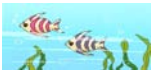
现在进入对统计推断的研究.先介绍一类重要的统计推断问题.

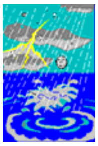
第2章 参数估计

估计新生儿的平均体重 

估计湖中鱼数 

估计废品率 

……

估计平均降雨量 

§2.1 参数估计问题

参数估计问题就是利用从总体抽样得到的信息，来估计总体的某些参数或者参数的某些函数.

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数.

一、点估计

例 已知某地区新生婴儿的体重 $X \sim N(\mu, \sigma^2)$，μ, σ^2未知，随机抽查100个婴儿，得100个体重数据

10, 7, 6, 6.5, 5, 5.2, …

全部信息就是这 100 个数. *如何据此估计μ和σ^2呢？*

问题是: *使用什么样的统计量去估计？*

对于待估参数μ：可用样本均值；也可用样本中位数等，

对于待估参数σ：可用样本方差等，

为估计θ的值，我们需要构造出适当的样本的函数$\hat{\theta}(X_1, X_2, \cdots X_n)$，对于确定的一组样本值，代入函数$\hat{\theta}(X_1, X_2, \cdots X_n)$中算出一个值$\hat{\theta}$，即得到 θ 的一个估计值.

点 估 计

定义1 设$X_1, X_2, \cdots X_n$是来自总体X的样本，$x_1, x_2, \cdots x_n$是其一组样本值，如果总体的分布类型已知，θ_i是总体的未知参数，则称用以估计参数θ_i的统计量$\hat{\theta}_i(X_1, X_2, \cdots X_n)$为参数$\theta_i$的点估计量，简称估计量. $\hat{\theta}_i(x_1, x_2, \cdots x_n)$称为$\theta_i$的点估计值，或估计值.

$i = 1, 2, \cdots, l$.

参数取值范围 —— 参数空间 Θ.

寻求估计量的方法
- 矩估计法
- 极大似然法
- 最小二乘法
- 贝叶斯方法

……

这里我们主要介绍前面两种方法.

二、区间估计

区间估计 —— 设法得到参数空间Θ的一个取值范围，使待估参数以较大的概率含于其内.

例如要估计一个参数θ，就是要建立两个统计量

$$\underline{\theta} = \underline{\theta}(X_1, X_2, \cdots, X_n),\quad \bar{\theta} = \bar{\theta}(X_1, X_2, \cdots, X_n)$$

使得待估参数θ以较大概率被区间$(\underline{\theta}, \bar{\theta}) \subset \Theta$ 覆盖. 则称$(\underline{\theta}, \bar{\theta})$为参数$\theta$的区间估计. —— 置信区间的方法.

若待估参数θ是多维的，这就是相应多维参数空间的一个区域.

例如，正态分布的未知参数 $\theta = (\mu, \sigma^2)$，

$$\Theta = \{(\mu, \sigma^2) \mid |\mu| < +\infty,\ \sigma > 0\}$$

先介绍进行点估计的重要方法：

§2.2 矩估计法和极大似然估计法

一、矩估计法

它是基于一种简单的“替换”思想建立起来的一种估计方法是英国统计学家 *K*. 皮尔逊最早提出的. 其基本思想是用样本矩估计总体矩.

理论依据 —— 辛钦大数定律

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu\right| < \varepsilon\right) = 1.$$

k 阶原点矩 A_k

$$\because X_i^k \text{ 仍是独立同分布的}, \therefore \lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i^k - E(X^k)\right| < \varepsilon\right) = 1.$$

采用相应的样本矩作为总体矩的估计量，进而由此确定待定参数的估计值.

这种用相应的样本矩去估计总体矩的估计方法就称为矩估计法

由此所得估计量称为矩估计量

求矩估计量的具体步骤

设总体 X 的分布函数 $F(x; \theta_1, \theta_2, \cdots, \theta_l)$中含有 l 个未知参数 $\theta_1, \theta_2, \cdots, \theta_l$，$X_1, X_2, \cdots, X_n$是总体 X 的样本，且总体的前 l 阶原点矩 $E(X^k)$ $(k=1, 2, \cdots, l)$存在，则它们应是这 l 个参数的函数：

$$E(X^k) = g_k(\theta_1, \cdots, \theta_l),\quad k=1, 2, \cdots, l$$

又样本$X_1, X_2, \cdots, X_n$的样本k阶原点矩为 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$，$k=1, 2, \cdots, l$

令总体矩$E(X^k)$等于其相应的样本矩A_k，即可得诸 $E(X^k)$的矩估计量：

$$\begin{cases} g_1(\theta_1, \cdots, \theta_l) = \frac{1}{n}\sum_{i=1}^{n} X_i, \\ g_2(\theta_1, \cdots, \theta_l) = \frac{1}{n}\sum_{i=1}^{n} X_i^2, \\ \cdots\cdots \\ g_l(\theta_1, \cdots, \theta_l) = \frac{1}{n}\sum_{i=1}^{n} X_i^l, \end{cases}$$

矩估计法

从这 l 个方程中可解得矩估计值为：

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(X_1, \cdots, X_n), \\ \hat{\theta}_2 = \hat{\theta}_2(X_1, \cdots, X_n), \\ \cdots\cdots \\ \hat{\theta}_l = \hat{\theta}_l(X_1, \cdots, X_n), \end{cases}$$

例1 设总体 X 的密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}, & 0\le x\le\theta;\\ 0, & 其它\end{cases}$$

θ 为未知参数，$X_1,\cdots,X_n$ 是总体 X 的一个样本，求 θ 的矩估计量.

