第六章　参数估计

第一节　点估计

第二节　估计量的评选标准

第三节　区间估计

第四节　单正态总体均值与方差的区间估计

第一节　　点估计

一、点估计问题的提法

二、估计量的求法

1. 矩估计法

2. 最大似然估计法

一、点估计问题的提法

设总体 X 的分布函数形式已知, 但它的一个或多个参数为未知. $X_1, X_2, \cdots, X_n$ 是 X 的一个样本, $x_1, x_2, \cdots, x_n$ 为相应的一个样本值. 设 θ 是待估参数.

借助于总体 X 的一个样本来估计总体未知参数的值的问题称为**点估计问题**.

构造一个适当的统计量 $\hat{\theta}(X_1, X_2, \cdots, X_n)$, 用它的观察值 $\hat{\theta}(x_1, x_2, \cdots, x_n)$ 来估计未知参数 θ.

$$\left.\begin{array}{l}\hat{\theta}(X_1, X_2, \cdots, X_n)\text{称为}\theta\text{的估计量.}\\ \hat{\theta}(x_1, x_2, \cdots, x_n)\text{称为}\theta\text{的估计值.}\end{array}\right\}\text{通称估计, 简记为}\hat{\theta}.$$

二、估计量的求法

由于估计量是样本的函数, 是随机变量, 故对不同的样本值, 得到的参数值往往不同, 如何求估计量是关键问题.

常用构造估计量的方法: (两种)

矩估计法和最大似然估计法.

1. 矩估计法

基本概念

设 $X_1, X_2, \cdots, X_n$ 是来自总体X的样本,

若 $E(X^k), \quad k = 1,2,\cdots$存在,

称它为总体X 的 k 阶原点矩,简称 k 阶矩.

称 样本 k 阶(原点)矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, k = 1, 2, \cdots;$

其观察值 $a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k, k = 1, 2, \cdots.$

结论: 若总体 X 的 k 阶矩 $E(X^k) \underline{\underline{\text{记成}}} \mu_k$ 存在,

则当 $n \to \infty$ 时, $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k, k = 1, 2, \cdots.$

证明: 因为 $X_1, X_2, \cdots, X_n$ 独立且与 X 同分布,

所以 $X_1^k, X_2^k, \cdots, X_n^k$ 独立且与 X^k 同分布,

故有 $E(X_1^k) = E(X_2^k) = \cdots = E(X_n^k) = \mu_k.$

再根据第四章辛钦大数定律知

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

设 X 为连续型随机变量,其概率密度为 $f(x;\theta_1,\theta_2,\cdots,\theta_m)$, 或 X 为离散型随机变量, 其分布律为 $P\{X=x\}=p(x;\theta_1,\theta_2,\cdots,\theta_m)$, 其中 $\theta_1,\theta_2,\cdots,\theta_m$ 为待估参数,

假设总体 X 的k阶原点矩存在,

$$E(X^k)=\mu_k=\mu_k(\theta_1,\theta_2,\cdots,\theta_m) \quad k=1,2,\cdots$$

均为 $\theta_1,\theta_2,\cdots,\theta_m$ 的函数.

若 $X_1,X_2,\cdots,X_n$ 为来自总体X 的样本,

矩估计法的核心思想：

用样本的k阶原点矩$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$来估计总体的k阶原点矩 $\mu_k=E(X^k)$

具体做法： 考虑1~m阶矩，得方程组：

根据总体的分布计算总体的1~m阶矩

$$\begin{cases} E(X)=\mu_1(\theta_1,\theta_2,\cdots,\theta_m) \\ E(X^2)=\mu_2(\theta_1,\theta_2,\cdots,\theta_m) \\ \cdots \\ E(X^m)=\mu_m(\theta_1,\theta_2,\cdots,\theta_m) \end{cases}$$

用样本的k阶矩来代替总体的k阶矩

$$\begin{cases} \mu_1(\theta_1,\theta_2,\cdots,\theta_m)=\frac{1}{n}\sum_{i=1}^{n}X_i \\ \mu_2(\theta_1,\theta_2,\cdots,\theta_m)=\frac{1}{n}\sum_{i=1}^{n}X_i^2 \\ \cdots \\ \mu_m(\theta_1,\theta_2,\cdots,\theta_m)=\frac{1}{n}\sum_{i=1}^{n}X_i^m \end{cases}$$

这是一个包含 m 个未知参数 $\theta_1,\theta_2,\cdots,\theta_m$ 的方程组,

解出其中 $\theta_1,\theta_2,\cdots,\theta_m$.

