第七章　参数估计

1. 参数的点估计
2. 估计量的评选标准
3. 参数的区间估计
4. 正态总体均值与方差的区间估计

- 统计推断
  - 估计问题
    - 参数估计
      - 点估计
        - 矩估计
        - 最大似然估计
      - 区间估计
    - 非参数估计
  - 假设检验问题

§1 参数的点估计

1. 矩估计法

2. 最大似然估计法

参数估计就是利用总体抽样得到的信息来估计总体的某些参数或参数的函数.

设总体X的分布函数的形式为已知的$F(x,\theta_1,\theta_2,\ldots,\theta_r)$,其中$x$是自变量, $\theta=(\theta_1,\theta_2,\ldots,\theta_r)$为未知参数(它可以是一个数,也可以是一个向量).
借助于总体X的样本$X_1, X_2, \ldots, X_n$,相应的样本观测值为$x_1, x_2, \ldots, x_n$,来估计未知参数θ的值.

点估计 构造适当的统计量$\hat{\theta}=\hat{\theta}(X_1, X_2, \ldots, X_n)$,用它的观测值$\hat{\theta}=\hat{\theta}(x_1, x_2, \ldots, x_n)$来估计未知参数$\theta$的值.
称统计量$\hat{\theta}(X_1, X_2, \ldots, X_n)$为$\theta$的**点估计量**,
$\hat{\theta}(x_1, x_2, \ldots, x_n)$为$\theta$的**点估计值**.

矩估计法

思想方法　**基于“替换”思想**

用样本 k 阶矩估计总体 k 阶矩, 建立含有待估参数的方程, 从而解出待估参数.

理论依据　**辛钦大数定律**

总体 k 阶矩　$\mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_r) = E(X^k),\quad k = 1, 2, \cdots r$

样本 k 阶矩　$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,\quad k = 1, 2, \cdots, r.$

当$n \to \infty$时，样本k阶矩$A_k \xrightarrow{P}$总体k阶矩μ_k

样本矩的连续函数$\xrightarrow{P}$总体矩的同一连续函数

设总体 X 的分布函数为 $F(x;\theta_1,\theta_2,\cdots,\theta_r)$，其中 $\theta_1,\theta_2,\cdots,\theta_r$ 为 r 个未知参数.

假设总体 X 的各阶原点矩 $E(X^k)$ $(k=1,2,\cdots,r)$ 存在, 则 $E(X^k)$ 是 $\theta_1,\theta_2,\cdots,\theta_r$ 的函数,即

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

步骤1.将总体k阶矩用未知参数表示

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

对于总体 X 的样本$(X_1, X_2, \ldots, X_n)$,样本的 k 阶原点矩为

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k, \quad k = 1, 2, \cdots, r.$$

根据辛钦大数定律 $A_k \xrightarrow{P} \mu_k \,(n \to \infty)$,令

$$A_k = \mu_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_r) = E\left(X^k\right), k = 1, 2, \cdots, r.$$

步骤2.用μ_k的估计量A_k分别替代μ_k,得方程(方程组)

即

$$\begin{cases} \mu_1(\theta_1, \theta_2, \cdots, \theta_r) = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i, \\ \mu_2(\theta_1, \theta_2, \cdots, \theta_r) = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ \mu_r(\theta_1, \theta_2, \cdots, \theta_r) = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^r. \end{cases}$$

步骤3.从r个方程中求解$\theta_k(k=1,\ldots r)$, 得矩估计量$\hat{\theta}_k=h_k(A_1,A_2,\cdots,A_r)$

从上述方程组中解出 $\theta_1,\theta_2,\cdots,\theta_r$,分别记作

$$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n),$$
$$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n),$$
$$\cdots\qquad\cdots\qquad\cdots\qquad\cdots$$
$$\hat{\theta}_r=\hat{\theta}_r(X_1,X_2,\cdots,X_n).$$

