概率论与数理统计

第七章　参数估计

练习：

1、设$X_1,X_2,\ldots X_n$是取自总体X的一个样本

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

求 θ的最大似然估计值和矩估计量.

解(1)　似然函数为

$$L(\theta)=\prod_{i=1}^{n}\theta x_i^{\theta-1}=\theta^n\left(\prod_{i=1}^{n}x_i\right)^{\theta-1}\qquad \begin{array}{c}(0<x_i<1)\\ 1\le i\le n\end{array}$$

对数似然函数为

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

对数似然函数为

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

求导并令其为0

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

从中解得

$$\hat{\theta} = -n\Big/\sum_{i=1}^{n}\ln x_i$$

即为θ的最大似然估计值．

解(2) $E(X)=\int_0^1 x\theta x^{\theta-1}dx=\frac{\theta}{\theta+1}=\mu_1$

$$\therefore \theta=\frac{\mu_1}{1-\mu_1}$$

$\therefore$ 矩估量 $\hat{\theta}=\frac{\overline{X}}{1-\overline{X}}$

2、设总体$X \sim b(n,p), 0<p<1, X_1, X_2, \cdots, X_n$为其样本，$n$及$p$的矩估计分别是______

解：　由$\mu_1 = E(X) = np$

$$\mu_2 = E(X^2) = D(X) + \left[E(X)\right]^2 = np(1-p) + (np)^2$$

得　$$p = 1 + \mu_1 - \frac{\mu_2}{\mu_1}, \quad n = \frac{\mu_1^2}{\mu_1 + \mu_1^2 - \mu_2},$$

故 $$\hat{p} = 1 + \overline{X} - \frac{\frac{1}{n}\sum_{i=1}^{n} X_i^2}{\overline{X}}, \quad \hat{n} = \frac{\overline{X}^2}{\overline{X} + \overline{X}^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2},$$

或者表达为：　$$\hat{p} = 1 - \frac{\frac{n-1}{n}S^2}{\overline{X}}, \quad \hat{n} = \frac{\overline{X}^2}{\overline{X} - \frac{n-1}{n}S^2}$$

3、设总体$X \sim U[0,\theta]$,$(X_1,X_2,\cdots,X_n)$是来自X的样本，则θ的极大似然估计量是________

解：　由 $f(x)=\begin{cases}\dfrac{1}{\theta} & 0\le x\le\theta \\ 0 & \text{其它}\end{cases}$

似然函数 $L(\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_i\le\theta, i=1,2,\cdots n \\ 0 & \text{其它}\end{cases}$

即　$L(\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le \min\limits_{1\le i\le n}\{x_i\}\le \max\limits_{1\le i\le n}\{x_i\}\le\theta, \\ 0 & \text{其它}\end{cases}$

$\therefore \hat{\theta}=\max\limits_{1\le i\le n}\{x_i\}$

注：为使$L(\theta)$达到极大，就必须使θ尽可能小，但是θ不能小于$\max\{x_i\}$，因而θ取$\max\{x_i\}$时使$L(\theta)$达到极大

4、设$X_1,X_2,\cdots,X_n$是取自总体X的一个简单样本，则$E(X^2)$的矩估计是______

（A）$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ （B）$S_n^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

（C）$S^2+\overline{X}^2$ （D）$S_n^2+\overline{X}^2$

解：由$\mu_2=E(X^2)$ 故$\widehat{E(X^2)}=A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$

而 $S_n^2+\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$

注：$S_n^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-2\frac{1}{n}\sum_{i=1}^{n}X_i\overline{X}+\overline{X}^2$

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

5、设总体X的概率密度为

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

其中 $\alpha>1$是未知参数 ,$X_1, X_2, \ldots, X_n$ 是取自 X 的样本,

求参数 α的矩估计.

