

概率论与数理统计期末复习

Probability theory and mathematical statistics

尹超

中国科学院大学，北京 100049

Carter Yin

University of Chinese Academy of Sciences, Beijing 100049, China

2025.1.4

## 序言

本笔记为概率论与数理统计的笔记整理

讲课教师:

- 李雷, 国科大教授, 中科院数学院南楼 511
- Office Hours: TBD
- telephone: 82541585
- email: lilei@ucas.ac.cn
- Website: <https://mooc.ucas.edu.cn/portal>

助教:

- 王钊, 博士研究生, 中科院数学院-email: wangke22@mails.ucas.ac.cn

# 目录

# 第1章 作业

## § 1.1 10.23

Gamma 分布  $\Gamma(\alpha, \lambda)$

$$f(x) = \begin{cases} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

Let  $Y_1, Y_2, \dots, Y_n$ , i.i.d.  $\sim \text{Exp}(\lambda)$ , then  $Y = \sum_{i=1}^n Y_i \sim \Gamma(n, \lambda)$

Gamma 函数:  $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} \lambda^\alpha e^{-\lambda x} dx$

Gamma 分布的概率密度函数:

$$f(x; \alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}, \quad x > 0$$

其中  $\alpha > 0$  是形状参数,  $\lambda > 0$  是率参数

卡方分布是自由度为  $k$  的独立标准正态随机变量平方和的分布:

$$X = Z_1^2 + Z_2^2 + \dots + Z_k^2, \quad Z_i \sim N(0, 1)$$

$$\text{则 } X \sim \chi^2(k)$$

卡方分布的概率密度函数:

$$f(x; k) = \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2}, \quad x > 0$$

对比 Gamma 分布的形式, 令  $\alpha = \frac{k}{2}, \lambda = \frac{1}{2}$

$$\begin{aligned} f(x; \frac{k}{2}, \frac{1}{2}) &= \frac{(1/2)^{k/2}}{\Gamma(k/2)} x^{k/2-1} e^{-x/2} \\ &= \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2} \end{aligned}$$

$$\text{因此 } \chi^2(k) \sim \Gamma(\frac{k}{2}, \frac{1}{2})$$

题目:

设  $X \sim \chi_n^2$ , 求  $E(X), \text{Var}(X)$  (可以利用和 Gamma 分布的关系)

设  $Z_1 \sim \chi_{n_1}^2, Z_2 \sim \chi_{n_2}^2, Z_1$  和  $Z_2$  独立, 说明  $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$

解:

(1) 求  $E(X)$  和  $\text{Var}(X)$ :

已知  $X \sim \chi_n^2$  意味着  $X \sim \Gamma(n/2, 1/2)$

对于 Gamma 分布, 有:

$$E(X) = \frac{\alpha}{\lambda}, \quad Var(X) = \frac{\alpha}{\lambda^2}$$

因此:

$$E(X) = \frac{n/2}{1/2} = n$$

$$Var(X) = \frac{n/2}{(1/2)^2} = 2n$$

(2) 证明  $Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$ :

先将问题转化为 Gamma 分布:

$$Z_1 \sim \chi_{n_1}^2 \text{ 意味着 } Z_1 \sim \Gamma(n_1/2, 1/2)$$

$$Z_2 \sim \chi_{n_2}^2 \text{ 意味着 } Z_2 \sim \Gamma(n_2/2, 1/2)$$

Gamma 分布有一个重要性质:

如果  $X_1 \sim \Gamma(\alpha_1, \lambda)$  且  $X_2 \sim \Gamma(\alpha_2, \lambda)$  且  $X_1, X_2$  独立,

$$\text{则 } X_1 + X_2 \sim \Gamma(\alpha_1 + \alpha_2, \lambda)$$

应用这个性质:

$$Z_1 + Z_2 \sim \Gamma\left(\frac{n_1}{2} + \frac{n_2}{2}, \frac{1}{2}\right) = \Gamma\left(\frac{n_1 + n_2}{2}, \frac{1}{2}\right)$$

由 Gamma 分布和卡方分布的关系可知:

$$\text{这等价于 } Z_1 + Z_2 \sim \chi_{n_1+n_2}^2$$

因此得证。

## § 1.2 10.28&10.30

定义: 对于伯努利试验, 事件发生概率为  $p$ , 在连续试验中第  $r+k$  次试验恰好出现第  $r$  次的概率  
即: 在前  $r+k-1$  次试验中出现  $r-1$  次, 且第  $r+k$  次试验中出现第  $r$  次

概率质量函数 (PMF):

$$P(X = k) = \binom{r+k-1}{r-1} p^r (1-p)^k$$

期望的推导:

$$\begin{aligned} E(X) &= \sum_{k=0}^{\infty} k \binom{r+k-1}{r-1} p^r (1-p)^k \\ &= \frac{r(1-p)}{p} \end{aligned}$$

方差的推导:

$$\begin{aligned} E(X^2) &= \sum_{k=0}^{\infty} k^2 \binom{r+k-1}{r-1} p^r (1-p)^k \\ &= \frac{r(1-p)(1 + (r-1)(1-p))}{p^2} \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X^2) - [E(X)]^2 \\ &= \frac{r(1-p)(1 + (r-1)(1-p))}{p^2} - \left(\frac{r(1-p)}{p}\right)^2 \\ &= \frac{r(1-p)}{p^2} \end{aligned}$$

## 1.2.1 1

1. 设  $X_1, \dots, X_n$  是抽自负二项分布的样本, 求  $p$  的矩估计与极大似然估计。

解:

首先回顾负二项分布的 PMF:

$$P(X = k) = \binom{r+k-1}{k} p^r (1-p)^k$$

## (1) 矩估计

由负二项分布的期望:

$$E(X) = \frac{r(1-p)}{p}$$

根据矩估计的原理, 用样本均值代替总体期望:

$$\bar{X} = \frac{r(1-p)}{p}$$

解这个方程:

$$\bar{X}p = r(1-p)$$

$$\bar{X}p = r - rp$$

$$\bar{X}p + rp = r$$

$$p(\bar{X} + r) = r$$

因此  $p$  的矩估计为:

$$\hat{p}_{moment} = \frac{r}{\bar{X} + r}$$

## (2) 极大似然估计

似然函数为:

$$L(p) = \prod_{i=1}^n \binom{r+x_i-1}{x_i} p^r (1-p)^{x_i}$$

取对数:

$$\ln L(p) = \sum_{i=1}^n \ln \binom{r+x_i-1}{x_i} + nr \ln p + \ln(1-p) \sum_{i=1}^n x_i$$

对  $p$  求导并令其等于 0:

$$\frac{\partial \ln L(p)}{\partial p} = \frac{nr}{p} - \frac{\sum_{i=1}^n x_i}{1-p} = 0$$

解这个方程:

$$\frac{nr}{p} = \frac{\sum_{i=1}^n x_i}{1-p}$$

$$nr(1-p) = p \sum_{i=1}^n x_i$$

$$nr - nrp = p \sum_{i=1}^n x_i$$

$$nr = p(nr + \sum_{i=1}^n x_i)$$

因此  $p$  的极大似然估计为:

$$\hat{p}_{MLE} = \frac{nr}{nr + \sum_{i=1}^n x_i} = \frac{r}{\bar{X} + r}$$

结论: 对于这个负二项分布, 矩估计和极大似然估计得到了相同的结果:

$$\hat{p} = \frac{r}{\bar{X} + r}$$

## 1.2.2 2

2. (a) 设  $a_1, \dots, a_n$  是  $n$  个实数, 定义函数  $h(a) = \sum_{i=1}^n |a_i - a|$ . 证明: 当  $a$  为  $a_1, \dots, a_n$  的样本中位数 (见 4.29 式) 时,  $h(a)$  达到最小值. (b) 设  $X_1, \dots, X_n$  为自具概率密度函数  $\frac{1}{2}e^{-|x-\theta|}$  中抽出的样本 (这个分布叫拉普拉斯分布), 求参数  $\theta$  的矩估计和最大似然估计.

解:

(a) 我们通过以下步骤证明样本中位数使  $h(a)$  达到最小值:

1) 首先考虑简单情况: 对于两个数  $c_1 < c_2$ , 函数  $g(a) = |c_1 - a| + |c_2 - a|$  在  $c_1 \leq a \leq c_2$  时达到最小值  $c_2 - c_1$ .

2) 将样本  $a_1, \dots, a_n$  按从小到大排序, 记为:

$$a_{(1)} \leq a_{(2)} \leq \dots \leq a_{(n)}$$

3) 将函数  $h(a)$  重新排列:

$$h(a) = \sum_{i=1}^n |a_{(i)} - a| = (|a_{(1)} - a| + |a_{(n)} - a|) + (|a_{(2)} - a| + |a_{(n-1)} - a|) + \dots$$

4) 根据步骤 1) 的结论, 要使此和达到最小值,  $a$  必须落在以下每一个区间之内:

$$[a_{(1)}, a_{(n)}], [a_{(2)}, a_{(n-1)}], [a_{(3)}, a_{(n-2)}], \dots$$

5) 分两种情况:

- 当  $n$  为奇数时, 适合这个条件的唯一的  $a$  是  $a_{(\frac{n+1}{2})}$
- 当  $n$  为偶数时, 区间  $[a_{(\frac{n}{2})}, a_{(\frac{n}{2}+1)}]$  中的任意数  $a$  都适合这个条件

6) 这正好说明样本中位数 (无论  $n$  是奇数还是偶数) 都会使  $h(a)$  达到最小值。

因此证明了样本中位数确实使  $h(a)$  达到最小值。

(b) 对于拉普拉斯分布, 似然函数为:

$$L(\theta) = \prod_{i=1}^n \frac{1}{2} e^{-|x_i - \theta|}$$

取对数得:

$$\ln L(\theta) = -n \ln 2 - \sum_{i=1}^n |x_i - \theta|$$

最大化  $\ln L(\theta)$  等价于最小化  $\sum_{i=1}^n |x_i - \theta|$ 。

根据 (a) 的结论, 此最小值在样本中位数处取得。因此,  $\theta$  的最大似然估计为样本中位数。

对于矩估计, 由拉普拉斯分布的性质可知:  $E(X) = \theta$  因此,  $\theta$  的矩估计为样本均值  $\bar{X}$ 。

## 1.2.3 3

3. 设  $X_1, \dots, X_n$  为抽自均匀分布  $R(\theta, 2\theta)$  的样本, 求  $\theta$  的矩估计与极大似然估计。

解:

(1) 矩估计:

总体均值为  $\frac{3\theta}{2}$ , 故矩估计为  $\frac{2\bar{X}}{3}$

(2) 极大似然估计:



样本  $(X_1, \dots, X_n)$  的似然函数为:

$$f(x_1, \dots, X_n, \theta) = \begin{cases} \theta^{-n}, & \text{当 } \theta \leq \min(X_i) \leq \max(X_i) \leq 2\theta \\ 0, & \text{其他情况} \end{cases}$$

可看出极大似然估计为  $\frac{1}{2} \max(X_1, \dots, X_n)$

#### 1.2.4 4

(a) 证明

$$f(x; a, \sigma) = (\sqrt{2\pi\sigma^3})^{-1} (x - a)^2 \exp\left(-\frac{1}{2\sigma^2} (x - a)^2\right)$$

$$-\infty < x < \infty$$

作为  $x$  的函数是概率密度, 其中  $a, \sigma$  为参数,  $-\infty < a < \infty, \sigma > 0$ .

