基础数理统计（研究生公共课）

肖柳青 教授 博士主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

主要内容

1.点估计的基本概念
2.两种基本的点估计方法
3.点估计的优良标准（无偏性及有效估计和C-R下界）
4.置信区间估计
5.最佳点估计（充分统计量）
6.贝叶斯估计

2.2 估计量的几个评选标准

在介绍估计量优良性的准则之前，我们必须强调指出：

评价一个估计量的好坏，不能仅仅依据一次试验的结果，而必须由多次试验结果来衡量 .

这是因为估计量是样本的函数，是随机变量. 所以，由不同的观测结果，就会求得不同的参数估计值.

因此一个好的估计，应在多次试验中体现出优良性.

常用的几条标准是：

1. 无偏性
2. 有效性

我们重点介绍前面两个标准.

3. 均方误差准则MSE
4. 一致性（相合性）

这些常用的标准源于均方误差这一概念！

均方误差准则(MSE)

定义 设 $\hat{\theta}$ 是未知参数 θ 的一个估计量，$\hat{\theta}$ 的均方误差定义为

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

$\hat{\theta}$ 的均方误差 $MSE(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = E\left[(\hat{\theta}-E\hat{\theta})+(E\hat{\theta}-\theta)\right]^2$

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

$$= D\hat{\theta} + (E\hat{\theta}-\theta)^2$$

1．无偏性(unbiasedness)

估计量是随机变量，对于不同的样本值会得到不同的估计值. 我们希望估计值在未知参数真值附近摆动, 即它的期望值等于未知参数的真值. 这就导致无偏性这个标准的产生.

真值

定义: 设 $\theta=\theta(X_1,\cdots,X_n)$ 是未知参数 θ 的一个估计量,

若 $E(\hat{\theta})=\theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计量, 否则称 $\hat{\theta}$ 为 θ 的有偏估计量.

直观上看, 所谓无偏估计量就是如果相互独立地多次用无偏估计量进行实际估计时， 所得的诸估计值的算术平均值与真值基本相同. 即没有系统性的偏差.

无偏性是对估计量的一个常见而重要的要求 .

例如，用样本均值作为总体均值的估计时，虽无法说明一次估计所产生的偏差, 但这种偏差随机地在 0 的周围波动, 对同一统计问题大量重复使用不会产生系统偏差 .

例1 设总体 X 的 k 阶原点矩 $E(X^k)=\mu_k\ (k\geq 1)$ 存在，$X_1,\cdots,X_n$ 是总体 X 的样本，证明：样本 k 阶原点矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$，是参数 μ_k 的无偏估计量.

证 由于样本 $X_1,\cdots,X_n$ 与总体 X 同分布，

$$\therefore\ E(X_i^k)=E(X^k)=\mu_k,\quad (k\geq 1,\ i=1,2,\cdots,n)$$

又 $$EA_k=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i^k\right)=\frac{1}{n}\cdot n\cdot\mu_k=\mu_k,$$

所以样本 k 阶原点矩 A_k 是参数 μ_k（总体 k 阶原点矩 $E(X^k)$）的无偏估计量.

特别地，只要总体X 的数学期望存在，样本均值 $\bar{X}$ 是总体均值 EX 的无偏估计量.

例2 设总体X的方差 $DX=\sigma^2$ 存在 $X_1, \cdots, X_n$ 是总体 X 的样本，

证明：样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 是 σ^2 的无偏估计量.

$B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$ 不是 σ^2 的无偏估计量.

证 $\because\ E(S^2)=\sigma^2,\quad EB_2=E(S_n^2)=\frac{n-1}{n}\sigma^2 \neq \sigma^2,$

所以样本方差 S^2 是总体方差 σ^2 的无偏估计量；

而样本方差的异型 S_n^2 是总体方差 σ^2 的有偏估计量 .

尽管 B_2 不是总体方差 σ^2 的无偏估计量，但当 $n\to\infty$ 时，总有

$$\lim_{n\to\infty} EB_2=\sigma^2,$$

我们称 B_2 是总体方差 σ^2 的渐近无偏估计量.

这表明样本容量很大时, 用 B_2 作为总体方差 σ^2 的估计量产生的偏差是很小的

无偏估计量的函数未必是无偏估计量

例3 设 $(X_1, X_2, \cdots, X_n)$ 是总体 $X \sim (\mu, \sigma^2)$ 的样本，试判别 σ 的估计量 $S^* = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$ 是否具有无偏性?

解 否,证明如下：已知 $\frac{(n-1)S^{*2}}{\sigma^2} \sim \chi^2(n-1)$,

$$E\left(\frac{\sqrt{n-1}S^*}{\sigma}\right) = \int_0^{\infty} \sqrt{x}\,\frac{1}{2^{\frac{n-1}{2}}\Gamma(\frac{n-1}{2})} x^{\frac{n-1}{2}-1} e^{-\frac{x}{2}} dx = \frac{\sqrt{2}\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}$$

$$E(S^*) = \sqrt{\frac{2}{n-1}}\,\frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \cdot \sigma \neq \sigma.$$

由例3可见:

虽有 $E(S^{*2})=\sigma^2$，但未必有 $E(S^*)=\sigma$.

一般若 $\hat{\theta}$ 是 θ 的无偏估计, $g(\theta)$ 不是 θ 的线性函数, 那么 $g(\hat{\theta})$ 不是 $g(\theta)$ 的无偏估计.

例4 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0,\\ 0 & x\le 0\end{cases}\qquad \theta>0$$

$(X_1,X_2,\cdots,X_n)$ 为 X 的一个样本

证明 $\overline{X}$ 与 $n\min\{X_1,X_2,\cdots,X_n\}$ 都是 θ 的无偏估计量.

