引 言

在数理统计学中，总体的分布是未知的。它包括两种情形：

1）总体分布的类型是已知的，但其中包含未知参数。我们的任务就是通过样本来估计这些未知参数。这就是参数估计问题。

2） 总体分布的类型是未知的。我们的任务就是通过样本来估计总体的分布。这就是非参数估计问题。

我们这里只讨论参数估计问题。

例:

设总体 X 是服从参数为λ的指数分布，其中参数λ未知，$\lambda>0$. $X_1,\cdots,X_n$是总体 X 的一个样本，我们的任务是根据样本，来估计λ的取值，从而估计总体的分布.

这是一个参数估计问题

第七章 参数估计

§1 点估计

§2 估计量的评选标准

§1 点估计

- 点估计问题
- 矩法
- 极大似然法

一、点估计问题

设总体X的分布函数$F(x;\theta)$的形式为已知，θ是待估参数。$X_1,\cdots,X_n$是X的一个样本，$x_1,\cdots,x_n$是相应的样本值。

构造一个适当的统计量$\hat{\theta}(X_1,\cdots,X_n)$，用它的观察值$\hat{\theta}(x_1,\cdots,x_n)$来估计未知参数$\theta$。

我们称$\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的<u>估计量</u>;

称$\hat{\theta}(x_1,\cdots,x_n)$为$\theta$的<u>估计值</u>。

这种对未知参数进行定值估计的问题就是点估计问题。

注意：

(1)估计量与估计值有着本质的不同：

估计量是统计量，因而它是随机变量(一维或多维)；而估计值则是一维或多维数组.

(2)在不引起混淆的情况下，我们统称估计量与估计值为未知参数θ的估计.

二、 矩估计法

设X为连续型随机变量，其概率密度为

$$f(x;\theta_1,\cdots,\theta_k),$$

X为离散型随机变量，其分布列为

$$P\{X=x\}=P(x;\theta_1,\cdots,\theta_k),$$

其中$\theta_1,\cdots,\theta_k$是待估参数，$X_1,\cdots,X_n$为来自X的样本.

设 $EX^l=\mu_l$ 存在， $l=1,2,\cdots,k$

则 $\mu_l=\mu_l(\theta_1,\cdots,\theta_k),l=1,2,\cdots,k.$

令 $A_l=\mu_l,\quad l=1,\cdots,k,$ 其中 $A_l=\dfrac{1}{n}\sum_{i=1}^{n}X_i^l$

这是包含k个未知参数$\theta_1,\cdots,\ \theta_k$的联立方程组，

$$\begin{cases} A_1 = \mu_1(\theta_1,\ \theta_2,\ \cdots,\ \theta_k) \\ A_2 = \mu_2(\theta_1,\ \theta_2,\ \cdots,\ \theta_k) \\ \cdots\ \cdots\ \cdots\ \cdots \\ A_k = \mu_k(\theta_1,\ \theta_2,\ \cdots,\ \theta_k) \end{cases}$$

从中解出方程组的解记为$\hat{\theta}_1,\cdots,\ \hat{\theta}_k$,即

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(X_1,\ X_2,\ \cdots,\ X_n) \\ \hat{\theta}_2 = \hat{\theta}_2(X_1,\ X_2,\ \cdots,\ X_n) \\ \cdots\ \cdots\ \cdots\ \cdots \\ \hat{\theta}_k = \hat{\theta}_k(X_1,\ X_2,\ \cdots,\ X_n) \end{cases}$$

用$\hat{\theta}_1,\cdots,\hat{\theta}_k$分别作为$\theta_1,\cdots,\theta_k$的估计量，

这种求估计量的方法称为矩估计法.

这种估计量称为矩估计量；矩估计量的观察值称为矩估计值。

矩法原理：由辛钦大数定律知

$$A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l \xrightarrow{P} \mu_l,\quad l=1,2,\cdots,k.$$

所以我们令 $A_l=\mu_l$，$l=1,\cdots,k$,用A_l估计μ_l.