解 注意到只有一个未知参数，由矩估计法知只需一个方程，

$$EX=\int_{-\infty}^{+\infty}xf(x;\theta)dx=\frac{\theta}{2},$$ 数学期望是一阶原点矩

令 总体原点矩 等于 样本矩 $\dfrac{\theta}{2}=\dfrac{1}{n}\sum_{i=1}^{n}X_i \Leftrightarrow \theta=2\overline{X},$

总体矩 样本矩

即得 θ 的矩估计量为 $\hat{\theta}=2\overline{X}.$

例2 设总体 X 的均值和方差分别为 μ 与 σ^2，(均未知)，$X_1,X_2,\cdots,X_n$ 是总体 X 的样本，若总体的一、二阶原点矩都存在，求 μ 与 σ^2 的矩估计量.

解 由矩估计法知需要两个方程，

$\because\ \mu=EX,\quad E(X^2)=DX+E^2X=\mu^2+\sigma^2,$

令 总体矩 等于 样本矩 $\begin{cases}\mu=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i,\\ \mu^2+\sigma^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2,\end{cases}$

解之得 $\begin{cases}\mu=\overline{X},\\ \sigma^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2,\end{cases}$

所以所求的矩估计量为 $\begin{cases}\hat{\mu}=\overline{X},\\ \hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2,\end{cases}$

例3 设 X 的分布函数为

$$F(x;\alpha,\beta)=\begin{cases}1-(\dfrac{\alpha}{x})^{\beta}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$$

其中参数 $\alpha>0,\beta>1$，$X_1,X_2,\cdots,X_n$ 是总体 X 的样本，求 $\alpha=1$ 时，未知参数 β 的矩估计量.

解 当 $\alpha=1$ 时，X 的概率密度为 $f(x;\beta)=\begin{cases}\dfrac{\beta}{x^{\beta+1}}, & x>1,\\ 0, & x\le 1,\end{cases}$

$$EX=\int_{-\infty}^{+\infty}xf(x;\beta)dx=\beta\int_{1}^{+\infty}x^{-\beta}dx=\frac{\beta}{\beta-1},$$

令 $\dfrac{\beta}{\beta-1}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$，解之得 $\beta=\dfrac{\overline{X}}{\overline{X}-1},$

$\therefore$ β 的矩估计量为 $\hat{\beta}=\dfrac{\overline{X}}{\overline{X}-1}.$

矩法的优点是简单易行，并不需要事先知道总体是什么分布 .

缺点是，当总体类型已知时，没有充分利用分布提供的信息 . 一般场合下，矩估计量不具有唯一性 .

其主要原因在于建立矩法方程时，选取那些总体矩用相应样本矩代替带有一定的随意性 .

二、极大似然法

—— 在总体分布类型已知条件下使用的一种参数估计方法 . 它首先由德国数学家高斯在1821年提出的.

然而, 这个方法常归功于英国统计学家费歇. 他在1922年重新发现了这一方法，并首先研究了这种方法的一些性质 .

极大似然法的基本思想

先看一个简单例子：某位同学与一位猎人一起外出打猎.

一只野兔从前方窜过. 只听一声枪响，野兔应声倒下.

如果让你推测是谁打中的，你会如何想呢？

一般会想：只一枪就打中了，而猎人命中的概率一般大于这位同学命中的概率. 这一枪应是猎人射中的.

此例所作的推断已经体现了极大似然法的基本思想.

例1 设袋中有黑球和白球共 4 个， 今有放回抽球 3 次，结果得到两次白球， 试问如何估计袋中白球的个数? p 是待估参数

解 设袋中有白球 m 个，则抽到白球的概率为 $p=m/4$，

记 X 为抽得的白球数，则 $X\sim B(3,p)$，$P(X=k)=C_3^k p^k(1-p)^{3-k}$，$k=0,1,2,3.$

袋中白球数	p \ 抽到白球数 X	$x=0$	$x=1$	$x=2$	$x=3$
0	0	1	0	0	0
1	1/4	27/64	27/64	9/64	1/64
2	2/4	8/64	24/64	24/64	8/64
3	3/4	1/64	9/64	27/64	27/64
4	1	0	0	0	1

对不同的 p，$B(n,p)$ 的分布列

对不同的 p，事件 $P(X=x)$ 发生的概率

上述估计思路体现的就是极大似然估计的思想方法：

在 p 所有可能的取值中选出能使样本观测值出现的概率为最大的那一个来作为它的估计值.

再如，设总体X服从$B(1,\theta)$分布，即其分布列为

$$f(x;\theta)=P(X=x)=\theta^x(1-\theta)^{1-x},\quad x=0,1$$

其中$\theta(0<\theta<1)$为未知参数，样本为$X_1,\cdots,X_n$，样本值为$x_1,\cdots,x_n$，则事件$(X_1=x_1, X_2=x_2,\cdots,X_n=x_n)$发生的概率为

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

$$=\prod_{i=1}^n f(x_i;\theta)=\prod_{i=1}^n \theta^{x_i}(1-\theta)^{1-x_i}=\theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i}$$

对总体分布中的未知参数θ进行估计时，既然观察结果$X_1=x_1, X_2=x_2,\cdots,X_n=x_n$出现了，理应选取$\theta$使得事件$(X_1=x_1, X_2=x_2,\cdots,X_n=x_n)$发生的概率为最大. 即用它作为$\theta$的估计值可使观察结果出现的可能性最大.

这种选择参数的估计量,使实验结果具有最大概率的思想就是极大似然法的基本思想. 即选取的估计量θ应满足$L(\theta)=\max L(\theta)$

下面给出似然函数的定义和极大似然估计的求法.