证 $X\sim E\left(\dfrac{1}{\theta}\right)$ $\quad E(X)=\theta$

$$E(\overline{X})=E(X)=\theta$$

$\overline{X}$ 是 θ 的无偏估计量.

$$Z = \min\{X_1, X_2, \cdots, X_n\}$$

$$F_Z(z) = 1 - P(X_1 > z, X_2 > z, \cdots, X_n > z)$$

$$= 1 - P(X_1 > z)P(X_2 > z)\cdots P(X_n > z)$$

$$= 1 - \prod_{i=1}^{n}(1 - P(X_i \le z)) = \begin{cases} 0 & z < 0 \\ 1 - e^{-\frac{nz}{\theta}} & z \ge 0 \end{cases}$$

$$f_Z(z) = \begin{cases} 0 & z < 0 \\ \frac{n}{\theta} e^{-\frac{nz}{\theta}} & z \ge 0 \end{cases}$$

$$Z \sim E\left(\frac{n}{\theta}\right) \qquad E(Z) = \frac{\theta}{n} \qquad E(nZ) = \theta$$

例5 设$(X_1, X_2, \cdots, X_m)$是总体 X 的一个样本，$X \sim B(n, p)$ $n > 1$，求 p^2 的无偏估计量.

解 由于子样矩是母体矩的无偏估计量以及数学期望的线性性质，只要将未知参数表示成母体矩的线性函数，然后用子样矩作为母体矩的估计量，这样得到的未知参数的估计量即为无偏估计量.

$$\mathrm{E}\bar{X} = E(X) = np$$

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m} X_i^2\right) = E(X^2) = (np)^2 + np(1-p)$$

故 $$(n^2-n)p^2=\frac{1}{m}\sum_{i=1}^{m}X_i^2-\overline{X}$$

因此, p^2 的无偏估计量为

$$\hat{p}^2=\frac{1}{n^2-n}\left(\frac{1}{m}\sum_{i=1}^{m}X_i^2-\overline{X}\right)$$

$$=\frac{\frac{1}{m}\sum_{i=1}^{m}X_i(X_i-1)}{n(n-1)}$$

存在这样的情形，对参数$g(\theta)$它的无偏估计不存在．请看下例:

例1. 设样本$X \sim$ 二项分布 $b(n,p)$, n已知而p未知．令$g(p)=1/p$,则参数$g(p)$的无偏估计不存在．

证 采用反证法: 若不然, $g(p)$有无偏估计$\hat{g}(X)$.由于X只取$0,1,\cdots,n$这些值, 令$\hat{g}(X)$的取值用$\hat{g}(i)=a_i$表示, $i=0,1,\cdots,n$.由$\hat{g}(X)$的无偏性,应有

$$E_p(\hat{g}(X))=\sum_{i=0}^{n} a_i \binom{n}{i} p^i(1-p)^{n-i}=1/p, \quad 0<p<1.$$

于是有

$$\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i}-1=0, \quad 0<p<1.$$

但上式左端是p的$n+1$次多项式, 它最多在$(0,1)$区间有$n+1$个实根, 可无偏性要求对$(0,1)$中的任一实数p上式都成立. 这个矛盾说明$g(p)=1/p$无的偏估计不存在.

今后我们把不存在无偏估计的参数除外. 参数的无偏估计若存在, 则此参数为**可估参数**; 若参数函数的无偏估计存在, 则称此函数为**可估函数** (Estimable function). 因此可估函数的无偏估计类是非空的.

2、有效性

我们知道,一个未知参数往往有不止一个无偏估计.

若 $\hat{\theta}_1$ 和 $\hat{\theta}_2$ 都是参数 θ 的无偏估计量，我们还可以通过比较

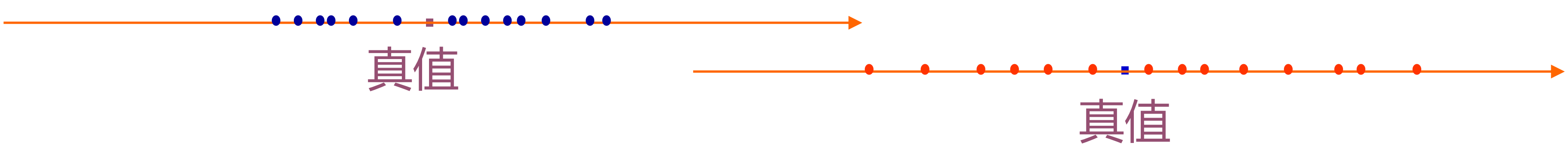
$$E(\hat{\theta}_1-\theta)^2 \quad 和 \quad E(\hat{\theta}_2-\theta)^2$$

的大小来决定二者谁更优．即比较 $D(\hat{\theta}_1)$ 和 $D(\hat{\theta}_2)$ 的大小.

显然无偏估计又以方差小者为好，这就产生了有效性这一概念.

定义：设 $\hat{\theta}_1=\hat{\theta}_1(X_1,\cdots,X_n)$ 和 $\hat{\theta}_2=\hat{\theta}_2(X_1,\cdots,X_n)$，都是未知参数 θ 的无偏估计量，若对任意样本容量的 n，总有 $D(\hat{\theta}_1)<D(\hat{\theta}_2)$，则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 有效.

稳定在真值附近，波动较小



蓝色是采用估计量 $\hat{\theta}_1$，用 14 组样本得到的 14 个估计值.

红色是采用估计量 $\hat{\theta}_2$，用 14 组样本得到的 14 个估计值.

例 6　设总体 X 的方差 $DX=\sigma^2$ 存在，$X_1,\cdots,X_n$ ($n>2$) 是总体 X 的样本，问总体均值 μ 的无偏估计量 X_i ($i=1, 2,\cdots, n$) 与其样本均值 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 哪一个更有效?

解　$\because\ D(X_i)=DX=\sigma^2,$

$$D\overline{X}=\frac{1}{n}DX=\frac{\sigma^2}{n},$$

$\therefore\ D\overline{X}<DX_i,$　　($i=1, 2,\cdots, n$)

故 $\overline{X}$ 作为 μ 的估计量比 X_i ($i=1, 2,\cdots, n$) 有效.

符合常识！！

例7　设总体 X 的均值和方差均存在，

$C_1, C_2, \cdots, C_n$ 为不全相同且满足 $\sum_{i=1}^{n} C_i = 1$ 的任一组常数,

证明:（1）样本的线性函数 $\sum_{i=1}^{n} C_i X_i$ 是总体均值 μ 的无偏估计量；

（2）总体均值的无偏估计量 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 较 $\sum_{i=1}^{n} C_i X_i$ 有效.