以此作为未知参数$\theta_1,\theta_2,\cdots,\theta_r$ 的估计量,称为矩估计量

如果样本观察值为$(x_1, x_2, \ldots,x_n)$,则得未知参数 $\theta_1,\theta_2,\cdots,\theta_r$的矩估计值为

$$\hat{\theta}_1=\hat{\theta}_1(x_1,x_2,\cdots,x_n),$$
$$\hat{\theta}_2=\hat{\theta}_2(x_1,x_2,\cdots,x_n),$$
$$\cdots\qquad\cdots\qquad\cdots$$
$$\hat{\theta}_r=\hat{\theta}_r(x_1,x_2,\cdots,x_n).$$

矩估计法计算步骤

步骤1. 将总体k阶矩用未知参数表示

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

步骤2. 用μ_k的估计量A_k分别替代μ_k，得方程(方程组)

步骤3. 从k个方程中求解θ_k，得矩估计量$\hat{\theta}_k = h_k(A_1, A_2, \cdots, A_r)$

例 已知某电话局在单位时间内收到用户的呼唤次数这个总体X服从参数为λ 的泊松分布, 其中$\lambda>0$ 未知. 今获得X的一组样本值$x_1, x_2, \ldots, x_n$, 求 λ 的矩估计量和矩估计值.

解 $X \sim \pi(\lambda)$, $\mu_1 = E(X) = \lambda$, 令$\mu_1 = A_1$, 即$\lambda = A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$,

得未知参数λ的矩估计量为$\hat{\lambda} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

代入样本值，得未知参数λ 的矩估计值为$\hat{\lambda} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

进而可以估计在单位时间内收到k 次呼唤的概率

$$\hat{P}\{X = k\} = \frac{\hat{\lambda}^k}{k!}e^{-\hat{\lambda}} = \frac{\bar{x}^k}{k!}e^{-\bar{x}}, k = 0,1,2\cdots$$

例 设总体 X 在区间 $[a, b]$ 上服从均匀分布,a与b为未知, $X_1, X_2, \ldots, X_n$ 是来自总体X的样本, 求a与b的矩估计量.

解 $\mu_1 = E(X) = \dfrac{a+b}{2}$,

$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4}$.

令 $\begin{cases} \mu_1 = A_1, \\ \mu_2 = A_2, \end{cases}$

即 $$\begin{cases} \dfrac{a+b}{2} = A_1 = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \\ \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4} = A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2, \end{cases}$$

整理得 $$\begin{cases} a+b = 2A_1, \\ b-a = \sqrt{12(A_2 - A_1^2)}. \end{cases}$$

$$A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = B_2$$

于是得到 a、b 的矩估计量为

$$\hat{a} = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}, \quad \hat{b} = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

例 设总体 X 的均值为μ,方差为$\sigma^2(\sigma>0)$且都未知,又设总体X 的一个样本为$X_1, X_2, \cdots, X_n$, 求μ 与σ^2 的矩估计量.

解 $\mu_1 = E(X) = \mu,$

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2 .$$

令 $\begin{cases} \mu_1 = A_1, \\ \mu_2 = A_2, \end{cases}$ 即 $\begin{cases} \mu = A_1 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i = \bar{X}, \\ \sigma^2 + \mu^2 = A_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2. \end{cases}$

解此方程组得到μ 与σ^2 的矩估计量为 $\hat{\mu} = A_1 = \bar{X},$

$$\hat{\sigma}^2 = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = B_2.$$

注1 总体μ和σ^2存在,总体均值μ的矩估计量为样本均值$\overline{X}$,

总体方差σ^2的矩估计量为样本二阶中心矩B_2,

无须知道其分布.

注2 矩估计量不唯一

例如总体$X \sim \pi(\lambda)$, 则 $E(X)=D(X)=\lambda$,

$$\begin{cases} \hat{\lambda}=\overline{X}, \\ \hat{\lambda}=B_2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2. \end{cases}$$

注3 矩估计方法由**Pearson**提出,直观简单,前提是原点矩存在.
(如**Cauchy**分布不适用).

最大似然估计法

思想方法 概率较大的事件在一次试验最有可能出现.