定义 设总体X的密度为$f(x;\theta)$(当X为离散型时$f(x;\theta)$为分布列), $\theta=(\theta_1,\theta_2,\cdots,\theta_l)$是总体的未知参数，$x_1,x_2,\cdots x_n$是样本$X_1,X_2,\cdots X_n$的一组样本值，则称

① $$L(\theta)=L(x_1x_2,\cdots,x_n;\theta)=\prod_{i=1}^n f(x_i;\theta)$$

求导麻烦!! 转向$\ln L(\theta)$!!

为样本的似然函数. 若存在某个$\hat\theta=(\hat\theta_1,\hat\theta_2,\cdots,\hat\theta_l)$，使得 ②

$$L(x_1x_2,\cdots,x_n;\hat\theta)=\max_{\theta\in\Theta} L(x_1x_2,\cdots,x_n;\theta)$$

即选取的估计量$\hat\theta$应满足

成立(其中Θ为参数空间)，则称

$$\hat\theta=\hat\theta(X_1,X_2,\cdots,X_n)=(\hat\theta_1(X_1,\cdots,X_n),\hat\theta_2(X_1,\cdots,X_n),\cdots,\hat\theta_l(X_1,\cdots,X_n)),$$

为θ的极大似然估计量. 称

$$\hat\theta=\hat\theta(x_1,x_2,\cdots,x_n)=(\hat\theta_1(x_1,\cdots,x_n),\hat\theta_2(x_1,\cdots,x_n),\cdots,\hat\theta_l(x_1,\cdots,x_n)),$$

为θ的极大似然估计值. ③ 如 何 求 ?

$L(\theta)$是θ的函数, 可用求导的方法找到使$L(\theta)$达到最大值的θ.

注意到$\ln L(\theta)$为单增函数，故$L(\theta)$与$\ln L(\theta)$达到最大值的自变量相同，故问题可转化为求$\ln L(\theta)$的最大值点 .

求极大似然估计的一般步骤是：

① 由总体分布建立似然函数$L(\theta)=\prod_{i=1}^n f(x_i;\theta)$ ——把自变量x看成常数，把未知参数$\theta=(\theta_1,\theta_2,\cdots,\theta_l)$看成自变量;

② 求似然函数$L(\theta)$的最大值点 —— 转化为求$\ln L(\theta)$的最大值点，即

1° 建立似然方程组：$\dfrac{\partial \ln L(\theta)}{\partial\theta_i}=0\quad(i=1,2,\cdots,l)$,

2° 解似然方程组得到$L(\theta)$的最大值点;

③ 将样本$X_1,X_2,\cdots X_n$代入最大值点的表达式中，就得未知参数的极大似然估计量$\hat\theta$，将样本值$x_1,x_2,\cdots x_n$代入最大值点的表达式中，就得未知参数的极大似然估计值$\hat\theta$.

- θ是实数时，似然方程组就是方程 $\dfrac{d\ln L(\theta)}{d\theta}=0$.

下面举例说明如何求极大似然估计

例1 设总体$X\sim B(1,p)$，其分布列为

$$P(x;p)=p^x(1-p)^{1-x},\quad x=0,1,$$

1, 0, 0, 1, 0, 0 是取自总体的一组样本值, 求参数p的极大似然估计.

解 样本的似然函数为: $L(p)=\prod_{i=1}^n P(x_i;p)=\prod_{i=1}^n p^{x_i}(1-p)^{1-x_i}$

对数似然函数为: $\ln L(p)=\sum_{i=1}^n x_i\ln p+\sum_{i=1}^n(1-x_i)\ln(1-p)$

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

对p求导并令其为0得似然方程: $\dfrac{d\ln L(p)}{dp}=\dfrac{1}{p}\sum_{i=1}^n x_i-\dfrac{1}{1-p}(n-\sum_{i=1}^n x_i)=0$,

解之得p的极大似然估计量: $\hat p=\dfrac{1}{n}\sum_{i=1}^n X_i=\bar X$,

代入样本值即得极大似然估计值为: $\hat p=\dfrac{1}{6}(1+0+0+1+0+0)=\dfrac{1}{3}$.

例2 设总体X的密度为 $f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & 其它,\end{cases}$

其中$\lambda>0$为未知参数, $X_1,X_2,\cdots,X_n$是取自总体X的一组样本, 求λ的极大似然估计量与矩估计量.

解(1) 样本的似然函数为

$$L(\lambda)=\prod_{i=1}^n f(x_i;\lambda)=\begin{cases}\lambda^n\prod_{i=1}^n e^{-\lambda x_i}, & x_i>0,\ i=1,2,\cdots,n;\\ 0, & 其他.\end{cases}$$

当$x_i>0$时, $L(\lambda)>0$, $1\le i\le n$,

故有对数似然函数: $\ln L(\lambda)=n\ln\lambda-\lambda\sum_{i=1}^n x_i$,

对λ求导并令其为0可得似然方程: $\dfrac{d\ln L(\lambda)}{d\lambda}=\dfrac{n}{\lambda}-\sum_{i=1}^n x_i=0$,

解得极大似然估计量: $\hat\lambda=n\Big/\sum_{i=1}^n X_i=\dfrac{1}{\bar X}$;

(2) $EX=\int_{-\infty}^{+\infty}xf(x;\lambda)dx=1/\lambda$, 令 $\dfrac{1}{\lambda}=\dfrac{1}{n}\sum_{i=1}^n X_i=\bar X$,

解得矩估计量: $\hat\lambda=\dfrac{1}{\bar X}$.