(b) 设  $X_1, \dots, X_n$  为抽自此总体的样本, 求  $a$  和  $\sigma^2$  的矩估计.

因为积分

$$\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma}} (x - a)^2 \exp\left(-\frac{1}{2\sigma^2} (x - a)^2\right) dx$$

是  $N(a, \sigma^2)$  的方差, 为  $\sigma^2$ , 故立即看出  $f(x; a, \sigma)$  为概率密度函数. 由对称性知此分布均值为  $a$ , 故  $a$  的矩估计为  $\bar{X}$ . 此分布的方差为  $3\sigma^2$ , 故得  $\sigma^2$  的矩估计为  $\frac{1}{3(n-1)} \sum_{i=1}^n (X_i - \bar{X})^2$ .

#### 1.2.5 5

### 题目

设随机变量  $X$  具有分布密度:

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

$X_1, X_2, \dots, X_n$  为从总体  $X$  的 i.i.d 样本

- 1) 求  $\theta$  的极大似然估计  $T_1$
- 2) 求  $\theta$  的矩估计  $T_2$
- 3) 计算  $T_1$  和  $T_2$  的均方误差  $E_\theta(T_1 - \theta)^2$  和  $E_\theta(T_2 - \theta)^2$

### 解答

1) 极大似然估计:

似然函数为:

$$L(\theta) = \prod_{i=1}^n p(x_i, \theta) = \prod_{i=1}^n e^{-(x_i - \theta)} I(x_i \geq \theta)$$

取对数得：

$$\ln L(\theta) = - \sum_{i=1}^n (x_i - \theta) I(x_i \geq \theta)$$

由似然函数的性质可知， $\theta$  必须小于等于所有样本值，否则似然函数为 0。在此条件下，似然函数关于  $\theta$  单调递增，因此：

$$T_1 = X_{(1)} = \min(X_1, \dots, X_n)$$

2) 矩估计：

对于指数分布，有：

$$E(X - \theta) = 1$$

因此：

$$E(X) = \theta + 1$$

用样本均值代替总体均值：

$$\bar{X} = \theta + 1$$

所以矩估计为：

$$T_2 = \bar{X} - 1$$

3) 均方误差计算：

对于  $T_1$ ：  $X_{(1)} - \theta$  服从指数分布参数为  $n$  的指数分布，因此：

$$E_{\theta}(T_1 - \theta)^2 = E_{\theta}(X_{(1)} - \theta)^2 = \frac{2}{n^2}$$

对于  $T_2$ ：

$$E_{\theta}(T_2 - \theta)^2 = E_{\theta}(\bar{X} - 1 - \theta)^2 = \text{Var}(\bar{X}) = \frac{1}{n}$$

因此，当  $n > 2$  时， $T_1$  的均方误差小于  $T_2$  的均方误差。

## 均方误差计算的详细过程

对于估计量  $T$  的均方误差定义：

$$MSE = E_{\theta}(T - \theta)^2 = \text{Var}(T) + [\text{Bias}(T)]^2$$

### 1. 计算 $T_1 = X_{(1)}$ 的均方误差

因为  $X_{(1)} - \theta$  服从参数为  $n$  的指数分布，所以：

$$\begin{aligned} E_{\theta}(T_1 - \theta)^2 &= \text{Var}(T_1) + [\text{Bias}(T_1)]^2 \\ &= \text{Var}(X_{(1)}) + [E(X_{(1)} - \theta)]^2 \\ &= \frac{1}{n^2} + \left(\frac{1}{n}\right)^2 \quad (\text{指数分布的方差和期望性质}) \\ &= \frac{2}{n^2} \end{aligned}$$

## 2. 计算 $T_2 = \bar{X} - 1$ 的均方误差

注意到对原始总体,  $X - \theta \sim \text{Exp}(1)$ , 因此  $\text{Var}(X) = 1$ 。

$$\begin{aligned}
 E_{\theta}(T_2 - \theta)^2 &= \text{Var}(T_2) + [\text{Bias}(T_2)]^2 \\
 &= \text{Var}(\bar{X}) + [E(\bar{X} - 1 - \theta)]^2 \\
 &= \text{Var}(\bar{X}) + [E(\bar{X}) - 1 - \theta]^2 \\
 &= \frac{\text{Var}(X)}{n} + [(\theta + 1) - 1 - \theta]^2 \quad (\text{因为 } E(X) = \theta + 1) \\
 &= \frac{1}{n} + 0^2 \\
 &= \frac{1}{n}
 \end{aligned}$$

## 结论

当  $n > 2$  时, 我们有:

$$E_{\theta}(T_1 - \theta)^2 = \frac{2}{n^2} < \frac{1}{n} = E_{\theta}(T_2 - \theta)^2$$

这说明虽然  $T_1$  是有偏估计, 但在样本量足够大时 ( $n > 2$ ), 其均方误差反而小于无偏估计  $T_2$  的均方误差。

## 补充说明

1.  $T_1$  的偏差为  $\frac{1}{n}$  (有偏估计) 2.  $T_2$  的偏差为 0 (无偏估计) 3. 均方误差的比较反映了方差-偏差权衡 (Variance-Bias Trade-off) 的典型例子。

## 样本均值方差的推导

设  $X_1, X_2, \dots, X_n$  是来自总体  $X$  的独立同分布样本。

### 1. 原始总体方差

对于原始总体:

$$\begin{aligned}
 X - \theta &\sim \text{Exp}(1) \\
 X &= (X - \theta) + \theta \\
 \text{Var}(X) &= \text{Var}(X - \theta) \quad (\text{常数 } \theta \text{ 的线性性质}) \\
 &= 1 \quad (\text{参数为 } 1 \text{ 的指数分布的方差})
 \end{aligned}$$

### 2. 样本均值的定义

样本均值定义为:

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

其中:

- 样本是独立同分布的 (i.i.d)
- 每个  $X_i$  都与总体  $X$  有相同的分布, 即  $\text{Var}(X_i) = 1$

### 3. 方差的基本性质

回顾方差的两个重要性质:

1. 对于独立随机变量, 其和的方差等于方差的和:

$$\text{Var}(X_1 + X_2) = \text{Var}(X_1) + \text{Var}(X_2) \quad \text{若 } X_1, X_2 \text{ 独立}$$

2. 常数与随机变量的线性性质:

$$\text{Var}(cX) = c^2 \text{Var}(X) \quad \text{对任意常数 } c$$

### 4. 样本均值方差的推导

根据以上性质, 我们可以推导样本均值的方差:

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &= \left(\frac{1}{n}\right)^2 \text{Var}\left(\sum_{i=1}^n X_i\right) \quad (\text{常数的线性性质}) \\ &= \left(\frac{1}{n}\right)^2 \sum_{i=1}^n \text{Var}(X_i) \quad (\text{独立性}) \\ &= \left(\frac{1}{n}\right)^2 \cdot n \cdot 1 \quad (n \text{ 个方差为 } 1 \text{ 的项}) \\ &= \frac{1}{n} \end{aligned}$$

### 结论

这个结果表明样本均值的方差等于总体方差除以样本量, 即:

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{1}{n}$$

这是大数定律的一个体现: 随着样本量  $n$  的增加, 样本均值的方差趋近于 0, 表明样本均值将越来越稳定地趋近于总体均值。

### 1.2.6 6

10. 设  $X_1, \dots, X_n$  为取自  $R(0, \theta)$  的样本。

1. 证明:  $\hat{\theta}_1 = \max(X_1, \dots, X_n) + \min(X_1, \dots, X_n)$  是  $\theta$  的一个无偏估计。
2. 证明: 对适当选择的参数  $c_n$ ,  $\hat{\theta}_2 = c_n \min(X_1, \dots, X_n)$  是  $\theta$  的无偏估计。但这个估计的方差比另外两个无偏估计  $\hat{\theta}_3 = \bar{X}$  和  $\hat{\theta}_4 = \frac{n+1}{n} \max(X_1, \dots, X_n)$  都大 (除非  $n = 1$ )。

解：依第二章 23 题， $\min(X_1, \dots, X_n)$  与  $\theta - \max(X_1, \dots, X_n)$  同分布。因此二者之均值相同，由此得

$$E[\min(X_1, \dots, X_n) + \max(X_1, \dots, X_n)] = \theta$$

这证明了 (a)。又由第二章 22 题知  $\min(X_1, \dots, X_n)$  的概率密度为

$$\frac{1}{\theta} n \left(1 - \frac{x}{\theta}\right)^{n-1} \quad (if 0 < x < \theta, else 0)$$

其均值为  $\theta/(n+1)$ 。由此可知，令  $C_n = n+1$ ，则  $C_n \min(X_1, \dots, X_n)$  为  $\theta$  的无偏估计。这证明了 (b)。为证 (c)，只须算出

$$\text{Var}(C_n \min(X_1, \dots, X_n)) = (n+1)^2 \int_0^\theta x^2 \left(1 - \frac{x}{\theta}\right)^{n-1} dx - \theta^2 = n\theta^2/(n+2)$$

与例 3.5 比较即得。

(问：由  $C_n \min(X_1, \dots, X_n)$  的方差表达式看出这个估计之不合理处，在什么地方？—— $n$  愈大，其方差非但不下降，反而上升，即样本愈多，估计误差愈大了)

## § 1.3 11.04

### 1.3.1 1

题目：

设  $Z = (Z_1, Z_2, \dots, Z_n)$  是从伯努利分布  $\text{Bernoulli}(\theta)$  中独立同分布抽取的样本，其中  $\theta$  是随机变量  $\Theta$  的一个实现值。 $X_n = \sum_{i=1}^n Z_i \sim \text{Bin}(n, \theta)$ 。设先验分布为  $\text{Beta}(\alpha, \beta)$ ：

$$\theta \sim \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

1) 证明后验分布是  $\text{Beta}(x + \alpha, n - x + \beta)$  2) 证明贝叶斯估计为：

$$E(\theta|X) = \frac{x + \alpha}{n + \alpha + \beta}$$

解答：

1) 证明后验分布是  $\text{Beta}(x + \alpha, n - x + \beta)$  分布：

根据贝叶斯定理：

$$P(\theta|X) \propto P(X|\theta)P(\theta)$$

似然函数为：

$$P(X|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

先验分布是  $\text{Beta}(\alpha, \beta)$ ：

$$P(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \theta^{\alpha-1} (1 - \theta)^{\beta-1}$$

因此，后验分布正比于：

$$\begin{aligned} P(\theta|X) &\propto \theta^x (1 - \theta)^{n-x} \cdot \theta^{\alpha-1} (1 - \theta)^{\beta-1} \\ &= \theta^{x+\alpha-1} (1 - \theta)^{n-x+\beta-1} \end{aligned}$$

这正是  $\text{Beta}(x + \alpha, n - x + \beta)$  分布的核心部分。

2) 求贝叶斯估计：

对于 Beta 分布, 我们知道:

$$E[\theta] = \frac{\alpha}{\alpha + \beta}$$

因此, 对于后验分布  $\text{Beta}(x + \alpha, n - x + \beta)$ :

$$\begin{aligned} E(\theta|X) &= \frac{x + \alpha}{(x + \alpha) + (n - x + \beta)} \\ &= \frac{x + \alpha}{n + \alpha + \beta} \end{aligned}$$

这样我们就证明了:

$$E(\theta|X) = \frac{x + \alpha}{n + \alpha + \beta}$$

**补充说明:**

似然函数的推导过程:

- 1) 首先, 根据题目条件: -  $Z_i$  是从伯努利分布中抽取的样本 -  $X_n = \sum_{i=1}^n Z_i \sim \text{Bin}(n, \theta)$
- 2) 对于伯努利试验: - 每次试验的概率为:  $P(Z_i = 1) = \theta, P(Z_i = 0) = 1 - \theta$  - 进行了  $n$  次独立试验 - 成功 (即  $Z_i = 1$ ) 的次数为  $x$
- 3) 由于: -  $X_n \sim \text{Bin}(n, \theta)$  是二项分布 -  $n$  次独立试验中恰好有  $x$  次成功的概率为:

$$P(X|\theta) = \binom{n}{x} \theta^x (1 - \theta)^{n-x}$$

这里: -  $\binom{n}{x}$  表示从  $n$  次试验中选择  $x$  次成功的方式数 -  $\theta^x$  表示  $x$  次成功的概率 -  $(1 - \theta)^{n-x}$  表示  $(n-x)$  次失败的概率

因此, 这就是为什么似然函数取这种形式 - 它代表了在参数  $\theta$  下观察到数据  $X$  的概率。

### 1.3.2 2

17. 设  $X_1, \dots, X_n$  为取自均匀分布  $R(0, \theta)$  中的样本. 证明: 对任给的  $1 - \alpha (0 < 1 - \alpha < 1)$ , 可找到常数  $c_n$ , 使  $[\max(X_1, \dots, X_n), c_n \max(X_1, \dots, X_n)]$  为  $\theta$  的一个置信系数  $1 - \alpha$  的区间估计.