把方程组的解 记作 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_m$,分别作为 $\theta_1,\theta_2,\cdots,\theta_m$ 的估计量,这个估计量称为矩估计量.

矩估计量的观察值称为矩估计值.

例6.1.3 设总体 X 在 $(0,\theta)$ 上服从均匀分布,其中 θ $(\theta>0)$ 未知, $(X_1, X_2,\cdots, X_n)$ 是来自总体 X 的样本, 求 θ 的矩估计量.

解 因为 $X \sim U(0,\theta)$， $E(X)=\dfrac{\theta}{2}$

根据矩估计法, 令 $E(X)=\dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$, 即 $\dfrac{\theta}{2}=\bar{X}$,

解得 $\hat{\theta}=2\bar{X}$ 即为所求 θ 的矩估计量.

补例1

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

求 θ 的矩估计量。

解

$$E(X)=\int_{-\infty}^{+\infty} xf(x)dx=\int_0^1 x(\theta+1)x^{\theta}dx$$

$$=\int_0^1(\theta+1)x^{\theta+1}dx=\frac{\theta+1}{\theta+2}$$

令 $E(X)=\overline{X}$ ，即 $\dfrac{\theta+1}{\theta+2}=\bar{X}$

解之得 θ 的矩估计量为 $\hat{\theta}=\dfrac{2\bar{X}-1}{1-\bar{X}}$

例6.1.1 设总体 X 的均值 $E(X)=\mu$ 和方差 $D(X)=\sigma^2$ $(\sigma>0)$都存在,但均未知, 又设 $X_1, X_2, \cdots, X_n$ 是一个样本, 求 μ 和 σ^2 的矩估计量.

解 $E(X)=\mu,$

$$E(X^2)=D(X)+[E(X)]^2=\sigma^2+\mu^2,$$

令
$$\begin{cases} \mu=\overline{X}, \\ \sigma^2+\mu^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2. \end{cases}$$

解方程组得到矩估计量分别为

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

上例表明:

不管总体服从什么类型的分布,

用样本均值$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$作为总体均值$E(X)$的矩估计,

用样本二阶中心矩 $B_2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$作为总体方差$D(X)$

的矩估计.

例6.1.2 设总体 X 服从参数为λ的泊松分布，$\lambda > 0$未知，8, 10, 11, 9, 10为从该总体抽取样本的一组观测值，试求λ的矩估计值.

解 因为 $E(X) = \lambda$,

根据矩估计法， 易知 $\hat{\lambda} = \bar{X}$ 为λ的矩估计量，

故λ的矩估计值为 $\hat{\lambda} = \bar{x} = \dfrac{8+10+11+9+10}{5}$=9.6

2. 最大似然估计法

最大似然思想：

引例 在一盒中放有红球和白球，且已知两种颜色的球的比例为1:9, 但不知何种球多, 一人从盒中有放回地随机抽取3个球, 发现都为红球, 试推断盒中红球多, 还是白球多?

分析：引进随机变量$X=\begin{cases}1, & \text{取得的球为红球}\\ 0, & \text{否则}\end{cases}$

则X就看作一个总体,显然它服从$0-1$分布,

它有一个未知参数p, p只能取0.1或0.9

现在我们进行了抽样, 得到一个容量为3的子样: 1,1,1. 那么这个样本值出现的概率是

$$P(X_1=1, X_2=1, X_3=1)=p^3$$

当p=0.1时其值为0.001, 而当p=0.9时其值为0.729.

显然此时我们认为p=0.9更合理.

