ,  $x, y \in R$ , 求  $\frac{X^2}{(Y-1)^2}$  的分布。

解:  $f(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + y^2 - 2y + 1)} = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2 + (y-1)^2)}$ , 所以根据二维正态分布的形式, 得到  $(X, Y) \sim (0, 1; 1, 1; 0)$ 。

即  $X \sim \Phi(x)$ ,  $Y - 1 \sim \Phi(x)$ ,  $\therefore X^2 \sim \chi^2(1)$ ,  $(Y - 1)^2 \sim \chi^2(1)$ ,  $\therefore \frac{X^2}{(Y - 1)^2} \sim F(1, 1)$ 。

## 2.4 函数分布

**例题：**设随机变量  $X \sim t(n)$ ,  $Y \sim F(1, n)$ , 常数  $C$  使得  $P\{X > C\} = 0.6$ , 求  $P\{Y > C^2\}$ 。

解:  $X \sim t(n)$ , 则  $X = \frac{X_1}{\sqrt{Y_1/n}} \sim t(n)$ , 其中  $X_1 \sim N(0, 1)$ ,  $Y_1 \sim \chi^2(n)$ 。

$$\therefore X^2 = \frac{X_1^2}{Y_1/n} = \frac{X_1^2/1}{Y_1/n} \sim \frac{\chi^2(1)/1}{\chi^2(n)/n} = F(1, n)。$$

$$\text{又 } P\{Y > C^2\} = 1 - P\{Y \leq C^2\}。 P\{X^2 > C^2\} = 1 - P\{X^2 \leq C^2\}。$$

$$\text{又 } P\{X^2 \leq C^2\} = P\{-C \leq X \leq C\}, \text{ 根据偶函数性质 } = 0.2。$$

$$\therefore P\{X^2 > C^2\} = 0.8。$$

## 3 参数估计

### 3.1 矩估计

基本方法就是  $EX = \frac{1}{n} \sum_{i=1}^n X_i$ 。

如果只有一个参数就使用一阶矩, 如果有两个参数就使用二阶矩, 一般不会超过两个未知数。

$$\text{即 } EX = \frac{1}{n} \sum_{i=1}^n X_i = \bar{X} = \hat{\mu}, \quad EX^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 = \hat{\sigma}^2 + \hat{\mu}^2。$$

$$\hat{\sigma}^2 = EX^2 - (EX)^2 = \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2。$$

#### 3.1.1 一阶矩

#### 3.1.2 二阶矩

**例题：**设  $X_i$  为来自区间  $[-a, a]$  上均匀分布的总体  $X$  的简单随机样本, 求  $a$  的矩估计量。

解: 首先矩估计就是  $E(X^k) = \frac{1}{n} \sum_{i=1}^n X_i^k$ 。

$$\text{又对于均匀分布 } X_i \sim U(-a, a), \quad EX = \frac{a+b}{2} = 0, \quad DX = \frac{(b-a)^2}{12} = \frac{a^2}{3}。$$

$$\text{所以 } EX \text{ 不含有 } a, \text{ 使用二阶矩 } EX^2 = DX + E^2X = \frac{a^2}{3} = \frac{1}{n} \sum_{i=1}^n X_i^2。$$

$$\text{解得 } a = \sqrt{\frac{3}{n} \sum_{i=1}^n X_i^2}。$$

## 3.2 最大似然估计

步骤：写出概率函数或密度函数；写出似然函数（代入观测值  $x_i$  并连乘）；两边取对数；求导数并令为 0 求出表达式。

### 3.2.1 离散型

例题：设总体  $X$  的概率分布为：

$X$	0	1	2	3
$P$	$\theta^2$	$2\theta(1-\theta)$	$\theta^2$	$1-2\theta$

其中  $\theta \in \left(0, \frac{1}{2}\right)$  为未知参数，从总体  $X$  中抽取容量为 8 的一组样本，其样本值为 3, 1, 3, 0, 3, 1, 2, 3。求  $\theta$  的最大似然估计值。

解：

根据样本值，可以得出：

$X$	0	1	2	3
次数	1	2	1	4

将所有的概率相乘： $L(\theta)l = (1-2\theta)^4[2\theta(1-\theta)]^2 \cdot \theta^2 \cdot \theta^2 = 4\theta^6(1-\theta)^2(1-2\theta)^4$ 。

对其求对数： $\ln L(\theta) = \ln 4 + 6 \ln \theta + 2 \ln(1-\theta) + 4 \ln(1-2\theta)$ 。

