概率论与数理统计
第四章

温灿红

wench@ustc.edu.cn

63607553

大纲

1. 点估计的优良准则
    1. 无偏性 $E\hat{\theta}=\theta$
    2. 有效性
        1. 最小方差无偏估计
    3. 相合性

均方误差

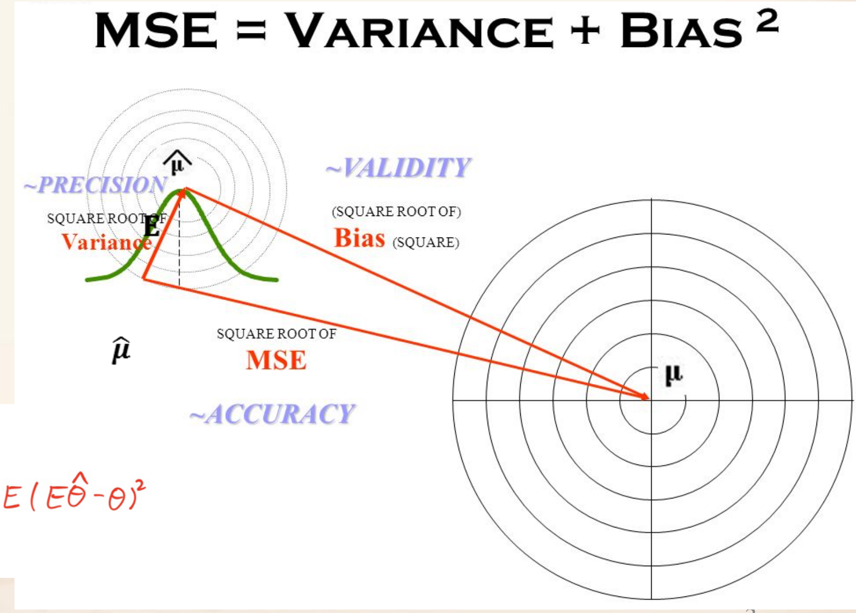
- 评价一个点估计的好坏：用点估计量$\hat{\theta}$和参数θ的距离来度量。我们用它们距离的平方来衡量，记为均方误差(MSE, Mean Squared Error)：

$$\mathrm{MSE}(\hat{\theta}) = E(\hat{\theta} - \theta)^2$$

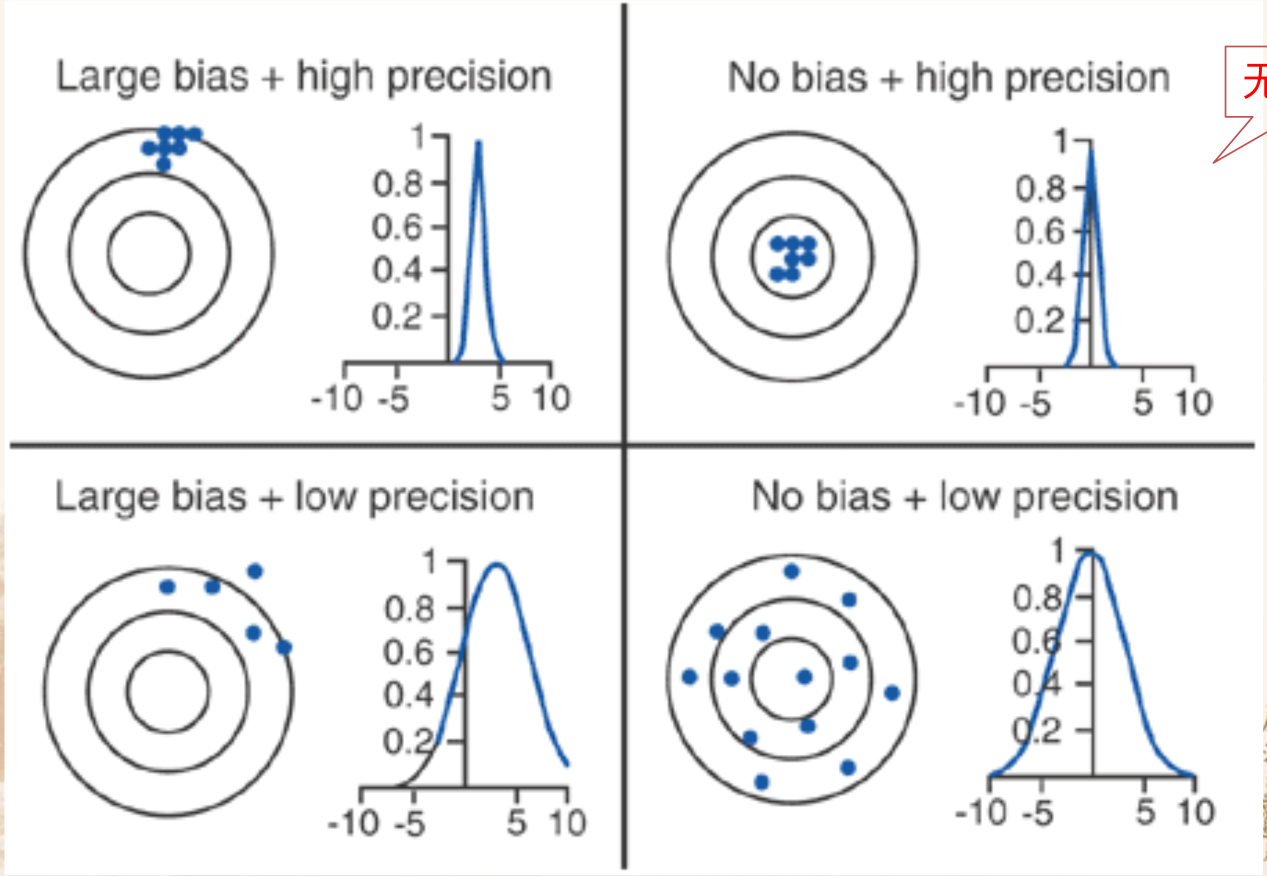
- $\mathrm{MSE}(\hat{\theta}) = \boxed{\mathrm{Var}(\hat{\theta})} + \boxed{(E\hat{\theta} - \theta)}^2$