最大似然原理: 一个E如有若干个可能结果A, B, C, ···, 在一次试验中, 结果A出现, 则一般认为试验条件对A出现有利, 即A出现的概率最大.

似然函数的定义

(1) 设总体 X 属离散型

设分布律 $P\{X=k\}=p(x;\theta)$, θ 为待估参数, $\theta\in\Theta$,

Θ 是 θ 的取值范围, $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本,

$x_1,x_2,\cdots,x_n$ 为该样本的一组观测值.

则事件 $\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}$ 发生的概率为

$$L(\theta)=L(\theta;x_1,x_2,\ldots,x_n)$$

$L(\theta)$称为样本似然函数.

$$=P(X_1=x_1,X_2=x_2,\ldots,X_n=x_n)=\prod_{i=1}^{n}p(x_i;\theta)$$

注意: 上式中$x_1,x_2,\cdots,x_n$是已知样本观测值, 视作已知常数, 未知变量只有θ.

(2) 设总体 X 属连续型

设概率密度为 $f(x;\theta)$, θ 为待估参数, $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本,

又设 $x_1, x_2, \cdots, x_n$ 为该样本的一组观测值

样本似然函数:

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

最大似然估计法(书定义6.1.2)

得到样本值 $x_1, x_2, \cdots, x_n$ 时,选取使似然函数 $L(\theta)$ 取得最大值的 $\hat{\theta}$ 作为未知参数 θ 的估计值,

即 $L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max\limits_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta)$.

($其中 Θ 是 θ 可能的取值范围)

这样得到的 $\hat{\theta}$ 与样本值 $x_1, x_2, \cdots, x_n$ 有关,记为

$\boxed{\hat{\theta}(x_1, x_2, \cdots, x_n)}$, 参数 θ 的最大似然估计值,

$\boxed{\hat{\theta}(X_1, X_2, \cdots, X_n)}$ 参数 θ 的最大似然估计量.

求最大似然估计量的步骤:

(一) 写出似然函数

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

(二) 取对数

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

(三) 对 θ 求导 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$, 并令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0}$, 对数似然方程

解方程即得未知参数 θ 的最大似然估计值 $\hat{\theta}$.

最大似然估计法也适用于分布中含有多个未知参数的情况. 此时只需令

$$\boxed{\frac{\partial}{\partial \theta_i}\ln L = 0, \qquad i = 1,2,\cdots,k.}$$ 对数似然方程组

解出由 k 个方程组成的方程组,即可得各未知参数 θ_i $(i = 1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例6.1.5 设X为某超市某种商品的月销售数，根据经验，X服从参数为λ的泊松分布，$\lambda > 0$未知，$x_1, x_2, \cdots, x_n$为该商品某n个月的销售数，求λ的最大似然估计.

解 X的分布律 $P\{X = x\} = \dfrac{\lambda^x}{x!} e^{-\lambda}$，$x = 0, 1, 2, \dots$

似然函数 $$L(\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!} e^{-\lambda} = \frac{\lambda^{\sum_{i=1}^{n} x_i}}{x_1! x_2! \cdots x_n!} e^{-n\lambda},$$

$$\ln L(\lambda) = \left(\sum_{i=1}^{n} x_i\right) \ln \lambda - n\lambda - \sum_{i=1}^{n} \ln(x_i!),$$

$$\ln L(\lambda)=\left(\sum_{i=1}^{n}x_i\right)\ln\lambda-n\lambda-\sum_{i=1}^{n}\ln(x_i\,!),$$

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\ln L(\lambda)=\frac{\sum\limits_{i=1}^{n}x_i}{\lambda}-n\quad=0$$

解得 λ 的最大似然估计值为 $\hat{\lambda}=\frac{1}{n}\sum\limits_{i=1}^{n}x_i=\bar{x}.$

λ 的最大似然估计量为 $\hat{\lambda}=\frac{1}{n}\sum\limits_{i=1}^{n}X_i=\bar{X}.$

这一估计量与矩估计量是相同的.