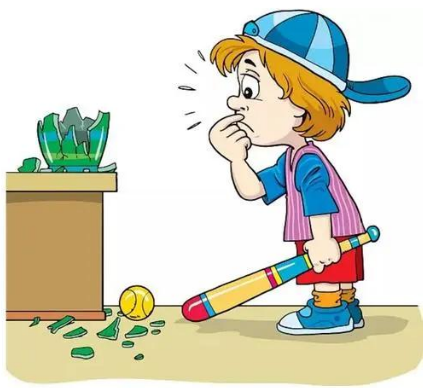
例如 有两外形相同的箱子,各装100个球,

甲箱 99个白球 1 个红球

乙箱 1 个白球 99个红球

现从两箱中任取一箱, 并从箱中任取一球,结果所取球是白球.问所取的球来自哪一箱?

答 甲箱.



设总体X分布类型已知,含有未知参数θ.

设 $X_1, X_2, \ldots, X_n$为来自 X 的样本, $x_1, x_2, \ldots, x_n$ 为样本观测值.

设总体X是离散型随机变量,其分布律为 $P\{X = x_k\} = p(x_k;\theta), \quad k = 1,2,\cdots$

样本联合概率分布

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

称$L(\theta)$为样本的似然函数.

似然函数的值的大小意味着该样本值出现的可能性的大小.概率最大的事件在一次试验最有可能出现.

目标 对于取定的样本值,选取适当的θ, 使得$L(\theta)$最大.

似然函数 $L(\theta)=\prod_{i=1}^{n}p(x_i;\theta).$

设 θ 取值范围为 Θ, 若有 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)\in\Theta$, 使得

$$L(\hat{\theta})\geq L(\theta),\quad \theta\in\Theta,$$

则称 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$ 为 θ 的**最大似然估计值,**

$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 θ 的**最大似然估计量.**

似然函数 $L(\theta)=\prod_{i=1}^{n} p(x_i;\theta).$

求θ的最大似然估计值方法

若$L(\theta)$可导, 令 $\dfrac{\mathrm{d}L(\theta)}{\mathrm{d}\theta}=0$, 称为**似然方程.**

常转化为对数似然函数,考虑 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0$, 称为**对数似然方程.**

$$\ln(XY)=\ln X+\ln Y$$

$$\ln X^a=a\ln X$$

若总体**X**是离散型随机变量,其分布律为 $P\{X=x\}=p(x;\theta)$,

似然函数 $L(\theta)=P\{X_1=x_1,X_2=x_2,\cdots,X_n=x_n\}=\prod_{i=1}^{n}p(x_i;\theta).$

若总体**X**是连续型随机变量,其概率密度为 $f(x;\theta)$,

似然函数 $L(\theta)=\prod_{i=1}^{n}f(x_i;\theta).$

最大似然估计法计算步骤

1.将$X_1, X_2, \cdots, X_n$代入样本值$x_1, x_2, \cdots, x_n$, 作为已知常数, 将参数θ作为自变量, 得似然函数

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

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

(1)对$L(\theta)$取对数得$\ln L(\theta)$,

(2)令$\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0.$

3.求解最大似然估计值(量)$\hat{\theta}$.

1．离散型总体

例 设总体 $X \sim B(m, p)$,其中$0 < p < 1$为未知参数, $X_1, X_2, \cdots, X_n$为 $\boldsymbol{X}$ 的样本,求 p 的矩估计量和最大似然估计量.

解 由 $\mu_1 = E(X) = mp$， 令 $\mu_1 = A_1$，即 $mp = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$,

得到 p 的矩估计量为 $\hat{p} = \dfrac{1}{m}\bar{X}$.

例 设总体 $X \sim B(m,p)$,其中$0<p<1$为未知参数, $X_1,X_2,\cdots,X_n$为 $\boldsymbol{X}$ 的样本,求 p 的矩估计量和最大似然估计量.