证（1）

$$\because\ E\left(\sum_{i=1}^{n} C_i X_i\right) = \sum_{i=1}^{n} C_i \cdot EX_i = \mu\sum_{i=1}^{n} C_i = \mu$$

$\therefore\ \sum_{i=1}^{n} C_i X_i$ 是 μ 的无偏估计量；

（2）由柯西—许瓦兹不等式 知

$$1 = \left(\sum_{i=1}^{n} C_i\right)^2 < \sum_{i=1}^{n} 1^2 \cdot \sum_{i=1}^{n} C_i^2 = n\sum_{i=1}^{n} C_i^2 \Rightarrow \sum_{i=1}^{n} C_i^2 > \frac{1}{n}$$

$$\therefore\ D\left(\sum_{i=1}^{n} C_i X_i\right) = \sum_{i=1}^{n} C_i^2 \cdot DX_i = \sigma^2\sum_{i=1}^{n} C_i^2 > \frac{\sigma^2}{n} = D\overline{X}.$$

这表明, 在 μ 的所有线性无偏估计量中，样本均值 $\overline{X}$ 是最有效的.

例如 $X \sim N(\mu,\sigma^2)$, (X_1, X_2) 是一子样.

$$\left.\begin{aligned}\hat{\mu}_1 &= \frac{2}{3}X_1 + \frac{1}{3}X_2 \\ \hat{\mu}_2 &= \frac{1}{4}X_1 + \frac{3}{4}X_2 \\ \hat{\mu}_3 &= \frac{1}{2}X_1 + \frac{1}{2}X_2\end{aligned}\right\}$$ 都是μ的无偏估计量

由例7 知 $\hat{\mu}_3$ 最有效.

有效性的思考？

定义 设 $\hat{\theta}_1, \hat{\theta}_2$ 都是 θ 的无偏估计量，若

$$D(\hat{\theta}_1) < D(\hat{\theta}_2)$$

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 有效.

若 θ 的所有二阶矩存在的无偏估计量中存在估计量 $\hat{\theta}_0$，使对任意无偏估计量 $\hat{\theta}$

$$D\hat{\theta}_0 \le D\hat{\theta}$$

$\hat{\theta}_0$ 是 θ 的一致最小方差无偏估计(量).

$\boxed{\hat{\theta}_0\text{的求法}}$要用到充分统计与完备统计.

（见下一讲）

问题

- 无偏估计的方差是否可以任意小？
- 若答案否定，那么它的下限是什么？
- 这个下限能否达到？

为此引入下面定理，即

罗-克拉美(C. R. *Rao* – *H.* *Cramer*)不等式

定理 设 Θ 为实数轴上的开区间，$\{f(x;\theta),\theta\in\Theta\}$ 是母体 X 的分布密度族，$(X_1,X_2,\cdots,X_n)$ 是 X 的子样，$\hat\theta=\hat\theta(X_1,X_2,\cdots,X_n)$ 是 θ 的无偏估计. 若以下三条件(即**正规条件**)满足:

(1) 集合 $S_\theta=\{x: f(x;\theta)\neq 0\}$ 与 θ 无关;

(2) $\dfrac{\partial f(x;\theta)}{\partial\theta}$ 存在，且对 Θ 中一切 θ 有

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} f(x;\theta)dx=\int_{-\infty}^{\infty}\frac{\partial f(x;\theta)}{\partial\theta}dx$$

和

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}(x_1,x_2,\cdots,x_n)L(x_1,x_2,\cdots,x_n;\theta)dx_1dx_2\cdots dx_n$$

$$=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\hat{\theta}(x_1,x_2,\cdots,x_n)\frac{\partial}{\partial\theta}L(x_1,x_2,\cdots,x_n;\theta)dx_1dx_2\cdots dx_n$$

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

(3) $E\left(\frac{\partial\ln f(X;\theta)}{\partial\theta}\right)^2=\int_{-\infty}^{\infty}\left(\frac{\partial\ln f(x;\theta)}{\partial\theta}\right)^2 f(x;\theta)dx>0$

则 $$\boxed{D\hat{\theta}\geq\frac{1}{nE\left(\frac{\partial\ln f(x;\theta)}{\partial\theta}\right)^2}\triangleq I_R}$$

罗-克拉美不等式

注1 若离散母体X的分布列为$P(x;\theta)$，且满足定理中的**正规条件**，则**罗-克拉美不等式**是

$$D\hat{\theta} \geq \frac{1}{n\sum_{x}\left(\frac{\partial \ln f(x;\theta)}{\partial \theta}\right)^{2} P(x;\theta)} \triangleq I_R$$

称 $D_0\theta$ 为无偏估计量方差的下界.

当 $D\hat{\theta}=D_0\theta$ 时, 称 $\hat{\theta}$ 为 θ 的优效估计量.

在满足**正规条件**的估计量族范围内优效估计是最小方差无偏估计.

设 θ 的无偏估计为 $\hat{\theta}$, 则称 $I_R/D\hat{\theta}$

$\hat{\theta}$ 的效率, 记为 $e(\hat{\theta})=I_R/D\hat{\theta}$

若估计量 $\hat{\theta}$ 满足

$$\lim_{n\to\infty} e(\hat{\theta})=1$$

则称 $\hat{\theta}$ 为 θ 的渐近优效估计量.