例6.1.6 设某电子元件的寿命T服从参数为 λ 的指数分布，$\lambda > 0$未知. 现任意抽取n个元件，测得它们的失效时间为 $x_1, x_2, \cdots, x_n$，求λ的最大似然估计值.

解 总体 X 的概率密度为 $f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0, \\ 0, & x \le 0, \end{cases}$

似然函数 $L(\lambda) = \prod_{i=1}^{n} f\left(x_i\right)$

$$= \begin{cases} \prod_{i=1}^{n} \lambda e^{-\lambda x_i}, & \forall x_i > 0, \\ 0, & \text{其他}. \end{cases} = \begin{cases} \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}, & \forall x_i > 0, \\ 0, & \text{其他}. \end{cases}$$

$\forall x_i > 0,\quad L(\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$

$$\ln L(\lambda) = n \ln \lambda - \lambda \left(\sum_{i=1}^{n} x_i \right),$$

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

解得 λ 的最大似然估计值为 $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} x_i} = \dfrac{1}{\bar{x}}.$

λ 的最大似然估计量为 $\hat{\lambda} = \dfrac{n}{\sum_{i=1}^{n} X_i} = \dfrac{1}{\bar{X}}.$

例6.1.7 设总体 $X \sim N(\mu,\sigma^2)$, μ,σ^2为未知参数, $x_1,x_2,\cdots,x_n$ 是来自 X 的一个样本值, 求 μ 和 σ^2 的最大似然估计量.

解 X的概率密度为 $f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$,

X 的似然函数为

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

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

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

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

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

由 $\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=0$ 解得 $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},$

由 $-\frac{n}{2\sigma^2}+\frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i-\mu)^2=0$ 解得

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

故 μ 和 σ^2 的最大似然估计量分别为

$\hat{\mu}=\bar{X},\quad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2.$ 它们与相应的矩估计量相同.

最大似然估计的不变性

设 θ 的函数 $u=u(\theta),\theta\in\Theta$ 具有单值反函数 $\theta=\theta(u),u\in U$, 又设 $\hat{\theta}$ 是 X 的概率密度函数 $f(x;\theta)$ (f 形式已知) 中的参数 θ 的最大似然估计, 则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计.

如例6.1.7中, σ^2 的最大似然估计为 $\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$,

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

故标准差 σ 的最大似然估计为 $\hat{\sigma}=\sqrt{\hat{\sigma}^2}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2}$.

例6.1.8 设总体 X 在 $[0, \theta]$ 上服从均匀分布,其中 θ 未知, $x_1, x_2, \cdots, x_n$ 是来自总体 X 的一个样本值,求 θ 的最大似然估计量.

解 X 的概率密度为 $f(x;\theta) = \begin{cases} \dfrac{1}{\theta}, & 0 \le x \le \theta, \\ 0, & \text{其他}. \end{cases}$

似然函数为

$$L(\theta) = \begin{cases} \dfrac{1}{\theta^n}, & 0 \le x_1, x_2, \cdots, x_n \le \theta, \\ 0, & \text{其他} \end{cases}$$

记 $x_{(1)}=\min(x_1,x_2,\cdots,x_n)$,

$$x_{(n)}=\max(x_1,x_2,\cdots,x_n),$$

因为$0\le x_1,x_2,\cdots,x_n\le\theta$等价于$0\le x_{(1)}\le x_{(n)}\le\theta$,

$$L(\theta)=\frac{1}{\theta^n}\le\frac{1}{(x_{(n)})^n},$$

即似然函数$L(\theta)$在 $\theta=x_{(n)}$ 时取到最大值$(x_{(n)})^{-n}$,

故θ的最大似然估计值 $\hat{\theta}=x_{(n)}=\max\limits_{1\le i\le n}x_i$,

θ的最大似然估计量 $\hat{\theta}=X_{(n)}=\max\limits_{1\le i\le n}X_i$.

补例2 设总体 X 在 $[a,b]$ 上服从均匀分布,其中 a, b 未知, $x_1, x_2, \cdots, x_n$ 是来自总体 X 的一个样本值, 求 a,b 的最大似然估计量.