17. 因为  $\max(X_1, \dots, X_n)$  (记为  $\hat{\theta}$ ) 的密度函数为  $nx^{n-1}/\theta^n$  (当  $0 < x < \theta$ , 此外为 0). 故

$$\begin{aligned} P_\theta(\hat{\theta} \leq \theta \leq c_n \hat{\theta}) &= P_\theta(\theta/c_n \leq \hat{\theta} \leq \theta) \\ &= \int_{\theta/c_n}^{\theta} nx^{n-1} dx / \theta^n = (\theta^n - (\theta/c_n)^n) / \theta^n = 1 - c_n^{-n} \end{aligned}$$

要此值等于  $1 - \alpha$ , 只须取  $c_n = \left(\frac{1}{1-\alpha}\right)^{1/n}$ .

(这里标答有误, 应为  $c_n = \left(\frac{1}{\alpha}\right)^{1/n}$ )

**补充说明:**

对于均匀分布  $R(0, \theta)$  中的样本  $X_1, \dots, X_n$ , 让我们逐步推导其最大值的概率密度函数:

- 1) 首先, 对于单个样本  $X_i$ , 其分布函数为:

$$F_X(x) = P(X \leq x) = \frac{x}{\theta}, \quad 0 \leq x \leq \theta$$

- 2) 相应的概率密度函数为:

$$f_X(x) = \frac{1}{\theta}, \quad 0 \leq x \leq \theta$$

3) 对于最大值  $Y = \max(X_1, \dots, X_n)$ , 其分布函数为:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(\max(X_1, \dots, X_n) \leq y) \\ &= P(X_1 \leq y, X_2 \leq y, \dots, X_n \leq y) \end{aligned}$$

4) 由于样本独立同分布:

$$\begin{aligned} F_Y(y) &= P(X_1 \leq y) \times P(X_2 \leq y) \times \dots \times P(X_n \leq y) \\ &= [F_X(y)]^n = \left(\frac{y}{\theta}\right)^n, \quad 0 \leq y \leq \theta \end{aligned}$$

5) 对分布函数求导得到概率密度函数:

$$\begin{aligned} f_Y(y) &= \frac{d}{dy}[F_Y(y)] = \frac{d}{dy}\left[\left(\frac{y}{\theta}\right)^n\right] \\ &= n\left(\frac{y}{\theta}\right)^{n-1} \times \frac{1}{\theta} \\ &= \frac{ny^{n-1}}{\theta^n}, \quad 0 < y < \theta \end{aligned}$$

因此, 最大值的密度函数为:

$$f_Y(y) = \begin{cases} \frac{ny^{n-1}}{\theta^n}, & 0 < y < \theta \\ 0, & \text{其他情况} \end{cases}$$

### 1.3.3 3

19. 设  $X_1, \dots, X_n$  是取自具参数  $\lambda_1$  的指数分布的样本,  $Y_1, \dots, Y_m$  是取自具参数为  $\lambda_2$  的指数分布的样本, 试求  $\lambda_2/\lambda_1$  的区间估计。

考虑

$$2\lambda_1 n \bar{X} / 2\lambda_2 m \bar{Y} = Z$$

分子分母独立, 分别服从卡方分布  $\chi_{2n}^2$  和  $\chi_{2m}^2$ 。故

$$P\left(F_{2n,2m}\left(1 - \frac{\alpha}{2}\right) \leq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \leq F_{2n,2m}\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha$$

此式可改写为

$$P\left(\frac{\bar{X}}{\bar{Y}} / F_{2n,2m}\left(\frac{\alpha}{2}\right) \leq \frac{\lambda_2}{\lambda_1} \leq \frac{\bar{X}}{\bar{Y}} / F_{2n,2m}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

即得  $\lambda_2/\lambda_1$  的置信区间。

**补充说明:**

19. 设  $X_1, \dots, X_n$  是取自具参数  $\lambda_1$  的指数分布的样本,  $Y_1, \dots, Y_m$  是取自具参数为  $\lambda_2$  的指数分布的样本, 试求  $\lambda_2/\lambda_1$  的区间估计。

**分布之间的关系:**

**1. 指数分布与卡方分布的关系**

若  $X_1, \dots, X_n$  独立同分布于指数分布  $E(\lambda)$ , 则:

$$2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2$$

特别地,

$$2\lambda n \bar{X} \sim \chi_{2n}^2$$

## 2. F 分布与卡方分布的关系

若  $U \sim \chi_m^2$ ,  $V \sim \chi_n^2$ , 且  $U$  和  $V$  独立, 则:

$$\frac{U/m}{V/n} \sim F_{m,n}$$

## 3. 本题中的应用

考虑统计量:

$$Z = \frac{2\lambda_1 n \bar{X}/2n}{2\lambda_2 m \bar{Y}/2m} \sim F_{2n,2m}$$

因为:

- $2\lambda_1 n \bar{X} \sim \chi_{2n}^2$
- $2\lambda_2 m \bar{Y} \sim \chi_{2m}^2$
- 两者独立

这就解释了为什么:

$$P\left(F_{2n,2m}\left(1 - \frac{\alpha}{2}\right) \leq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \leq F_{2n,2m}\left(\frac{\alpha}{2}\right)\right) = 1 - \alpha$$

此式可改写为

$$P\left(\frac{\bar{X}}{\bar{Y}}/F_{2n,2m}\left(\frac{\alpha}{2}\right) \leq \frac{\lambda_2}{\lambda_1} \leq \frac{\bar{X}}{\bar{Y}}/F_{2n,2m}\left(1 - \frac{\alpha}{2}\right)\right) = 1 - \alpha$$

即得  $\lambda_2/\lambda_1$  的置信区间。

**补充说明:**

1. 指数分布的期望为  $\frac{1}{\lambda}$ , 方差为  $\frac{1}{\lambda^2}$
2. 卡方分布  $\chi_n^2$  的期望为  $n$ , 方差为  $2n$
3.  $F_{m,n}$  分布的期望为  $\frac{n}{n-2}$  (当  $n > 2$ )

## § 1.4 11.06

设  $X_1, X_2, \dots, X_n$  i.i.d  $\sim \text{Poisson}(\lambda)$ ,  $n$  充分大, 试求  $\lambda$  的置信水平为  $1 - \alpha$  的置信区间。

解: (1) 由泊松分布的性质知:  $E(X_i) = \lambda$ ,  $\text{Var}(X_i) = \lambda$

(2) 令  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ , 则:

$$E(\bar{X}) = \lambda, \text{Var}(\bar{X}) = \frac{\lambda}{n}$$

(3) 当  $n$  充分大时, 根据中心极限定理:

$$\frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \sim N(0, 1)$$

(4) 对于置信水平  $1 - \alpha$ , 有:

$$P(-z_{\alpha/2} \leq \frac{\bar{X} - \lambda}{\sqrt{\lambda/n}} \leq z_{\alpha/2}) = 1 - \alpha$$

(5) 整理不等式:



$$P(\bar{X} - z_{\alpha/2}\sqrt{\lambda/n} \leq \lambda \leq \bar{X} + z_{\alpha/2}\sqrt{\lambda/n}) = 1 - \alpha$$

(6) 将  $\lambda$  代入方差项, 用  $\bar{X}$  估计  $\lambda$ , 得近似置信区间:

$$(\bar{X} - z_{\alpha/2}\sqrt{\bar{X}/n}, \bar{X} + z_{\alpha/2}\sqrt{\bar{X}/n})$$

因此,  $\lambda$  的置信水平为  $1 - \alpha$  的近似置信区间为:

$$(\bar{X} - z_{\alpha/2}\sqrt{\bar{X}/n}, \bar{X} + z_{\alpha/2}\sqrt{\bar{X}/n})$$

## § 1.5 11.11

### 1.5.1 1

## 统计假设检验问题

设  $X$  为取自正态总体  $N(\theta, \sigma^2)$  中的样本 (样本大小为 1)。  $\sigma$  已知,  $a, b$  都是给定常数,  $a < b$ 。要找原假设  $H_0: a \leq \theta \leq b$  的水平  $\alpha$  检验。

### 解答

1° 从直观考虑,  $H_0$  的接受域应取为  $C_1 \leq X \leq C_2$ , 即当  $C_1 \leq X \leq C_2$  时接受  $H_0$ , 不然就否定  $H_0$ 。功率函数为  $\beta(\theta)$ 。

其中功率函数的具体形式为:

$$\beta(\theta) = 1 - \left[ \Phi\left(\frac{C_2 - \theta}{\sigma}\right) - \Phi\left(\frac{C_1 - \theta}{\sigma}\right) \right]$$

$\Phi$  为标准正态  $N(0, 1)$  的分布函数。

2° 找出常数  $C_1, C_2$  使 1° 中找出  $\beta(\theta)$  满足:

$$\beta(a) = \beta(b) = \alpha$$

这归结为方程组:

$$\begin{aligned} \Phi\left(\frac{C_2 - a}{\sigma}\right) - \Phi\left(\frac{C_1 - a}{\sigma}\right) &= 1 - \alpha \\ \Phi\left(\frac{C_2 - b}{\sigma}\right) - \Phi\left(\frac{C_1 - b}{\sigma}\right) &= 1 - \alpha \end{aligned}$$

求解这个方程组的步骤如下: (a) 由于  $\beta(\theta)$  的对称性, 我们有  $C_1 + C_2 = a + b$  (b) 令  $C_1^0$  为  $C_1$  的初始值, 取  $C_1^0 < (a + b)/2$  (c) 由  $C_2^0 = (a + b) - C_1^0$  定出  $C_2$  的初始值 (d) 代入方程 (1) 和 (2) 验证, 若右边小于  $1 - \alpha$ , 说明  $C_1^0$  选得太大, 否则太小 (e) 通过迭代调整直到满足方程组

3° 证明由 1°, 2° 决定的检验确是  $H_0$  的水平  $\alpha$  检验, 即证明当  $a \leq \theta \leq b$  时,  $\beta(\theta) \leq \alpha$ 。

