数理统计

-----CH3　参数估计

$$
\text{估计}\begin{cases}\text{非参数估计（经验分布函数，频率直方图）}\\ \text{参数估计}\begin{cases}\text{点估计}\\ \text{区间估计}\end{cases}\end{cases}
$$

第三章：　参数估计

– 矩法估计

– 极大似然估计

– 点估计的评价

– 区间估计

数理统计的基本问题是如何根据样本所提供的信息，对于总体的分布以及分布的某些数字特征进行统计推断

根据样本，对于分布中的未知参数进行估计，这类问题常称为参数估计问题

$$\text{参数估计}\begin{cases} \text{点估计——估计参数的值} \\ \text{区间估计——估计参数的范围} \end{cases}$$

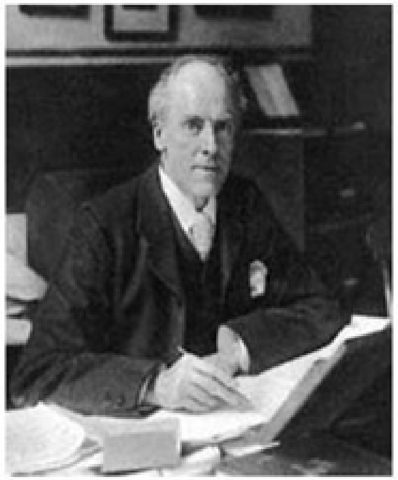
设θ为未知参数，找统计量 $T = T(X_1, X_2 \cdots, X_n)$

将$T = T(X_1, X_2 \cdots, X_n)$作为$\theta$的估计，记为:

$\hat{\theta} = T(X_1, X_2 \cdots, X_n).---\theta$的点估计

矩法估计

矩法估计是一种最早给出的、也是最容易想到的点估计的方法，这种方法是由卡尔皮尔逊首先提出的



卡尔皮尔逊（Karl Pearson，1857-1936），现代数理统计学的创始人。英国人，1879年毕业于剑桥大学。1884年，被聘为英国伦敦大学教授，担任数学和力学课程的教学，不久之后，开始对生物学产生兴趣，尝试用数学方法，特别是用概率论的方法，研究生物学，在研究中逐步建立起一套理论和方法，奠定了现代数理统计学的基础，1902年，皮尔逊在Biometrika上发表了一篇论文，在论文中首次提出了矩法估计的思想。

矩法估计

----用样本矩替换总体矩

矩法估计的基本思想：总体矩是从总体分布计算出来的，其中必然包含总体分布中的未知参数，所以，可以让总体矩的估计等于样本矩，列出若干个灯饰，组成一个含有未知参数的方程组。一般来说，只要一个方程组中等式的个数等于未知数的个数，就可以从这个方程组中把未知数求解出来。

求矩法估计的步骤为：

(1)计算总体分布的矩$E(\xi^k) = f_k(\theta_1, \theta_2, ..., \theta_m), k = 1, 2, ..., m,$计算到$m$ 阶矩为止

（m 是总体分布中未知参数的个数）

(2)列方程

$$\begin{cases} f_1(\theta_1, \theta_2, ..., \theta_m) = E\xi = \overline{X} \\ f_2(\theta_1, \theta_2, ..., \theta_m) = E(\xi^2) = \overline{X^2} \\ \vdots \\ f_m(\theta_1, \theta_2, ..., \theta_m) = E(\xi^m) = \overline{X^m} \end{cases}$$

从方程中解出$\hat{\theta}_1, \hat{\theta}_2, ..., \hat{\theta}_m$，它们就是未知参数$\theta_1, \theta_2, ..., \theta_m$ 的矩法估计

例1. 设$X_1, X_2, \cdots, X_n$为来自总体X的样本，

已知总体 X的期望和方差均存在， 但未知

令$EX = \mu, DX = \sigma^2$。求μ、σ^2与σ的矩估计。

解：解方程组：$\begin{cases} EX = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i \\ EX^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases}$ 即 $\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 + \hat{\mu}^2 = \overline{X^2} \end{cases}$