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

似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_1, x_2, \cdots, x_n \le b,\\ 0, & \text{其他}\end{cases}$$

记 $x_{(1)} = \min(x_1, x_2, \cdots, x_n)$, $x_{(n)} = \max(x_1, x_2, \cdots, x_n)$,

因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(1)}, x_{(n)} \le b$,

于是对于满足条件 $a \le x_{(1)}, b \ge x_{(n)}$的任意$a,b$有

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

即$L(a,b)$在$a=x_{(1)}, \ b=x_{(n)}$时取到最大值$(x_{(n)}-x_{(1)})^{-n}$.

故a,b的最大似然估计值

$$\hat{a}=x_{(1)}=\min_{1\le i\le n} x_i, \qquad \hat{b}=x_{(n)}=\max_{1\le i\le n} x_i,$$

a,b的最大似然估计量

$$\hat{a}=X_{(1)}=\min_{1\le i\le n} X_i, \qquad \hat{b}=X_{(n)}=\max_{1\le i\le n} X_i.$$

补例3：书P167习题6.2

2. 设总体 X 具有分布律

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中 $0<\theta<1$ 未知. 已知取得了一组样本观测值 $x_1=1, x_2=2, x_3=1$, 试求 θ 的矩估计值和最大似然估计值.

$$E(X)=1\times\theta^2+2\times 2\theta(1-\theta)+3\times(1-\theta)^2=-2\theta+3$$

令 $-2\theta+3=\bar{X}$ 解得 $\hat{\theta}=\dfrac{3-\bar{X}}{2}$ 为所求θ的矩估计量.

将$\bar{x}=\dfrac{4}{3}$代入得, θ的矩估计值：$\hat{\theta}=\dfrac{5}{6}$.

补例3：书P167习题6.2

2. 设总体 X 具有分布律

X	1	2	3
P	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

其中 $0<\theta<1$ 未知. 已知取得了一组样本观测值 $x_1=1, x_2=2, x_3=1$, 试求 θ 的矩估计值和最大似然估计值.

似然函数 $L(\theta)=P(X_1=1, X_2=2, X_3=1)$

$$=\theta^2\cdot 2\theta(1-\theta)\cdot\theta^2=2(\theta^5-\theta^6)$$

$$\frac{\mathrm{d}L(\theta)}{\mathrm{d}\theta}=2(5\theta^4-6\theta^5)=0$$

解得 $\hat{\theta}=\dfrac{5}{6}$ 为θ的最大似然估计值.

练习册§6.1-6.2 P23

3．设总体X的分布律是：

$$P\{X=-1\}=\theta^2,\quad P\{X=0\}=2\theta(1-\theta),\qquad P\{X=1\}=(1-\theta)^2$$

从总体中抽取的一组样本观测值为：$x_1 = x_4 = -1,\qquad x_3 = x_6 = x_7 = 0,\qquad x_2 = x_5 = 1$

(1) 求θ的最大似然估计$\hat{\theta}$;

(2) 设 $X_1, X_2, \cdots, X_n$是来自总体X的样本，在$\theta = \hat{\theta}$时，由中心极限定理，计算

$$P\left\{\left|\frac{X_1 + X_2 + \cdots + X_{50}}{50}\right| \le 0.1\right\}$$

Ex：生产线上生产的产品，次品率是$p\ (0\le p\le 1)$，随机检查100件产品，结果是96个合格品4个次品，

(1) 次品记为1，合格品记为0，写出总体X的分布律；

(2) 求次品率p的矩估计值和最大似然估计值.

X	0	1
P	$1-p$	p

$E(X)=p$

p的矩估计值 $\hat{p}=\bar{x}=0.04$

总体X的分布律为 $P\{X=x\}=p^x(1-p)^{1-x}\quad (x=0,1)$

$$L(p)=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}=p^4(1-p)^{96}$$

$L'(p)=0$ 得p的最大似然估计值 $\hat{p}=\bar{x}=0.04$

第二节　估计量的评选标准

一、无偏性

二、有效性

三、相合性

问题的提出

从前一节可以看到, 对于同一个参数, 用不同的估计方法求出的估计量可能不相同, 如第一节的例6.1.3和例6.1.8. 而且, 很明显, 原则上任何统计量都可以作为未知参数的估计量.