矩法求估计量的步骤：

1) 求 $\mu_1 = EX\ ,\mu_2 = EX^2 \ldots;$

2) 令 $A_1 = \mu_1\ , A_2 = \mu_2 \ldots;$

3) 解上面方程（组），得

$$\hat{\theta}_1 = \hat{\theta}_1(X_1, \cdots, X_n)$$

$$\hat{\theta}_2 = \hat{\theta}_2(X_1, \cdots, X_n)$$

…

例 1 设某炸药厂一天中发生着火现象的次数X服从参数为λ的泊松分布，λ未知，有以下样本值；试估计参数λ（用矩法）。

着火的次数k	0	1	2	3	4	5	6	
发生k次着火天数n_k	75	90	54	22	6	2	1	$\sum = 250$

解： $\mu_1 = EX = \lambda,\qquad A_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

令 $\overline{X} = \lambda,$

则 $\hat{\lambda} = \overline{x} = \dfrac{1}{250}(0\times 75 + 1\times 90 + \cdots + 6\times 1) = 1.22$

所以估计值$\hat{\lambda} = 1.22$。

例2

设总体$X \sim U[a,b]$，a,b未知，$X_1,\cdots,X_n$是一个样本，求：a,b的矩估计量。

解：

$$\mu_1 = EX = \frac{a+b}{2},$$

$$\mu_2 = EX^2 = DX + (EX)^2 = \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4}$$

令 $\dfrac{a+b}{2} = A_1$ 即 $a+b = 2A_1,$

$$\frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} = A_2 \qquad b-a = \sqrt{12(A_2 - A_1^2)}$$

例2（续）

即 $a+b=2A_1$，$b-a=\sqrt{12(A_2-A_1^2)}$

解得：

$$\hat{a}=A_1-\sqrt{3(A_2-A_1^2)}=\bar{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$\hat{b}=A_1+\sqrt{3(A_2-A_1^2)}=\bar{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

$$A_2-A_1^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=\frac{1}{n}\left(\sum_{i=1}^{n}X_i^2-n\bar{X}^2\right)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

例3 设总体X的均值μ，方差σ^2都存在，且$\sigma^2>0$，但μ，σ^2未知，又设$X_1,\cdots,X_n$是一个样本；求：μ,σ^2的矩估计量。

解：$\mu_1 = EX = \mu,$

$$\mu_2 = EX^2 = DX + (EX)^2 = \sigma^2 + \mu^2$$

令 $\mu_1 = A_1,\quad \mu_2 = A_2,$

即 $\mu = A_1,\quad \sigma^2 + \mu^2 = A_2,$

所以 $\hat{\mu} = A_1 = \overline{X},$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

特别，若 $X \sim N(\mu, \sigma^2)$，μ, σ^2 未知；

则 $\hat{\mu} = \overline{X}$，$\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$

例4 设总体 X 服从参数为 λ 的指数分布，其中 $\lambda > 0$ 未知，$X_1, X_2, \cdots, X_n$ 是从该总体中抽取的一个样本，试求参数 λ 的矩估计.

解： 总体 X 的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0. \end{cases}$$

所以，$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\cdot\lambda e^{-\lambda x}dx=\dfrac{1}{\lambda}$

令

$$\overline{X}=\frac{1}{\lambda},$$

得参数λ的矩估计量为

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

例5 设总体 X 的密度函数为

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

其中 $\alpha>0$ 为未知参数，试求参数 α 的矩估计.

解：

$$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{1}x\cdot(\alpha+1)x^{\alpha}dx=\frac{\alpha+1}{\alpha+2}$$

令 $\overline{X}=\dfrac{\hat{\alpha}+1}{\hat{\alpha}+2}$

由此得 α 的矩估计量为 $\hat{\alpha}=\dfrac{2\overline{X}-1}{1-\overline{X}}$.

例6 设总体 X 服从 $\Gamma-$分布，其密度函数为

$$f(x)=\begin{cases}\dfrac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x} & x>0\\ 0 & x\le 0\end{cases}$$

其中 $\alpha>0$，$\beta>0$ 为未知参数，试求参数 α 与 β 的矩估计.

解：

$$EX=\int_{-\infty}^{+\infty}xf(x)dx=\int_{0}^{+\infty}x\cdot\frac{\beta^{\alpha}}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx$$

$$=\frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)}\int_{0}^{+\infty}\frac{\beta^{\alpha+1}}{\Gamma(\alpha+1)}x^{(\alpha+1)-1}e^{-\beta x}dx$$

$$=\frac{\Gamma(\alpha+1)}{\beta\Gamma(\alpha)}=\frac{\alpha\Gamma(\alpha)}{\beta\Gamma(\alpha)}=\frac{\alpha}{\beta}$$

例6（续）

$$EX^2=\int_{-\infty}^{+\infty}x^2f(x)dx=\int_0^{+\infty}x^2\cdot\frac{\beta^\alpha}{\Gamma(\alpha)}x^{\alpha-1}e^{-\beta x}dx$$

$$=\frac{\Gamma(\alpha+2)}{\beta^2\Gamma(\alpha)}\int_0^{+\infty}\frac{\beta^{\alpha+2}}{\Gamma(\alpha+2)}x^{(\alpha+2)-1}e^{-\beta x}dx$$

$$=\frac{\Gamma(\alpha+2)}{\beta^2\Gamma(\alpha)}=\frac{\alpha(\alpha+1)\Gamma(\alpha)}{\beta^2\Gamma(\alpha)}=\frac{\alpha(\alpha+1)}{\beta^2}$$