所求矩估计为 $\hat{\mu} = \overline{X}$, $\hat{\sigma}^2 = \overline{X^2} - (\overline{X})^2$,

$$\hat{\sigma} = \sqrt{\overline{X^2} - (\overline{X})^2}$$

注意：期望、方差的矩估计没有涉及总体的分布。

已知分布类型，参数未知。求参数的矩估计

例2. 设 $X_1, X_2, \cdots$，X_n为来自总体X的样本，

已知总体 $X \sim E(\lambda)$，求λ的矩估计。

解： $EX = \dfrac{1}{\hat{\lambda}} = \overline{X}$，

$$\Rightarrow \hat{\lambda} = \frac{1}{\overline{X}}$$

即λ的矩估计为 $\dfrac{1}{\overline{X}}$。

例3．设 $X_1, X_2, \cdots, X_n$为来自总体X的样本,

已知总体X服从[0，θ]上的均匀分布,$\theta > 0$

未知，求θ 的矩估计。

解：$EX = \dfrac{\hat{\theta}}{2} = \overline{X}, \quad \Rightarrow \hat{\theta} = 2\overline{X}$

即θ的矩估计为 $2\overline{X}$。

若（x_1, x_2, x_3, x_4, x_5）=（1,2,3,5,9),

$$\overline{X} = \frac{1+2+3+5+9}{5} = 4$$

$$\Rightarrow \hat{\theta} = 2\overline{X} = 8$$

例4. 设总体ξ的密度函数为:

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

求参数λ, μ的矩法估计

解: $$E\xi=\int_{-\infty}^{+\infty}xp(x)dx=\int_{\mu}^{+\infty}x\lambda e^{-\lambda(x-\mu)}dx=\mu+\frac{1}{\lambda}$$

$$E\xi^2=\int_{-\infty}^{+\infty}x^2p(x)dx=\int_{\mu}^{+\infty}x^2\lambda e^{-\lambda(x-\mu)}dx=(\mu+\frac{1}{\lambda})^2+\frac{1}{\lambda^2}$$

令: $$\begin{cases}E\xi=\overline{X}\\ E\xi^2=\overline{X^2}\end{cases}$$ 即 $$\begin{cases}\mu+\frac{1}{\lambda}=\overline{X}\\ (\mu+\frac{1}{\lambda})^2+\frac{1}{\lambda^2}=\overline{X^2}\end{cases}$$

解得: $$\begin{cases}\hat{\lambda}=\dfrac{1}{\sqrt{\overline{X^2}-\overline{X}^2}}=\dfrac{1}{S}\\ \hat{\mu}=\overline{X}-\dfrac{1}{\hat{\lambda}}=\overline{X}-S\end{cases}$$

思考题

总体X的分布律为:

X	-1	0	1
P	a	1-2a	a

求a的矩法估计

提示：因 EX=0，可由 $EX^2=\overline{X^2}$ 解出a

矩估计的优缺点

优点：计算简单。

缺点：

（1）总体的矩不一定存在。故矩估计不一定可行。

（2）可能会有不同的矩估计。

规定：尽量使用低阶矩。

（3）可能会得到不合理的解。

极大似然估计

极大似然估计的思想一般认为是高斯(Gauss)首先提出的,也有学者认为极大似然估计方法是Fisher(下图)提出的,应归功于Fisher的工作.



费希尔（R. A. Fisher，1890-1962）数理统计学创始人之一，英国人。1909年，费希尔因数学成绩优秀，得到奖学金资助，如剑桥大学数学系学习。在大学学习期间，对数理统计产生了浓厚兴趣，课余经常阅读皮尔逊主编的Biometrika杂志。1912年，费希尔在读大学三年级时，提出了极大似然估计的思想。

极大似然原理

一个随机试验如果有若干个可能的结果 $A, B, C, \cdots$，而在一次试验中，结果 A 出现了，则一般认为试验的机理最有利于结果 A 出现,即使得 A 出现的可能性最大。

例5：从一批产品中抽出5件检查，发现有前2件是次品，问这批产品的次品率是多少？

设产品的次品率为 $\boldsymbol{p}$，$\boldsymbol{p \in (0, 1)}$

令 $\xi_i = \begin{cases} 1 & \text{第i次抽到次品} \\ 0 & \text{否则} \end{cases} \quad (i = 1, 2, 3, 4, 5),$

$$\boldsymbol{P(\xi_1 = 1, \xi_2 = 1, \xi_3 = 0, \xi_4 = 0, \xi_5 = 0) = p^2(1 - p)^3}$$