例3 设总体$X\sim N(\mu,\sigma^2)$，其中μ,σ^2均未知，设$X_1,X_2,\ldots X_n$是取自X的一个样本. 求μ与σ^2的极大似然估计量.

解 样本的似然函数为

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

故有对数似然函数 $\ln L(\mu,\sigma^2)=-\dfrac{n}{2}\ln 2\pi\sigma^2-\dfrac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2$

对μ和σ分别求偏导并令其为0得似然方程组:

$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\sum_{i=1}^n(x_i-\mu)=0,\\[2mm] \dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma}=-\dfrac{n}{\sigma}+\dfrac{1}{\sigma^3}\sum_{i=1}^n(x_i-\mu)^2=0,\end{cases}$$

注意到σ^2是σ的函数！

估计的不变性

$\hat\sigma^2=(\hat\sigma)^2$

解之得极大似然估计量 $\begin{cases}\hat\mu=\dfrac{1}{n}\sum_{i=1}^n X_i=\bar X,\\ \hat\sigma^2=\dfrac{1}{n}\sum_{i=1}^n(X_i-\hat\mu)^2=B_2,\end{cases}$ $\hat\sigma=\sqrt{\dfrac{1}{n}\sum_{i=1}^n(x_i-\hat\mu)^2}$,

正态分布：极大似然估计量 === 矩估计量

可证明极大似然估计具有下面的单调函数不变性：

若$\hat{\theta}$为未知参数θ的极大似然估计量，而$g(\theta)$为θ的单调函数，则$g(\hat{\theta})$也是$g(\theta)$的极大似然估计量.

例4　一罐中装有白球和黑球，有放回地抽取一个容量为n的样本，其中有k个白球，求罐中黑球与白球之比R的极大似然估计.

解　设$X_1, X_2, \cdots, X_n$为所抽样本，$X_i=\begin{cases}1, & \text{取到白球}\\ 0, & \text{取到黑球}\end{cases}\quad i=1,\cdots,n$

则$X_1,\cdots,X_n \sim B(1,p)$，$p$是每次抽取时取到白球的概率，且$p$未知.

容易求得p的极大似然估计为：$\hat{p}=\dfrac{k}{n}$，　$\because R=\dfrac{1-p}{p}$，

由极大似然估计的不变性知R的极大似然估计是$\hat{R}=\dfrac{1-\hat{p}}{\hat{p}}=\dfrac{n}{k}-1$.

上述解法是应用微积分中的技巧求似然函数$L(\theta)$的最大值点. 但当似然函数$L(\theta)$不可微或偏导数不为零时，就不能用上述求导方法求未知参数的极大似然估计了. 这时要用极大似然原则来求.

例5　设总体X服从均匀分布$U[0,\theta]$，为θ未知参数，$X_1, X_2, \cdots, X_n$是总体X的一组样本，求θ的极大似然估计量.

解　样本的似然函数为

不必建立
对数似然方程

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_i\le\theta,\ i=1,2,\cdots,n;\\ 0, & \text{其他}.\end{cases}$$

$$\frac{dL(\theta)}{d\theta}=-n\theta^{-(n+1)},$$

无法用求导建立似然方程的方法确定其极大似然估计量!!

用极大似然原则来求：即用其他方法求似然函数$L(\theta)$的最大值点.

显然，似然函数$L(\theta)$的值随θ的减小而增大，故应取θ的值尽量地小；

另一方面，θ必须满足条件$0\le x_i\le\theta$（$i=1,2,\cdots,n$），

而事件$\{0\le X_i\le\theta,\ i=1,2,\cdots,n\}=\{\max\limits_{1\le i\le n}\{X_i\}\le\theta\}$

故可取极大似然估计量为$\hat{\theta}=\max\limits_{1\le i\le n}\{X_i\}\le\theta$.

例6　设X的分布函数为　$F(x;\alpha,\beta)=\begin{cases}1-(\frac{\alpha}{x})^{\beta}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$

其中$\alpha>0, \beta>1$，$X_1,\cdots,X_n$是总体X的样本，求$\alpha=1$时，未知参数β的

(1) 求$\alpha=1$时，未知参数β的极大似然估计量　β的矩估计量为$\hat{\beta}=\dfrac{\overline{X}}{\overline{X}-1}$.

(2) 求$\beta=2$时，未知参数α的极大似然估计量.

解（1）$\alpha=1$时，　$f(x;\beta)=\begin{cases}\dfrac{\beta}{x^{\beta+1}}, & x>1,\\ 0, & x\le 1,\end{cases}$

$$L(\beta)=\prod_{i=1}^{n} f(x_i;\beta)=\begin{cases}\prod\limits_{i=1}^{n}\dfrac{\beta}{x_i^{\beta+1}}, & x_i>1,\ i=1,\cdots,n;\\ 0, & \text{其他}.\end{cases}$$

当$x_i>1$时，$L(\beta)>0$，$1\le i\le n$，

$$\ln L(\beta)=n\ln\beta-(\beta+1)\sum_{i=1}^{n}\ln x_i,\quad \text{令}\ \frac{d\ln L(\beta)}{d\beta}=\frac{n}{\beta}-\sum_{i=1}^{n}\ln x_i=0,$$

解得β的极大似然估计量：$\hat{\beta}=n\Big/\sum\limits_{i=1}^{n}\ln X_i$.

例6　设X的分布函数为$F(x;\alpha,\beta)=\begin{cases}1-(\frac{\alpha}{x})^{\beta}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$

其中$\alpha>0, \beta>1$，$X_1,\cdots,X_n$是总体X的样本，

(2) 求$\beta=2$时，未知参数α的极大似然估计量.