解 $$\mu_1=E(X)=\int_0^1 x(\alpha+1)x^{\alpha}dx=(\alpha+1)\int_0^1 x^{\alpha+1}dx=\frac{\alpha+1}{\alpha+2}$$

解得 $$\alpha=\frac{2\mu_1-1}{1-\mu_1}$$

总体矩

故α 的矩估计量为 $$\hat{\alpha}=\frac{2\overline{X}-1}{1-\overline{X}}$$

样本矩

§2 基于截尾样本的最大似然估计

- 定时截尾样本的最大似然估计
- 定数截尾样本的最大似然估计

在研究产品可靠性时, 需要研究产品寿命T的各种特征. 产品寿命T是一个随机变量, 它的分布称为寿命分布. 为了对寿命分布进行统计推断, 就需要通过对产品的寿命试验, 以取得寿命数据.

一种典型的寿命试验是, 将随机抽取的n个产品在时间t=0时, 同时投入试验, 直到每个产品都失效. 记录每个产品的失效时间, 这样得到的样本(即由所有产品的失效时间$0 \le t_1 \le t_2 \le \ldots \le t_n$所组成的样本)叫完全样本. 但产品的寿命往往较长, 我们不可能得到完全样本, 于是就考虑截尾寿命试验.

截尾寿命试验常用的有两种：一种是定时截尾寿命试验. 假设将随机抽取的n个产品在时间$t=0$时同时投入试验, 试验进行到事先规定的截尾时间t_0停止.

如试验截止时共有m个产品失效, 它们的失效时间分别为 $0\le t_1\le t_2\le\ldots\le t_m\le t_0$,

此时m是一个随机变量, 所得的样本 $t_1,t_2,\ldots,t_m$称为定时截尾样本.

另一种是定数截尾寿命试验. 假设将随机抽取的n个样本在时间$t=0$时同时投入试验, 试验进行到有m个(m是事先规定的, $m<n$)产品失效时停止.

m个失效产品的失效时间分别为

$$0 \le t_1 \le t_2 \le \ldots \le t_m,$$

这里t_m是第m个产品的失效时间. 所得的样本$t_1, t_2, \ldots, t_m$称为定数截尾样本. 用截尾来进行统计推断是可靠性研究中常见的问题.

设产品的寿命分布是指数分布,其概率密度为

$$f(t)=\begin{cases}\dfrac{1}{\theta}e^{-t/\theta}, & t>0,\\ 0, & t\le 0,\end{cases}$$

$\theta>0$未知. 设有n个产品投入定数截尾试验,截尾数为m, 得定数截尾样本$0\le t_1\le t_2\le\ldots\le t_m$, 现在要利用这一样本来估计未知参数$\theta$ (即产品的平均寿命). 在时间区间$[0,t_m]$有m个产品失效,而有$n-m$个产品在t_m时尚未失效, 即有$n-m$个产品寿命超过t_m.

用最大似然估计法估计θ，求上述观察结果的概率. 一个产品在$[t_i, t_i+\mathrm{d}t_i]$失效的概率近似为

$f(t_i)\,\mathrm{d}t_i = \dfrac{1}{\theta}\mathrm{e}^{-t_i/\theta}\,\mathrm{d}t_i, i=1,2,\cdots,m$, 其余$n-m$个产品寿命超过$t_m$的概率为

$$\left(\int_{t_m}^{\infty}\frac{1}{\theta}\mathrm{e}^{-t/\theta}\,\mathrm{d}t\right)^{n-m}=\left(e^{-t_m/\theta}\right)^{n-m},$$

故上述观察结果出现的概率近似地为

$$C_n^m\left(\frac{1}{\theta}\mathrm{e}^{-t_1/\theta}\,\mathrm{d}t_1\right)\left(\frac{1}{\theta}\mathrm{e}^{-t_2/\theta}\,\mathrm{d}t_2\right)\cdots\left(\frac{1}{\theta}\mathrm{e}^{-t_m/\theta}\,\mathrm{d}t_m\right)\left(\mathrm{e}^{-t_m/\theta}\right)^{n-m}$$

$$C_n^m\left(\frac{1}{\theta}\mathrm{e}^{-t_1/\theta}\mathrm{d}t_1\right)\left(\frac{1}{\theta}\mathrm{e}^{-t_2/\theta}\mathrm{d}t_2\right)\cdots\left(\frac{1}{\theta}\mathrm{e}^{-t_m/\theta}\mathrm{d}t_m\right)\left(\mathrm{e}^{-t_m/\theta}\right)^{n-m}$$