P取何值时上式有极大值 **?**

上式对**p** 求导**,**令导数为零**,**得**p=0.4**

设$X_1, X_2, \cdots, X_n$为来自总体X的一组样本，

$\boxed{\text{样本值出现可能性}}$

(1)若总体 X是离散型随机变量，

$$\boxed{P\{X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n\}} = \prod_{i=1}^{n} P\{X_i = x_i\}$$

(2)若总体X是连续型随机变量，概 率密度为$\varphi(x)$,

与$\boxed{\varphi^*(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} \varphi(x_i)}$成正比。

（1）、（2）中的分布概率或密度函数均含有未知参数。

问题转化：

当参数取何值时，样本分布概率或概率密度函数值最大？

称以上样本的联合概率分布或联合密度为极大似然函数，

记为L。

似然函数：

离散型

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i)$$

连续型

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

极大似然估计：寻找统计量$\hat{\theta}$，使其满足

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

为了方便乘积求导，常常改求

$$\ln L(\hat{\theta}) = \max_{\theta \in \Theta} \ln L(\theta)$$

一般步骤：

1：写出似然函数 $L(\theta)$。

2：将似然函数取对数，求得对数似然函数 $\ln L(\theta)$。

3：$\ln L(\theta)$ 关于 θ 求导（若 θ 为多个参数，则分别关于每个参数求偏导），并令其为 **0**，得对数似然方程组。

4：解上述似然方程组，求得 θ 的极大似然估计 $\hat{\theta}$。

例 6　设总体$\xi \sim P(\lambda)$，具有分布列

$$P(\xi = k) = \frac{e^{-\lambda} \cdot \lambda^k}{k!},\ k = 0,1,2,\cdots$$

$\lambda > 0$为一未知参数。求λ的极大似然估计。

解　设 $(X_1, X_2, \cdots, X_n)$ 为一个样本，

$(x_1, x_2, \cdots, x_n)$为一组观测值。似然函数为

$$L(\lambda) = \prod_{i=1}^{n} P(X_i = x_i) = \prod_{i=1}^{n} \left(\frac{e^{-\lambda} \cdot \lambda^{x_i}}{x_i!} \right)$$

两边取对数，得 $\ln L(\lambda) = \sum_{i=1}^{n}(-\lambda + x_i \ln\lambda - \ln x_i!)$

上式两侧关于 λ 求导，有 $\dfrac{d\ln L}{d\lambda} = \sum_{i=1}^{n}(-1 + \dfrac{x_i}{\lambda}) \overset{令}{=} 0$

解上式，可得 $\hat{\lambda} = \overline{x}$

由于 $\left.\dfrac{d^2 \ln L}{d\lambda^2}\right|_{\lambda=\overline{x}} < 0$，可知 $\hat{\lambda}$ 使 L 达到最大，从而得出 λ 的极大似然估计为 $\hat{\lambda} = \overline{X}$。

例 7 设$(X_1, X_2, \cdots, X_n)$为取自正态总体$N(\mu, \sigma^2)$的一个样本，其中μ, σ^2为未知参数，试求参数μ与σ^2的极大似然估计。

解 设$(x_1, x_2, \cdots, x_n)$为样本的一组观测值，似然函数为

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} p(x_i) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}\right)$$

$$\ln L(\mu,\sigma^2)=\sum_{i=1}^{n}\{-\frac{1}{2}\ln(2\pi)-\frac{1}{2}\ln\sigma^2-\frac{(x_i-\mu)^2}{2\sigma^2}\}$$

关于μ,σ^2求偏导，并令其为**0**，得似然方程组

$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0\\[2ex]\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2\sigma^2}=0\end{cases}$$

解此方程组，可得μ,σ^2的极大似然估计为

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},\qquad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2=S^2$$

当无法通过求导求极值时：

写出似然函数$L(\theta)$。

求得θ的极大似然估计$\hat{\theta}$，使得

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta) \qquad \text{(考虑}\theta\text{取值的边界)}$$