方差

偏差
bias



$$\begin{aligned} E(\hat{\theta}-\theta)^2 &= E((\hat{\theta}-E\hat{\theta})+(E\hat{\theta}-\theta))^2 \\ &= E(\hat{\theta}-E\hat{\theta})^2 + 2E[(\hat{\theta}-E\hat{\theta})(E\hat{\theta}-\theta)] + E(E\hat{\theta}-\theta)^2 \\ &= \mathrm{Var}(\hat{\theta}) + (E\hat{\theta}-\theta)^2 \end{aligned}$$


Large bias + high precision
1
0.8
0.6
0.4
0.2
-10 -5 5 10
No bias + high precision
1
0.8
0.6
0.4
0.2
-10 -5 5 10
Large bias + low precision
1
0.8
0.6
0.4
0.2
-10 -5 5 10
No bias + low precision
1
0.8
0.6
0.4
0.2
-10 -5 5 10

无偏性

无偏性

- 设$\hat{g}(X_1, \dots, X_n)$为待估参数函数$g(\theta)$的一个估计量，若对于任意的$\theta \in \Theta$，都有

$$\mathrm{E}[\hat{g}(X_1, \dots, X_n)] = g(\theta)$$

- 则称$\hat{g}(X_1, \dots, X_n)$为$g(\theta)$的无偏估计量。

- 两个含义：
  - 没有系统性的偏差；
  - 在多次取平均时，能以接近于100%的把握无限逼近被估计量。

例子

- 设$X_1, \dots, X_n$为从某总体抽取的样本，其均值为μ，方差为σ^2，则样本均值$\bar{X}$和样本方差S^2分别是μ和σ^2的无偏估计。

- S是σ的无偏估计吗？

$\hat{\mu} = \bar{X}$

$\hat{\sigma}_1^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$

$E\hat{\mu} \overset{?}{=} \mu \qquad E\hat{\sigma}_1^2 \overset{?}{=} \sigma^2$

$E\hat{\mu} = E\bar{X} = \frac{1}{n}\sum_{i=1}^{n} EX_i = \mu$

$E\hat{\sigma}_1^2 = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2]$

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

$E(X_i - \bar{X})^2 = E(X_i^2 - 2X_i\bar{X} + \bar{X}^2) = EX_i^2 - 2EX_i\bar{X} + E\bar{X}^2$