其中$\mathbf{d}\boldsymbol{t}_1,\ldots,\mathbf{d}\boldsymbol{t}_m$为常数. 因忽略一个常数因子不影响$\boldsymbol{q}$的最大似然估计, 故可取似然函数为

$$L(\theta)=\frac{1}{\theta^m}\mathrm{e}^{-\frac{1}{\theta}[t_1+t_2+\cdots+t_m+(n-m)t_m]}.$$

对数似然函数为

$$\ln L(\theta)=-m\ln\theta-\frac{1}{\theta}[t_1+t_2+\cdots+t_m+(n-m)t_m].$$

$$\ln L(\theta) = -m\ln\theta - \frac{1}{\theta}[t_1 + t_2 + \cdots + t_m + (n-m)t_m].$$

$$令 \frac{\mathrm{d}}{\mathrm{d}\theta}\ln L(\theta) = -\frac{m}{\theta} +$$

$$\frac{1}{\theta^2}[t_1 + t_2 + \cdots + t_m + (n-m)t_m] = 0.$$

于是得到θ 的最大似然估计为 $\hat{\theta} = \frac{s(t_m)}{m}.$

其中$s(t_m)=t_1+t_2+\ldots+t_m+(n-m)t_m$称为总试验时间，它表示直至时刻$t_m$为止$n$个产品的试验时间总和。

对于定时截尾样本

$$0 \le t_1 \le t_2 \le \ldots \le t_m \le t_0$$

(其中t_0是截尾时间), 与上面的讨论类似,

可得似然函数为

$$L(\theta) = \frac{1}{\theta^m} \mathrm{e}^{-\frac{1}{\theta}[t_1+t_2+\cdots+t_m+(n-m)t_0]},$$

θ的最大似然估计为

$$\hat{\theta} = \frac{s(t_0)}{m},$$

其中$s(t_0)=t_1+t_2+\ldots+t_m+(n-m)t_0$称为总试验时间, 它表示直至时刻$t_0$为止$n$个产品的试验时间的总和.

例　设电池的寿命服从指数分布, 其概率密度为

$$f(t)=\begin{cases}\dfrac{1}{\theta}\mathrm{e}^{-t/\theta}, & t>0,\\ 0, & t\le 0,\end{cases}$$

θ >0未知. 随机地取50只电池投入寿命试验, 规定试验进行到其中有15只失效时结束试验, 测得失效时间(小时)为

115,119,131,138,142,147,148,155,

158,159,163,166,167,170,172

试求电池的平均寿命估计.

解 n=50, m=15, $s(t_{15})$=115+119+...+172+(50-15)×172=8270, 得θ 的最大似然估计为

$$\hat{\theta}=\frac{8270}{15}=551.33(\text{小时})$$

§3 估计量的评选标准

- 无偏性
- 有效性
- 相合性

$X \sim N(\mu, \sigma^2)$

样本均值是否是 μ 的一个好的估计量?

样本方差是否是 σ^2 的一个好的估计量?

这就需要讨论以下几个问题:

(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) 如何求得合理的估计量?

估计量的评选标准

在介绍估计量的评选标准之前，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 . 因此，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．相合性

这里我们重点介绍前面两个标准．

一、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，而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．

设 $\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数θ 的估计量，若

$$E(\hat{\theta})=\theta$$

则称 $\hat{\theta}$为 θ 的无偏估计．

无偏性是对估计量的一个常见而重要的要求.

无偏性的实际意义是指没有系统性的偏差.