问题

(1)对于同一个参数究竟采用哪一个估计量好?

(2)评价估计量的标准是什么?

下面介绍几个常用标准.

一、无偏性

若 $X_1, X_2, \cdots, X_n$ 为总体 X 的一个样本，$\theta \in \Theta$ 是包含在总体 X 的分布中的待估参数，(Θ 是 θ 的取值范围)

若估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ 的数学期望 $E(\hat{\theta})$ 存在，且对于任意 $\theta \in \Theta$ 有 $\boxed{E(\hat{\theta}) = \theta,}$ 则称 $\hat{\theta}$ 是 θ 的无偏估计量.

无偏估计的实际意义：无系统误差.

结论 对于均值 μ，方差 $\sigma^2>0$ 都存在的总体，若 μ，σ^2 均为未知，$X_1, X_2, \cdots, X_n$ 为一组样本，可构造出

$$\hat{\mu}=\bar{X}, \qquad \hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2, \quad S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$$

则：

$E(\bar{X})=\mu$，$\bar{X}$ 是 μ 的无偏估计量；

$E(S^2)=\sigma^2$，$S^2=\dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是 σ^2 的无偏估计量；

$E(\hat{\sigma}^2)\neq\sigma^2$，$\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2$ 是 σ^2 的有偏估计量.

设总体X（不管服从什么分布，只要均值和方差存在）的均值$E(X)=\mu$，方差 $D(X)=\sigma^2$；$X_1, X_2, \cdots, X_n$是来自总体X的样本，$\overline{X}, S^2$分别是样本均值和样本方差，则有

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

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

$$E(S^2)=\sigma^2$$

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

$$=\frac{1}{n}\left[\sum_{i=1}^{n}E(X_i^2)-nE(\bar{X}^2)\right]$$

$$=\frac{1}{n}\left\{\sum_{i=1}^{n}\left[D(X_i)+\left(EX_i\right)^2\right]-n\left[D(\bar{X})+\left(E\bar{X}\right)^2\right]\right\}$$

$$=\frac{1}{n}\left[n(\sigma^2+\mu^2)-n(\frac{\sigma^2}{n}+\mu^2)\right]=\frac{n-1}{n}\sigma^2\neq\sigma^2$$ （有偏）

$$E\left(S^2\right)=E\left[\frac{1}{n-1}\left(\sum_{i=1}^{n}X_i^{\,2}-n\bar{X}^{2}\right)\right]=E\left(\frac{n}{n-1}\hat{\sigma}^2\right)=\frac{n}{n-1}E\left(\hat{\sigma}^2\right)=\sigma^2$$

（无偏）

练习册P22 § 6.1-6.2

2. 设 $X_1, X_2, \cdots, X_n$是来自总体$N(0, \sigma^2)$的样本，下列四个估计量是σ^2的无偏估计的是________.

① $\frac{1}{n}\sum_{i=1}^{n} {X_i}^2$ ② $\frac{1}{n}\sum_{i=1}^{n} {X_i}^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$ ③ $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ ④ $\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2$

5. 设 $X_1, X_2, \cdots, X_n$是来自于总体$X \sim N(\mu, 1)$的样本，则$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) =$______________,

$E(\bar{X} - \mu)^2 =$______________, $E[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2] =$___________.