注2 Fisher信息量的两种形式证明：

证：$\int_{-\infty}^{+\infty} f(x;\theta)dx = 1$，$\dfrac{\partial}{\partial\theta}\int_{-\infty}^{+\infty} f(x;\theta)dx = \int_{-\infty}^{+\infty} f'(x;\theta)dx = 0$

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

$$= \int_{-\infty}^{+\infty}\left(\frac{f''(x;\theta)}{f(x;\theta)}\right)\cdot f(x;\theta)dx - E\left(\frac{\partial \ln f(x;\theta)}{\partial\theta}\right)^2$$

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

即 $I(\theta) = E\left[\dfrac{\partial \ln f(X;\theta)}{\partial\theta}\right]^2 = -E\left[\dfrac{\partial^2 \ln f(X;\theta)}{\partial\theta^2}\right] > 0$。

统计思想：总体分布参数的最小方差无偏估计的方差若能达到 C-R 下界，则与 $I(\theta)$ 成反比。

$I(\theta)$ 越大，则最小方差无偏估计的方差越小，总体分布参数就可以越精确地估计出来，即

样本包含的关于总体分布参数的信息越多，因而称 $I(\theta)$ 为 Fisher 信息量。

例 6 总体 $X \sim P(\lambda)$，$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，求未知参数 λ 的无偏估计的方差下界。

解：$f(x;\lambda) = P(X = x) = \dfrac{\lambda^x}{x!} e^{-\lambda}$，$\ln f(x;\lambda) = x \ln \lambda - \ln x! - \lambda, x = 0,1,2,\cdots$

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

C-R 下界 $L = \dfrac{1}{nI(\lambda)} = \dfrac{\lambda}{n}$，$\lambda$ 的一个无偏估计 $\hat{\lambda} = \bar{X}$，$D\hat{\lambda} = D\bar{X} = \dfrac{\lambda}{n}$，$\hat{\lambda} = \bar{X}$ 是达到 C-R 下界的无偏估计量，即 $\hat{\lambda} = \bar{X}$ 是 λ 的最小方差无偏估计量，也是有效估计。

例 7 总体 $X \sim N(\mu, \sigma^2)$， $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，求未知参数 μ, σ^2 的无偏估计的方差下界。

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

$I(\mu) = -E\left[\dfrac{\partial^2 \ln f(X;\mu,\sigma^2)}{\partial\mu^2}\right] = -E\left[-\dfrac{1}{\sigma^2}\right] = \dfrac{1}{\sigma^2}$，$\mu$ 的无偏估计的 C-R 下界 $L = \dfrac{1}{nI(\mu)} = \dfrac{\sigma^2}{n}$，

μ 的无偏估计 $\overline{X}$ 的方差 $D\overline{X} = \dfrac{\sigma^2}{n}$，即 $\overline{X}$ 是 μ 的最小方差无偏估计，也是有效估计。

$$I(\sigma^2) = -E\left[\frac{\partial^2 \ln f(X;\mu,\sigma^2)}{\partial(\sigma^2)^2}\right] = -E\left[\frac{1}{2\sigma^4} - \frac{1}{\sigma^6}(X-\mu)^2\right] = -\frac{1}{2\sigma^4} + \frac{DX}{\sigma^6} = \frac{1}{2\sigma^4}$$

σ^2 的无偏估计的 C-R 下界 $L = \dfrac{1}{nI(\sigma^2)} = \dfrac{2\sigma^4}{n}$ 。

σ^2 是无偏估计 S^2 的方差 $DS^2 = \dfrac{\sigma^4}{(n-1)^2} D\left[\dfrac{(n-1)S^2}{\sigma^2}\right] = \dfrac{\sigma^4}{(n-1)^2}\cdot 2(n-1) = \dfrac{2\sigma^4}{n-1} > \dfrac{2\sigma^4}{n} = L$

S^2 不是 σ^2 的有效估计，但是 σ^2 的渐近有效估计。$e(\hat{\sigma}^2) = e(S^2) = \dfrac{n-1}{n} \xrightarrow{n\to\infty} 1$

定理证明：（孙荣恒书上P. 38）

证　由于 $X_1, \cdots, X_n$ 为 i.i.d. 样本, 故有 $f(\boldsymbol{x}, \theta) = \prod_{i=1}^{n} f(x_i, \theta)$. 记

$$S(\boldsymbol{x}, \theta) = \frac{\partial \log f(\boldsymbol{x}, \theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(x_i, \theta)}{\partial \theta},$$

因此由正则条件 (3) 和 (4) 可知

$$E_\theta\{S(\boldsymbol{X}, \theta)\} = \sum_{i=1}^{n} E_\theta\left\{\frac{\partial \log f(X_i, \theta)}{\partial \theta}\right\} = \sum_{i=1}^{n} \int \frac{1}{f(x_i, \theta)} \frac{\partial f(x_i, \theta)}{\partial \theta} \cdot f(x_i, \theta) dx_i$$

$$= \sum_{i=1}^{n} \int \frac{\partial f(x_i, \theta)}{\partial \theta} dx_i = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \int f(x_i, \theta) dx = 0.$$

由 $\hat{g}(\boldsymbol{x})$ 为 $g(\theta)$ 的无偏估计和正则条件 (5) 和 (6) 可知

$$\begin{aligned}
\mathrm{Cov}_\theta(\hat{g}(\boldsymbol{X}), S(\boldsymbol{X},\theta)) &= E_\theta\{\hat{g}(\boldsymbol{X})\cdot S(\boldsymbol{X},\theta)\}\\
&= \int\cdots\int \hat{g}(\boldsymbol{x})\left[\frac{1}{f(\boldsymbol{x},\theta)}\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}\right] f(\boldsymbol{x},\theta)d\boldsymbol{x}\\
&= \int\cdots\int \hat{g}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}d\boldsymbol{x}\\
&= \frac{\partial}{\partial\theta}\int\cdots\int \hat{g}(\boldsymbol{x})f(\boldsymbol{x},\theta)d\boldsymbol{x} = \frac{\partial g(\theta)}{\partial\theta} = g'(\theta),
\end{aligned}$$

$$\begin{aligned}
D_\theta(S(\boldsymbol{X},\theta)) &= \sum_{i=1}^{n} D_\theta\left\{\frac{\partial\log f(X_i,\theta)}{\partial\theta}\right\}\\
&= \sum_{i=1}^{n} E_\theta\left\{\frac{\partial\log f(X_i,\theta)}{\partial\theta}\right\}^2 = nI(\theta).
\end{aligned}$$

由 Cauchy-Schwartz 不等式, 得

$$D_\theta\{\hat{g}(\boldsymbol{X})\}\cdot D_\theta\{S(\boldsymbol{X},\theta)\} \geqslant [\mathrm{Cov}_\theta(\hat{g}(\boldsymbol{X}), S(\boldsymbol{X},\theta))]^2 = [g'(\theta)]^2,$$

将式 (3.5.4) 代入上式得

$$D_\theta[\hat{g}(\boldsymbol{X})] \geqslant \frac{[g'(\theta)]^2}{nI(\theta)}, \quad 一切\ \theta\in\Theta.$$

例8 设母体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0, \\ 0 & x\le 0\end{cases}\qquad \theta>0$$

由例4可知，$\bar{X}$ 与 $n\min\{X_1,X_2,\cdots,X_n\}$ 都是θ的无偏估计量,问哪个估计量有效?