证明: (1) 首先计算  $\beta(\theta)$  的导数: 记  $\Phi'(x) = \varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ , 则:

$$\beta'(\theta) = \frac{1}{\sigma} \left[ \varphi\left(\frac{C_2 - \theta}{\sigma}\right) - \varphi\left(\frac{C_1 - \theta}{\sigma}\right) \right]$$

(2) 分析导数的符号: - 当  $\theta < \frac{C_1 + C_2}{2}$  时,  $\beta'(\theta) < 0$  - 当  $\theta > \frac{C_1 + C_2}{2}$  时,  $\beta'(\theta) > 0$

(3) 因此  $\beta(\theta)$  在  $\theta = \frac{C_1+C_2}{2}$  处取得最小值 - 由于  $\beta(a) = \beta(b) = \alpha$  - 且  $a < \frac{C_1+C_2}{2} < b$  - 所以在区间  $[a, b]$  内,  $\beta(\theta) \leq \alpha$

4° 证明渐近性质:  $\beta(\theta) \rightarrow 1$ , 当  $|\theta| \rightarrow \infty$

证明: 由标准正态分布函数的性质:

$$\lim_{x \rightarrow -\infty} \Phi(x) = 0, \quad \lim_{x \rightarrow \infty} \Phi(x) = 1$$

当  $\theta \rightarrow \infty$  时:

$$\frac{C_2 - \theta}{\sigma} \rightarrow -\infty, \quad \frac{C_1 - \theta}{\sigma} \rightarrow -\infty$$

因此  $\beta(\theta) \rightarrow 1$

当  $\theta \rightarrow -\infty$  时:

$$\frac{C_2 - \theta}{\sigma} \rightarrow \infty, \quad \frac{C_1 - \theta}{\sigma} \rightarrow \infty$$

因此也有  $\beta(\theta) \rightarrow 1$

这说明当  $\theta$  的真值与原假设距离愈来愈远时, 本检验以愈来愈确定的把握否定  $H_0$ 。

5° 对于多个样本的情况: 若  $X_1, \dots, X_n$  为取自  $N(\theta, \sigma^2)$  的样本,  $\sigma$  已知, 则:

(1) 由于  $\bar{X} \sim N(\theta, \frac{\sigma^2}{n})$ , 我们可以构造检验统计量:

$$\bar{X} \sim N(\theta, \frac{\sigma^2}{n})$$

(2) 将上述单个样本的结果应用到  $\bar{X}$  上, 接受域为:

$$C_1 \leq \bar{X} \leq C_2$$

(3) 相应的功率函数为:

$$\beta(\theta) = 1 - \left[ \Phi\left(\frac{C_2 - \theta}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{C_1 - \theta}{\sigma/\sqrt{n}}\right) \right]$$

(4)  $C_1, C_2$  的值可以通过相同的方程组求解, 只需将  $\sigma$  替换为  $\sigma/\sqrt{n}$ :

$$\Phi\left(\frac{C_2 - a}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{C_1 - a}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

$$\Phi\left(\frac{C_2 - b}{\sigma/\sqrt{n}}\right) - \Phi\left(\frac{C_1 - b}{\sigma/\sqrt{n}}\right) = 1 - \alpha$$

注意: 随着样本量  $n$  的增加, 检验的功率会提高, 这体现在  $\sigma/\sqrt{n}$  的减小上。

## 1.5.2 2

2. 设  $X_1, \dots, X_n$  是取自指数分布总体的样本,  $0 < a < b, a, b$  为已知常数. 要检验原假设  $H_0: a \leq \lambda \leq b$ . 描述一下 (不须详细推导) 用解第 1 题的思想来解这个问题的过程.

解答:

Step 1. 基本性质

(1) 样本  $X_1, \dots, X_n$  来自指数分布, 其密度函数为:

$$f(x; \lambda) = \lambda e^{-\lambda x}, x > 0, \lambda > 0$$

(2) 似然函数为:

$$L(\lambda) = \prod_{i=1}^n \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}$$

(3) 对数似然函数:

$$\ln L(\lambda) = n \ln \lambda - \lambda \sum_{i=1}^n x_i$$

(4) 由充分性定理可知,  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  是  $\lambda$  的充分统计量

Step 2. 分布性质推导

(1) 对于指数分布, 有性质:  $2\lambda X_i \sim \chi_2^2$

(2) 由独立性:  $2\lambda \sum_{i=1}^n X_i \sim \chi_{2n}^2$

(3) 因此:  $2n\lambda\bar{X} \sim \chi_{2n}^2$

Step 3. 检验统计量构造

(1) 对于原假设  $H_0: a \leq \lambda \leq b$ , 考虑:

$$\frac{2na\bar{X}}{2n} \leq \frac{2n\lambda\bar{X}}{2n} \leq \frac{2nb\bar{X}}{2n}$$

(2) 简化为:

$$a\bar{X} \leq \frac{\chi_{2n}^2}{2n} \leq b\bar{X}$$

Step 4. 检验方法确定

依据似然比检验的思想, 构造检验形式为:

$$\text{当 } c_1 \leq \bar{X} \leq c_2 \text{ 时接受 } H_0$$

Step 5. 临界值确定

(1) 由  $2n\lambda\bar{X} \sim \chi_{2n}^2$ , 令:

$$P(c_1 \leq \bar{X} \leq c_2 | \lambda = a) = P(c_1 \leq \bar{X} \leq c_2 | \lambda = b) = \alpha$$

(2) 当  $\lambda = a$  时:

$$P(2nac_1 \leq \chi_{2n}^2 \leq 2nac_2) = 1 - \alpha$$

(3) 当  $\lambda = b$  时:

$$P(2nbc_1 \leq \chi_{2n}^2 \leq 2nbc_2) = 1 - \alpha$$

(4) 解得:

$$c_1 = \frac{\chi_{2n}^2(1-\alpha/2)}{2na}, \quad c_2 = \frac{\chi_{2n}^2(\alpha/2)}{2nb}$$

Step 6. 检验效能分析

(1) 检验的功效函数:

$$\beta(\lambda) = P(c_1 \leq \bar{X} \leq c_2 | \lambda)$$

(2) 第一类错误概率:

$$\alpha = 1 - P(c_1 \leq \bar{X} \leq c_2 | \lambda), \quad \lambda = a \text{ 或 } b$$

(3) 检验的置信水平:  $1 - \alpha$

结论:

- 检验统计量:  $\bar{X}$
- 接受域:  $[c_1, c_2]$
- 临界值:  $c_1 = \frac{\chi_{2n}^2(1-\alpha/2)}{2na}, c_2 = \frac{\chi_{2n}^2(\alpha/2)}{2nb}$
- 显著性水平:  $\alpha$

## § 1.6 11.13

### 1.6.1 1

### 问题 6

设  $X_1, \dots, X_n$  为抽自具参数为  $\lambda_1$  的指数分布的样本,  $Y_1, \dots, Y_m$  为抽自具参数为  $\lambda_2$  的指数分布的样本。作出原假设  $H_0: \lambda_1 \leq \lambda_2$  的水平  $\alpha$  的检验。

### 解答

(1) 首先, 对于指数分布, 我们知道:

$$\bar{X} \sim \frac{1}{2n\lambda_1} \chi^2(2n), \quad \bar{Y} \sim \frac{1}{2m\lambda_2} \chi^2(2m)$$

(2) 根据题目提示, 可以利用:

$$\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \sim F_{2n, 2m}$$

(3) 在原假设  $H_0: \lambda_1 \leq \lambda_2$  下, 有:

$$\frac{\bar{X}}{\bar{Y}} \leq F_{2n, 2m}(\alpha)$$

(4) 因此, 检验统计量为:

$$T = \frac{\bar{X}}{\bar{Y}}$$

(5) 拒绝域为:

$$W = \{T : T > F_{2n, 2m}(\alpha)\}$$

其中,  $F_{2n, 2m}(\alpha)$  为自由度为  $(2n, 2m)$  的  $F$  分布的上  $\alpha$  分位数。

(6) 结论: 当观察值  $t > F_{2n, 2m}(\alpha)$  时, 拒绝原假设  $H_0$ 。当观察值  $t \leq F_{2n, 2m}(\alpha)$  时, 接受原假设  $H_0$ 。

补充说明:

$$F_{2n, 2m}(1 - \alpha) = \frac{1}{F_{2m, 2n}(\alpha)}$$

## 拒绝域的推导过程

### 1. 基本假设

首先, 我们有:

- 原假设  $H_0: \lambda_1 \leq \lambda_2$
- 备择假设  $H_1: \lambda_1 > \lambda_2$

### 2. 已知条件

题目给出的重要关系:

$$\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \sim F_{2n, 2m}$$

### 3. 推导过程

在原假设  $H_0: \lambda_1 \leq \lambda_2$  条件下:

$$\frac{\lambda_2}{\lambda_1} \geq 1$$

因此:

$$\frac{\bar{X}}{\bar{Y}} = \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \cdot \frac{\lambda_2}{\lambda_1} \geq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}}$$

这意味着在原假设下:

$$\begin{aligned} P\left(\frac{\bar{X}}{\bar{Y}} > F_{2n,2m}(\alpha)\right) &\leq P\left(\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} > F_{2n,2m}(\alpha)\right) \\ &= \alpha \end{aligned}$$

### 4. 拒绝域的确定

基于上述推导, 我们可以确定拒绝域为:

$$W = \{T : T > F_{2n,2m}(\alpha)\}$$

其中:

- $T = \frac{\bar{X}}{\bar{Y}}$  为检验统计量
- $F_{2n,2m}(\alpha)$  为自由度为  $(2n, 2m)$  的  $F$  分布的上  $\alpha$  分位数

### 5. 拒绝域的合理性

这个拒绝域的构造具有以下特点:

1. 检验的显著性水平为  $\alpha$
2. 当原假设为真时, 犯第一类错误的概率不超过  $\alpha$
3. 拒绝域的方向与备择假设  $H_1: \lambda_1 > \lambda_2$  相一致

### 6. 关键点说明

构造这个拒绝域的关键在于利用了不等式:

$$\frac{\bar{X}}{\bar{Y}} \geq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \quad (\text{当 } \lambda_1 \leq \lambda_2 \text{ 时成立})$$

这个不等式保证了在原假设下, 我们构造的检验统计量的分布是可控的, 从而可以准确控制第一类错误的概率。

## 概率不等式推导过程

### 1. 基本条件

在原假设  $H_0: \lambda_1 \leq \lambda_2$  下, 我们已经得到:

$$\frac{\bar{X}}{\bar{Y}} \geq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}}$$

## 2. 概率不等式基本性质

对于任意随机变量  $X, Y$  和常数  $c$ , 若  $X \geq Y$ , 则:

$$P(X > c) \leq P(Y > c)$$

这是概率的单调性性质。

## 3. 应用到本题

将上述性质应用到我们的问题中:

- 令  $X = \frac{\bar{X}}{\bar{Y}}$
- 令  $Y = \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}}$
- 令  $c = F_{2n, 2m}(\alpha)$

## 4. 已知分布

根据题目给出的条件:

$$\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} \sim F_{2n, 2m}$$

由  $F$  分布的定义, 对其上  $\alpha$  分位点  $F_{2n, 2m}(\alpha)$ , 有:

$$P\left(\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} > F_{2n, 2m}(\alpha)\right) = \alpha$$