例 **8** 设总体$\xi \sim U[0,\theta]$， $\theta > 0$为未知参数。

试求θ的极大似然估计。

解 设$(X_1, X_2, \cdots, X_n)$为一个样本，$(x_1, x_2, \cdots, x_n)$

为一组观测值。似然函数为

$$L(\theta) = \prod_{i=1}^{n} \left(\frac{1}{\theta}\right) = \frac{1}{\theta^n}。$$

两边取对数，得对数似然函数

$$\ln L(\theta) = -n \ln \theta$$

关于θ求导，可得$\dfrac{d\ln L(\theta)}{d\theta}=\dfrac{-n}{\theta}<0,\quad(\theta>0)$

$\ln L(\theta)$关于θ严格单调递减，即θ越小$\ln L(\theta)$越大.但

$$0\le x_1\le\theta,0\le x_2\le\theta,\ldots,0\le x_n\le\theta$$

可知$\theta\ge\max\limits_{1\le i\le n}x_i$，即$\theta=\max\limits_{1\le i\le n}x_i$时，$L(\theta)$取到最大值。

因此θ的极大似然估计为：$\hat{\theta}=\max\limits_{1\le i\le n}x_i$

如果$\hat{\theta}$是θ的极大似然估计，则对任何函数$g(\theta)$，

其极大似然估计为$g(\hat{\theta})$ --“不变性”

例 9 设$(X_1, X_2, \cdots X_n)$是来自正态总体$N(\mu, \sigma^2)$

的样本，求标准差σ以及变异系数$\gamma \overset{\Delta}{=} \sigma / \mu$的极大似然估计。

解 根据前例中已求的参数 μ 与 σ^2 的极大似然估计

$$\hat{\mu} = \bar{x}，\quad \hat{\sigma}^2 = S^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2，$$

所以σ以及$\gamma \overset{\Delta}{=} \sigma / \mu$的极大似然估计为

$$\hat{\sigma} = S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \hat{\gamma} = \frac{\hat{\sigma}}{\hat{\mu}} = \frac{S}{\bar{x}}$$

课堂练习:

设随机变量X分布律如下(其中 $0<a<1$)

X	-1	0	1
P	a/2	1-a	a/2

样本观测值为(0,0,1,-1,1)

求参数a的矩法估计和极大似然估计

解答:

因 EX=0（不含参数a），

因此不能用 $EX = \bar{X}$ 来求a的矩法估计,而要用 $EX^2 = \overline{X^2}$

$$EX^2 = (-1)^2 \tfrac{a}{2} + 0^2 (1-a) + 1^2 \tfrac{a}{2} = a$$

$$\overline{X^2} = \frac{0^2 + 0^2 + 1^2 + (-1)^2 + 1^2}{5} = \frac{3}{5}$$

由 $EX^2 = \overline{X^2}$, 得a的矩法估计 $\hat{a} = \dfrac{3}{5}$

极大似然估计:

$$L(a) = \mathrm{P}\{X_1 = 0, X_2 = 0, X_3 = 1, X_4 = -1, X_5 = 1\}$$

$$= (1-a)^2 (\tfrac{a}{2})^3$$

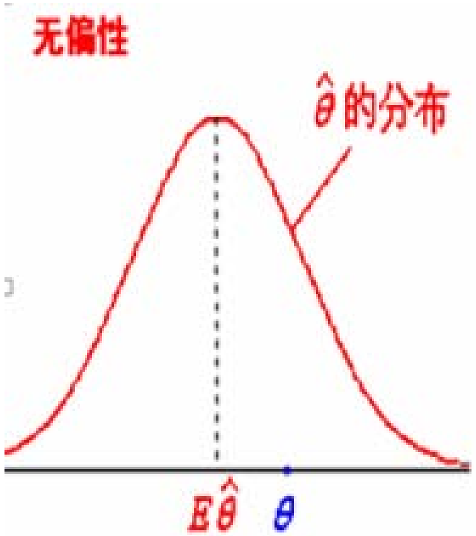
$$\frac{dL(a)}{da} = 0 \quad \Rightarrow \quad \hat{a} = \frac{3}{5} \text{ (极大似然估计)}$$

点估计的评价

问题：参数的点估计有不同的方法，得到的结果可能是不同的.

那么如何来评价点估计的优劣?

