Ch 10 参数估计

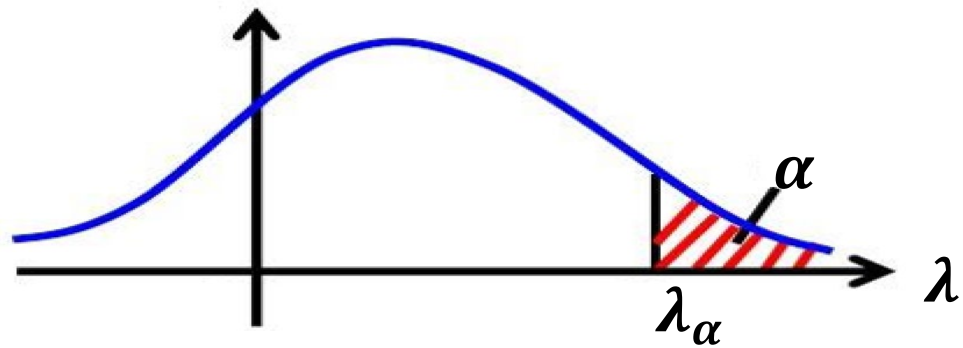
五大正态分布的抽样分布定理

- 设$X_1,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的样本, 则有$\bar{X}\sim N(\mu,\sigma^2/n)$
- 设$X_1,X_2,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的样本,其样本均值$\bar{X}$和修正样本方差S^2相互独立, 且$\frac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$
- 设$X_1,X_2,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的样本,其样本均值$\bar{X}$和修正样本方差S^2, 且$\frac{\bar{X}-\mu}{\sqrt{S^2/n}}\sim t(n-1)$
- 设$X_1,\cdots,X_m$和$Y_1,\cdots,Y_n$分别来自总体$N(\mu_X,\sigma_X^2)$和$N(\mu_Y,\sigma_Y^2)$的样本, 修正样本方差分别为S_X^2和S_Y^2, 则$\frac{S_X^2/\sigma_X^2}{S_Y^2/\sigma_Y^2}\sim F(m-1,n-1)$.
- 设$X_1,\cdots,X_m$和$Y_1,\cdots,Y_n$分别来自总体$N(\mu_X,\sigma^2)$和$N(\mu_Y,\sigma^2)$的样本, 其样本均值分别$\bar{X}$和$\bar{Y}$, 修正样本方差分别为S_X^2和S_Y^2, 则

$$\frac{\bar{X}-\bar{Y}-(\mu_X-\mu_Y)}{\sqrt{\frac{(m-1)S_X^2+(n-1)S_Y^2}{m+n-2}}\sqrt{\frac{1}{m}+\frac{1}{n}}}\sim t(m+n-2)$$

分位数(点)

对给定$\alpha \in (0,1)$ 和随机变量X，称满足$P(X > \lambda_\alpha) = \alpha$的实数$\lambda_\alpha$为上侧$\alpha$分位数(点)



- 正态分布$X \sim N(0,1)$上侧α分位点为μ_α
- χ^2分布$X \sim \chi^2(n)$上侧α分位点为$\chi^2_\alpha(n)$
- t-分布$X \sim t(n)$上侧α分位点为$t_\alpha(n)$
- F-分布$X \sim F(m,n)$上侧α分位点为$F_\alpha(m,n)$

参数估计

问题：如何依据样本$X_1, X_2, \cdots, X_n$估计参数θ，或θ的函数$g(\theta)$， 此类问题称为 **参数估计问题**

矩估计法：样本矩估计总体矩或样本中心矩去估计总体中心矩求参数

总体X的分布函数F包含m个未知参数$\theta_1, \theta_2, \cdots, \theta_m$

- 计算总体X的k阶矩: $a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m) = E[X^k]\ k \in [m]$ (a_k一般为$\theta_1, \theta_2, \cdots, \theta_m$的函数)

- 计算样本的k阶矩: $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

- 令样本矩等于总体矩：

$$A_k = a_k = a_k(\theta_1, \theta_2, \cdots, \theta_m)\ k \in [m]$$

得到m个关于$\theta_1, \theta_2, \cdots, \theta_m$的方程组

- 求解方程组得到估计量$\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_m$

离散型随机变量的似然函数

设$X_1, X_2, \cdots, X_n$是来自总体X的一个样本. 若总体X为离散型随机变量，其分布列为$P(X = x) = P(X = x; \theta)$，则样本$X_1, X_2, \cdots, X_n$的分布列为

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

这里$L(\theta)$表示样本$X_1 = x_1, \cdots, X_n = x_n$发生的概率

称**$L(\boldsymbol{\theta})$为样本$\boldsymbol{X_1, X_2, \cdots, X_n}$的似然函数**

连续型随机变量的似然函数

若总体X为连续型随机变量，其概率密度为$f(x;\theta)$，则$X_1=x_1,X_2=x_2,\cdots,X_n=x_n$的联合概率密度为

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

根据概率密度定义可知$L(\theta)$ 越大，样本$X_1,X_2,\cdots,X_n$落入$x_1,x_2,\cdots,x_n$的邻域内概率越大

称$\boldsymbol{L(\theta)}$**为样本**$\boldsymbol{X_1,X_2,\cdots,X_n}$**的似然函数**

最大似然估计

综合离散和连续随机变量, 统称**$L(\theta)$为样本$X_1, X_2, \cdots, X_n$的似然函数**, 可以发现$L(\theta)$是θ的函数，若

$$\hat{\theta} = \mathrm{argmax}_{\theta} L(x_1, x_2, \cdots, x_m; \theta)$$

则称 **$\hat{\theta}$为θ的最大似然估计量**

直觉而言：最大似然估计量$\hat{\theta}$是使观测值$X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$出现的概率最大