因此有 $\begin{cases}EX=\dfrac{\alpha}{\beta},\\EX^2=\dfrac{\alpha(\alpha+1)}{\beta^2}.\end{cases}$ 令 $\begin{cases}A_1=\dfrac{\alpha}{\beta},\\A_2=\dfrac{\alpha(\alpha+1)}{\beta^2}.\end{cases}$

例6（续）

解此方程组，得

$$\begin{cases} \hat{\alpha} = \dfrac{A_1^{\ 2}}{A_2 - A_1^{\ 2}}, \\ \\ \hat{\beta} = \dfrac{A_1}{A_2 - A_1^{\ 2}}. \end{cases}$$

即

$$\begin{cases} \hat{\alpha} = \dfrac{\overline{X}^{\ 2}}{B_2}, \\ \\ \hat{\beta} = \dfrac{\overline{X}}{B_2}. \end{cases}$$

其中 $B_2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ 为样本的二阶中心矩.

三、 极大似然法

例1 如果一个射手击中目标的概率可能是

$$p=\frac{1}{5},\frac{8}{15},\frac{4}{5}.$$

现在让他打三发子弹，在不同的命中目标的次数下，我们应该如何取 p 的估计值$\hat{p}$?

解： 用 X 表示命中目标的次数， 则 $X \sim B(3, p)$,即

$$P\{X=x)=C_3^x p^x(1-p)^{3-x}=P(x;p).$$

计算结果列表如下：

例1（续）

命中次数 x	0	1	2	3
$P(x;\frac{1}{5})$	$\frac{1728}{3375}$	$\frac{1296}{3375}$	$\frac{324}{3375}$	$\frac{27}{3375}$
$P(x;\frac{8}{15})$	$\frac{343}{3375}$	$\frac{1176}{3375}$	$\frac{1344}{3375}$	$\frac{512}{3375}$
$P(x;\frac{4}{5})$	$\frac{27}{3375}$	$\frac{324}{3375}$	$\frac{1296}{3375}$	$\frac{1728}{3375}$

因为 $P(0;\frac{1}{5}) > P(0;\frac{8}{15}) > P(0;\frac{4}{5})$,

这表明, $p=\frac{1}{5}$ 使得打三发子弹命中次数为0的概率最大.

由实际推断原理知，此时应取 $\hat{p}=\frac{1}{5}$.

例1（续）

命中次数 x	0	1	2	3
$P(x;\frac{1}{5})$	$\frac{1728}{3375}$	$\frac{1296}{3375}$	$\frac{324}{3375}$	$\frac{27}{3375}$
$P(x;\frac{8}{15})$	$\frac{343}{3375}$	$\frac{1176}{3375}$	$\frac{1344}{3375}$	$\frac{512}{3375}$
$P(x;\frac{4}{5})$	$\frac{27}{3375}$	$\frac{324}{3375}$	$\frac{1296}{3375}$	$\frac{1728}{3375}$

因此，由上表可得下面的结论：

打三发命中次数x=1时，命中率 p 的合理估计 $\hat{p}=\frac{1}{5}$;

打三发命中次数x=2时，命中率 p 的合理估计 $\hat{p}=\frac{8}{15}$;

打三发命中次数x=3时，命中率 p 的合理估计 $\hat{p}=\frac{4}{5}$.

1) 若总体X属离散型，其分布律

$$P\{X = x\} = p(x;\theta), \theta \in \Theta$$

的形式为已知，θ 为待估参数，Θ是θ 可能取值的范围。设$X_1, \cdots, X_n$是来自X的样本,则$X_1, \cdots, X_n$的联合分布律：

$$\prod_{i=1}^{n} p(x_i;\theta)$$

又设$x_1, \cdots, x_n$是$X_1, \cdots, X_n$的一个样本值,

易知样本$X_1, \cdots, X_n$取$x_1, \cdots, x_n$的概率，亦即

事件$\{X_1 = x_1, \cdots, X_n = x_n\}$发生的概率为：

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

它是θ的函数. $L(\theta)$称为样本的似然函数

极大似然法原理：

固定$x_1, \cdots, x_n$，挑选使概率$L(x_1, \cdots, x_n; \theta)$达到最大的参数$\hat{\theta}$，作为$\theta$的估计值，即取$\hat{\theta}$使得：

$$L(x_1, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, \cdots, x_n; \theta)$$