无 偏 性



参数θ的点估计$\hat{\theta}$是一个随机变量，它有自己的分布，从分布可以求出一个平均值，即数学期望$E\hat{\theta}$，既然$\hat{\theta}$是θ的估计，我们自然希望它的数学期望（均值）正好等于θ这就是无偏性。

定义：设$\hat{\theta}$是参数θ的估计，如果有$E\hat{\theta}=\theta$，则称$\hat{\theta}$是θ的无偏估计。

例10 设总体$\xi \sim N(\mu,\sigma^2)$，前面我们已求得μ 的估计$\hat{\mu}=\bar{X}$ ，σ^2 的估计$\hat{\sigma}^2=S^2$ ，问$\hat{\mu}=\bar{X}$ 和$\hat{\sigma}^2=S^2$是不是μ 和σ^2 的无偏估计？

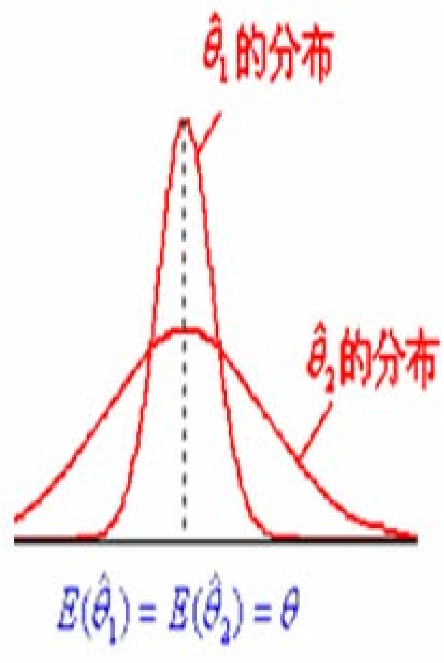
解: $E\hat{\mu}=E\bar{X}=E\xi=\mu$,所以$\hat{\mu}=\bar{X}$是$\mu$的无偏估计。

而$E(\hat{\sigma}^2)=\mathrm{E}(\mathrm{S}^2)=\dfrac{n-1}{n}D\xi=\dfrac{n-1}{n}\sigma^2\neq\sigma^2$,所以$\hat{\sigma}^2=S^2$不是$\sigma^2$ 的无偏估计

但是，只要对它稍作修正，用修正样本方差$S^{*2}=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2=\dfrac{n}{n-1}S^2$ 代替S^2作为σ^2的估计，由于$E(S^{*2})=D\xi=\sigma^2$,所以S^{*2}是σ^2的无偏估计。

有 效 性

同样是无偏估计，也可以比较好坏，无偏而且方差小显然要比无偏但是方差大来得好。



定义：设$\hat{\theta}_1$和$\hat{\theta}_2$都是参数θ的无偏估计，如果有$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，则称$\hat{\theta}_1$比$\hat{\theta}_2$有效

例11设总体$\xi \sim N(\mu,\sigma^2)$,(X_1, X_2)是ξ 的一个样本，证明

$\hat{\mu}_1 = \frac{2}{3}X_1 + \frac{1}{3}X_2$，$\hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{2}X_2$都是$\mu$的无偏估计。并比较哪一个估计更有效。

解 因为

$$E\hat{\mu}_1 = \frac{2}{3}EX_1 + \frac{1}{3}EX_2 = \frac{2}{3}E\xi + \frac{1}{3}E\xi = E\xi = \mu,$$

$$E\hat{\mu}_2 = \frac{1}{2}EX_1 + \frac{1}{2}EX_2 = \frac{1}{2}E\xi + \frac{1}{2}E\xi = E\xi = \mu$$

所以$\hat{\mu}_1$，$\hat{\mu}_2$都是μ的无偏估计。

因为

$$D\hat{\mu}_1 = \frac{4}{9}DX_1 + \frac{1}{9}DX_2 = \frac{4}{9}D\xi + \frac{1}{9}D\xi = \frac{5}{9}D\xi = \frac{5}{9}\sigma^2,$$

$$D\hat{\mu}_2 = \frac{1}{4}DX_1 + \frac{1}{4}DX_2 = \frac{1}{4}D\xi + \frac{1}{4}D\xi = \frac{1}{2}D\xi = \frac{1}{2}\sigma^2,$$