解 设样本观测值 $x_1,x_2,\cdots,x_n$, $P\{X=x\}=C_m^x p^x(1-p)^{m-x}, x=0,1,\ldots,m.$

似然函数 $L(p)=\prod_{i=1}^{n} C_m^{x_i} p^{x_i}(1-p)^{m-x_i}=\prod_{i=1}^{n} C_m^{x_i} p^{\sum_{i=1}^{n} x_i}(1-p)^{mn-\sum_{i=1}^{n} x_i}$,

取对数 $\ln L(p)=\sum_{i=1}^{n}\ln C_m^{x_i}+(\sum_{i=1}^{n}x_i)\ln p+(mn-\sum_{i=1}^{n}x_i)\ln(1-p).$

令

$$\frac{\mathrm{d}\ln L(p)}{\mathrm{d}p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{(mn-\sum_{i=1}^{n}x_i)}{1-p}=0,$$

得 $\boldsymbol{p}$ 的最大似然估计值及量为 $\hat{p}=\frac{\sum_{i=1}^{n}x_i}{mn}=\frac{\bar{x}}{m}$ 和 $\hat{p}=\frac{\sum_{i=1}^{n}X_i}{mn}=\frac{\bar{X}}{m}$.

2．连续型总体

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

其中λ 为未知参数,$(X_1, X_2, \dots, X_n)$为样本,求 λ的最大似然估计量.

解 设样本值为$(x_1, x_2, \dots, x_n)(x_i>0, i=1,2,\dots,n)$,

似然函数为 $L(\lambda)=\prod_{i=1}^{n} f(x_i,\lambda)=\prod_{i=1}^{n}\lambda \mathrm{e}^{-\lambda x_i}=\lambda^n \mathrm{e}^{-\lambda\sum_{i=1}^{n}x_i}$,

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

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

得λ 的最大似然估计值为$\hat{\lambda}=\frac{1}{\overline{x}}$, 最大似然估计量为 $\hat{\lambda}=\frac{1}{\overline{X}}$.

例 设总体 X 的概率密度为 $f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1,\\ 0, & \text{其它},\end{cases}$ 其中θ>-1为未知参数,又设$X_1, X_2, \dots, X_n$为X的样本,求θ的矩估计量与最大似然估计量.

解 由于 $\mu_1=E(X)=\int_{-\infty}^{+\infty}xf(x)\mathrm{d}x=\int_0^1 x(\theta+1)x^{\theta}\mathrm{d}x=\dfrac{\theta+1}{\theta+2}$,

令 $\mu_1=A_1$, 即

$$\frac{\theta+1}{\theta+2}=\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X},$$

解得θ的矩估计量为

$$\hat{\theta}=\frac{2\bar{X}-1}{1-\bar{X}}.$$

例 设总体 X 的概率密度为 $f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1,\\ 0, & 其它,\end{cases}$ 其中θ>-1为未知参数,又设$X_1, X_2, \ldots, X_n$为X的样本,求θ的矩估计量与最大似然估计量.

解 设样本值为$x_1, x_2, \ldots, x_n$ $(0<x_i<1)$,

似然函数为 $L(\theta)=\prod_{i=1}^{n}(\theta+1)x_i^{\theta}=(\theta+1)^n\left(\prod_{i=1}^{n}x_i\right)^{\theta}$,

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

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

得θ的最大似然估计值为$\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln x_i}-1$, 最大似然估计量为$\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln X_i}-1$.

3. 总体的分布中含有多个未知参数的情形

此时θ是未知向量,求解似然方程组.

设总体 X 的分布中含有r个未知参数$\theta_1, \theta_2, \ldots\theta_r$, 则似然函数是这些参数的函数

$$L = L(\theta_1, \theta_2, \cdots, \theta_r),$$

求出L 或$\ln L$关于θ_i的偏导数,并令它等于零,得到似然方程组或对数似然方程组

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k$$

或

$$\frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, k$$

解得r个未知参数的最大似然估计值.

例 设总体 $X \sim N(\mu,\sigma^2)$，μ 与 σ^2 未知，$(X_1,X_2,\ldots,X_n)$ 为总体 X 的样本，求 μ 与 σ^2 的最大似然估计量.