$\hat{\theta}$与$x_1, \cdots, x_n$有关，记为$\hat{\theta}(x_1, \cdots, x_n)$;

称其为参数θ的极大似然估计值

$\hat{\theta}(X_1,\cdots,X_n)$称为参数$\theta$的极大似然估计量

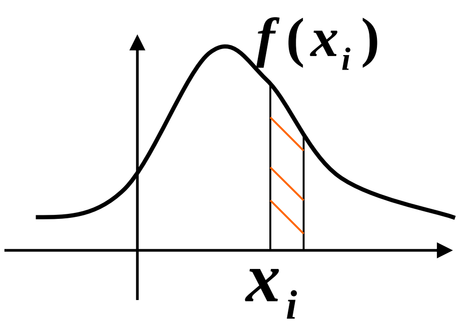
这种求未知参数θ的方法称为极大似然法

2) 若总体X属连续型，其概率密度$f(x;\theta),\theta\in\Theta$的形式已知，$\theta$为待估参数.则$X_1,\cdots,X_n$的联合密度：

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

设$x_1,\cdots,x_n$是相应$X_1,\cdots,X_n$的一个样本值，则随机点$(X_1,\cdots,X_n)$落在$(x_1,\cdots,x_n)$的邻域（边长分别为$dx_1,\cdots,dx_n$的n维立方体）内的概率近似为：

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



在得到观测值$(x_1, \quad x_2, \quad \cdots, \quad x_n)$的前提下，自然应当选取使得

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

达到最大的θ值作为未知参数θ的估计值.

因为当未知参数θ等于这个值时，出现给定的那个样本观测值的可能性最大.

但$\prod_i dx_i$不随θ而变，故只需考虑：

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

的最大值，这里$L(\theta)$称为样本的似然函数.

若 $L(x_1,\cdots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta} L(x_1,\cdots,x_n;\theta)$

则称$\hat{\theta}(x_1,\cdots,x_n)$为$\theta$的<u>极大似然估计值</u>

称$\hat{\theta}(X_1,\cdots,X_n)$为$\theta$的<u>极大似然估计量</u>。

一般，$p(x;\theta), f(x;\theta)$关于θ可微，故θ可由下式求得：

$$\frac{dL(\theta)}{d\theta}=0$$ ------似然方程

又因$L(\theta)$与$\ln L(\theta)$在同一θ处取到极值，因此θ的极大似然估计θ也可从下述方程解得：

$$\frac{d}{d\theta}\ln L(\theta)=0$$ ------对数似然方程

若总体的分布中包含多个参数,

即可令 $\dfrac{\partial L}{\partial \theta_i}=0, i=1,\cdots,k$ ------似然方程组

或 $\dfrac{\partial \ln L}{\partial \theta_i}=0, i=1,\cdots,k$ ------对数似然方程组

解k个方程组求得$\theta_1,\cdots,\theta_k$的极大似然估计值。

即可得$\theta_1,\cdots,\theta_k$的极大似然估计量。

一般来讲,极大似然估计优于矩估计,因而在应用中,我们应当尽可能地使用极大似然估计

极大似然法求估计量的步骤：（一般情况下）

1) 构造似然函数 $L(\theta)$：

$$L(\theta)=\prod_{i=1}^{n}P(x_i)\ (\text{离散型}),\qquad L(\theta)=\prod_{i=1}^{n}f(x_i)\ (\text{连续型});$$

2) 取对数：$\ln L(\theta)$;

3) 令 $\dfrac{d\ln L}{d\theta}=0$;

4) 解似然方程得 θ 的极大似然估计量 $\hat{\theta}$.

说明：若似然方程（组）无解，或似然函数不可导，此法失效，改用其它方法。

例2 设$X \sim B(1,p)$; $X_1,\cdots,X_n$是来自X的一个样本，试求参数 p 的极大似然估计量。

解：设$x_1,\cdots,x_n$是一个样本值。X的分布律为：

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

故似然函数为

$$L(p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum\limits_{i=1}^{n}x_i}(1-p)^{n-\sum\limits_{i=1}^{n}x_i},$$

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

例2（续）

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

令 $\frac{d}{dp}\ln L(p) = 0$,即 $\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0.$

解得p的极大似然估计值 $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$

p的极大似然估计量为

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$$

--------它与矩估计量是相同的。

例3 设$X \sim N(\mu,\sigma^2)$; μ,已知，σ^2为未知参数，$x_1,\cdots,x_n$ 是来自X的一个样本值，求σ^2的极大似然估计量

解： X的概率密度为：

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

似然函数为:

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

$$=\left(2\pi\sigma^2\right)^{-\frac{n}{2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

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

例3（续）

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

$$\frac{d\ln L}{d\sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$$

令：$\dfrac{d\ln L}{d\sigma^2} = 0$，

得似然方程 $-\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum_{i=1}^{n}\left(x_i-\mu\right)^2 = 0$

解此方程，得 $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(x_i-\mu\right)^2$，

因此 σ^2 的极大似然估计量为 $\hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$.