## 5. 最终结论

结合以上条件, 我们可以得到:

$$\begin{aligned} P\left(\frac{\bar{X}}{\bar{Y}} > F_{2n, 2m}(\alpha)\right) &\leq P\left(\frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}} > F_{2n, 2m}(\alpha)\right) \\ &= \alpha \end{aligned}$$

## 6. 结论说明

这个不等式的意义在于:

- 在原假设  $H_0$  下, 检验统计量  $\frac{\bar{X}}{\bar{Y}}$  超过临界值  $F_{2n, 2m}(\alpha)$  的概率不超过  $\alpha$
- 这保证了我们构造的检验的显著性水平确实为  $\alpha$
- 这验证了我们构造的拒绝域是合理的

## 7. 关键点总结

整个推导的核心是利用了两个重要性质:

1. 在原假设下的不等式关系:  $\frac{\bar{X}}{\bar{Y}} \geq \frac{\lambda_1 \bar{X}}{\lambda_2 \bar{Y}}$
2. 概率的单调性: 当  $X \geq Y$  时,  $P(X > c) \leq P(Y > c)$

这两个性质共同保证了我们能够准确控制第一类错误的概率。

## 1.6.2 2

## 7

## 题目

设  $X_1, \dots, X_n$  是取自均匀分布  $R(0, \theta)$  的样本, 给定  $\theta_0 > 0$ . 作出原假设  $H_0: \theta \leq \theta_0$  的水平  $\alpha$  检验。

## 解答

1. 记  $T = \max(X_1, \dots, X_n)$ . 从直观上看,  $\theta$  愈大,  $T$  也愈倾向于取大值。故一个合理的检验为: 当  $T \leq C$  时接受  $H_0$ , 不然就否定  $H_0$ 。
2. 为确定  $C$ , 计算其功效函数:

$$\begin{aligned}\beta(\theta) &= P(T > C) \\ &= 1 - P(T \leq C) \\ &= 1 - P(\max(X_1, \dots, X_n) \leq C) \\ &= 1 - P(X_1 \leq C, \dots, X_n \leq C) \\ &= 1 - \prod_{i=1}^n P(X_i \leq C) \\ &= 1 - (C/\theta)^n\end{aligned}$$

3. 它是  $\theta$  的增函数, 故为使  $\beta(\theta) \leq \alpha$  当  $\theta \leq \theta_0$ , 只须使  $\beta(\theta_0) = \alpha$  即可。
4. 由此得出:

$$\begin{aligned}\beta(\theta_0) &= \alpha \\ 1 - (C/\theta_0)^n &= \alpha \\ (C/\theta_0)^n &= 1 - \alpha \\ C &= (1 - \alpha)^{1/n} \theta_0\end{aligned}$$

5. 因此, 检验的形式为:

- 当  $T \leq (1 - \alpha)^{1/n} \theta_0$  时, 接受  $H_0$
- 当  $T > (1 - \alpha)^{1/n} \theta_0$  时, 否定  $H_0$

6. 这个检验的功效函数为:

$$\beta(\theta) = 1 - \left( \frac{(1 - \alpha)^{1/n} \theta_0}{\theta} \right)^n = 1 - (1 - \alpha) \left( \frac{\theta_0}{\theta} \right)^n$$

## 1.6.3 3

**问题 8.** 设  $X_1, \dots, X_n$  是从有下述密度函数的总体中抽出的样本:

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

给定常数  $\theta_0$ ，作出原假设  $H_0: \theta \leq \theta_0$  的水平  $\alpha$  检验。

**解答:** 从  $f(x, \theta)$  的图形可以看出, 观察值  $X_1, \dots, X_n$  落在  $\theta$  附近的可能性大, 所以  $T = \min(X_1, \dots, X_n)$  接近  $\theta$  且包含了  $\theta$  较多的信息。显然, 当  $\theta$  大时,  $T$  倾向于大。故  $H_0$  的一个直观上合理的检验是: 当  $T \leq C$  时接受  $H_0$ , 不然就否定  $H_0$ 。

为要根据水平  $\alpha$  决定  $C$ , 要算出  $T$  的分布。令  $X'_i = X_i - \theta, i = 1, \dots, n$ 。则易见  $X'_i$  有指数密度  $e^{-x}$  当  $(x > 0, x \leq 0 \text{ 时为 } 0)$ 。从此出发用第二章第 22 题, 易得  $T' = \min(X'_1, \dots, X'_n)$  的密度函数为  $ne^{-nx}$  (当  $x > 0, x \leq 0$  时为 0)。

由于  $T = T' + \theta$ , 得出  $T$  的密度函数  $g(x, \theta)$  为

$$g(x, \theta) = \begin{cases} ne^{-n(x-\theta)}, & x > \theta \\ 0, & x \leq \theta \end{cases}$$

因此上述检验的功效函数为

$$\beta(\theta) = P_\theta(T > C) = \int_{\max(c, \theta)}^{\infty} ne^{-n(x-\theta)} dx = e^{-n(\max(c, \theta) - \theta)}$$

此为  $\theta$  的增函数 (何故?) 故为使  $\beta(\theta) \leq \alpha$  当  $\theta \leq \theta_0$ , 只须使  $\beta(\theta_0) = \alpha$ 。这定出

$$C = \theta_0 + \frac{1}{n} \log \left( \frac{1}{\alpha} \right)$$

**推导过程:** 求  $T' = \min(X'_1, \dots, X'_n)$  的密度函数

1) 首先, 我们知道每个  $X'_i$  的密度函数为:

$$f_{X'_i}(x) = \begin{cases} e^{-x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

2) 计算  $T'$  的分布函数。对于任意  $t > 0$ :

$$\begin{aligned} F_{T'}(t) &= P(T' \leq t) \\ &= 1 - P(T' > t) \\ &= 1 - P(\min(X'_1, \dots, X'_n) > t) \\ &= 1 - P(X'_1 > t, \dots, X'_n > t) \\ &= 1 - \prod_{i=1}^n P(X'_i > t) \\ &= 1 - \prod_{i=1}^n [1 - P(X'_i \leq t)] \\ &= 1 - \prod_{i=1}^n [1 - (1 - e^{-t})] \\ &= 1 - (e^{-t})^n \\ &= 1 - e^{-nt} \end{aligned}$$

3) 求密度函数, 对分布函数求导:

$$\begin{aligned} f_{T'}(t) &= \frac{d}{dt} F_{T'}(t) \\ &= \frac{d}{dt} (1 - e^{-nt}) \\ &= ne^{-nt} \end{aligned}$$



4) 因此,  $T'$  的密度函数为:

$$f_{T'}(t) = \begin{cases} ne^{-nt}, & t > 0 \\ 0, & t \leq 0 \end{cases}$$

注: 这个结果显示  $T'$  服从参数为  $n$  的指数分布。这是合理的, 因为最小值的指数分布的参数应该是各个独立指数分布参数的和。

## § 1.7 11.18

### 1.7.1 1

### 题目

设  $X_1, \dots, X_n$  和  $Y_1, \dots, Y_m$  分别是来自正态总体  $N(a, \sigma^2)$  和  $N(b, \sigma^2)$  的样本,  $a, b, \sigma^2$  都未知。试仿照两样本  $t$  检验的做法, 构造出原假设  $H_0$ :

$a = cb$  的一个水平  $\alpha$  检验。这里  $c \neq 0$  为已知常数。

已知  $\bar{X} - c\bar{Y} \sim N(a - cb, \frac{m+nc^2}{mn}\sigma^2)$  仿照两样本  $t$  检验的得出过程, 作统计量

$$T = \frac{\bar{X} - c\bar{Y}}{\sqrt{\frac{m+nc^2}{mn(m+n-2)} [\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2]}}$$

而得出当  $H_0$  成立时  $T \sim t_{m+n-2}$ 。由此得出检验: 当  $|T| \leq t_{m+n-2}(\alpha/2)$  时接受  $H_0$ , 不然就否定  $H_0$ 。

### 解答

#### 解析过程:

1. 题目给出的条件:

- 两个独立样本:

$$X_1, \dots, X_n \sim N(a, \sigma^2)$$

$$Y_1, \dots, Y_m \sim N(b, \sigma^2)$$

- 参数  $a, b, \sigma^2$  均未知

- $c \neq 0$  为已知常数

2. 原假设:

$$H_0 : a = cb$$

3. 已知条件:

$$\bar{X} - c\bar{Y} \sim N(a - cb, \frac{m+nc^2}{mn}\sigma^2)$$

4. 构造检验统计量:

$$T = \frac{\bar{X} - c\bar{Y}}{\sqrt{\frac{m+nc^2}{mn(m+n-2)} [\sum_{i=1}^n (X_i - \bar{X})^2 + \sum_{j=1}^m (Y_j - \bar{Y})^2]}}$$

5. 在原假设  $H_0$  成立时:

$$T \sim t_{m+n-2}$$

6. 检验法则:

- 当  $|T| \leq t_{m+n-2}(\alpha/2)$  时, 接受  $H_0$
- 当  $|T| > t_{m+n-2}(\alpha/2)$  时, 否定  $H_0$

说明:

这是一个双侧检验问题。检验统计量  $T$  的构造遵循了经典的两样本  $t$  检验的思路:

- 分子是  $\bar{X} - c\bar{Y}$ , 表示样本均值之差
- 分母是标准差的估计, 用于标准化处理
- 在  $H_0$  成立时, 该统计量服从自由度为  $m+n-2$  的  $t$  分布
- 临界值  $t_{m+n-2}(\alpha/2)$  对应双侧检验的显著性水平  $\alpha$

## t 分布与正态分布的关系

### 1. 基本定义

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

$$T = \frac{Z}{\sqrt{Y/n}}$$

的分布称为自由度为  $n$  的  $t$  分布, 记为  $T \sim t(n)$ 。

### 2. 主要性质

1.  $t$  分布的概率密度函数:

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

2. 基本特征:

- 对称性:  $t$  分布是关于  $y$  轴对称的
- 均值:  $E(T) = 0$  (当  $n > 1$  时)
- 方差:  $Var(T) = \frac{n}{n-2}$  (当  $n > 2$  时)

3. 与正态分布的关系:

- 当  $n \rightarrow \infty$  时,  $t(n)$  分布趋近于标准正态分布  $N(0, 1)$
- $t$  分布比标准正态分布有更重的尾部
- 当自由度  $n$  较小时,  $t$  分布的尾部更厚, 随着  $n$  增大, 尾部逐渐变薄

### 3. 在统计推断中的应用

1. 单个正态总体均值的检验：当总体标准差  $\sigma$  未知时，检验统计量

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

其中， $\bar{X}$  为样本均值， $S$  为样本标准差。

2. 两个正态总体均值差的检验：当总体标准差相等但未知时，检验统计量

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

其中， $S_p$  为合并样本标准差。

3. 置信区间的构造：总体均值  $\mu$  的  $(1 - \alpha)$  置信区间为

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

### 4. 重要结论

1. t 分布是对正态分布的一种修正，用于处理小样本情况
2. 当样本量增大时，t 分布逐渐接近正态分布
3. t 分布比正态分布更保守，这体现在：
  - 置信区间更宽
  - 检验的把握度更大
  - 对数据分布的偏离较不敏感

#### 1.7.2 2

13. 设样本  $X \sim B(n_1, p_1)$ ,  $Y \sim B(n_2, p_2)$ 。要检验假设  $H_0: p_1 = p_2$ 。设  $n_1$  和  $n_2$  都充分大，试作出  $H_0$  的水平  $\alpha$  的大样本检验。