$E(\bar{X})^2 = \frac{1}{n}E(\Sigma X_i)^2 = \frac{1}{n}E(\sum_{i=1}^{n} X_i^2 + \sum_{j \neq i} X_i X_j) = \frac{1}{n^2}(n(\mu^2+\sigma^2) + n(n-1)\mu^2) = \frac{1}{n}\sigma^2 + \mu^2$

$EX_i\bar{X} = \frac{1}{n}\sum_{j=1}^{n} EX_i X_j = \frac{1}{n}(EX_i^2 + \sum_{j \neq i} EX_i X_j) = \frac{1}{n}[\mu^2 + \sigma^2 + (n-1)\mu^2] = \frac{1}{n}\sigma^2 + \mu^2$

$\therefore E\hat{\sigma}_1^2 = \frac{1}{n-1}\sum_{i=1}^{n} EX_i^2 - 2\sum_{i=1}^{n} EX_i\bar{X} + nE(\bar{X})^2$

$= \frac{1}{n-1}(n(\mu^2+\sigma^2) - 2n(\frac{1}{n}\sigma^2 + \mu^2) + n(\frac{1}{n}\sigma^2 + \mu^2))$

$= \frac{n}{n-1}(1 - \frac{1}{n})\sigma^2$

例子

- 设$X_1, \dots, X_n$为从均匀分布总体$U(0,\theta)$中抽取的样本，则
  - θ的矩估计是多少？
  - 极大似然估计呢？
  - 他们是无偏估计吗？

$f(x) = \frac{1}{\theta} \quad 0<x<\theta$

$EX = \int x f(x) dx = \int_0^\theta \frac{1}{\theta} x dx = \frac{\theta}{2}$

$\Rightarrow \hat{\theta}_1 = 2\bar{X} \qquad E\hat{\theta}_1 = \frac{2}{n}\sum_{i=1}^n EX_i = \frac{2}{n}\sum_{i=1}^n \frac{\theta}{2} = \theta$ 矩估计

$\Rightarrow \hat{\theta}_2 = X_{(n)}$ $X_{(n)}$的概率密度函数为 $g(x) = n(F(x))^{n-1} f(x) = n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta} = \frac{nx^{n-1}}{\theta^n} \quad 0<x<\theta$

$E\hat{\theta}_2 = EX_{(n)} = \int x g(x) dx = \int_0^\theta x \cdot n\left(\frac{x^{n-1}}{\theta^n}\right) dx = \frac{n}{\theta^n}\frac{1}{n+1}x^{n+1}\Big|_0^\theta = \frac{n}{n+1}\theta$ 极大似然估计

$\hat{\theta}_2$不是无偏估计 调整使之为 $\frac{n+1}{n}\hat{\theta}_2 = \frac{n+1}{n}X_{(n)}$ 为无偏估计

从有偏到无偏

- 通过乘以一个调整因子c_n(通常与n有关)，使得修正后的估计量是无偏的。

- 如正态总体的方差σ^2的MLE为

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

- 是有偏的，而$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$，那么调整因子为 $c_n = \frac{n}{n-1}$。

问题

- 设$X_1, \dots, X_n$为从正态总体$N(\mu, \sigma^2)$中抽取的样本，则μ的无偏估计有：
  - 样本均值$\bar{X}$； $(E\hat{\theta}-\theta)^2 = \frac{\sigma^2}{n}$
  - X_1 σ^2
- 哪个更好呢？

创寰宇学府 育天下英才
严济慈 一九八八年五月 题

有效性

- 设$\hat{g}_1(X_1,\dots,X_n)$和$\hat{g}_2(X_1,\dots,X_n)$为待估参数函数$g(\theta)$的两个不同的无偏估计量，若对于任意的$\theta \in \Theta$，都有

$$Var\big(\hat{g}_1(X_1,\dots,X_n)\big) \le Var\big(\hat{g}_2(X_1,\dots,X_n)\big)$$