解(2)　$\beta=2$时，$f(x;\alpha)=\begin{cases}\dfrac{2\alpha^2}{x^3}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$

$$L(\alpha)=\prod_{i=1}^{n} f(x_i;\alpha)=\begin{cases}\prod\limits_{i=1}^{n}\dfrac{2^n\alpha^{2n}}{(x_1x_2\cdots x_n)^3}, & x_i>\alpha,\ i=1,2,\cdots,n;\\ 0, & \text{其他}.\end{cases}$$

因为$L(\alpha)$随α的增大而增大，故应取α的值尽量地大；

另一方面，由于$L(\alpha)=0$不可能是最大值，故α必须满足$x_i>\alpha$，

而事件$\{X_i>\alpha,\ i=1,2,\cdots,n\}=\{\min\limits_{1\le i\le n}\{X_i\}\ge\alpha\}$　$(1\le i\le n)$，

故α的极大似然估计量为$\hat{\alpha}=\min\limits_{1\le i\le n}\{X_i\}\ge\alpha$.

我们介绍了参数点估计，给出了寻求估计量最常用的方法——矩法和极大似然法.

计算简单
无需总体分布
不唯一、不够稳定

使用了总体分布，质量好
计算困难

参数点估计是用一个确定的值去估计未知的参数. 看来似乎精确，实际上把握不大.

自然要问：*样本均值、样本方差是否是一个好的估计量？*
多个估计量时，哪一个估计量更好？

这就需要讨论以下几个问题：估计量的评选标准

(1) 我们希望一个“好的”估计量具有什么特性？

(2) 怎样决定一个估计量是否比另一个估计量“好”？

(3) 应该如何寻求一个合理的估计量？

§2.3　估计量的几个评价标准

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
3. 一致性（相合性）
4. 最小均方误差准则

1. 无偏性

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.

定义 设$\hat\theta=\hat\theta(X_1,\cdots,X_n)$是未知参数$\theta$的一个估计量，若$E(\hat\theta)=\theta$, 则称$\hat\theta$为$\theta$的无偏估计量, 否则称$\hat\theta$为$\theta$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 $E(X^k)=\mu_k\ (k\ge1)$ 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,

$$\therefore\ E(X_i^k)=E(X^k)=\mu_k,\ (k\ge1,\ i=1,2,\cdots,n)$$

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

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$\sigma^2$的无偏估计量.

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

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 的渐进无偏估计量.

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 设总体为$N(\mu,\sigma^2)$，$x_1, x_2,\ldots, x_n$是样本，则s^2是σ^2的无偏估计，且可求出

$$Es=\sqrt{\frac{2}{n-1}}\cdot\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\cdot\sigma\equiv\frac{\sigma}{c_n}$$

这说明 s 不是σ的无偏估计.

利用修正技术可得 $c_n s$ 是σ的无偏估计，其中

$$c_n=\sqrt{\frac{n-1}{2}}\cdot\frac{\Gamma((n-1)/2)}{\Gamma(n/2)}$$ 是修偏系数.

可以证明，当$n\to\infty$时，有$c_n\to1$.

这说明 s 是 σ 的渐近无偏估计。

例4：设总体X服从均匀分布$U(0,\theta)$，其中θ为未知，证明：θ的矩估计是θ的无偏估计，而θ的极大似然估计是θ的渐进无偏估计，并将其修正成无偏估计.

解：θ的矩估计$\hat\theta_1=2\bar X$，θ的极大似然估计$\hat\theta_2=X_{(n)}$.

$$E\hat\theta_1=2E\bar X=\theta,$$

$$f_{X_{(n)}}(x)=n\frac{x^{n-1}}{\theta^n},0\le x\le\theta,$$

$$E\hat\theta_2=EX_{(n)}=\int_0^\theta x\cdot n\frac{x^{n-1}}{\theta^n}dx=\frac{n}{n+1}\frac{x^{n+1}}{\theta^n}\bigg|_0^\theta=\frac{n}{n+1}\theta,$$

故$\hat\theta_2$修正的无偏估计量为$\frac{n+1}{n}\hat\theta_2$.

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

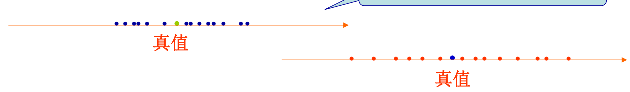
$$E(\hat\theta_1-\theta)^2 \quad 和 \quad E(\hat\theta_2-\theta)^2$$

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$\theta$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3 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 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 哪一个更有效？

解 $\because\ D(X_i)=DX=\sigma^2,$

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

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

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

符合常识！！

例4 设总体 X 的均值和方差均存在，$X_1,\cdots,X_n$ 是总体X的样本，$C_1,C_2,\cdots,C_n$ 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 $\sum_{i=1}^{n}C_iX_i$ 是总体均值 μ 的无偏估计量；

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

证 (1) $\because\ E(\sum_{i=1}^{n}C_iX_i)=\sum_{i=1}^{n}C_i\cdot EX_i=\mu\sum_{i=1}^{n}C_i=\mu$

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

(2) 由柯西—许瓦兹不等式知

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

$$\therefore\ D(\sum_{i=1}^{n}C_iX_i)=\sum_{i=1}^{n}C_i^2\cdot DX_i=\sigma^2\sum_{i=1}^{n}C_i^2>\frac{\sigma^2}{n}=D\bar{X}.$$

这表明，在μ的所有线性无偏估计量中，样本均值$\bar{X}$是最有效的.