例6. 2. 1

设$X_1, X_2, \cdots, X_n$是来自总体X的样本，作为总体均值μ的估计量有

$$T_1 = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad T_2 = X_1, \quad T_3 = \sum_{i=1}^{n} a_i X_i$$

其中$a_i > 0 (i = 1, 2 \cdots n)$，且$\sum_{i=1}^{n} a_i = 1$

试证T_1, T_2, T_3都是μ的无偏估计量；

证明 $E(T_1) = E(\overline{X}) = \mu \quad E(T_2) = E(X_1) = \mu$

$$E(T_3) = E\left(\sum_{i=1}^{n} a_i X_i\right) = \sum_{i=1}^{n} a_i E(X_i) = \mu\left(\sum_{i=1}^{n} a_i\right) = \mu$$

例6.2.2 设总体 X 在 $[0,\theta]$上服从均匀分布,参数$\theta>0$,

$X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本，试验证$\hat{\theta}_1=2\bar{X}$ 和

$\hat{\theta}_2=X_{(n)}=\max(X_1, X_2, \cdots, X_n)$是否是$\theta$的无偏估计量.

证 因$E(\hat{\theta}_1)=E(2\bar{X})=2E(\bar{X})=2E(X)=2\times\dfrac{\theta}{2}=\theta$,

所以$\hat{\theta}_1=2\bar{X}$ 是 θ 的无偏估计量.

因 $X_{(n)}=\max(X_1, X_2, \cdots, X_n)$的概率密度

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

所以 $E(X_{(n)}) = \int_0^{\theta} x \cdot \frac{nx^{n-1}}{\theta^n} \mathrm{d}x = \frac{n}{n+1}\theta \neq \theta$

故 $\hat{\theta}_2 = X_{(n)}$不是 θ 的无偏估计量.

令 $\hat{\theta}_3 = \frac{n+1}{n} X_{(n)}$,

无偏化过程

则有$E\left(\hat{\theta}_3\right) = E\left(\frac{n+1}{n} X_{(n)}\right) = \frac{n+1}{n} E\left(X_{(n)}\right) = \theta$,

故$\hat{\theta}_3 = \frac{n+1}{n} X_{(n)}$ 也是θ的无偏估计量.

练习册P23 § 6.1-6.2

3. 设总体服从指数分布，概率密度函数为：$f(x)=\begin{cases}\frac{1}{\theta}e^{-x/\theta}, & x\geq 0\\ 0, & x<0\end{cases}$， $X_1, X_2, \cdots, X_n$是取自总体的样本，

(1) 求θ的最大似然估计$\hat{\theta}_1$;

(2) 记$X_{(1)}=\min(X_1, X_2, \cdots\cdots X_n)$, 求$X_{(1)}$的概率密度函数;

(3) 证明：$\hat{\theta}_1$和 $\hat{\theta}_2=nX_{(1)}$ 都是θ的无偏估计.

练习册P23 T3

设总体 X 服从参数为 θ 的指数分布, 概率密度

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

其中参数 $\theta>0$, 又设 $X_1, X_2, \cdots, X_n$ 是来自总体 X 的样本, 试证 $\bar{X}$ 和 $nX_{(1)}=n[\min(X_1, X_2, \cdots, X_n)]$ 都是 θ 的无偏估计.

证明 因为 $E(\bar{X})=E(X)=\theta$,

所以 $\bar{X}$ 是 θ 的无偏估计量.

$X_{(1)}$的分布函数 $F_{\min}(x)=1-[1-F(x)]^n=\begin{cases}1-\mathrm{e}^{-\frac{nx}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$

$X_{(1)}$的密度函数 $f_{\min}(x)=\begin{cases}\dfrac{n}{\theta}\mathrm{e}^{-\frac{nx}{\theta}}, & x>0,\\ 0, & \text{其他}.\end{cases}$

即$X_{(1)}=\min(X_1, X_2, \ldots, X_n)$服从参数为$\dfrac{n}{\theta}$的指数分布.

故知 $E(X_{(1)})=\dfrac{\theta}{n}$, $\quad E(nX_{(1)})=\theta$,

所以$nX_{(1)}$也是θ的无偏估计量.

由以上例子可知,一个参数可以有不同的无偏估计量.

二、有效性

比较参数θ的两个无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，如果在样本容量n相同的情况下，$\hat{\theta}_1$的观察值在真值θ的附近较$\hat{\theta}_2$更密集，则认为$\hat{\theta}_1$较$\hat{\theta}_2$有效.

