第七章　参数估计

§1　点估计

~~§2~~　基于截尾样本的最大似然估计

§3　估计量的评选标准

§4　区间估计

§5　正态总均值与方差的区间估计

§6　(0--1)分布参数的区间估计

§7　单侧置信区间

第七章　参数估计

从本章开始，讨论数理统计学的基本问题---统计推断。

- 统计推断：利用样本提供的信息对总体的某些统计特性进行估计或判断，从而认识总体。
- 统计推断分为两大类：

  (1) 参数估计(第七章)　(2) 假设检验(第八章)

§7.1　点估计

设总体X的分布函数的类型为已知，但是它的某些参数是未知的，**通过总体的一个样本来估计总体未知参数的值的问题称为参数的点估计问题.**

例 在某炸药制造厂，一天中发生着火现象的次数 X 是一个随机变量，假设它服从以 $\lambda > 0$ 为参数的泊松分布，参数 λ 为未知，现有以下的样本值，试估计参数 λ.

着火次数 k	0	1	2	3	4	5	6	≥ 7	
发生 k 次着火的天数 n_k	75	90	54	22	6	2	1	0	$\Sigma = 250$

解 因为 $X \sim \pi(\lambda)$, 所以 $\lambda = E(X)$.

用样本均值来估计总体的均值 $E(X)$.

$$\bar{x} = \frac{\sum_{k=0}^{6} k n_k}{\sum_{k=0}^{6} n_k} = \frac{1}{250}(0 \times 75 + 1 \times 90 + 2 \times 54 + 3 \times 22$$

$$+ 4 \times 6 + 5 \times 2 + 6 \times 1) = 1.22.$$

故 $E(X) = \lambda$ 的估计为 1.22 .

➢点估计问题的一般提法：

设总体X的分布函数为$F(x,\theta)$，其中θ为待估计的参数．X_1, $X_2,\ldots,X_n$是X的一个样本, $x_1, x_2, \ldots, x_n$是相应的样本值.

点估计：用样本$X_1, X_2, \ldots, X_n$构造一个**适当**的统计量

$$\hat{\theta}(X_1, X_2, \ldots, X_n),$$

用它的观察值 $\hat{\theta}(x_1, x_2, \ldots, x_n)$

作为未知参数θ的近似值．称$\hat{\theta}(X_1, X_2, \ldots, X_n)$为$\theta$ 的<u>估计量</u>.

称 $\hat{\theta}(x_1, x_2, \ldots, x_n)$ 为的θ <u>估计值</u>.

估计量和估计值统称估为<u>估计</u>，并都简记为 $\hat{\theta}$.

[注]参数θ的估计量 $\hat{\theta}$是样本$X_1, X_2,\ldots,X_n$ 的函数.

➢点估计常用方法：矩估计法；极大似然估计法.

一、矩估计法

k阶样本矩 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k$

矩估计法的基本思想是用样本矩估计总体矩. 因为由大数定律知，样本的k阶矩依概率收敛于总体的k阶矩. 这种用**样本(原点)矩作为总体(原点)矩的估计量的方法**称为**矩估计法**.

设总体X的分布函数为$F(x;\theta_1,\theta_2,\ldots,\theta_k)$，其中$\theta_1,\theta_2,\ldots,\theta_k$为待估参数，如果 $\mu_i=E(X^i)$ $(i=1,2,\ldots,k)$存在， μ_i为$\theta_1,\theta_2,\ldots,\theta_k$的函数，记$\mu_i=\mu_i(\theta_1,\theta_2,\ldots,\theta_k)$ $(i=1,2,\ldots,k)$，$X_1,X_2,\ldots,X_n$为总体X的样本，**用A_i来估计$E(X^i)$，建立 k 个方程**:

$$\begin{cases} A_1=\mu_1(\theta_1,\theta_2,\ldots,\theta_k)\\ A_2=\mu_2(\theta_1,\theta_2,\ldots,\theta_k)\\ \cdots\cdots\cdots\cdots\cdots\\ A_k=\mu_k(\theta_1,\theta_2,\ldots,\theta_k)\end{cases} \Longrightarrow \begin{cases} \hat{\theta}_1=\theta_1(A_1,A_2,\ldots,A_k)\\ \hat{\theta}_2=\theta_2(A_1,A_2,\ldots,A_k)\\ \cdots\cdots\cdots\cdots\cdots\\ \hat{\theta}_k=\theta_k(A_1,A_2,\ldots,A_k)\end{cases}$$

用$\hat{\theta}_i$作为θ_i的估计量------<u>矩估计量</u>.