解 $D(\bar{X})=\dfrac{\theta^2}{n}$ $\quad D(n\min\{X_1,X_2,\cdots,X_n\})=\theta^2$

$\bar{X}$ $\quad n\min\{X_1,X_2,\cdots,X_n\}$有效.

例9 设母体 X 的密度函数为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}} & x>0, \\ 0 & x\le 0\end{cases}\qquad \theta>0$$

$(x_1,x_2,\cdots,x_n)$ 为 X 的一个子样值.

似然估计量, 并判断它是否 θ 的优效估计.

解 由似然函数

$$L(\theta)=\frac{1}{\theta^n}e^{-\frac{\sum_{i=1}^{n}x_i}{\theta}} \Rightarrow \ln L(\theta)=-n\ln\theta-\frac{\sum_{i=1}^{n}x_i}{\theta}$$

$$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta) = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2} \overset{令}{=} 0$$

$$\longrightarrow \quad \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$$

θ的最大似然估计量为 $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$

它是θ的无偏估计量.

$$D(\hat{\theta}) = D\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{\theta^2}{n}$$

$$\ln f(x,\theta)=-\ln\theta-\frac{x}{\theta}\Rightarrow\left(\frac{\partial\ln f(x,\theta)}{\partial\theta}\right)^2=\left(-\frac{1}{\theta}+\frac{x}{\theta^2}\right)^2$$

$$E\left(\frac{\partial\ln f(X,\theta)}{\partial\theta}\right)^2=E\left(-\frac{1}{\theta}+\frac{X}{\theta^2}\right)^2=\frac{1}{\theta^2}$$

$$\Rightarrow\quad I_R=\frac{1}{nE\left(\frac{\partial\ln f(X,\theta)}{\partial\theta}\right)^2}=\frac{\theta^2}{n}=D(\overline{X})$$

故 $\overline{X}$ 是 θ 的优效估计.

例10 设正态母体 $X \sim N(\mu, \sigma^2)$，问 $\overline{X}$ 和 S^{*2} 分别是 μ 和 σ^2 的优效估计吗？

解 对于 μ

$$\int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x)}{\partial \mu}\right)^2 f(x)dx = \int_{-\infty}^{\infty}\left(\frac{x-\mu}{\sigma^2}\right)^2 \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

$$\overset{y=\frac{x-\mu}{\sigma}}{=} \frac{1}{\sigma^2}\int_{-\infty}^{\infty} y^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = \frac{1}{\sigma^2} \Rightarrow I_R = \frac{\sigma^2}{n}$$

$$E\overline{X} = \mu, \quad D\overline{X} = \sigma^2 / n = I_R$$

$\overline{X}$ 是 μ 的优效估计.

对于 σ^2

$$\int_{-\infty}^{\infty}\left(\frac{\partial \ln f(x)}{\partial \sigma^2}\right)^2 f(x)dx = \int_{-\infty}^{\infty}\left(\frac{(x-\mu)^2}{2\sigma^4}-\frac{1}{2\sigma^2}\right)^2 \frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$y=\frac{x-\mu}{\sigma} \quad \frac{1}{4\sigma^4}\int_{-\infty}^{\infty}\left(y^2-1\right)^2\frac{1}{\sqrt{2\pi}}e^{-\frac{y^2}{2}}dy = \frac{1}{2\sigma^4}$$

σ^2 的罗-克拉美下界 $I_R=\dfrac{2\sigma^4}{n}$

$$ES^{*2}=\sigma^2, \qquad DS^{*2}=\frac{2}{n-1}\sigma^4 > I_R$$

所以 S^{*2} 不是 σ^2 的优效估计.

例11 设总体 $X \sim N(\mu, \sigma^2)$, $(X_1, X_2, \cdots, X_n)$ 为 X 的一个样本,常数 k 取何值可使 $k\sum_{i=1}^{n}|X_i - \overline{X}|$ 为 σ 的无偏估计量

解 $$E\left(k\sum_{i=1}^{n}|X_i - \overline{X}|\right) = k\left(\sum_{i=1}^{n}E|X_i - \overline{X}|\right)$$

注意到 $X_i - \overline{X}$ 是 $X_1, X_2, \ldots, X_n$ 的线性函数,

$$X_i - \overline{X} = \frac{1}{n}\left(-X_1 - X_2 \cdots + (n-1)X_i - \cdots - X_n\right)$$

$$E(X_i - \overline{X}) = 0, \quad D(X_i - \overline{X}) = \frac{n-1}{n}\sigma^2$$

$$X_i - \overline{X} \sim N\left(0, \frac{n-1}{n}\sigma^2\right)$$

$$E(|X_i - \overline{X}|) = \int_{-\infty}^{+\infty} |z| \frac{1}{\sqrt{2\pi}\sqrt{\frac{n-1}{n}}\sigma} e^{-\frac{z^2}{2\frac{n-1}{n}\sigma^2}} dz$$

$$= 2\int_0^{+\infty} z \frac{1}{\sqrt{2\pi}\sqrt{\frac{n-1}{n}}\sigma} e^{-\frac{z^2}{2\frac{n-1}{n}\sigma^2}} dz$$