解 X 的概率密度为 $f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$，设 $x_1,x_2,\ldots,x_n$ 为样本值，

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

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

令

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

$$\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x},$$

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

解得μ与 σ^2的最大似然估计值为

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

μ与σ^2的最大似然估计量为

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

由此可得则标准差σ的最大似然估计量为 $\hat{\sigma}=\sqrt{\hat{\sigma}^2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}.$

注 最大似然估计的性质

设$\boldsymbol{u}(\theta)$是关于未知参数θ的函数, $\theta\in\Theta$, $\boldsymbol{u}(\theta)$具有单值反函数,又设$\hat{\theta}$是总体分布中所含参数θ的最大似然估计,则$\hat{u}=u(\hat{\theta})$是 $\boldsymbol{u}$ 的最大似然估计.

4. 似然方程无效

此时$L(\theta)$不是θ的连续函数, 若参数空间有界, 利用定义分析.

例 设总体 X 在区间$[a, b]$上服从均匀分布,其中a, b未知,$X_1, X_2,\dots, X_n$为总体 X 的样本,求a、b的最大似然估计.

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

设样本值为 $x_1, x_2, \dots, x_n$ ($a\le x_i\le b$),似然函数为 $L(a,b)=\dfrac{1}{(b-a)^n}$.

对数似然函数$\ln L(a,b)=-n\ln(b-a)$. $\begin{cases}\dfrac{\partial \ln L(a,b)}{\partial a}=\dfrac{n}{b-a},\\ \dfrac{\partial \ln L(a,b)}{\partial b}=\dfrac{-n}{b-a}.\end{cases}$

似然方程无效

因为$L(a, b)$是关于a的单调递增函数, a越大, $L(a, b)$就越大,但 a 不能大于$x_{(1)} = \min\{x_1, x_2, \ldots, x_n\}$;

又因为$L(a, b)$是关于b的单调递减函数, b 越小, $L(a, b)$就越大, 但b 不能小于 $x_{(n)} = \max\{x_1, x_2, \ldots, x_n\}$.

当$a=x_{(1)}$, $b=x_{(n)}$时$L(a, b)$ 取得最大值 $\dfrac{1}{[x_{(n)} - x_{(1)}]^n}$.

a, b的最大似然估计值为

$$\hat{a} = x_{(1)} = \min\{x_1, x_2, \cdots, x_n\}, \hat{b} = x_{(n)} = \max\{x_1, x_2, \cdots, x_n\}.$$

a,b的最大似然估计量为

$$\hat{a} = X_{(1)} = \min\{X_1, X_2, \cdots, X_n\}, \hat{b} = X_{(n)} = \max\{X_1, X_2, \cdots, X_n\}.$$

例 设总体X的概率分布如下,其中$\theta(0<\theta<\dfrac{1}{2})$是未知参数,

X	0	1	2	3
p_k	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

利用总体X的如下样本值 3 , 1 , 3 , 0 , 3 , 1 , 2 , 3,
求θ的矩估计值和最大似然估计值.

解 1)由 $\mu_1=E(X)=0\times\theta^2+1\times2\theta(1-\theta)+2\times\theta^2+3\times(1-2\theta)=3-4\theta$,

令$\mu_1=A_1$,$3-4\theta=\bar{X}$, θ的矩估计量为$\hat{\theta}=\dfrac{3-\bar{X}}{4}$.

$\bar{x}=2$, θ的矩估计值为$\hat{\theta}=\dfrac{1}{4}$.

2)对于给定样本值,似然函数为

$$L(\theta)=\theta^2\left[2\theta(1-\theta)\right]^2\theta^2(1-2\theta)^4=4\theta^6(1-\theta)^2(1-2\theta)^4$$

$$L(\theta)=4\theta^6(1-\theta)^2(1-2\theta)^4$$

取对数$\ln L(\theta)=\ln 4+6\ln\theta+2\ln(1-\theta)+4\ln(1-2\theta)$

求导数, 令$\dfrac{d\ln L(\theta)}{d\theta}=\dfrac{6}{\theta}-\dfrac{2}{1-\theta}-\dfrac{8}{1-2\theta}=\dfrac{6-28\theta+24\theta^2}{\theta(1-\theta)(1-2\theta)}=0$

注意到已知$0<\theta<\dfrac{1}{2}$,

解得θ的最大似然估计值为$\hat{\theta}=\dfrac{7-\sqrt{13}}{12}\approx 0.2809.$

§2 估计量的评选标准

1. 无偏性
2. 有效性
3. 一致性

估计量是样本的函数,它是一个随机变量,由不同的方法得到的估计量可能相同也可能不同. 而对同一估计量,由不同的样本观察值得到参数的估计值也可能不同.