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 服从参数为 θ 的指数分布，其概率密度为

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-x/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

其中 $\theta>0$ 为未知，$X_1,X_2,\dots X_n$ 是取自总体的一个样本，

试证 $\bar{X}$ 和 $nZ=n\min(X_1,\dots,X_n)$ 都是参数 θ 的无偏估计量.

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

所以 $\overline{X}$ 是参数 θ 的无偏估计量.而

$Z=\min(X_1,\ldots,X_n)$ 具有概率密度

$$f_{\min}(x;\theta)=\begin{cases}\dfrac{n}{\theta}e^{-nx/\theta}, & x>0,\\ 0, & \text{其它},\end{cases}$$

故知 $E(Z)=\dfrac{\theta}{n}$, $E(nZ)=\theta$

即 nZ 也是参数 θ 的无偏估计量.

一个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，我们可以比较 $E(\hat{\theta}_1-\theta)^2$ 和 $E(\hat{\theta}_2-\theta)^2$ 的大小来决定二者谁更优．

由于

$$D(\hat{\theta}_1)=E(\hat{\theta}_1-\theta)^2$$

$$D(\hat{\theta}_2)=E(\hat{\theta}_2-\theta)^2$$

所以无偏估计以方差小者为好，这就引进了**有效性**这一概念．

二、有效性

设 $\hat{\theta}_1=\hat{\theta}_1(X_1,\cdots,X_n)$ 和 $\hat{\theta}_2=\hat{\theta}_2(X_1,\cdots,X_n)$

都是参数 θ 的无偏估计量，若对任意 $\theta\in\Theta$，

$$\mathrm{D}(\hat{\theta}_1)\le\mathrm{D}(\hat{\theta}_2)$$

且至少对于某个 $\theta\in\Theta$ 上式中的不等号成立，

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

例2 (续例1) 试证 当 n > 1 时 θ 的无偏估计量 $\bar{X}$ 较 $nZ=\min(X_1,\ldots,X_n)$有效.

证 $D(X)=\theta^2$,

故有 $D(\bar{X})=D(\frac{1}{n}\sum_{i=1}^{n}X_i)=\frac{1}{n^2}\sum_{i=1}^{n}D(X_i)=\frac{\theta^2}{n}$

而 $D(Z)=\frac{\theta^2}{n^2}$, 故有 $D(nZ)=\theta^2$.

当 $n>1$ 时, $D(nZ)>D(\bar{X})$, 故$\bar{X}$较 nZ 有效.

三、相合性

设 $\hat{\theta}(X_1,\ldots,X_n)$ 是参数 θ 的估计量，若对于任意 $\theta\in\Theta$， 当 $n\to\infty$ 时 $\hat{\theta}(X_1,\ldots,X_n)$ 依概率收敛于 θ，则称 $\hat{\theta}$ 为 θ 的相合估计量.

$\hat{\theta}$ 为 θ 的相合估计量

$\Leftrightarrow$ 对于任意 $\varepsilon>0$，有

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

由辛钦定理

若总体 X 的数学期望 $E(X)=\mu$ 有限, 则有

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X^k)=\mu_k \quad (k=1,2,\cdots)$$

$$g(A_1,A_2,\cdots,A_k) \xrightarrow{P} g(\mu_1,\mu_2,\cdots,\mu_k)$$

其中 g 为连续函数 .

故

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \text{ 为 } E(X^k) = \mu_k \ (k = 1, 2, \cdots) \text{ 的相合}$$

估计量.

若 g 为连续函数, 则有

$g(A_1, A_2, \cdots, A_k)$ 为 $g(\mu_1, \mu_2, \cdots, \mu_k)$ 的相合估计量.

小结

对于一个未知参数可以提出不同的估计量，因此自然提出比较估计量的好坏的问题，这就需要给出评定估计量好坏的标准.

在本节中,介绍了评定估计量好坏的三个标准:无偏性、有效性、和相合性.

作业：课后习题 **10、11、12、14**