$$= \frac{2}{\sqrt{2\pi}}\sqrt{\frac{n-1}{n}}\sigma$$

$$E\left(k\sum_{i=1}^{n}|X_i-\overline{X}|\right)=k\left(\sum_{i=1}^{n}E|X_i-\overline{X}|\right)$$

$$=kn\frac{2}{\sqrt{2\pi}}\sqrt{\frac{n-1}{n}}\sigma \overset{令}{=} \sigma$$

$$\Longrightarrow \quad k=\sqrt{\frac{\pi}{2n(n-1)}}$$

注 解题中最易发生错误是 $D(X_i-\overline{X})=DX_i+D\overline{X}=\frac{n+1}{n}\sigma^2$

X_i 与 $\overline{X}$ 不独立

3. 均方误差估计最小（MSE）

定义 8 若 $E(\hat{\theta}_1-\theta)^2 \le E(\hat{\theta}_2-\theta)^2, \theta\in\Theta$，至少有一个 $\theta_0\in\Theta$ 使得不等式成立，则称在均方误差意义下 $\hat{\theta}_1$ 优于 $\hat{\theta}_2$。

注：① 若 $\hat{\theta}$ 是无偏估计，即 $E\hat{\theta}=\theta$，则 $MSE(\hat{\theta})=D\hat{\theta}$。有效估计是均方误差最小的估计

② 均方误差一致达到最小的最优估计不存在。$\theta\in\Theta$ 变化 $\theta=\theta_0$ 时取 $\hat{\theta}\equiv\theta_0$

均方误差反映了估计量 $\hat{\theta}$ 与被估参数 θ 的平均（平方）误差。均方误差越小说明估计的效果越好，均方误差越大则说明估计的效果越差。

例 8 总体 $X\sim N(\mu,\sigma^2)$，样本 $(X_1,X_2,\cdots,X_n)$，方差 σ^2 的估计

$$\hat{\sigma}_1^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2=S^2,\quad \hat{\sigma}_2^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=S_n^2,\quad \hat{\sigma}_3^2=\frac{1}{n+1}\sum_{i=1}^{n}(X_i-\overline{X})^2$$

解：$E\hat{\sigma}_1^2=ES^2=\sigma^2$，$\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$，$D\left[\dfrac{(n-1)S^2}{\sigma^2}\right]=2(n-1)$，$DS^2=\dfrac{2\sigma^4}{n-1}$

$$MSE(\hat{\sigma}_1^2)=D\hat{\sigma}_1^2=DS^2=\frac{2\sigma^4}{n-1}$$

$$MSE(\hat{\sigma}_2^2)=D\hat{\sigma}_2^2+(E\hat{\sigma}_2^2-\sigma^2)^2=DS_n^2+(ES_n^2-\sigma^2)^2=D\left(\frac{n-1}{n}S^2\right)+(ES_n^2-\sigma^2)^2$$

$$=\frac{(n-1)^2}{n^2}DS^2+\left(\frac{n-1}{n}\sigma^2-\sigma^2\right)^2=\frac{(n-1)^2}{n^2}\cdot\frac{2\sigma^4}{n-1}+\frac{\sigma^4}{n^2}=\frac{(2n-1)\sigma^4}{n^2}$$

$$MSE(\hat{\sigma}_3^2)=D\hat{\sigma}_3^2+(E\hat{\sigma}_3^2-\sigma^2)^2=D\left(\frac{n-1}{n+1}S^2\right)+\left[E\left(\frac{n-1}{n+1}S^2\right)-\sigma^2\right]$$

$$=\frac{(n-1)^2}{(n+1)^2}DS^2+\left(\frac{n-1}{n+1}\sigma^2-\sigma^2\right)^2=\frac{(n-1)^2}{(n+1)^2}\cdot\frac{2\sigma^4}{n-1}+\frac{4\sigma^4}{(n+1)^2}=\frac{2\sigma^4}{n+1}$$

$MSE(\hat{\sigma}_1^2)>MSE(\hat{\sigma}_2^2)>MSE(\hat{\sigma}_3^2)$，在均方误差意义下，$\hat{\sigma}_3^2$ 优于 $\hat{\sigma}_2^2$，$\hat{\sigma}_2^2$ 优于 $\hat{\sigma}_1^2$。

4、相合性（一致性（consistency））

定义 设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$是母体参数$\theta$的估计量. 若对于任意的$\theta\in\Theta$, 当$n\to\infty$时,

$\hat{\theta}$ 依概率收敛于θ, 即 $\forall\varepsilon>0,$

$$\lim_{n\to\infty}P(\left|\hat{\theta}-\theta\right|\geq\varepsilon)=0$$

$\hat{\theta}$ 是母体参数θ的相合(或一致)估计量.

相合性估计量仅在样本容量 n 足够大时, 才显示其优越性.

关于相合性的两个常用结论

1. 子样 k 阶矩是母体 k 阶矩的相合估计量.

由大数定律证明

2. 设 $\hat{\theta}$ 是 θ 的无偏估计量，且 $\lim\limits_{n\to\infty} D(\hat{\theta}) = 0$，则 $\hat{\theta}$ 是 θ 的相合估计量.

用切贝雪夫不等式证明

矩法得到的估计量一般为相合估计量

在一定条件下，极大似然估计具有相合性

（2）相合估计的判断方法

定理 设$\hat{\theta}_n$是θ的一个估计量，若$\lim\limits_{n\to\infty} E\hat{\theta}_n=\theta$，$\lim\limits_{n\to\infty} D\hat{\theta}_n=0$，则$\hat{\theta}_n$是$\theta$的相合估计。

证：$\forall\varepsilon>0$，$0\le P\left(\left|\hat{\theta}_n-\theta\right|\ge\varepsilon\right)=\int\limits_{\left|\hat{\theta}_n-\theta\right|\ge\varepsilon} dF(x)\le\int\limits_{\left|\hat{\theta}_n-\theta\right|\ge\varepsilon}\frac{(\hat{\theta}_n-\theta)^2}{\varepsilon^2}dF(x)\le\frac{1}{\varepsilon^2}\int_{-\infty}^{+\infty}(\hat{\theta}_n-\theta)^2 dF(x)$

$$=\frac{1}{\varepsilon^2}E(\hat{\theta}_n-\theta)^2=\frac{1}{\varepsilon^2}\left[D\hat{\theta}_n+(E\hat{\theta}_n-\theta)^2\right]\xrightarrow{n\to\infty}0$$