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$与$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$都是$\theta$的无偏估计量，若有$\boxed{D(\hat{\theta}_1)\le D(\hat{\theta}_2)}$,

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

例6.2.1(续)

设$X_1,X_2,\cdots,X_n$是来自总体X的样本,作为总体均值μ的估计量有

$$T_1=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i,\quad T_2=X_1,\quad T_3=\sum_{i=1}^{n}a_iX_i$$

其中$a_i>0(i=1,2\cdots n)$,且$\sum_{i=1}^{n}a_i=1$

试证T_1,T_2,T_3都是μ的无偏估计量;

设总体X的方差$D(X)$存在,试问T_1,T_2,T_3哪个更有效?

证明 $E(T_1)=E(\overline{X})=\mu \qquad E(T_2)=E(X_1)=\mu$

$$E(T_3)=E\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i E(X_i)=\mu\left(\sum_{i=1}^{n} a_i\right)=\mu$$

$$D(T_1)=D(\overline{X})=\frac{1}{n}D(X)=\frac{\sigma^2}{n} \qquad D(T_2)=D(X_1)=\sigma^2$$

$$D(T_3)=D\left(\sum_{i=1}^{n} a_i X_i\right)=\sum_{i=1}^{n} a_i^2 D(X_i)=\sigma^2\left(\sum_{i=1}^{n} a_i^2\right)$$

注意 $\displaystyle\sum_{i=1}^{n} a_i^2 \geq \frac{1}{n}$

所以 $T_1=\overline{X}$ 是三个无偏估计量中最有效的估计量

例6.2.2续

设总体 X 在 $[0,\theta]$上服从均匀分布,参数$\theta>0$, $X_1,X_2,\cdots,X_n$ 是来自总体 X 的样本，试证明$\hat{\theta}_1=2\bar{X}$ 和$\hat{\theta}_3=\dfrac{n+1}{n}X_{(n)}$都是$\theta$的无偏估计.

现证当$n\geq 2$时，比较$\hat{\theta}_1$和$\hat{\theta}_3$哪一个更有效.

证明 由于 $D(\hat{\theta}_1)=D(2\bar{X})=4D(\bar{X})=\dfrac{4}{n}D(X)=\dfrac{\theta^2}{3n}$,

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

又因为 $X_{(n)} \sim f(x) = \begin{cases} \dfrac{nx^{n-1}}{\theta^n}, & 0 \le x \le \theta \\ 0, & \text{其他} \end{cases},$

$$E(X_{(n)}) = \frac{n}{n+1}\theta, \quad E(X_{(n)}^2) = \int_0^\theta x^2 \frac{nx^{n-1}}{\theta^n}\mathrm{d}x = \frac{n}{n+2}\theta^2,$$

$$D(X_{(n)}) = E(X_{(n)}^2) - [E(X_{(n)})]^2$$

$$= \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 = \frac{n}{(n+1)^2(n+2)}\theta^2,$$

故 $D(\hat{\theta}_3) = \left(\dfrac{n+1}{n}\right)^2 D\left(X_{(n)}\right) = \dfrac{1}{n(n+2)}\theta^2 < D(\hat{\theta}_1) = \dfrac{\theta^2}{3n},$

当$n \ge 2$, 所以 $D(\hat{\theta}_3) < D(\hat{\theta}_1)$, $\hat{\theta}_3$ 较 $\hat{\theta}_1$ 有效.

练习册P22 § 6.1-6.2

1. 设X_1, X_2, X_3是来自总体$X \sim U(0,\theta)$的样本，

(1) 验证$\hat{\theta}_1 = (2X_1 + 4X_3)/3$是$\theta$的无偏估计；

(2) 验证$\hat{\theta}_2 = \frac{4}{3}\max\{X_1, X_2, X_3\}$是$\theta$的无偏估计；

(3) 上述哪个估计更有效?

三、相合性

若$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$为参数$\theta$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$\theta$，则称$\hat{\theta}$为$\theta$的相合估计量.

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

注：矩估计量具有相合性