例4 设$X \sim N(\mu,\sigma^2)$; μ,σ^2为未知参数，$x_1,\cdots,x_n$是来自X的一个样本值,求μ,σ^2的极大似然估计量

解： X的概率密度为：

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

似然函数为：

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

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

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

例4（续）

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

令 $\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = 0 \\ \dfrac{\partial \ln L}{\partial \sigma^2} = 0 \end{cases}$ 即：$\begin{cases} \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu) = 0 \\ -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases}$

解得：$\hat{\mu} = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i = \bar{x}$ ， $\hat{\sigma}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2$

故μ,σ^2的极大似然估计量为：

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

例5 设总体 X 的密度函数为

$$f(x)=\begin{cases}\theta x^{\theta-1}, & 0<x<1,\\ 0, & \text{其它}.\end{cases}$$

其中θ未知，$\theta>1$， X_1， $\cdots$， X_n 是从该总体抽取的一个样本．试求θ的极大似然估计．

解： 似然函数为

$$L(\theta)=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1},$$

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

例5（续）

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

$$\frac{d\ln L}{d\theta} = \frac{n}{\theta} + \sum_{i=1}^{n}\ln x_i$$

令：$\dfrac{d\ln L}{d\theta} = 0$，

得似然方程为 $\dfrac{n}{\theta} + \sum_{i=1}^{n}\ln x_i = 0,$

解得 $\hat{\theta} = -\dfrac{n}{\sum_{i=1}^{n}\ln x_i},$

因此θ的极大似然估计量为 $\hat{\theta} = -\dfrac{n}{\sum_{i=1}^{n}\ln X_i}.$

例6 设$X \sim U[a,b]$;a,b未知，$x_1,\cdots,x_n$是一个样本值，求：a,b的极大似然估计量。

分析：X 的概率密度为：$f(x;a,b)=\begin{cases}\dfrac{1}{b-a}, & a \le x \le b; \\ 0, & 其它\end{cases}$

似然函数为

$$L(a,b)=\frac{1}{(b-a)^n} \quad \left(a < x_i < b，\quad i=1,\quad 2,\quad \cdots,\quad n\right)$$

$$\ln L(a,b)=-n\ln(b-a),$$

$$\frac{\partial}{\partial a}\ln L(a,b)=\frac{n}{b-a}=0;\quad \frac{\partial}{\partial b}\ln L(a,b)=-\frac{n}{b-a}=0,$$

显然，似然方程组无解，但这不能说明不存在极大似然估计量，只是不能由似然方程组求解。

例6（续）

解：将$x_1,\cdots,x_n$按从小到大顺序排列成

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)},$$

则 $$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_{(1)} \le \cdots \le x_{(n)} \le b; \\ 0, & \text{其它}\end{cases}$$

对于满足$a \le x_{(1)} \le \cdots \le x_{(n)} \le b$的任意$a,b$有

$$L(a,b)=\frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)}-x_{(1)})^n}$$

例6（续）

即： $L(a,b)$在$a=x_{(1)},b=x_{(n)}$时，

取最大值$(x_{(n)}-x_{(1)})^{-n}$

故a,b的极大似然估计值为：

$$\hat{a}=x_{(1)}=\min x_i,\quad \hat{b}=x_{(n)}=\max x_i,$$

故a,b的极大似然估计量为：

$$\hat{a}=\min X_i,\quad \hat{b}=\max X_i.$$

例7 一个罐子里装有黑球和白球，有放回地抽取 *n* 个球，发现有 *k* 个黑球。试求罐子里黑球数与白球数之比 *R* 的极大似然估计量。

解：设罐中装有 *a* 只黑球 *b* 只白球，则 $R=\dfrac{a}{b}$.