定义 设 $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，$\hat{\theta}$ 是未知参数 θ 的一个估计量，若对 θ 的任一无偏估计量 $\hat{\theta}'$，有

$$D\hat{\theta}\leqslant D\hat{\theta}'$$

则称 $\hat{\theta}$ 为 θ 的最小方差无偏估计。

记为 MVUE (minimum variance unbiased estimator的缩写)

Cramer-Rao 不等式

设总体 X 的密度函数为 $f(x;\theta)$，$\theta\in\Theta=(a,b)$，a,b 为常数，a 可以是 $-\infty$，b 可以是 $+\infty$，$X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，$T=T(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的任一无偏估计，若下列条件成立：

(1) $\frac{\partial f(x;\theta)}{\partial\theta}$ 存在，且

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}f(x;\theta)\,\mathrm{d}x=\int_{-\infty}^{+\infty}\frac{\partial f(x;\theta)}{\partial\theta}\mathrm{d}x,\ \theta\in\Theta;$$

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

(3) $$\frac{\partial}{\partial\theta}\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\prod_{i=1}^{n}f(x_i;\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n;$$

(4) $g'(\theta)$ 存在，且

$$g'(\theta)=\int_{-\infty}^{+\infty}\cdots\int_{-\infty}^{+\infty}T(x_1,\cdots,x_n)\frac{\partial}{\partial\theta}\prod_{i=1}^{n}f(x_i;\theta)\,\mathrm{d}x_1\cdots\mathrm{d}x_n,$$

则有

$$DT\geqslant\frac{[g'(\theta)]^2}{nI(\theta)}.$$

特别地，当$g(\theta)=\theta$, $Var_\theta(T)\geq\frac{1}{nI(\theta)}$.

$I(\theta)$的另一等价表达式：

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

定义 设 $T=T(X_1,X_2,\cdots,X_n)$ 是 $g(\theta)$ 的无偏估计，比值

$$e(T)=\frac{[g'(\theta)]^2}{nI(\theta)}/DT$$

称为 T 的有效率，显然有 $0<e(T)\leqslant1$，若 $e(T)=1$ 即 T 达到 C－R 下界时，则称 T 为 $g(\theta)$ 的优效估计，

若 $e(T)\neq1$，但 $\lim_{n\to\infty}e(T)=1$，则称 T 为 $g(\theta)$ 的渐近优效估计.

例 设总体 $X \sim B(1,p)$，$p \in (0,1)$，试求 p 的无偏估计 C－R 下界.

解 概率函数

$$f(x)=P(X=x)=p^{x}(1-p)^{1-x},x=0,1,$$

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

$$=E\left[\frac{\partial \ln p^{X}(1-p)^{1-X}}{\partial p}\right]^{2}$$

$$=E\left[\frac{\partial(X\ln p+(1-X)\ln(1-p)}{\partial p}\right]^{2}$$

$$=E\left[\frac{\partial(X\ln p+(1-X)\ln(1-p))}{\partial p}\right]^{2}$$

$$=E\left[\frac{\partial(X\ln p+(1-X)\ln(1-p))}{\partial p}\right]^{2}$$

$$=E\left[\frac{X}{p}-\frac{1-X}{1-p}\right]^{2}=\frac{1}{p^{2}(1-p)^{2}}E(X-p)^{2}$$

$$=\frac{1}{p^{2}(1-p)^{2}}DX=\frac{1}{p(1-p)},$$

所以 p 的无偏估计 C－R 下界为

$$\frac{1}{nI(p)}=\frac{p(1-p)}{n},$$

所以 p 的无偏估计 C－R 下界为

$$\frac{1}{nI(p)}=\frac{p(1-p)}{n},$$

设样本均值为 $\bar{X}$，则

$$E\bar{X}=p,D\bar{X}=\frac{p(1-p)}{n}=\frac{1}{nI(p)}$$

因此 $\bar{X}$ 是 p 的优效估计，从而 $\bar{X}$ 也是 p 的 MVUE.

例 设总体 $X \sim N(\mu,\sigma^{2})$，$-\infty<\mu<+\infty$，$\sigma>0$. 试验证样本均值 $\bar{X}$，样本方差 S^{2} 分别是否是 μ,σ^{2} 的优效估计？

解 已证明 $\bar{X}$ 是 μ 的无偏估计。

$$f(x,\mu)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^{2}}(x-\mu)^{2}\right]$$

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

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

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

$$=\frac{1}{\sigma^{4}}\cdot DX=\frac{1}{\sigma^{2}}$$

所以 μ 的无偏估计 C－R 下界为

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

因为

$$E\bar{X}=\mu,\quad D\bar{X}=\frac{\sigma^{2}}{n}=\frac{1}{nI(\mu)},$$

所以 $\bar{X}$ 是 μ 的优效估计，从而 $\bar{X}$ 也是 μ 的 MVUE

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

$$=E\left[\frac{(X-\mu)^{4}}{4\sigma^{8}}-\frac{(X-\mu)^{2}}{2\sigma^{6}}+\frac{1}{4\sigma^{4}}\right]$$

$$=\left[\frac{1}{4\sigma^{4}}E\left(\frac{X-\mu}{\sigma}\right)^{4}-\frac{DX}{2\sigma^{6}}+\frac{1}{4\sigma^{4}}\right].$$