**解：**当  $n_1, n_2$  充分大时有

$$(X - n_1 p_1) / \sqrt{n_1 p_1 (1 - p_1)} \sim N(0, 1)$$

$$(Y - n_2 p_2) / \sqrt{n_2 p_2 (1 - p_2)} \sim N(0, 1)$$

故近似地有：

$$X/n_1 \sim N(p_1, p_1(1-p_1)/n_1)$$

$$Y/n_2 \sim N(p_2, p_2(1-p_2)/n_2)$$

因而近似地也有：

$$Z \equiv X/n_1 - Y/n_2 \sim N(p_1 - p_2, \sigma^2)$$

其中  $\sigma^2 = p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$ 。如  $\sigma^2$  已知，则检验  $p_1 - p_2 = 0$  相当于检验正态变量  $Z$  之均值为 0，其否定域应取为：

$$|Z| > \sigma u_{\alpha/2}$$

现  $\sigma^2$  未知, 可以用

$$\hat{\sigma}^2 = \hat{p}_1(1 - \hat{p}_1)/n_1 + \hat{p}_2(1 - \hat{p}_2)/n_2$$

去估计之,  $\hat{p}_1 = X/n_1$ ,  $\hat{p}_2 = Y/n_2$ 。

最后得出  $H_0: p_1 = p_2$  的大样本检验的否定域为:

$$|X/n_1 - Y/n_2| > u_{\alpha/2} [(X/n_1)(1 - X/n_1)/n_1 + (Y/n_2)(1 - Y/n_2)/n_2]^{1/2}$$

### 1.7.3 3

**题目 16.** 设变量  $X$  取 1, 2, 3, 4 等值. 有一种理论认为,  $X$  取这 4 个值的概率呈等比级数, 即

$$\frac{P(X=2)}{P(X=1)} = \frac{P(X=3)}{P(X=2)} = \frac{P(X=4)}{P(X=3)}$$

为验证此理论是否正确, 对  $X$  进行  $n$  次观察, 发现  $X$  取 1, 2, 3, 4 为值分别有  $n_1, n_2, n_3, n_4$  次. 试作拟合优度检验, 描述步骤即可以, 不必去解方程.

记题中之公共比值为  $\theta$ , 则易见

$$P(X=i) = \frac{\theta^{i-1}}{1 + \theta + \theta^2 + \theta^3}, i = 1, 2, 3, 4$$

于是得似然函数

$$L(\theta) = \prod_{i=1}^4 [P(X=i)]^{n_i} = \frac{\theta^{n_2+2n_3+3n_4}}{(1 + \theta + \theta^2 + \theta^3)^n}$$

由此得到决定  $\theta$  值的方程  $\frac{d(\log L(\theta))}{d\theta} = 0$ , 即

$$(n_2 + 2n_3 + 3n_4)/\theta - n(1 + 2\theta + 3\theta^2)/(1 + \theta + \theta^2 + \theta^3) = 0$$

遍乘  $\theta(1 + \theta + \theta^2 + \theta^3)$ , 得到  $\theta$  的一个 3 次方程, 它有公式求解. 如有多于一个实根, 还须逐一代入  $L(\theta)$  中, 看哪一个达到最大, 这一个就取为  $\theta$  的估计值  $\hat{\theta}$ . 因只有一个参数  $\theta$ , 自由度应为  $4 - 1 - 1 = 2$ .

**自由度计算:**

- 有 4 个类别:  $X$  可取值  $\{1, 2, 3, 4\}$
- 概率和限制:  $\sum_{i=1}^4 P(X=i) = 1$  损失 1 个自由度
- 参数约束: 需估计参数  $\theta$ , 再损失 1 个自由度

因此, 自由度为  $4 - 1 - 1 = 2$ 。

## § 1.8 11.27

### 1.8.1 1

#### 题目

在模型 (2.6) 中, 假设 (2.5) 成立, 仍记残差为  $\delta_1, \dots, \delta_n$ . 证明以下各点:

- (a)  $E(\delta_i) = 0, i = 1, \dots, n$ .

### 1.8.2 解答

(a) 利用  $\hat{\beta}_0, \hat{\beta}_1$  的无偏性, 因为:

$$\begin{aligned}\delta_i &= Y_i - \hat{Y}_i \\ &= \beta_0 + \beta_1(X_i - \bar{X}) + e_i - (\hat{\beta}_0 + \hat{\beta}_1(X_i - \bar{X})) \\ &= (\beta_0 - \hat{\beta}_0) + (\beta_1 - \hat{\beta}_1)(X_i - \bar{X}) + e_i\end{aligned}$$

且  $E(e_i) = 0$ , 即得  $E(\delta_i) = 0$ 。

### 1.8.3 背景知识

回顾本题中涉及的重要概念:

- 线性回归模型:  $Y_i = \beta_0 + \beta_1(X_i - \bar{X}) + e_i, i = 1, \dots, n$
- 误差项假设:  $E(e_i) = 0, \text{Var}(e_i) = \sigma^2, i = 1, \dots, n$
- 残差定义:  $\delta_i = Y_i - \hat{Y}_i$
- 中心化处理: 使用  $(X_i - \bar{X})$  代替  $X_i$

### 1.8.4 证明要点说明

本题的核心证明步骤是:

1. 首先写出残差  $\delta_i$  的完整表达式
2. 利用  $\hat{\beta}_0, \hat{\beta}_1$  的无偏性
3. 利用误差项  $e_i$  的期望为 0 的性质
4. 综合上述条件得到结论

这个证明展示了线性回归中残差的一个重要性质: 残差的期望为 0, 这说明残差是对真实误差的无偏估计。

### 1.8.5 模型介绍与中心化处理

考虑一元线性回归模型:

$$Y_i = b_0 + b_1 X_i + e_i, i = 1, \dots, n$$

其中:

- $b_0$  为常数项 (截距)
- $b_1$  为回归系数
- $e_i$  为随机误差, 满足:  $E(e_i) = 0, \text{Var}(e_i) = \sigma^2$

将此模型进行中心化处理, 得到:

$$Y_i = \beta_0 + \beta_1(X_i - \bar{X}) + e_i, i = 1, \dots, n$$

其中:

$$\beta_1 = b_1, \beta_0 = b_0 + b_1 \bar{X}$$

## § 1.9 12.02

### 1.9.1 1

### 题目

Show that in simple linear regression,

$$\text{Coefficient of Determination} = (\text{Coefficient of correlation})^2$$

Where Coefficient of Determination is defined as:

$$\text{Coefficient of Determination} = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

证明在简单线性回归中，

$$\text{决定系数} = (\text{相关系数})^2$$

其中决定系数定义为：

$$\text{决定系数} = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

证明. 首先回顾一些基本定义：

First, let's recall some basic definitions:

1) 相关系数 (Correlation coefficient):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

2) 在简单线性回归中，回归线方程为 (In simple linear regression, the regression line equation is):

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

其中 (where):

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

现在我们开始证明 (Now let's begin the proof):

$$SS_{\text{total}} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$SS_{\text{err}} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

根据决定系数的定义 (According to the definition of coefficient of determination):

$$R^2 = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}}$$

可以证明 (It can be shown that):

$$SS_{\text{total}} - SS_{\text{err}} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

因此 (Therefore):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

代入回归方程 (Substituting the regression equation):

$$R^2 = \frac{\hat{\beta}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

代入  $\hat{\beta}_1$  的表达式 (Substituting the expression for  $\hat{\beta}_1$ ):

$$R^2 = \frac{[\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})]^2}{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}$$

这正好是相关系数的平方 (This is exactly the square of the correlation coefficient):

$$R^2 = (r_{xy})^2$$

因此, 我们证明了决定系数等于相关系数的平方。

Thus, we have proved that the coefficient of determination equals the square of the correlation coefficient.  $\square$

**背景知识:**

1)

$$S_{xy}$$

是 x 和 y 的离差乘积和 (sum of cross-products of deviations):

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

2)

$$S_{xx}$$

是 x 的离差平方和 (sum of squares of deviations for x):

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

类似地, 也有:

3)

$$S_{yy}$$

是 y 的离差平方和:

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$$

使用这些符号, 我们可以更简洁地表示一些重要公式:

1) 相关系数 (correlation coefficient):

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

2) 回归系数 (regression coefficient):

$$\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$$

3) 决定系数 (coefficient of determination):

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

这些符号在统计推断和回归分析中经常使用，能让公式表达更加简洁明了。需要注意的是，这些都是针对样本的计算公式，而不是总体参数。

**决定系数 ( $R^2$ ) 的等价表示:**

1) 使用平方和 (Using sum of squares):

$$R^2 = \frac{SS_{\text{reg}}}{SS_{\text{total}}} = \frac{SS_{\text{total}} - SS_{\text{err}}}{SS_{\text{total}}} = 1 - \frac{SS_{\text{err}}}{SS_{\text{total}}}$$

2) 使用相关系数 (Using correlation coefficient):

$$R^2 = r_{xy}^2 = \left( \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}} \right)^2$$

3) 使用实际值与预测值 (Using actual and predicted values):

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

4) 使用回归系数 (Using regression coefficient):

$$R^2 = \hat{\beta}_1^2 \cdot \frac{S_{xx}}{S_{yy}}$$

其中 (where):

$$\begin{aligned} SS_{\text{total}} &= \sum_{i=1}^n (y_i - \bar{y})^2 = S_{yy} \\ SS_{\text{reg}} &= \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \\ SS_{\text{err}} &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \end{aligned}$$

**注:** 所有这些表达式在简单线性回归中都是等价的。不同的表示形式在不同的分析场景下可能更为方便。

## 1.9.2 2

### 题目

假设  $Y_i = b_0 + b_1 x_i + e_i$ ，其中  $e_1, e_2, \dots, e_n$  独立同分布，且  $e_i \sim N(0, \sigma^2)$ 。  
找到  $b_0$  和  $b_1$  的最大似然估计并验证它们是最小二乘估计。

### 解答过程

首先，写出似然函数：

$$L(b_0, b_1, \sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(Y_i - b_0 - b_1 x_i)^2}{2\sigma^2}\right)$$



取对数得到对数似然函数：

$$\ell(b_0, b_1, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i)^2$$

对  $b_0$  和  $b_1$  求偏导数并令其等于零：

$$\frac{\partial \ell}{\partial b_0} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial \ell}{\partial b_1} = \frac{1}{\sigma^2} \sum_{i=1}^n (Y_i - b_0 - b_1 x_i) x_i = 0$$

解这两个方程可以得到  $b_0$  和  $b_1$  的最大似然估计：

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{Y} - b_1 \bar{x}$$

这正是最小二乘估计的公式，因此最大似然估计和最小二乘估计是一致的。

## § 1.10 12.04

### 1.10.1 1

在回归分析中， $SS_{model}$ （也称为  $SS_{reg}$  或 SSR, Sum of Squares due to Regression）表示由模型解释的方差平方和。 $SS_{reg}$  反映了自变量对因变量的解释程度，即回归方程对观测值的拟合程度。让我们从多个角度来理解它：

1. 数学表达式：

$$SS_{model} = SS_{reg} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

其中：

- $\hat{y}_i$  是模型预测值（回归值）
- $\bar{y}$  是因变量的平均值
- $SS_{reg}$  度量了回归值与因变量均值之间的变异

2. 几何解释：

- $SS_{reg}$  代表回归直线上的点与因变量均值之间的偏差平方和
- 它衡量了回归模型能够解释的变量
- 越大的  $SS_{reg}$  意味着模型解释了更多的变异
- $SS_{reg}$  反映了自变量对因变量变化的影响程度