- 而且至少存在某个$\theta_0 \in \Theta$使得严格不等式成立，则称$\hat{g}_1$比$\hat{g}_2$有效。

- 方差为小者为优！

例子

- 设$X_1,\ldots,X_n$为从正态总体$N(\mu,\sigma^2)$中抽取的样本，则μ的无偏估计$\bar{X}$比另一个无偏估计X_1有效。

- 设$X_1,\ldots,X_n$为从均匀分布总体$U(0,\theta)$中抽取的样本，则θ的矩估计和修正后的极大似然估计哪个更有效？

$F_{X_{(n)}}(x) = P(X_{(n)} \le x) = P(X_1 \le x \dots X_n \le x) = \prod_{i=1}^{n} P(X_i \le x) = (F(x))^n$

$\therefore f(x) = F'_{X_{(n)}}(x) = n(F(x))^{n-1} \cdot F'(x) = n(F(x))^{n-1} f(x)$

- 设$X_1, \dots, X_n$为从均匀分布总体$U(0,\theta)$中抽取的样本，则θ的矩估计和修正后的极大似然估计哪个更有效？

矩估计：$\hat{\theta}_1 = 2\bar{X}$ $\quad E\hat{\theta}_1 = \theta$

修正后的极大似然估计 $\hat{\theta}_2 = \frac{n+1}{n} X_{(n)}$ $\quad E\hat{\theta}_2 = \theta$ $\quad [EX_{(n)} = \frac{n}{n+1}\theta]$

$Var(\hat{\theta}_1) = Var(2\bar{X}) = \frac{4}{n} Var(X_1) = \frac{4}{n} \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n}$

$X_{(n)}$的概率密度函数为 $f(x) = n\left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta}$ $\quad 0<x<\theta$

$EX_{(n)} = \frac{n}{n+1}\theta$

$EX_{(n)}^2 = \int_0^\theta n\left(\frac{x}{\theta}\right)^{n-1} \frac{1}{\theta} x^2 dx = \frac{n}{\theta^n} \cdot \frac{1}{n+2} x^{n+2} \Big|_0^\theta = \frac{n}{n+2}\theta^2$

$Var(X_{(n)}) = EX_{(n)}^2 - (EX_{(n)})^2 = \frac{n}{(n+2)(n+1)^2}\theta^2$

$\therefore Var(\hat{\theta}_2) = \left(\frac{n+1}{n}\right)^2 Var(X_{(n)})^2 = \frac{1}{n(n+1)}\theta^2$

当$n>1$时 $Var(\hat{\theta}_2) < Var(\hat{\theta}_1)$ 则$\hat{\theta}_2$比$\hat{\theta}_1$有效

最小方差无偏估计（MVUE)

- 设$\hat{g}(X_1,\dots,X_n)$为$g(\theta)$的无偏估计，若对于$g(\theta)$的任意的无偏估计$\hat{g}_2(X_1,\dots,X_n)$，都有

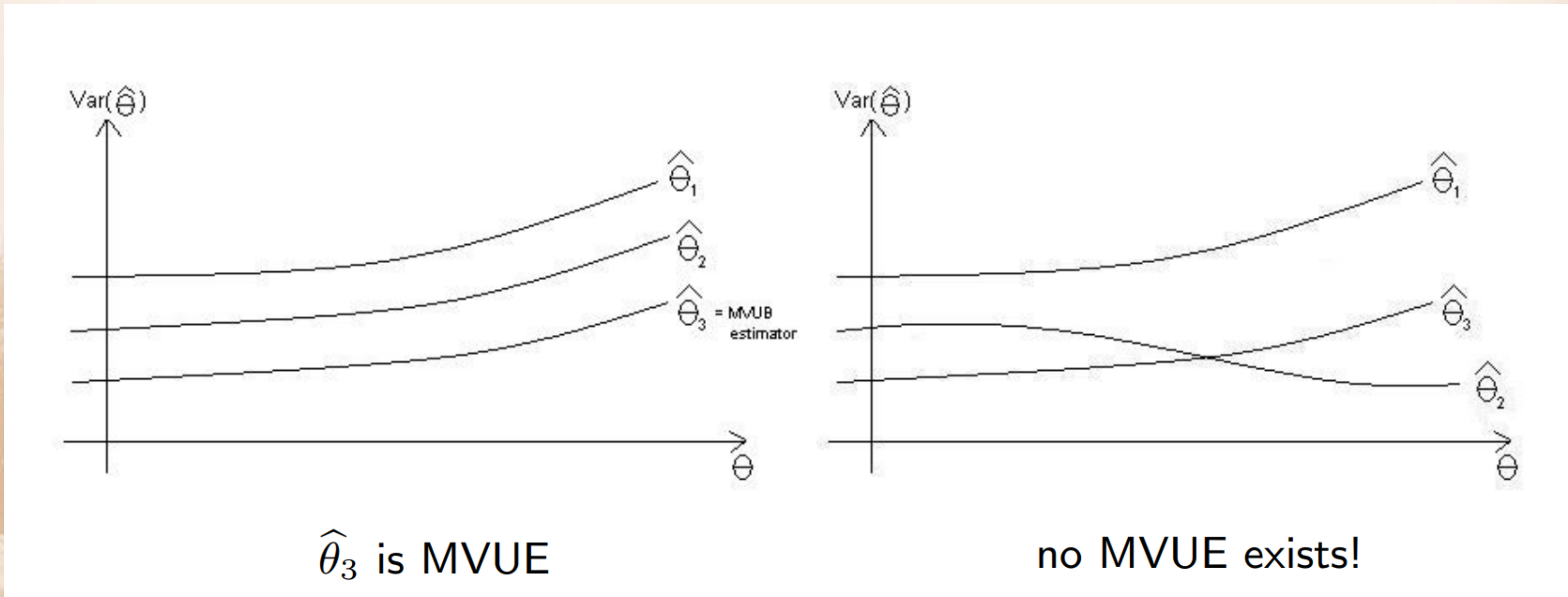
$$Var\big(\hat{g}(X_1,\dots,X_n)\big) \le Var\big(\hat{g}_2(X_1,\dots,X_n)\big)$$