无偏性

定义 设 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 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(量).

注 定义的合理性

我们不可能要求每一次由样本得到的估计值与真值都相等,但可以从平均意义上要求这些估计值与真值相等.

例 设总体X的均值为μ,(X_1, X_2, X_3)是总体 X 的样本,证明下列两个估计量都是μ 的无偏估计,

$$\hat{\mu}_1 = X_2, \quad \hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{6}X_2 + \frac{1}{3}X_3.$$

证 由于 $E(\hat{\mu}_1) = E(X_2) = E(X) = \mu,$

$$E(\hat{\mu}_2) = \frac{1}{2}E(X_1) + \frac{1}{6}E(X_2) + \frac{1}{3}E(X_3) = \left(\frac{1}{2} + \frac{1}{6} + \frac{1}{3}\right)\mu = \mu.$$

所以$\hat{\mu}_1$与$\hat{\mu}_2$ 都是μ 的无偏估计.

注 只需 $k_1 + k_2 + \cdots + k_n = 1$,则 $\hat{\mu} = k_1 X_1 + k_2 X_2 + \cdots + k_n X_n$就是$\mu$的无偏估计.

例 设总体 X 的 k 阶矩 $\mu_k = E(X^k)$ 存在，$X_1, X_2, \cdots, X_n$ $(n > 1)$为来自总体 X 的样本,证明不论 X 服从什么分布, $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 μ_k 的无偏估计量.

证明 由于 $E(X_i^k) = \mu_k, i = 1, 2, \cdots, n$, 因而

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

注 样本均值 $\bar{X}$ 是总体期望 $E(X)$ 的无偏估计量

样本二阶原点矩 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ 是总体二阶原点矩 $\mu_2 = E(X^2)$ 的无偏估计量.

注 样本均值 $\overline{X}$ 是总体期望 $E(X)$ 的无偏估计量.

样本方差 $S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是总体方差 σ^2 的无偏估计量.

样本二阶中心矩 $B_2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 是总体方差 σ^2 的有偏估计量,

但为渐进无偏估计量.

若 $\hat{\theta}$ 是 $\boldsymbol{\theta}$ 的无偏估计量, $g(\hat{\theta})$ 未必是 $\boldsymbol{g(\theta)}$ 的无偏估计量.

如 $X\sim N(\mu,\sigma^2), E(\overline{X})=\mu,$

$$E(\overline{X}^2)=D(\overline{X})+(E(\overline{X}))^2=\frac{\sigma^2}{n}+\mu^2\neq\mu^2.$$

例　设总体X的均值为μ, 方差为σ^2, 分别独立地从总体X 中抽取样本$X_1,\cdots,X_m$及$X_1',\cdots,X_n'$样本均值分别为$\overline{X}$及$\overline{X}'$，令$\hat{\mu}=k\overline{X}+k'\overline{X}'$,

(1)当k和k'满足什么条件时，$\hat{\mu}$ 是μ 的无偏估计？

(2)当k和k'为何值时，(1)中μ 的无偏估计量$\hat{\mu}$ 的方差最小？

解　(1)当$k+k'=1$时，$\hat{\mu}$ 是μ 的无偏估计.

$$(2)D(\hat{\mu})=k^2D(\overline{X})+k'^2D(\overline{X}')=(\frac{k^2}{m}+\frac{k'^2}{n})\sigma^2=(\frac{k^2}{m}+\frac{(1-k)^2}{n})\sigma^2,$$

$$令0=\frac{\mathrm{d}D(\hat{\mu})}{\mathrm{d}k}=2\sigma^2(\frac{nk-m+mk}{mn}),$$

$$k=\frac{m}{m+n},\ k'=\frac{n}{m+n}.$$

有效性

定义 设 $\hat{\theta}_1=\theta_1(X_1,X_2,\cdots,X_n)$和 $\hat{\theta}_2=\theta_2(X_1,X_2,\cdots,X_n)$都是总体参数$\boldsymbol{\theta}$的**无偏估计量**, 且

$$D(\hat{\theta}_1)<D(\hat{\theta}_2)$$

则称 $\hat{\theta}_1$ 比$\hat{\theta}_2$更**有效**.