3. 在简单线性回归中的特殊形式：

$$SS_{reg} = \hat{b}_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$$

4. 在总变差分解中的作用:

$$SS_{total} = SS_{reg} + SS_{error}$$

其中  $SS_{error}$  表示残差平方和, 衡量了未被模型解释的变异。

5. 与  $R^2$  的关系:

$$R^2 = \frac{SS_{reg}}{SS_{total}}$$

这个关系表明  $SS_{reg}$  直接决定了模型的拟合优度,  $R^2$  值越接近 1, 表示模型的解释能力越强。

6. 在 F 检验中的应用:

$$F = \frac{SS_{reg}/df_{reg}}{SS_{error}/df_{error}}$$

这里  $SS_{reg}$  用于构建检验整个模型显著性的 F 统计量, 其中  $df_{reg}$  为回归自由度,  $df_{error}$  为误差自由度。

理解  $SS_{reg}$  对于掌握回归分析中的模型评估和假设检验非常重要, 它是衡量模型解释能力的基础指标之一。 $SS_{reg}$  越大, 表明回归方程对观测数据的拟合程度越好, 模型的预测能力越强。

## 题目描述

在简单线性回归模型中, 证明 ANOVA 中的 F 统计量是用于检验  $H_0: b_1 = 0$  与  $H_1: b_1 \neq 0$  的 T 统计量的平方。

## 证明

在简单线性回归中, 模型为:

$$Y_i = b_0 + b_1 x_i + e_i$$

假设检验  $H_0: b_1 = 0$  与  $H_1: b_1 \neq 0$ , T 统计量定义为:

$$t = \frac{b_1}{SE(b_1)}$$

其中,  $SE(b_1)$  是  $b_1$  的标准误。

ANOVA 中的 F 统计量定义为:

$$F = \frac{MS_{reg}}{MS_{err}}$$

其中,  $MS_{reg}$  是回归均方,  $MS_{err}$  是误差均方。

在简单线性回归中,  $MS_{reg}$  和  $MS_{err}$  可以表示为:

$$MS_{reg} = \frac{SS_{reg}}{1}$$

$$MS_{err} = \frac{SS_{err}}{n-2}$$

其中,  $SS_{reg}$  是回归平方和,  $SS_{err}$  是误差平方和。

由于  $SS_{reg} = b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2$ , 我们有:

$$F = \frac{b_1^2 \sum_{i=1}^n (x_i - \bar{x})^2}{\frac{SS_{err}}{n-2}}$$

注意到  $SE(b_1) = \sqrt{\frac{SS_{\text{err}}}{(n-2)\sum_{i=1}^n(x_i - \bar{x})^2}}$ , 我们可以得到:

$$t^2 = \left( \frac{b_1}{SE(b_1)} \right)^2 = \frac{b_1^2}{\frac{SS_{\text{err}}}{(n-2)\sum_{i=1}^n(x_i - \bar{x})^2}} = F$$

因此, ANOVA 中的 F 统计量是用于检验  $H_0: b_1 = 0$  与  $H_1: b_1 \neq 0$  的 T 统计量的平方。

## § 1.11 12.09

### 1.11.1 1

#### 题目

在 Deming 回归中, 当  $\delta = 1$  时, 请证明残差平方和

$$RSS = \frac{1}{1 + \beta_1^2} \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2$$

注意: 除了笔记中的方法, 也可以用直角三角形面积方法。

#### 解答

设数据点为  $(x_i, y_i), i = 1, \dots, n$ 。在 Deming 回归中, 我们需要找到每个观测点  $(x_i, y_i)$  到拟合直线  $y = \beta_1 x + \beta_0$  的垂直距离。

由于  $\delta = 1$ , 这意味着  $x$  和  $y$  的测量误差具有相同的方差。在这种情况下, 从点到直线的垂直距离可以表示为:

$$d_i = \frac{|y_i - \beta_1 x_i - \beta_0|}{\sqrt{1 + \beta_1^2}}$$

这是因为对于直线  $y = \beta_1 x + \beta_0$ , 其法向量为  $(\beta_1, -1)$ , 单位法向量为

$$\frac{1}{\sqrt{1 + \beta_1^2}}(\beta_1, -1)$$

因此, 残差平方和为:

$$\begin{aligned} RSS &= \sum_{i=1}^n d_i^2 \\ &= \sum_{i=1}^n \frac{(y_i - \beta_1 x_i - \beta_0)^2}{1 + \beta_1^2} \\ &= \frac{1}{1 + \beta_1^2} \sum_{i=1}^n (y_i - \beta_1 x_i - \beta_0)^2 \end{aligned}$$

得证。

**备注:** 这个结果也可以通过直角三角形面积方法得到。对于每个点到直线的垂直距离, 可以构造一个直角三角形, 其中斜边长度为  $|y_i - \beta_1 x_i - \beta_0|$ , 而垂直距离  $d_i$  与斜边的关系可以通过角度  $\theta = \arctan(\beta_1)$  得到:

$$d_i = |y_i - \beta_1 x_i - \beta_0| \cos(\theta) = \frac{|y_i - \beta_1 x_i - \beta_0|}{\sqrt{1 + \beta_1^2}}$$

这与代数方法得到的结果一致。

## 背景知识

在 Deming 回归中,  $\delta$  表示  $x$  和  $y$  方向测量误差方差之比:

$$\delta = \frac{\sigma_{\varepsilon}^2}{\sigma_{\eta}^2}$$

其中:

- $\sigma_{\varepsilon}^2$  是  $y$  方向测量误差的方差
- $\sigma_{\eta}^2$  是  $x$  方向测量误差的方差

因此, 当  $\delta = 1$  时, 这意味着:

$$\sigma_{\varepsilon}^2 = \sigma_{\eta}^2$$

这表示:

1.  $x$  和  $y$  方向的测量误差具有相同的方差
2. 在几何意义上, 这意味着我们需要最小化观测点到拟合直线的垂直距离平方和
3. 这种情况下, Deming 回归等价于正交回归 (Orthogonal Regression)

这就是为什么在这种情况下, 残差平方和公式中会出现  $\frac{1}{1+\beta_1^2}$  这个项, 它反映了到直线的垂直距离。

### 1.11.2 2

#### 题目描述

设  $\rho$  和  $r$  分别为相关系数和皮尔逊相关系数。使用 Fisher 变换, 我们已经得到了  $z(\rho) = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$  的渐近  $1-\alpha$  置信区间为  $[\zeta_{\text{lower}}, \zeta_{\text{upper}}]$ , 其中

$$\zeta_{\text{lower}} = z(r) - z_{(1-\alpha/2)} \sqrt{n - \frac{1}{3}}, \quad \zeta_{\text{upper}} = z(r) + z_{(1-\alpha/2)} \sqrt{n - \frac{1}{3}}$$

请进一步证明  $\rho$  的渐近置信区间为:

$$\left[ \frac{\exp\{2\zeta_{\text{lower}}\} - 1}{\exp\{2\zeta_{\text{lower}}\} + 1}, \frac{\exp\{2\zeta_{\text{upper}}\} - 1}{\exp\{2\zeta_{\text{upper}}\} + 1} \right]$$

## 背景知识

设  $X$  和  $Y$  是两个随机变量:

- $\rho$  表示总体相关系数 (population correlation coefficient)
- $r$  表示样本皮尔逊相关系数 (sample Pearson correlation coefficient)
- $n$  表示样本量 (sample size)

Fisher 变换定义为:

$$z(\rho) = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$$

该变换具有如下性质:

- 当  $n$  较大时,  $z(r)$  近似服从正态分布
- $E[z(r)] \approx z(\rho)$
- $Var[z(r)] \approx \frac{1}{n-3}$

## 题目描述

设  $\rho$  和  $r$  分别为相关系数和皮尔逊相关系数。使用 Fisher 变换, 我们已经得到了  $z(\rho)$  的渐近  $1 - \alpha$  置信区间为  $[\zeta_{\text{lower}}, \zeta_{\text{upper}}]$ , 其中

$$\zeta_{\text{lower}} = z(r) - z_{(1-\alpha/2)} \sqrt{\frac{1}{n-3}}, \quad \zeta_{\text{upper}} = z(r) + z_{(1-\alpha/2)} \sqrt{\frac{1}{n-3}}$$

请进一步证明  $\rho$  的渐近置信区间为:

$$\left[ \frac{\exp\{2\zeta_{\text{lower}}\} - 1}{\exp\{2\zeta_{\text{lower}}\} + 1}, \frac{\exp\{2\zeta_{\text{upper}}\} - 1}{\exp\{2\zeta_{\text{upper}}\} + 1} \right]$$

## 证明

证明. 根据 Fisher 变换, 我们有:

$$z(\rho) = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$$

我们已经得到了  $z(\rho)$  的置信区间  $[\zeta_{\text{lower}}, \zeta_{\text{upper}}]$ 。

为了得到  $\rho$  的置信区间, 我们需要求  $z(\rho)$  的反函数。记为  $z^{-1}$ 。

设  $z = z(\rho)$ , 则:

$$z = \frac{1}{2} \log_e \frac{1+\rho}{1-\rho}$$

因此:

$$2z = \log_e \frac{1+\rho}{1-\rho}$$

$$e^{2z} = \frac{1+\rho}{1-\rho}$$

$$e^{2z}(1-\rho) = 1+\rho$$

$$e^{2z} - e^{2z}\rho = 1+\rho$$

$$e^{2z} + 1 = \rho(e^{2z} + 1)$$

解得:

$$\rho = \frac{e^{2z} - 1}{e^{2z} + 1}$$

这就是 Fisher 变换的反函数。将  $z$  分别替换为  $\zeta_{\text{lower}}$  和  $\zeta_{\text{upper}}$ , 即可得到  $\rho$  的置信区间:

$$\left[ \frac{\exp\{2\zeta_{\text{lower}}\} - 1}{\exp\{2\zeta_{\text{lower}}\} + 1}, \frac{\exp\{2\zeta_{\text{upper}}\} - 1}{\exp\{2\zeta_{\text{upper}}\} + 1} \right]$$

□

## § 1.12 12.11

### 1.12.1 1

#### 题目描述

假设  $X_1$  和  $X_2$  是方差为  $\sigma^2$  的不相关随机变量，使用矩阵方法证明  $Y = X_1 + X_2$  和  $Z = X_1 - X_2$  是不相关的。

#### 证明

设  $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ ，则  $\mathbf{Y} = \begin{pmatrix} Y \\ Z \end{pmatrix}$  可以表示为：

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

其中， $\mathbf{A} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ 。

首先，计算  $\mathbf{X}$  的协方差矩阵  $\Sigma_{\mathbf{X}}$ ：

$$\Sigma_{\mathbf{X}} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$

然后，计算  $\mathbf{Y}$  的协方差矩阵  $\Sigma_{\mathbf{Y}}$ ：

$$\Sigma_{\mathbf{Y}} = \mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T$$

计算  $\mathbf{A}\Sigma_{\mathbf{X}}$ ：

$$\mathbf{A}\Sigma_{\mathbf{X}} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & -\sigma^2 \end{pmatrix}$$

然后计算  $\mathbf{A}\Sigma_{\mathbf{X}}\mathbf{A}^T$ ：

$$\Sigma_{\mathbf{Y}} = \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & -\sigma^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}$$