- 对于任意的$\theta \in \Theta$都成立，则称$\hat{g}(X_1,\dots,X_n)$ 为$g(\theta)$的最小方差无偏估计（Minimum Variance Unbiased Estimate）。

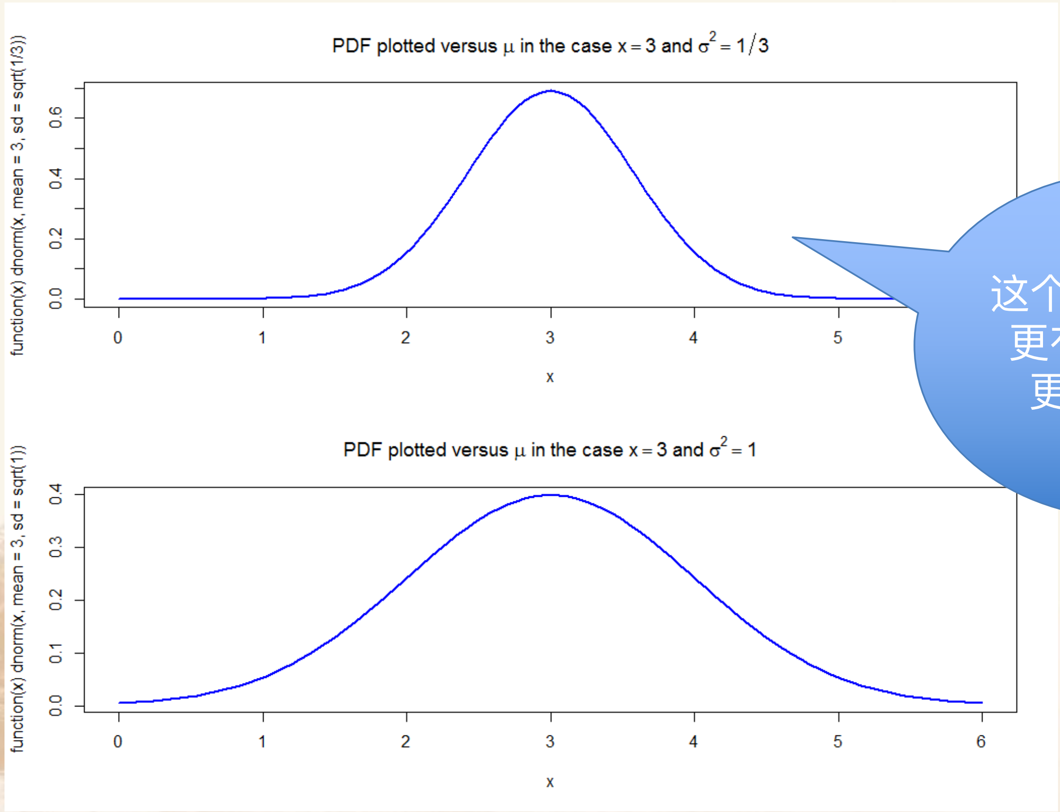
- MVUE一定存在吗?
- 如何找MVUE？

MVUE一定存在吗?