对其求导： $\frac{d \ln L(\theta)}{d\theta} = \frac{6}{\theta} - \frac{2}{1-\theta} - \frac{8}{1-2\theta} = 0$ 。解得  $\theta = \frac{7 \pm \sqrt{13}}{12}$ 。

$0 < \theta < \frac{1}{2}$ ，舍去正值，得到  $\hat{\theta} = \frac{7 - \sqrt{13}}{12}$ 。

### 3.2.2 连续型

例题：设随机变量  $X$  在区间  $[0, \theta]$  上服从均匀分布， $X_1, X_2, \dots, X_n$  是来自  $X$  的简单随机样本，求  $\theta$  的最大似然估计量  $\hat{\theta}$

解： $X \sim U(0, \theta)$ ， $f(x) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0, & \text{其他} \end{cases}$ ， $L(\theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta \\ 0, & \text{其他} \end{cases}$ 。

求  $\hat{\theta}$  即求  $L(\theta)$  的最大值， $\theta$  的最小值。又必然  $0 < x_i < \theta$ 。

所以  $\hat{\theta} = \max_{1 \leq i \leq n} x_i$ ，即  $\theta$  的最大似然估计为  $\max_{1 \leq i \leq n} X_i$ 。

（取最大值而不是最小值是因为为保证所有  $x_i$  都在定义域上， $0 < x_i < \theta$ ，所以要求  $\theta > \max x_i$ ）

例题：设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的简单随机样本， $X$  的概率密度函数  $f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}$ ， $x \in R$ ， $\lambda > 0$ ，求  $\lambda$  的最大似然估计量  $\hat{\lambda}$ 。

解:  $\because f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}, \therefore L(\lambda) = \prod_{i=1}^n \frac{1}{2\lambda} e^{-\frac{|x_i|}{\lambda}} = \left(\frac{1}{2\lambda}\right)^n e^{-\frac{1}{\lambda} \sum_{i=1}^n |x_i|}.$

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

令  $\frac{d \ln L(\lambda)}{d \lambda} = 0$ , 则  $\frac{n}{\lambda} = \frac{1}{\lambda^2} \sum_{i=1}^n |x_i|$ , 解得  $\lambda = \frac{1}{n} \sum_{i=1}^n |x_i|.$

即  $\hat{\lambda} = \frac{1}{n} \sum_{i=1}^n |X_i|.$

## 4 估计量评价标准

### 4.1 无偏性

$$E\hat{\theta} = \theta.$$

### 4.2 有效性

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

### 4.3 一致性

## 5 置信区间

### 5.1 方差已知

**例题:** 一批零件的长度服从正态分布  $N(\mu, \sigma^2)$ , 其中  $\mu, \sigma^2$  均未知。现从中随机抽取 16 个零件, 测得样本均值  $\bar{x} = 20cm$ , 样本标准差为  $s = 1cm$ , 求  $\mu$  的置信水平为 0.90 的置信区间。

解:  $\sigma$  未知, 所以使用  $s$  来求置信空间。

$$\text{置信空间为 } \left( \bar{X} - t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}}, \bar{X} + t_{\frac{\alpha}{2}}(n-1) \frac{S}{\sqrt{n}} \right).$$

$$\text{已知 } \bar{x} = 20, s = 1, n = 16, \alpha = 1 - 0.90 = 0.1.$$

$$\text{所以置信空间为 } \left( 20 - \frac{1}{4} t_{0.05}(15), 20 + \frac{1}{4} t_{0.05}(15) \right).$$

### 5.2 方差未知

**例题:** 设某群人的年龄  $X \sim N(\mu, \sigma^2)$ , 随机了解到五个人的年龄: 39, 54, 61, 72, 59, 求均值  $\mu$  的置信度为 0.95 的置信区间。

解：由于  $\sigma$  未知，所以使用样本方差， $\frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ 。

其中置信区间为  $\left(\bar{X} - \frac{S}{\sqrt{n}}t_{0.025}(n-1), \bar{X} + \frac{S}{\sqrt{n}}t_{0.025}(n-1)\right)$ 。

又  $\bar{x} = \frac{1}{5}(39 + 54 + 61 + 72 + 59) = 57$ ,  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^5 (x_i - \bar{x})^2} = 12$ 。