即$\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n}X_i^k-\mu_k\right|\ge\varepsilon\right)=0$，$\hat{\theta}_n$是$\theta$的相合估计。

例 9 样本均值 $\bar{X}$ 是总体均值 EX 的相合估计，样本矩 A_k 是总体矩 EX^k 的相合估计。

解：$E\bar{X}=EX$，$\lim\limits_{n\to\infty} D\bar{X}=\lim\limits_{n\to\infty}\frac{DX}{n}=0$，即 $\bar{X}$ 是 EX 的相合估计。

$$EA_k=E\left[\frac{1}{n}\sum_{i=1}^{n}X_i^k\right]=\frac{1}{n}\sum_{i=1}^{n}EX_i^k=\frac{1}{n}\sum_{i=1}^{n}EX^k=EX^k$$

$$DA_k=D\left[\frac{1}{n}\sum_{i=1}^{n}X_i^k\right]=\frac{1}{n^2}\sum_{i=1}^{n}DX_i^k=\frac{1}{n^2}\sum_{i=1}^{n}DX^k=\frac{DX^k}{n}\xrightarrow{n\to\infty}0$$

即样本矩 A_k 是总体矩 EX^k 的相合估计。

（3）不变原则

$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n$是$\theta_1,\theta_2,\cdots,\theta_n$的相合估计，$g(\theta_1,\theta_2,\cdots,\theta_n)$连续，则$g(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n)$是$g(\theta_1,\theta_2,\cdots,\theta_n)$的相合估计。

注：相合估计具有线性性，即若$\hat{\theta}_1$是θ_1的相合估计，$\hat{\theta}_2$是θ_2的相合估计，则$a\hat{\theta}_1+b\hat{\theta}_2$是$a\theta_1+b\theta_2$的相合估计。

例 10 总体X，样本$(X_1,X_2,\cdots,X_n)$，$\nu_k=E(X-EX)^k$，$B_k=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$，则$B_k$是$\nu_k$的相合估计。

证：$\nu_k=E(X-EX)^k=E(X-\mu_1)^k=\sum_{i=0}^{k}C_k^i\mu_i(-\mu_1)^{k-i}$， $\mu_0=1$

$$B_k=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k=\frac{1}{n}\sum_{i=1}^{n}(X_i-A_1)^k=\frac{1}{n}\sum_{i=1}^{n}\sum_{j=0}^{k}C_k^jX_i^j(-A_1)^{k-j}$$
$$=\sum_{j=0}^{k}C_k^j\left(\frac{1}{n}\sum_{i=1}^{n}X_i^j\right)(-A_1)^{k-j}=\sum_{j=0}^{k}C_k^jA_j(-A_1)^{k-j}$$
， $A_0=1$

由不变原则，B_k是ν_k的相合估计。特别地，$S_n^2=B_2$是$\sigma^2=\nu_2$的相合估计。

例 $X \sim f(x;\theta) = \begin{cases} \dfrac{1}{\theta} e^{-\frac{x}{\theta}} & x > 0, \\ 0 & x \le 0 \end{cases}$ $\theta > 0$

则 $\overline{X}$ 是θ的无偏、优效、相合估计量.

证 由前例知 $\overline{X}$是θ的无偏、优效估计量.

$$\lim_{n\to\infty} D(\overline{X}) = \lim_{n\to\infty} \frac{\theta^2}{n} = 0$$

所以$\overline{X}$ 是 θ 的相合估计量, 证毕.

例 11 $X_1, X_2, \cdots, X_n$ 是来自均匀分布 $U[0,\theta]$ 的样本，$\theta > 0$。

（1）求 θ 的矩法估计 $\hat{\theta}_1$ 和极大似然估计 $\hat{\theta}_2$；

（2）证明矩法估计 $\hat{\theta}_1$ 是 θ 的无偏、相合估计；

（3）极大似然估计 $\hat{\theta}_2$ 是否为 θ 无偏、相合估计？若不是，构造 θ 的无偏估计 $\hat{\theta}_3$；

（4）比较 $\hat{\theta}_1, \hat{\theta}_3$ 的有效性；（5）比较 $\hat{\theta}_1, \hat{\theta}_2$ 的均方误差。

解：（1）$EX = \dfrac{\theta}{2}$，$\bar{X} = \dfrac{\hat{\theta}_1}{2}$，$\hat{\theta}_1 = 2\bar{X}$，$\theta$ 的矩法估计 $\hat{\theta}_1 = 2\bar{X}$，

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & 0 < x_i < \theta, i = 1, 2, \cdots, n \\ 0, & otherwise \end{cases}$$

要使似然函数 $L(\theta)$ 达到最大值，θ 应尽量小，且 $0 < x_i < \theta, i = 1, 2, \cdots, n$，即 $\hat{\theta}_2 = x_{(n)}$ 时 $L(\theta)$ 最大，故 θ 的 MLE 为 $\hat{\theta}_2 = X_{(n)}$。

（2） $E\hat{\theta}_1 = 2E\overline{X} = 2EX = 2\cdot\frac{\theta}{2} = \theta$，所以 $\hat{\theta}_1 = 2\overline{X}$ 是 θ 的无偏估计.

$\overline{X}$ 是 $\frac{\theta}{2}$ 的相合估计，则 $\hat{\theta}_1 = 2\overline{X}$ 是 θ 的相合估计.