设 $X_i=\begin{cases}1, & \text{第}i\text{次摸到黑球；}\\ 0, & \text{第}i\text{次摸到白球.}\end{cases}\quad i=1,\cdots,n.$

则 $X_1,\cdots,X_n$ **是总体** $X\sim b(1,p)$ **的样本,**

其中 $p=P\{X_i=1\}=\dfrac{a}{a+b}=\dfrac{R}{1+R}.$

$$p=\frac{R}{1+R}$$

似然函数 $L(R)=\prod_{i=1}^{n}p(x_i)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}$

$$=\prod_{i=1}^{n}\left(\frac{R}{1+R}\right)^{x_i}\left(1-\frac{R}{1+R}\right)^{1-x_i}=\left(\frac{R}{1+R}\right)^{\sum_{i=1}^{n}x_i}\left(1-\frac{R}{1+R}\right)^{n-\sum_{i=1}^{n}x_i},$$

则 $\ln L=\sum_{i=1}^{n}x_i\ln\left(\frac{R}{1+R}\right)+\left(n-\sum_{i=1}^{n}x_i\right)\ln\left(\frac{1}{1+R}\right)$

$$=\sum_{i=1}^{n}x_i[\ln R-\ln(1+R)]\quad-\left(n-\sum_{i=1}^{n}x_i\right)\ln(1+R)]$$

令 $\frac{d\ln L}{dR}=0$, 则 $\sum_{i=1}^{n}x_i\left(\frac{1}{R}-\frac{1}{1+R}\right)-\left(n-\sum_{i=1}^{n}x_i\right)\frac{1}{1+R}=0$

解出 $\hat{R}=\dfrac{\sum_{i=1}^{n}x_i}{n-\sum_{i=1}^{n}x_i}=\dfrac{k}{n-k}.$

极大似然估计性质：

设θ的函数$u=u(\theta),\theta\in\Theta$具有单值反函数，$\hat{\theta}$是$\theta$的极大似然估计，则$\hat{u}=u(\hat{\theta})$是$u(\theta)$的极大似然估计。

例：$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$是$\sigma^2$的极大似然估计，

$u=u(\sigma^2)=\sqrt{\sigma^2}$有单值反函数$\sigma^2=u^2,(u\geq 0)$

故$\hat{\sigma}=\sqrt{\hat{\sigma}^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$是$\sigma$的极大似然估计.

例8 设$X \sim N(\mu,\sigma^2)$，μ,σ^2未知，求使$P\{X > A\} = 0.05$的点A的极大似然估计量

解： $P\{X > A\} = 1 - \Phi(\dfrac{A-\mu}{\sigma}) = 0.05$

查表有 $\dfrac{A-\mu}{\sigma} = 1.645$，所以 $A = \mu + 1.645\sigma.$

由前面知μ和σ^2的极大似然估计量分别为

$$\hat{\mu} = \overline{X}, \quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

所以A的极大似然估计量为

$$\hat{A} = \hat{\mu} + 1.645\hat{\sigma} = \overline{X} + 1.645\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}.$$

§2 估计量的的评选标准

- 无偏性
- 有效性
- 一致性

我们注意到，在上一节中对于同一个未知参数，用不同方法可以得到不同的估计量。究竟采用哪个为好呢？这就涉及到用什么标准来评价估计量的问题。我们介绍三个常用的标准：

1）无偏性；

2）有效性；

3）一致性。

一、无偏性

若$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$的数学期望存在，且$E\hat{\theta}=\theta$，则称$\hat{\theta}$是$\theta$的无偏估计量。

例1 $X_1,\cdots,X_n$是总体$X\sim N(\mu,\sigma^2)$的样本，μ,σ^2均未知.

因为$E\overline{X}=\mu$， 所以$\hat{\mu}=\overline{X}$是μ的无偏估计量

而 $ES^2=\sigma^2$，

所以$\hat{\sigma}^2=S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$是$\sigma^2$的无偏估计量.

考察 $B_2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$

由于 $E\left(B_2\right)=E\left\{\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2\right\}$

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

$$=\frac{n-1}{n}E\left(S^2\right)=\frac{n-1}{n}\sigma^2$$

因此，$B_2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2$ 是总体方差 σ^2 的有偏估计.

例2 设总体 X 存在 m 阶矩，并设

$$EX^k = \mu_k, \quad (k = 1, 2, \cdots, m)$$

$X_1, \cdots, X_n$是总体X的样本，又设$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

是样本的k阶原点矩，$(k = 1, 2, \cdots, m)$

由于 $E(A_k) = E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^k\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i^k) = \frac{1}{n}\sum_{i=1}^{n}\mu_k = \mu_k$

因此 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是总体的k阶原点矩μ_k的无偏估计.

$(k = 1, 2, \cdots, m)$

例3 设总体 X 服从区间$[0, \theta]$上的均匀分布，其中$\theta > 0$为未知参数，$X_1, \cdots, X_n$是从该总体中抽取的一个样本.求θ的矩估计，并验证是否是无偏估计.

解： $EX = \dfrac{\theta}{2}$,

令 $\overline{X} = \dfrac{\theta}{2}$，得$\theta$的矩估计量为$\hat{\theta} = 2\overline{X}$.