- 假设有三个无偏估计：$\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$



如何找MVUE?



这个方差更小，更有可能得到更精确的解

- 似然函数的弯曲程度决定了我们能够有多精确估计参数。
- 弯曲程度可以用曲率来度量：

$$-\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta}$$

- 但这是跟数据有关系的，是个随机变量，因此我们可以通过对随机变量取平均得到平均曲率：

$$-E_X\left\{\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta}\right\}$$

中国科学技术大学 University of Science and Technology of China

Cramer-Rao不等式

- 在一定的条件下，对θ的任一无偏估计$\hat{\theta}$，都有

$$\mathrm{Var}\left(\hat{\theta}(X_1,\dots,X_n)\right) \geq \frac{1}{n\,I(\theta)}$$

Cramer-Rao下界

- 其中$I(\theta)$ 为Fisher信息量，其定义如下：

$$I(\theta) = -E\left[\frac{\partial^2 \ln f(x;\theta)}{\partial^2\theta}\right] = E\left[\frac{\partial \ln f(x;\theta)}{\partial\theta}\right]^2.$$

- $I(\theta)$越大，总体分布中包含未知参数θ的信息越大。

用Cramer-Rao下界来得到MVUE

$$I(\theta) = -E_x\left[\frac{\partial^2 \ln f}{\partial\theta^2}\right] = -E\left[\frac{-\left(\frac{\partial f}{\partial\theta}\right)^2 + \frac{\partial^2 f}{\partial\theta^2}\cdot f}{f^2}\right] = E\left(\frac{\partial \ln f}{\partial\theta}\right)^2 - E\left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2}\right]$$

$$E\left[\frac{1}{f}\frac{\partial^2 f}{\partial\theta^2}\right] = \int \frac{1}{f}\cdot\frac{\partial^2 f}{\partial\theta^2}\cdot f\,dx = \int \frac{\partial^2 f}{\partial\theta^2}\,dx = \frac{\partial^2}{\partial\theta^2}\int f(x)\,dx = 0 \qquad \left[E g(x) = \int g(x) f(x)\,dx\right]$$

中国科学技术大学
University of Science and Technology of China

例子

- 设$X_1,\ldots,X_n$为从正态总体$N(\mu,\sigma^2)$中抽取的样本，其中σ^2已知，试证明μ的无偏估计$\bar{X}$是μ的MVUE。

$E\bar{X}=\mu \qquad Var(\bar{X})=\frac{\sigma^2}{n} \qquad f(x,\mu)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad x\in R$

，试

$\ln f(x;\mu)=-\ln(\sqrt{2\pi}\sigma)-\frac{1}{2\sigma^2}(x-\mu)^2$

$\frac{\partial \ln f(x;\mu)}{\partial \mu}=-\frac{1}{2\sigma^2}2(\mu-x)=\frac{1}{\sigma^2}(x-\mu)$

$\frac{\partial^2 \ln f(x;\mu)}{\partial \mu^2}=-\frac{1}{\sigma^2}$

$\therefore I(\theta)=-E\left[\frac{\partial^2 \ln f(x;\mu)}{\partial \mu^2}\right]=\frac{1}{\sigma^2}$

C-R下界为 $\frac{1}{nI(\theta)}=\frac{\sigma^2}{n}=Var(\bar{X})$ 因此$\bar{X}$是MVUE

育天下英才 創寰宇學府

嚴濟慈題 一九八八年五月

- 设$X_1, \dots, X_n$为从泊松分布$P(\lambda)$中抽取的样本，试证明λ的无偏估计$\bar{X}$是λ的MVUE。

$E\bar{X} = \lambda \qquad Var(\bar{X}) = \frac{1}{n}\lambda$

$f(k, \lambda) = \frac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0, 1, 2 \dots)$