（3） $F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{\theta}, & 0 \le x \le \theta \\ 1, & x > \theta \end{cases}$， $p(x) = \begin{cases} \frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & otherwise \end{cases}$

$$F_n(x) = F^n(x),\quad p_n(x) = nF^{n-1}(x)p(x) = \begin{cases} n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}, & 0 \le x \le \theta \\ 0, & otherwise \end{cases}$$

$$E\hat{\theta}_2 = EX_{(n)} = \int_0^\theta x\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx = \frac{n}{\theta^n}\int_0^\theta x^n dx = \frac{n}{n+1}\frac{x^{n+1}}{\theta^n}\Bigg|_0^\theta = \frac{n}{n+1}\theta \ne \theta$$

所以 $\hat{\theta}_2 = X_{(n)}$ 不是 θ 的无偏估计， $\hat{\theta}_3 = \frac{n+1}{n}X_{(n)}$ 是 θ 的无偏估计。

$$E\hat{\theta}_2^2 = EX_{(n)}^2 = \int_0^\theta x^2\cdot n\left(\frac{x}{\theta}\right)^{n-1}\frac{1}{\theta}dx = \frac{n}{\theta^n}\int_0^\theta x^{n+1}dx = \frac{n}{n+2}\frac{x^{n+2}}{\theta^n}\Bigg|_0^\theta = \frac{n}{n+2}\theta^2$$

$$D\hat{\theta}_2 = E\hat{\theta}_2^2 - \left[E\hat{\theta}_2\right]^2 = \frac{n}{n+2}\theta^2 - \left[\frac{n}{n+1}\theta\right]^2 = \frac{n}{(n+1)^2(n+2)}\theta^2$$

$$D\hat{\theta}_3 = D\left(\frac{n+1}{n}X_{(n)}\right) = \frac{(n+1)^2}{n^2}DX_{(n)} = \frac{(n+1)^2}{n^2}\frac{n}{(n+1)^2(n+2)}\theta^2 = \frac{1}{n(n+2)}\theta^2$$

$\hat{\theta}_2 = X_{(n)}$， $\hat{\theta}_3 = \frac{n+1}{n}X_{(n)}$ 是 θ 的相合估计。

（4）$D\hat{\theta}_1 = D(2\bar{X}) = 4D\bar{X} = 4\cdot\frac{DX}{n} = 4\cdot\frac{1}{n}\cdot\frac{\theta^2}{12} = \frac{\theta^2}{3n}$， $D\hat{\theta}_3 = \frac{1}{n(n+2)}\theta^2$

$D\hat{\theta}_1 \geq D\hat{\theta}_3$， $n \geq 2$ 时 $D\hat{\theta}_1 > D\hat{\theta}_3$， $\hat{\theta}_3$ 比 $\hat{\theta}_1$ 有效。

（5）$MSE(\hat{\theta}_1) = E(\hat{\theta}_1 - \theta)^2 = D\hat{\theta}_1 = \frac{\theta^2}{3n}$

$$MSE(\hat{\theta}_2) = E(\hat{\theta}_2 - \theta)^2 = D\hat{\theta}_2 + (E\hat{\theta}_2 - \theta)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2 + \left(\frac{n}{n+1}\theta - \theta\right)^2$$

$$= \frac{n}{(n+1)^2(n+2)}\theta^2 + \frac{1}{(n+1)^2}\theta^2 = \frac{2}{(n+1)(n+2)}\theta^2$$

$MSE(\hat{\theta}_1) \geq MSE(\hat{\theta}_2)$， $n \geq 3$ 时 $MSE(\hat{\theta}_1) > MSE(\hat{\theta}_2)$， $\hat{\theta}_2$ 优于 $\hat{\theta}_1$。

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

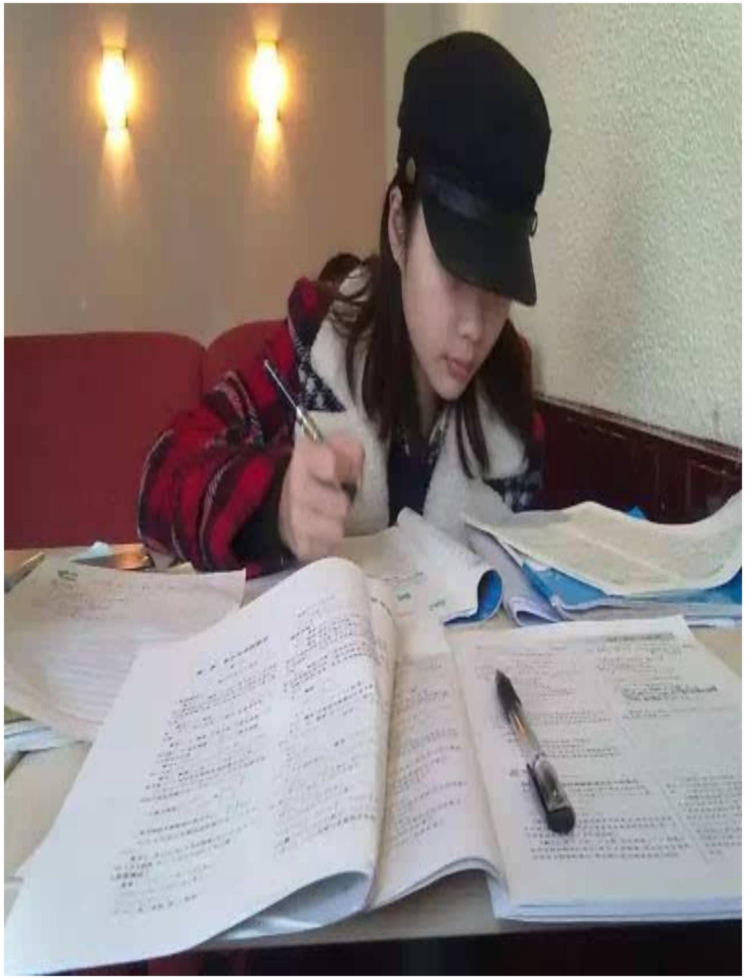
(1) 若 是$\boldsymbol{\theta}$的无偏估计，则 ；
这说明用方差考察无偏估计有效性是合理的。

(2) 当 不是$\boldsymbol{\theta}$的无偏估计时，就要看其均方误差 $MSE(\hat{\theta})$，

下面的例子说明：在均方误差的含义下有些有偏估计优于无偏估计。

第5次作业:

- 孙上p. 59 习题二
- 16、19、21、22、25、28.



谢 谢！