因为 $Z=\frac{X-\mu}{\sigma}\sim N(0,1)$

所以 $E\left(\frac{X-\mu}{\sigma}\right)^{4}=EZ^{4}=\int_{-\infty}^{+\infty}z^{4}\frac{1}{\sqrt{2\pi}}e^{-\frac{z^{2}}{2}}dz$

$$=3,$$

$$I(\sigma^2)=\frac{3}{4\sigma^4}-\frac{\sigma^2}{2\sigma^6}+\frac{1}{4\sigma^4}=\frac{1}{2\sigma^4},$$

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

因为 $\dfrac{(n-1)S^2}{\sigma^2}\sim\chi^2(n-1)$

$$D\frac{(n-1)S^2}{\sigma^2}=2(n-1),\quad DS^2=\frac{2\sigma^4}{n-1}>\frac{2\sigma^4}{n}=\frac{1}{nI(\sigma^2)},$$

因此 S^2 不是 σ^2 的优效估计. 但 $\lim\limits_{n\to\infty}\dfrac{\frac{1}{nI(\sigma^2)}}{DS^2}=1$,

所以 S^2 是 σ^2 的渐近优效估计.

例. 设样本 $X_1,\cdots,X_n$，iid. 服从均值为 θ 的指数分布，

则 $f(x;\theta)=\frac{1}{\theta}e^{-\frac{x}{\theta}},x>0.$

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

$$E\left[\frac{\partial\ln f(X;\theta)}{\partial\theta}\right]^2=E\left(-\frac{1}{\theta}+\frac{X}{\theta^2}\right)^2=\frac{1}{\theta^4}E[X-\theta]^2$$

$$=\frac{1}{\theta^4}Var(X)=\frac{1}{\theta^2}$$

$$g'(\theta)=\theta'=1\qquad\therefore\ \frac{[g'(\theta)]^2}{n\cdot E\left[\frac{\partial\ln f(X;\theta)}{\partial\theta}\right]^2}=\frac{\theta^2}{n}=Var(\bar X)$$

$\because$ $\bar X$ 是 θ 的无偏估计，$\therefore$ $\bar X$ 是 θ 的MVUE。

例. 设 总体 $X\sim P(\lambda)$，$X_1,\cdots,X_n$ 是来自该总体的样本，试证 $\hat\lambda=\bar X$ 为 λ 的MVUE估计。

证　由 $E(\hat\lambda)=E(\bar X)=\lambda$,, 知 $\hat\lambda=\bar X$ 为 λ 的无偏估计量。而 $Var(\hat\lambda)=Var(\bar X)=\dfrac{\lambda}{n}$.

又 $f(x,\lambda)=\dfrac{\lambda^x}{x!}e^{-\lambda}$, $\ln f(x,\lambda)=-\lambda+x\ln\lambda-\ln(x!)$,

$$\therefore\quad E\left[\frac{\partial^2\ln f(X,\lambda)}{\partial\lambda^2}\right]=E\left[-\frac{X}{\lambda^2}\right]=-\frac{1}{\lambda^2}E(X)=-\frac{1}{\lambda}$$

$$g'(\theta)=\lambda'=1\quad\therefore\ \frac{[g'(\theta)]^2}{-nE\left[\frac{\partial^2\ln f(X,\lambda)}{\partial\lambda^2}\right]}=\frac{\lambda}{n}=Var(\hat\lambda)$$

$\therefore$　$\hat\lambda=\bar X$　即为 λ 的MVUE.

3、相合性(一致性)

若 $\hat\theta=\hat\theta(X_1,X_2,\cdots,X_n)$ 为参数 θ 的估计量, 若对于任意 $\theta\in\Theta$, 当 $n\to\infty$ 时, $\hat\theta(X_1,X_2,\cdots,X_n)$ 依概率收敛于 θ, 则称 $\hat\theta$ 为 θ 的相合（一致）估计量.

设 $\hat\theta$ 是 θ 的一致估计量, $g(x)$ 是连续函数, $g(\hat\theta)$ 是 $g(\theta)$ 的一致估计量.

例如　由样本 $k(k\ge1)$ 阶矩是总体 X 的 k 阶矩 $\mu_k=E(X^k)$ 的相合估计量,

进而若待估参数 $\theta=g(\mu_1,\mu_2,\cdots,\mu_n)$, 其中 g 为连续函数, 则 θ 的矩估计量 $\hat\theta=g(\hat\mu_1,\hat\mu_2,\cdots,\hat\mu_n)=g(A_1,A_2,\cdots,A_n)$ 是 θ 的相合估计量 .

例　试证: 样本均值 $\bar X$ 是总体均值 μ 的相合估计量, 样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar X)^2$ 及样本的二阶中心矩 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar X)^2$ 都是总体方差 σ^2 的相合估计量.

证明　由大数定律知,

$$\forall\varepsilon>0,\ \text{有}\ \lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|<\varepsilon\right\}=1,$$

所以 $\bar X=\frac{1}{n}\sum_{i=1}^{n}X_i$ 是 μ 的相合估计量.

又 $B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}(X_i{}^2-2X_i\overline{X}+\overline{X}^2)$

$$=\frac{1}{n}\sum_{i=1}^{n}X_i{}^2-\overline{X}^2=A_2-\overline{X}^2,$$

(A_2是样本二阶原点矩)

由大数定律知,

$A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$依概率收敛于$E(X^2)$,

$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$依概率收敛于$E(X)$,

故 $B_2=A_2-\overline{X}^2$

依概率收敛于 $E(X^2)-[E(X)]^2=\sigma^2$,

所以 B_2 是 σ^2 的相合估计量.

又 $\lim\limits_{n\to\infty}\frac{n}{n-1}=1$,

所以 $S^2=\frac{n}{n-1}B_2$ 也是 σ^2 的相合估计量.