$\ln f(k;\lambda) = k\ln\lambda - \ln k! - \lambda$

$\frac{\partial \ln f(k;\lambda)}{\partial \lambda} = \frac{k}{\lambda} - 1 \qquad \frac{\partial^2 \ln f(k;\lambda)}{\partial \lambda^2} = -\frac{k}{\lambda^2}$

$\therefore I(\theta) = E\left[\frac{\partial \ln f(k;\lambda)}{\partial \lambda}\right]^2 = E\left[\frac{k}{\lambda} - 1\right]^2 = \frac{1}{\lambda^2}E(X-\lambda)^2 = \frac{1}{\lambda^2}Var(X) = \frac{1}{\lambda}$

$= -E\left[\frac{\partial^2 \ln f(k;\lambda)}{\partial \lambda^2}\right] = \frac{1}{\lambda^2}EX = \frac{1}{\lambda}$

C-R下界为 $\frac{1}{nI(\theta)} = \frac{\lambda}{n} = Var(\bar{X})$ 因此$\bar{X}$是MVUE

总结

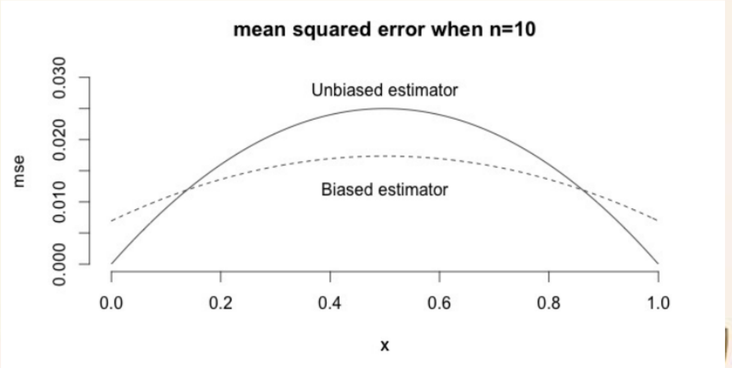
- 无偏性和有效性都是小样本性质。接下来要讨论的相合性则是大样本性质。

- MVUE不一定是使得MSE最小的估计，它是使得MSE最小的无偏估计。

- 有偏估计有可能MSE更小。见下例

- X服从二项分布$B(n,\theta)$，两个估计，哪个更好？

$$\hat{\theta}_1 = \frac{X}{n}, \quad \text{矩估计} \qquad \hat{\theta}_2 = \frac{X+1}{n+2} \quad \text{贝叶斯估计}$$



$EX = n\theta \qquad Var(X) = n\theta(1-\theta)$

$E\hat{\theta}_1 = \theta \qquad E\hat{\theta}_2 = \frac{n\theta+1}{n+2} \neq \theta$

$Var(\hat{\theta}_1) = \frac{1}{n^2} Var(X) = \frac{1}{n}\theta(1-\theta) \Rightarrow MSE(\hat{\theta}_1) = Var(\hat{\theta}_1) + (E\hat{\theta}_1 - \theta)^2 = \frac{\theta(1-\theta)}{n}$

$Var(\hat{\theta}_2) = \frac{1}{(n+2)^2} Var(X) = \frac{n}{(n+2)^2}\theta(1-\theta) \Rightarrow MSE(\hat{\theta}_2) = Var(\hat{\theta}_2) + (E\hat{\theta}_2 - \theta)^2 = \frac{1}{(n+2)^2}[(4-n)\theta^2 + (n-4)\theta + 1]$

相合性

- 设$X_1, \dots, X_n$为从某个以来于参数θ的总体中抽取的样本， $\hat{g}(X_1, \dots, X_n)$为待估参数函数$g(\theta)$的一个估计量。如果对于任意的$\epsilon > 0$，有

$$\lim_{n \to \infty} P_\theta(|\hat{g}(X_1, \dots, X_n) - g(\theta)| \geq \epsilon) = 0,$$

- 对所有的$\theta \in \Theta$都成立。则称$\hat{g}(X_1, \dots, X_n)$是$g(\theta)$的一个相合估计。记为$\hat{g}(X_1, \dots, X_n) \xrightarrow{p} g(\theta)$。

Thank You!

創寰宇學府
育天下英才
嚴濟慈題
一九八八年五月