其中  $t_{0.025}(n-1) = t_{0.025}(4) = 2.7764$ ，所以代入得到  $(42.13, 71, 87)$ 。

## 6 假设检验

**例题：**设考试成绩服从正态分布，随机抽取 36 位考生成绩，平均分为 66.5 分，标准差为 15 分。在显著性水平 0.05 下是否可以认为这次考试的平均水平为 70 分。

解：首先提出假设  $H_0: \mu = 70$ ,  $H_1: \mu \neq 70$ 。

将  $X$  使用样本标准差进行标准化： $T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$ 。

给定显著性水平 0.05，写出拒绝域  $T < -t_{\frac{\alpha}{2}}(n-1)$  或  $T > t_{\frac{\alpha}{2}}(n-1)$ 。

代入计算统计量， $|T| = \left| \frac{\bar{X} - \mu}{S/\sqrt{n}} \right| = \left| \frac{66.5 - 70}{15/\sqrt{36}} \right| = 1.4$ 。

又  $t_{\frac{\alpha}{2}}(n-1) = t_{0.05}(35) = 2.0301 > 1.4$  不在拒绝域内，所以接受原假设。

即可以认为平均水平为 70 分。

**例题：**设  $X_1, X_2, \dots, X_{36}$  是取自正态总体  $N(\mu, 0.04)$  的简单随机样本，其中  $\mu$  为未知参数，即  $\bar{X} = \frac{1}{36} \sum_{i=1}^{36} X_i$ ，若对于检验问题  $H_0: \mu \leq 0.5$ ,  $H_1: \mu > 0.5$  在显著性水平  $\alpha = 0.05$ ，取得检验拒绝域  $D = \{(x_1, x_2, \dots, x_{36}) : \bar{x} > C\}$ ，求  $C$ 。

解：当  $H_0$  成立，则  $X \sim N(0.5, 0.04)$ ,  $\bar{X} \sim N(0.5, 0.04 \div 36) = N\left(\frac{1}{2}, \frac{1}{900}\right)$ 。

$\alpha = 0.05 = P\{\text{拒绝 } H_0 | H_0 \text{ 成立}\} = P\{\bar{X} > C\} = 1 - P\{\bar{X} \leq C\} = 1 - \Phi((C - 0.5) \times 30) = 1 - \Phi(30C - 15)$ 。

$\therefore \Phi(30C - 15) = 0.95 = \Phi(1.645)$ ，即  $30C - 15 = 1.645$ ,  $C = 0.5548$ 。

**例题：**已知某机器生产出来的零件长度  $X$  (单位:  $cm$ ) 服从正态分布  $N(\mu, \delta^2)$ ，现从中随意抽取容量为 16 的一个样本，测得样本均值  $\bar{x} = 10$ ，样本方差  $s^2 = 0.16$ ， $t_{0.025}(15) = 2.132$ 。

(1) 求总体均值  $\mu$  置信水平为 0.95 的置信区间。

(2) 在显著性水平 0.05 下检验假设  $H_0: \mu = 9.7$ ,  $H_1: \mu \neq 9.7$ 。

(1) 解：根据公式直接解出置信空间  $(10 - 0.1t_{0.025}(15), 10 + 0.1t_{0.025}(15)) = (9.7868, 10.2132)$ 。

(2) 解：根据假设  $H_0$ ，得到拒绝域  $(-\infty, 9.4868] \cup [9.9132, +\infty)$ 。



又  $\bar{X} = 10$  在拒绝域  $[9.9132, +\infty)$  上, 所以假设  $H_0$  拒绝。

## 7 两类错误

**例题:** 假定  $X$  是连续型随机变量,  $U$  是对  $X$  的一次观测值, 关于其概率密度  $f(x)$  有如下假设:

$$H_0: f(x) = \begin{cases} \frac{1}{2}, & 0 \leq x \leq 2 \\ 0, & \text{其他} \end{cases}, \quad H_1: f(x) = \begin{cases} \frac{x}{2}, & 0 \leq x \leq 2 \\ 0, & \text{其他} \end{cases}。$$

检验规则: 当事件  $V = \left\{ U > \frac{3}{2} \right\}$  出现时, 否定假设  $H_0$ , 接受  $H_1$ , 求犯第一类错误概率和第二类错误概率  $\alpha\beta$ 。

$$\text{解: } \alpha = P \left\{ U > \frac{3}{2} \middle| H_0 \right\} = \int_{\frac{3}{2}}^2 \frac{1}{2} dx = \frac{1}{4}。$$

$$\beta = P \left\{ U \leq \frac{3}{2} \middle| H_1 \right\} = \int_0^{\frac{3}{2}} \frac{x}{2} dx = \frac{9}{16}。$$