而$\frac{1}{2}\sigma^2 < \frac{5}{9}\sigma^2$,即$D\hat{\mu}_2 < D\hat{\mu}_1$,所以$\hat{\mu}_2$比$\hat{\mu}_1$更有效

例 11 设总体 ξ 服从 $[0,\theta]$ 上的均匀分布，$\theta > 0$为未知参数，$X_1, X_2, ..., X_n$是一个样本。

（1） 试验证$\hat{\theta} = 2\overline{X}$ 和 $\hat{\theta}^* = \frac{n+1}{n}\max\limits_i X_i$ 都是 θ的无偏估计

（2） 它们哪一个更有效？

解 （1）因为$\xi \sim U(0,\theta)$,

X的概率密度为

$$\phi(x) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{其他} \end{cases}$$

$\hat{\theta}_L = \max\limits_i X_i$的分布函数为

$$F_{\max}(x) = P\{\max_i X_i \le x\} = \begin{cases} 0 & x \le 0 \\ x^n / \theta^n & 0 < x \le \theta \\ 1 & x > \theta \end{cases}$$

$$p(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n} & 0 < x \le \theta \\ 0 & 其他 \end{cases}$$

$$E(\hat{\theta}^*) = E(\frac{n+1}{n}\hat{\theta}_L) = \frac{n+1}{n}E(\hat{\theta}_L) = \frac{n+1}{n} \cdot \frac{n\theta}{n+1} = \theta$$

$$E(\hat{\theta}) = E(2\bar{X}) = 2E\bar{X} = 2E\xi = 2\cdot\frac{\theta}{2} = \theta$$

$\hat{\theta}$，$\hat{\theta}^*$都是θ的无偏估计

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

$$= \frac{n^3+2n^2+n-n^3-2n^2}{(n+2)(n+1)^2}\theta^2$$

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

$$D(\hat{\theta}^*) = D\left(\frac{n+1}{n}\max_i X_i\right) = D\left(\frac{n+1}{n}\hat{\theta}_L\right) = \frac{(n+1)^2}{n^2}D(\hat{\theta}_L) = \frac{\theta^2}{n(n+2)}$$

因为$n \geq 1, n+2 \geq 3, \dfrac{\theta^2}{n(n+2)} \leq \dfrac{\theta^2}{3n}$,即$D(\hat{\theta}^*) \leq D(\hat{\theta})$，所以

$\hat{\theta}^* = \dfrac{n+1}{n}\max\limits_i X_i$比$\hat{\theta} = 2\bar{X}$更有效。

相合性

定义　设 $\hat{\theta}$ 是参数 θ 的估计，n 是样本容量，如果任何 $\varepsilon > 0$，都有

$$\lim_{n\to\infty} P\{\,|\,\hat{\theta}-\theta\,| < \varepsilon\,\} = 1\ ,$$

则称 $\hat{\theta}$ 是 θ 的**相合估计（一致估计）**。

结论：（1）矩法估计都是相合估计
（2）大多数极大似然估计是相合估计

均 方 误 差

一个综合了无偏性与有效性的评价指标---均方误差

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

思 考 题：

要利用总体X一个容量为n的样本来估计总体分布中含未知参数a，已知张三和李四给出了两种不同的估计，经验证两种估计都是无偏的，且是同样有效的，你能否给出参数a的一种更好的估计？

解答：已知：$E\hat{a}_1 = a \quad E\hat{a}_2 = a$

$$D\hat{a}_1 = D\hat{a}_2$$

令 $\hat{a}_3 = \frac{1}{2}(\hat{a}_1 + \hat{a}_2)$

则 $E\hat{a}_3 = a$

$$D\hat{a}_3 = \frac{1}{4}D(\hat{a}_1 + \hat{a}_2) = \frac{1}{4}[D\hat{a}_1 + D\hat{a}_2 + 2\mathrm{cov}(\hat{a}_1, \hat{a}_2)]$$

$$\leq \frac{1}{4}[2D\hat{a}_1 + 2D\hat{a}_2] = D\hat{a}_1$$

$\boxed{D(\hat{a}_1 - \hat{a}_2) \geq 0}$

即组合后的估计量更有效

其实还可进一步证明上式中的“不大于”可改为“<”