因此， $\Sigma_{\mathbf{Y}}$  是对角矩阵，这表明  $Y$  和  $Z$  是不相关的。

#### 题目描述

假设  $X_1$  和  $X_2$  是方差为  $\sigma^2$  的不相关随机变量，使用矩阵方法证明  $Y = X_1 + X_2$  和  $Z = X_1 - X_2$  是不相关的。

#### 解答

证明. (1) 首先，我们可以将  $Y$  和  $Z$  用矩阵形式表示：

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$$

(2) 令  $A = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$ ， $\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$

(3) 由于  $X_1$  和  $X_2$  不相关且方差相同，其协方差矩阵为：

$$\text{Cov}(\mathbf{X}) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} = \sigma^2 I$$

(4) 根据协方差矩阵的性质，有：

$$\text{Cov}(\mathbf{AX}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}^T$$

(5) 计算  $\mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}^T$ ：

$$\begin{aligned} \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}^T &= \sigma^2 \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \sigma^2 \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \end{aligned}$$

(6) 由上述结果可知， $Y$  和  $Z$  的协方差为 0，即它们是不相关的。

因此，我们证明了  $Y = X_1 + X_2$  和  $Z = X_1 - X_2$  是不相关的。 □

## 结论

这个证明展示了矩阵方法在处理随机变量线性组合的优越性。通过矩阵运算，我们可以清晰地看到两个新随机变量  $Y$  和  $Z$  的不相关性。同时也可以注意到，它们的方差都是原始随机变量方差的 2 倍，即  $2\sigma^2$ 。

## 补充说明：为什么对角矩阵表示不相关性

让我们详细解释为什么得到对角矩阵就意味着随机变量是不相关的。

### 1. 随机变量不相关的定义

两个随机变量  $X$  和  $Y$  不相关的充要条件是它们的协方差为零：

$$\text{Cov}(X, Y) = 0$$

### 2. 协方差矩阵的含义

对于随机向量  $\begin{pmatrix} Y \\ Z \end{pmatrix}$ ，其协方差矩阵的一般形式为：

$$\begin{pmatrix} \text{Var}(Y) & \text{Cov}(Y, Z) \\ \text{Cov}(Z, Y) & \text{Var}(Z) \end{pmatrix}$$

其中：

- 对角线元素是各个随机变量的方差
- 非对角线元素是随机变量之间的协方差
- 由于  $\text{Cov}(Y, Z) = \text{Cov}(Z, Y)$ ，该矩阵是对称矩阵

### 3. 对角矩阵的意义

在我们的证明中，我们得到了协方差矩阵：

$$\sigma^2 \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

这是一个对角矩阵，其特点是：

- 非对角元素都为 0，即  $\text{Cov}(Y, Z) = \text{Cov}(Z, Y) = 0$
- 对角元素  $2\sigma^2$  表示  $Y$  和  $Z$  的方差

#### 4. 几何解释

对角协方差矩阵具有重要的几何意义：

- 它表明随机变量在统计意义上是“正交的”
- 一个随机变量的变化不会线性地影响到另一个随机变量
- 这种正交性与向量空间中的正交概念类似

#### 5. 结论

因此，当我们通过矩阵运算得到对角矩阵时：

$$\text{Cov}(Y, Z) = 0$$

这直接证明了  $Y$  和  $Z$  是不相关的随机变量。这个结果不仅在数学上是严格的，而且在直观上也是合理的，因为它表明了这两个新的随机变量之间没有线性关系。

**注记：**需要特别指出的是，不相关性仅表示随机变量之间没有线性关系，这比独立性的要求更弱。两个独立的随机变量一定不相关，但不相关的随机变量不一定独立。

### 协方差矩阵计算的详细推导

让我们详细说明为什么协方差矩阵是这样计算的。

#### 1. 基本设定

我们有线性变换：

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \underbrace{\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}}_A \underbrace{\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}}_X$$

#### 2. 协方差矩阵的性质

对于线性变换，协方差矩阵满足以下性质：

$$\text{Cov}(AX) = A\text{Cov}(X)A^T$$

这个性质源于协方差的线性性质。

#### 3. 原始随机变量的协方差矩阵

由于  $X_1$  和  $X_2$  不相关且方差均为  $\sigma^2$ ：

$$\text{Cov}(X) = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$$



#### 4. 详细计算步骤

$$\begin{aligned} ACov(\mathbf{X}) &= \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \\ &= \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & -\sigma^2 \end{pmatrix} \end{aligned}$$

然后：

$$\begin{aligned} ACov(\mathbf{X})A^T &= \begin{pmatrix} \sigma^2 & \sigma^2 \\ \sigma^2 & -\sigma^2 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix} \end{aligned}$$

#### 5. 结果解释

最终得到的矩阵：

$$\begin{pmatrix} 2\sigma^2 & 0 \\ 0 & 2\sigma^2 \end{pmatrix}$$

表明：

- $\text{Var}(Y) = \text{Var}(Z) = 2\sigma^2$
- $\text{Cov}(Y, Z) = 0$
- 方差变为原来的 2 倍是因为我们对两个方差为  $\sigma^2$  的随机变量进行了加减运算

**注记：**这个计算过程展示了矩阵方法在处理多维随机变量线性变换时的优越性，它使得复杂的协方差计算变得系统和清晰。

#### 线性变换下协方差矩阵性质的证明

对于随机向量  $\mathbf{X}$  和线性变换矩阵  $A$ ，有：

$$\text{Cov}(A\mathbf{X}) = A\text{Cov}(\mathbf{X})A^T$$

证明. 让我们从协方差的定义开始逐步推导：

##### 1. 协方差矩阵的定义

对于随机向量  $\mathbf{X}$ ，其协方差矩阵定义为：

$$\text{Cov}(\mathbf{X}) = E[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T]$$

##### 2. 线性变换后的期望

根据期望的线性性质：

$$E[A\mathbf{X}] = AE[\mathbf{X}]$$

### 3. 线性变换后的协方差

$$\begin{aligned}
 \text{Cov}(A\mathbf{X}) &= E[(A\mathbf{X} - E[A\mathbf{X}])(A\mathbf{X} - E[A\mathbf{X}])^T] \\
 &= E[(A\mathbf{X} - AE[\mathbf{X}])(A\mathbf{X} - AE[\mathbf{X}])^T] \\
 &= E[A(\mathbf{X} - E[\mathbf{X}])(A(\mathbf{X} - E[\mathbf{X}]))^T] \\
 &= E[A(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T A^T] \\
 &= AE[(\mathbf{X} - E[\mathbf{X}])(\mathbf{X} - E[\mathbf{X}])^T] A^T \\
 &= ACov(\mathbf{X})A^T
 \end{aligned}$$

关键步骤说明：

- 第一步使用协方差矩阵的定义
- 第二步利用期望的线性性质
- 第三步将  $A$  提出来
- 第四步使用矩阵转置的性质  $(AB)^T = B^T A^T$
- 第五步利用期望的线性性质，将常数矩阵  $A$  提出
- 最后一步使用协方差矩阵的定义

□

### 重要性说明

这个性质非常重要，因为它：

1. 提供了计算线性变换后随机向量协方差的简便方法
2. 保持了协方差矩阵的对称性
3. 使得我们能够追踪线性变换对随机变量之间相关性的影响
4. 在多变量统计分析中广泛应用

### 1.12.2 2

#### 题目描述

考虑拟合曲线  $y = b_0x + b_1x^2$  到点  $(x_i, y_i)$ ，其中  $i = 1, \dots, n$ 。

1. 使用矩阵形式找到  $b_0$  和  $b_1$  的最小二乘估计的表达式。
2. 拟合值是什么？

## 解答

1. 使用矩阵形式找到  $b_0$  和  $b_1$  的最小二乘估计的表达式。

设  $\mathbf{y}$  是观测值向量,  $\mathbf{X}$  是设计矩阵,  $\mathbf{b}$  是参数向量,  $\mathbf{e}$  是误差向量, 则有:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$$

其中,

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \\ \vdots & \vdots \\ x_n & x_n^2 \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_0 \\ b_1 \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix}$$

最小二乘估计  $\mathbf{b}$  满足:

$$\mathbf{b} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

2. 拟合值是什么?

拟合值  $\hat{\mathbf{y}}$  为:

$$\hat{\mathbf{y}} = \mathbf{X}\mathbf{b} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

## 解答

1. 首先, 我们用矩阵形式表示最小二乘问题:

设

$$\mathbf{X} = \begin{bmatrix} x_1 & x_1^2 \\ x_2 & x_2^2 \\ \vdots & \vdots \\ x_n & x_n^2 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$$

最小二乘估计通过最小化残差平方和得到:

$$\min_{\mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b})$$

对  $\mathbf{b}$  求导并令其等于零:

$$\frac{\partial}{\partial \mathbf{b}} (\mathbf{y} - \mathbf{X}\mathbf{b})^T (\mathbf{y} - \mathbf{X}\mathbf{b}) = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\mathbf{b} = \mathbf{0}$$

因此, 最小二乘估计为:

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

2. 拟合值可以表示为:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\mathbf{b}} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

令

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$$

则

$$\hat{\mathbf{y}} = \mathbf{H}\mathbf{y}$$

这里的  $\mathbf{H}$  被称为帽子矩阵 (hat matrix), 因为它将观测值  $\mathbf{y}$  转换为拟合值  $\hat{\mathbf{y}}$ 。

## 拟合值的具体表达式推导

首先，我们展开  $\mathbf{X}^T \mathbf{X}$ :

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} \sum x_i^2 & \sum x_i^3 \\ \sum x_i^3 & \sum x_i^4 \end{bmatrix}$$

同样地，

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

令

$$(\mathbf{X}^T \mathbf{X})^{-1} = \frac{1}{|\mathbf{X}^T \mathbf{X}|} \begin{bmatrix} \sum x_i^4 & -\sum x_i^3 \\ -\sum x_i^3 & \sum x_i^2 \end{bmatrix}$$

其中行列式

$$|\mathbf{X}^T \mathbf{X}| = (\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2$$

因此，最小二乘估计的具体表达式为：

$$\begin{bmatrix} \hat{b}_0 \\ \hat{b}_1 \end{bmatrix} = \frac{1}{|\mathbf{X}^T \mathbf{X}|} \begin{bmatrix} \sum x_i^4 & -\sum x_i^3 \\ -\sum x_i^3 & \sum x_i^2 \end{bmatrix} \begin{bmatrix} \sum x_i y_i \\ \sum x_i^2 y_i \end{bmatrix}$$

即：

$$\hat{b}_0 = \frac{\sum x_i^4 \sum x_i y_i - \sum x_i^3 \sum x_i^2 y_i}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2}$$

$$\hat{b}_1 = \frac{-\sum x_i^3 \sum x_i y_i + \sum x_i^2 \sum x_i^2 y_i}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2}$$

因此，对于任意点  $x$ ，其拟合值为：

$$\hat{y}(x) = \hat{b}_0 x + \hat{b}_1 x^2$$

将  $\hat{b}_0$  和  $\hat{b}_1$  的表达式代入，得到拟合值的完整表达式：

$$\hat{y}(x) = \frac{\sum x_i^4 \sum x_i y_i - \sum x_i^3 \sum x_i^2 y_i}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2} x + \frac{-\sum x_i^3 \sum x_i y_i + \sum x_i^2 \sum x_i^2 y_i}{(\sum x_i^2)(\sum x_i^4) - (\sum x_i^3)^2} x^2$$

这就是拟合值的具体表达式，其中所有的求和都是从  $i = 1$  到  $n$ 。