由于

$$E(\hat{\theta}) = E(2\overline{X}) = 2E(\overline{X}) = 2E(X) = 2 \cdot \frac{\theta}{2} = \theta,$$

因此$\hat{\theta} = 2\overline{X}$ 是未知参数θ的无偏估计.

例4 设总体 $X \sim N(\mu, \ \sigma^2)$，其中$\mu$已知，而$\sigma^2 > 0$为未知参数，$X_1, \cdots, X_n$是从该总体中抽取的一个样本.

则由§7.1例3知，未知参数σ^2 的极大似然估计为

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

由于 $\dfrac{X_i - \mu}{\sigma} \sim N(0, \ 1), i = 1, \cdots, n,$ 且它们独立，

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu)^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu}{\sigma} \right)^2 \sim \chi^2(n)$$

例4（续）

因此，
$$E\left(\hat{\sigma}^2\right)=E\left(\frac{\sigma^2}{n}\sum_{i=1}^{n}\frac{\left(X_i-\mu\right)^2}{\sigma^2}\right)$$

$$=\frac{\sigma^2}{n}E\left(\sum_{i=1}^{n}\frac{\left(X_i-\mu\right)^2}{\sigma^2}\right)=\frac{\sigma^2}{n}\times n=\sigma^2$$

这表明，$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\mu\right)^2$ 是总体方差σ^2的无偏估计.

说明:

如果未知参数θ有两个不同的无偏估计$\hat{\theta}_1$与$\hat{\theta}_2$，则θ一定有无穷多个无偏估计.

这是因为，对任意的实数α，

$$\alpha\hat{\theta}_1+(1-\alpha)\hat{\theta}_2$$

一定是未知参数θ的无偏估计.

二、有效性

若$\hat{\theta}_1=\hat{\theta}_1(X_1,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,\cdots,X_n)$都是$\theta$的无偏估计量，且$D(\hat{\theta}_1)\le D(\hat{\theta}_2)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$有效.

$D(\hat{\theta}_1)=E(\hat{\theta}_1-\mathrm{E}\hat{\theta}_1)^2=E(\hat{\theta}_1-\theta)^2$ 表示$\hat{\theta}_1$与θ的偏离程度.

例5 设总体$X\sim N\left(\mu,\ \sigma^2\right)$，其中$\mu$未知，

X_1,X_2,X_3是从该总体中抽取的一个样本.

试验证：$\hat{\mu}_1=\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3$；

$\hat{\mu}_2=\frac{1}{3}X_1+\frac{1}{4}X_2+\frac{5}{12}X_3$；　$\hat{\mu}_3=\frac{1}{3}X_1+\frac{1}{6}X_2+\frac{1}{2}X_3$

都是未知参数μ的无偏估计，并指出在这三个μ的估计中，哪一个最有效？

例5（续）解： 由于

$$E(\hat{\mu}_1)=E\left(\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3\right)$$

$$=\frac{1}{5}E(X_1)+\frac{3}{10}E(X_2)+\frac{1}{2}E(X_3)=\frac{1}{5}\mu+\frac{3}{10}\mu+\frac{1}{2}\mu=\mu$$

$$E(\hat{\mu}_2)=E\left(\frac{1}{3}X_1+\frac{1}{4}X_2+\frac{5}{12}X_3\right)$$

$$=\frac{1}{3}E(X_1)+\frac{1}{4}E(X_2)+\frac{5}{12}E(X_3)=\frac{1}{3}\mu+\frac{1}{4}\mu+\frac{5}{12}\mu=\mu$$

$$E(\hat{\mu}_3)=E\left(\frac{1}{3}X_1+\frac{1}{6}X_2+\frac{1}{2}X_3\right)$$

$$=\frac{1}{3}E(X_1)+\frac{1}{6}E(X_2)+\frac{1}{2}E(X_3)=\frac{1}{3}\mu+\frac{1}{6}\mu+\frac{1}{2}\mu=\mu$$

例5（续）

这表明，$\hat{\mu}_1$，$\hat{\mu}_2$，$\hat{\mu}_3$ 都是未知参数μ的无偏估计.