求解最大似然估计量的步骤

- 计算对数似然函数$\ln\big(L(x_1, x_2, \cdots, x_m; \theta)\big)$
- 求对数似然函数中参数θ的一阶偏导, 令其等于零
- 求解方程组得到最大似然估计量$\hat{\theta}$

例题

设$X_1, X_2, \cdots, X_n$是取自总体$X \sim B(1, p)$的样本，求参数p的最大似然估计

最大似然估计不可变性

定理： 设$\mu = \mu(\theta)$为θ的函数，且存在反函数$\theta \ = \theta(\mu)$. 若$\hat{\theta}$是θ的最大似然估计，则$\hat{\mu} = \mu\left(\hat{\theta}\right)$是$\mu$的最大似然估计称为**最大似然估计的不变性**

例题

设$X_1, X_2, \cdots, X_n$为总体$X \sim N(\mu, \sigma^2)$的样本，求μ和$\sigma > 0$的最大似然估计

例题

设总体X的概率密度函数

$$f(x)=\begin{cases}(\alpha+1)x^{\alpha} & x\in(0,1)\\ 0 & \text{其它}\end{cases}$$

设$X_1,X_2,\cdots,X_n$是来自总体X的样本, 求α的最大似然估计

例题

设$X_1, X_2, \cdots, X_n$是总体$X \sim U(a, b)$的样本, 求a和b的最大似然估计

例题

设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 以及总体X的密度函数为

$$f(x) = \begin{cases} \frac{1}{\theta}\, e^{-(x-\mu)/\theta} & x \geq \mu \\ 0 & \text{其它} \end{cases}$$

其中$\theta > 0$, 求μ和θ的最大似然估计

估计量的评价标准

不同的估计方法可能得到不同的估计值,

问题: 采用哪一种估计量更好, 或更好的标准是什么呢?

常用标准:

- 无偏性
- 有效性
- 一致性

无偏性

设$X_1, X_2, \cdots, X_n$来自总体X的样本，令$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的一个估计量, 若

$$E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}\right] = E_{X_1,X_2,\cdots,X_n}\left[\hat{\theta}(X_1, X_2, \cdots, X_n)\right] = \theta$$

则称**$\hat{\theta}$为θ的无偏估计**

无偏估计不要求估计值$\hat{\theta}$在任意情况下都等于θ，但在期望情形下有$E(\hat{\theta}) = \theta$成立.

其意义在于无系统性偏差，无偏性是一种对估计量常见而且重要的标准.

例题

样本k阶原点矩为总体k阶原点矩的无偏估计

设$X_1, X_2, \cdots, X_n$是总体X的样本, 若$E[X^k]$存在, 则$A_k = \sum_{i=1}^{n} X_i^k / n$是总体$a_k = E[X^k]$的无偏估计

设$X_1, X_2, \cdots, X_n$来自总体X的样本, 期望μ, 方差σ^2, 则:

1) $S_0^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 / n$是$\sigma^2$的有偏估计;

2) $S^2 = \sum_{i=1}^{n}(X_i - \bar{X})^2 / (n-1)$是$\sigma^2$的无偏估计.

无偏估计不具有函数不变性

注意$\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$是$\theta$的无偏估计，但并不一定有$g(\hat{\theta})$是$g(\theta)$的无偏估计, 因为

$$E[\hat{\theta}] = \theta \text{并不能推导出} E\left[g\left(\hat{\theta}\right)\right] = g(\theta)$$

例如：$E[\bar{X}] = E[X] = \mu$但$E[\bar{X}^2] \neq \mu^2$

例题

设$X_1, X_2, \cdots, X_n$是总体X的样本, 及总体X的概率密度为

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

证明: $\bar{X}=\sum_{i=1}^{n}X_i/n$和$n\min\{X_1, X_2, \cdots, X_n\}$ 均是θ的无偏估计

有效性

参数可能存在多个无偏估计，若$\hat{\theta}_1$和$\hat{\theta}_2$都是θ的无偏估计，则可以比较方差

$$\mathrm{Var}(\hat{\theta}_1) = E\left[(\hat{\theta}_1 - \theta)^2\right]$$

$$\mathrm{Var}(\hat{\theta}_2) = E\left[(\hat{\theta}_2 - \theta)^2\right]$$

一般而言: 方差越小, 无偏估计越好

设$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n)$ 和$\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n)$是$\theta$的两个无偏估计, 若

$$\mathrm{Var}(\hat{\theta}_1) \leq \mathrm{Var}(\hat{\theta}_2)$$

则称$\boldsymbol{\theta_1}$**比**$\boldsymbol{\theta_2}$**有效**

例题

设$X_1, X_2, \cdots, X_n$是来自总体X的样本, 且X的概率密度为

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

证: 当$n > 1$时$\bar{X} = \sum_{i=1}^{n} X_i / n$比$n \min\{X_1, X_2, \cdots, X_n\}$有效

例题

设$X_1, X_2, \cdots, X_n$是总体X的样本,且$E(X) = \mu$和$\mathrm{Var}\,(X) = \sigma^2$. 常数$c_1, c_2, \cdots, c_n \geq 0$满足$\sum_{i=1}^{n} c_i = 1, c_i \neq 1/n$, 求证: $\bar{X}$比$\sum_{i=1}^{n} c_i X_i$有效