例 设总体X的均值和方差分别为μ和σ^2, X_1, X_2, X_3 是总体 X 的样本,比较下列μ的无偏估计哪个更有效.

$$\hat{\mu}_1 = X_2, \quad \hat{\mu}_2 = \frac{1}{2}X_1 + \frac{1}{6}X_2 + \frac{1}{3}X_3.$$

解 设$D(X) = \sigma^2$

$$D(\hat{\mu}_1) = D(X_2) = D(X) = \sigma^2,$$

$$D(\hat{\mu}_2) = \frac{1}{2^2}D(X_1) + \frac{1}{6^2}D(X_2) + \frac{1}{3^2}D(X_3) = \left(\frac{1}{4} + \frac{1}{36} + \frac{1}{9}\right)\sigma^2 = \frac{14}{36}\sigma^2.$$

故 $\hat{\mu}_2$比 $\hat{\mu}_1$有 效

例 设总体X的均值和方差分别为μ和σ^2, $X_1, X_2,\ldots, X_n$ 是总体 X 的样本,比较下列μ的无偏估计哪个更有效.

$$\hat{\mu}_1 = \bar{X}, \quad \hat{\mu}_2 = \sum_{i=1}^{n} k_i X_i (\sum_{i=1}^{n} k_i = 1).$$

证 由于 $\left(\sum_{i=1}^{n} k_i\right)^2 \le n \sum_{i=1}^{n} k_i^2,$

$$D(\hat{\mu}_1) = D(\bar{X}) = \frac{\sigma^2}{n} = \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} k_i\right)^2 = \frac{\sigma^2}{n}\left(\sum_{i=1}^{n} k_i\right)^2$$

$$< \sigma^2\left(\sum_{i=1}^{n} k_i^2\right) = D(\sum_{i=1}^{n} k_i X_i) = D(\hat{\mu}_2)$$

可见$\bar{X}$较一切$\sum_{i=1}^{n} k_i X_i$有效.

一致性

定义 设 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)(n = 1, 2, \cdots)$ 是总体参数θ的估计序列.

若对于任意的$\theta \in \Theta$,当$n \to \infty$时, $\hat{\theta}_n$依概率收敛于θ,即对于任意给定的正数ε,

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

则称$\hat{\theta}_n$为θ的**一致估计量(相合估计量)**.

注 一致性估计量仅在样本容量 n 足够大时,才显示其优越性.

注 **根据大数定律,**

样本均值 $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ **是总体均值** μ **的一致估计量.**

样本k阶原点矩$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$是总体同阶原点矩$\mu_k = E(X^k)$ $(k = 1, 2, \dots)$

的一致估计量.

例 设θ是总体X分布中的未知参数, $X_1, X_2, \cdots, X_n$是来自总体的样本, $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 是θ 的无偏估计量,且 $\lim\limits_{n\to\infty} D(\hat{\theta}_n) = 0$, 证明$\hat{\theta}_n$为$\theta$ 的一致估计量.

证明 因为$E(\hat{\theta}_n) = \theta$, 根据切比雪夫不等式,对于任意的$\varepsilon > 0$, 有

$$1 - \frac{D(\hat{\theta}_n)}{\varepsilon^2} \le P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} \le 1$$

又因为 $\lim\limits_{n\to\infty} D(\hat{\theta}_n) = 0$, 所以 $\lim\limits_{n\to\infty} P\left\{\left|\hat{\theta}_n - \theta\right| < \varepsilon\right\} = 1$,

故 $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ 为 θ 的一致估计量.