相合估计量未必是无偏的!

例 设 $\hat{\theta}$ 是 θ 的无偏估计,且 $\lim\limits_{n\to+\infty}D\hat{\theta}=0$,

则 $\hat{\theta}$ 是 θ 的一致估计量.

证明 对于任意的 $\varepsilon>0$,由切比雪夫不等式得

$$1\geqslant P(|\hat{\theta}-\theta|<\varepsilon)\geqslant 1-\frac{D\hat{\theta}}{\varepsilon^2}\to 1\ (n\to+\infty),$$

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

所以 $\hat{\theta}$ 是 θ 的一致估计量.

例 设总体 X 的密度函数为

$$f(x)=\mathrm{e}^{-(x-\theta)},x\geqslant\theta,-\infty<\theta<+\infty,$$

用 $\hat{\theta}$ 表示 θ 的极大似然估计,试判别 $\hat{\theta}$ 是否是 θ 的一致估计?

解 先求 $\hat{\theta}$,似然函数

$$L(\theta)=\mathrm{e}^{-\sum_{i=1}^{n}x_i+n\theta},x_i\geqslant\theta,i=1,2,\cdots,n$$

$$=\mathrm{e}^{-\sum_{i=1}^{n}x_i+n\theta},x_{(1)}\geqslant\theta,$$

所以 θ 取 $X_{(1)}$ 时 $L(\theta)$ 达到最大,即 $\hat{\theta}=X_{(1)}$.

$X_{(1)}$ 的密度函数为

$$f_{X_{(1)}}(x)=n[1-F(x)]^{n-1}f(x)=n\mathrm{e}^{-n(x-\theta)},x\geqslant\theta,$$

对任意的 $\varepsilon>0$,有

$$P(|\hat{\theta}-\theta|\geqslant\varepsilon)=P(X_{(1)}\geqslant\theta+\varepsilon)$$

$$=\int_{\theta+\varepsilon}^{+\infty}n\mathrm{e}^{-n(x-\theta)}\mathrm{d}x=\mathrm{e}^{-n\varepsilon}\to 0\ (n\to\infty),$$

所以 $\hat{\theta}$ 是 θ 的一致估计.

4 最小均方误差准则

假设用 $\hat{\theta}$ 估计 θ，评价该估计好坏的一 个自然度量是 $|\hat{\theta}-\theta|$,由于 θ 是未知的，样本又具有 随机性，直接使用这种自然度量 在实际中是不可行的，为排除样本随机性的影 响，可以对它求期望，由于数学处理上的方便 考虑，最常用的标准是由下式给出的均方误 差。

定义 设 θ 为一个一维未知参数，$\hat{\theta}$ 为 θ 的一个估计，$\hat{\theta}$ 的均方误差定义为

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

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

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

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

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

$$\because \quad 2E\left[\hat{\theta}-E(\hat{\theta})\right]\left[E(\hat{\theta})-\theta\right]$$

$$=2\left[E(\hat{\theta})-E(\hat{\theta})\right]\left[E(\hat{\theta})-\theta\right]=0$$

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

$MSE(\hat{\theta})$是由两个量迭加而成,一个是估计量的方差,另一个是估计的偏差的平方.如果$\hat{\theta}$是θ的无偏估计,则后一项为0.则有 $MSE(\hat{\theta})=Var(\hat{\theta})$

在θ的所有的估计量中均方误差达到最小的通常是不存在的.

例 设$X_1,X_2,\cdots,X_n$来自正态总体$N(\mu,\sigma^2)$的样本,试在σ^2的估计类$\left\{T_c = c\sum_{i=1}^{n}(X_i-\bar{X})^2, c>0\right\}$中,求具有最小均方误差的$\sigma^2$的估计量.

解 因为 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\bar{X})^2 \sim \chi^2(n-1)$,

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

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

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

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

$$D\sum_{i=1}^{n}(X_i-\bar{X})^2 = 2(n-1)\sigma^4,$$

$$\mathrm{MSE}(T_c) = DT_c + (ET_c-\sigma^2)^2$$

$$=2c^2(n-1)\sigma^4+[c(n-1)\sigma^2-\sigma^2]^2$$

$$=2c^2(n-1)\sigma^4+[c(n-1)-1]^2\sigma^4,$$

令 $g(c)=2c^2(n-1)+[c(n-1)-1]^2$,

$$g'(c)=4c(n-1)+2[c(n-1)-1](n-1)=0,$$

解得 $c=\frac{1}{n+1}$.

因为 $g''(c)\Big|_{c=\frac{1}{n+1}}=4(n-1)+2(n-1)^2>0$,

所以当$c=\frac{1}{n+1}$时,$g(c)$取得最小值.

所以在此估计类中具有最小均方误差的σ^2的估计量为

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

小结

两种求点估计的方法: $\begin{cases}\text{矩估计法}\\ \text{极大似然估计法}\end{cases}$

在统计问题中往往先使用极大似然估计法,在最大似然估计法使用不方便时,再用矩估计法.

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

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

估计量的评选的四个标准 $\begin{cases}\text{无偏性}\\ \text{有效性}\\ \text{相合性}\\ \text{最小均方误差准则}\end{cases}$

相合性是对估计量的一个基本要求,不具备相合性的估计量是不予以考虑的.

由最大似然估计法得到的估计量,在一定条件下也具有相合性.估计量的相合性只有当样本容量相当大时,才能显示出优越性,这在实际中往往难以做到,因此,在工程中往往使用无偏性和有效性这两个标准.

- 作业：P58

5， 11, 12， 14,
20, 23, 24, 26