$$D(\hat{\mu}_1) = D\left(\frac{1}{5}X_1 + \frac{3}{10}X_2 + \frac{1}{2}X_3\right)$$

$$= \frac{1}{25}D(X_1) + \frac{9}{100}D(X_2) + \frac{1}{4}D(X_3)$$

$$= \frac{1}{25}\sigma^2 + \frac{9}{100}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{684}{1800}\sigma^2$$

$$D(\hat{\mu}_2) = D\left(\frac{1}{3}X_1 + \frac{1}{4}X_2 + \frac{5}{12}X_3\right)$$

$$= \frac{1}{9}D(X_1) + \frac{1}{16}D(X_2) + \frac{25}{144}D(X_3)$$

$$= \frac{1}{9}\sigma^2 + \frac{1}{16}\sigma^2 + \frac{25}{144}\sigma^2 = \frac{625}{1800}\sigma^2$$

例5（续）

$$D(\hat{\mu}_3) = D\left(\frac{1}{3}X_1 + \frac{1}{6}X_2 + \frac{1}{2}X_3\right)$$

$$= \frac{1}{9}D(X_1) + \frac{1}{36}D(X_2) + \frac{1}{4}D(X_3)$$

$$= \frac{1}{9}\sigma^2 + \frac{1}{36}\sigma^2 + \frac{1}{4}\sigma^2 = \frac{700}{1800}\sigma^2$$

由于 $D(\hat{\mu}_2) < D(\hat{\mu}_1) < D(\hat{\mu}_3)$,

所以在 $\hat{\mu}_1$、$\hat{\mu}_2$、$\hat{\mu}_3$ 这三个估计量中, $\hat{\mu}_2$ 最有效.

例6 设总体 X 存在二阶矩，并设

$$EX=\mu,\qquad DX=\sigma^2,$$

$X_1,\cdots,X_n$是总体X的一个样本，又设

$a_i\geq 0,\ \ i=1,2,\cdots,n,$ 且 $\sum_{i=1}^{n}a_i=1$.

试证：

(1) $\sum_{i=1}^{n}a_iX_i$ 是 μ 的无偏估计；

(2) 在 μ 的所有形如上述的$\sum_{i=1}^{n}a_iX_i$ 估计中，$\overline{X}$ 的方差最小.

例6（续）

证明:

(1) 由于

$$E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \sum_{i=1}^{n} a_i \mu = \mu$$

所以，$\sum_{i=1}^{n} a_i X_i$ 是 μ 的无偏估计.

(2) 由 *Schwarz* 不等式，

$$\left(\sum_{i=1}^{n} u_i v_i\right)^2 \le \left(\sum_{i=1}^{n} u_i^2\right) \cdot \left(\sum_{i=1}^{n} v_i^2\right)$$

在上面的不等式中，令:

例6（续）

$$u_i = 1, i = 1,2,\cdots,n;$$

$$v_i = a_i, i = 1,2,\cdots,n.$$

则有

$$\left(\sum_{i=1}^{n} a_i\right)^2 \le n \cdot \left(\sum_{i=1}^{n} a_i^2\right)$$

所以，$D\overline{X} = \dfrac{DX}{n} = \dfrac{DX}{n} \cdot \left(\sum_{i=1}^{n} a_i\right)^2$ （因为$\sum_{i=1}^{n} a_i = 1$）

$$\le \frac{DX}{n} \cdot n\left(\sum_{i=1}^{n} a_i^2\right) = DX \cdot \left(\sum_{i=1}^{n} a_i^2\right)$$

例6（续）

$$D\overline{X} \le DX \cdot \left(\sum_{i=1}^{n} a_i^2\right) = \sum_{i=1}^{n} a_i^2 DX$$

$$= \sum_{i=1}^{n} a_i^2 DX_i = \sum_{i=1}^{n} D(a_i X_i) = D\left(\sum_{i=1}^{n} a_i X_i\right)$$

这表明，在μ的所有形如$\sum_{i=1}^{n} a_i X_i$的无偏估计中，$\overline{X}$的方差为最小，即$\overline{X}$为最有效.

三、一致性

若$\hat{\theta}=\hat{\theta}_n(X_1,\cdots,X_n)$为参数$\theta$的估计量,如果

对于任意$\theta\in\Theta$，当$n\longrightarrow\infty$时$\hat{\theta}_n\xrightarrow{P}\theta$.

则称$\hat{\theta}$是θ的一致估计。

例7 若$X_1,\cdots,X_n$是总体X的样本，$EX=\mu$,

$EX^k=\mu_k$. 由辛钦大数定律，知

对于任意给定的$\varepsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geq\varepsilon\right\}=0$$

因此，样本均值$\overline{X}$是总体期望μ的一致估计量.

例7（续）

还有 $$\lim_{n\to\infty} P\left\{ \left| \frac{1}{n}\sum_{i=1}^{n} X_i^k - \mu_k \right| \geq \varepsilon \right\} = 0$$

因此，样本的k阶原点矩$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$是总体$k$阶原点矩$\mu_k = EX^k$的一致估计量.