◆求矩估计的方法

k价样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

设总体X的分布函数为$F(x;\theta_1,\theta_2,\ldots,\theta_k)$，其中$\theta_1,\theta_2,\ldots,\theta_k$为待估参数，

(1) 求总体X的前 k 阶矩

$$\mu_i = E(X^i) = \mu_i(\theta_1,\theta_2,\ldots,\theta_k),\quad i=1,2,\ldots,k$$

(2) 令

$$A_i = \mu_i(\theta_1,\theta_2,\ldots,\theta_k),\quad i=1,2,\ldots,k$$

(3) 解出

$$\hat{\theta}_i = \theta_i(A_1,A_2,\ldots,A_k),\quad i=1,2,\ldots,k$$

$\hat{\theta}_i$ 为θ_i 的 矩估计量.

例1 设总体X服从$[a，b]$上的均匀分布，$a，b$未知，$X_1, X_2, \ldots, X_n$为来自总体X的样本, 试求$a，b$的矩估计量.

解
$$\begin{cases} \mu_1 = E(X) = \dfrac{a+b}{2}, \\ \mu_2 = E(X^2) = D(X) + [E(X)]^2 = \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4} \end{cases}$$

由矩估计法, 令

$$\begin{cases} \dfrac{a+b}{2} = A_1 \\ \dfrac{(b-a)^2}{12} + \dfrac{(a+b)^2}{4} = A_2 \end{cases}$$

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

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

例2 设总体X服从参数为 θ 的指数分布, $X_1, X_2, \ldots, X_n$ 为来自总体X的样本，试用矩估计法求θ 的估计值.

解 因为 $\mu_1 = E(X) = \theta,$

由矩估计法, 令 $\mu_1 = A_1 = \overline{X}$

所以θ 矩估计量为 $\hat{\theta} = \overline{X}$

故θ 的矩估计值为 $\hat{\theta} = \overline{x}$

例3 设总体X的均值$E(X)=\mu$，方差$D(X)=\sigma^2$ 都存在，且$\sigma^2>0$. 但 μ, σ^2 均为未知. $X_1, X_2, \ldots, X_n$为来自总体X的样本，求μ, σ^2 的矩估计量.

解

$$\begin{cases} \mu_1 = E(X) = \mu \\ \mu_2 = E(X^2) = D(X) + [E(X)]^2 = \sigma^2 + \mu^2 \end{cases}$$

由矩估计法，令 $\begin{cases} \mu = A_1, \\ \sigma^2 + \mu^2 = A_2 \end{cases}$

$$\Rightarrow \begin{cases} \hat{\mu} = A_1 = \overline{X} \\ \hat{\sigma}^2 = A_2 - A_1^2 \end{cases} \Rightarrow \boxed{\begin{cases} \hat{\mu} = \overline{X}, \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \end{cases}}$$

上述结果表明，总体均值与方差的矩估计量的表达式不因不同的总体分布而不同.

➢常见分布的参数矩估计量

(1)若总体X~$b(1, p)$, 则未知参数 p 的矩估计量为

$$\hat{p} = \overline{X}$$

(2)若总体X~$b(N, p)$, 则未知参数p , N的矩估计量为

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

(3)若总体X~ N(μ,σ²), 则未知参数μ,σ² 的矩估计量为

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

(4)若总体X~ π(λ), 则未知参数λ 的矩估计量为

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

点估计的矩法是由皮尔逊提出的，它直观、简便，特别对总体期望和方差进行估计时不需要知道总体的分布。但它要求总体原点矩存在，而有些随机变量的原点矩不存在，就不能用此法进行参数估计。此外，矩估计有时不唯一；再者它没有利用总体分布函数所提供的信息，因此很难保证它有优良的性质。

二、最大似然估计法

最大似然估计法是求点估计的另一种方法。它最早是由德国数学家高斯于1821年所提出，后来为英国统计学家费希尔在1912年重新提出并做了进一步的研究。这是目前仍然得到最广泛应用的一种方法。它是建立在极大似然原理的基础上的一个统计方法。

最大似然法原理的直观想法:“概率最大的事件最可能出现”.例如有一个事件,若知道它出现的概率只能是0.01或0.99,而在一次观测中,此事件出现,此时自然会说它的概率应为0.99.因此,参数估计的极大似然法是要**选取这样的值来作为参数的估计值,使得当参数取这一数值时,观测结果出现的可能性为最大.**

例4 设在罐中放有许多白球和黑球, 已知两种球的数目之比为1:3, 但不知哪种颜色的球多, 若采用有放回方式从罐中取3个球, 发现有一只黑球, 问在此情况下应估计哪种颜色的球多?

解: 设 p=黑球所占比例= $\dfrac{\text{罐中黑球数}}{\text{罐中全部球的数目}}$

则 p=1/4或 p=3/4

又设X= "取出的3只球中黑球的数目", 则 $X \sim b(3, p)$

$$P\{X=1, p=1/4\}=C_3^1 \frac{1}{4}(\frac{3}{4})^2=\frac{27}{64},$$

$$P\{X=1, p=3/4\}=C_3^1 \frac{3}{4}(\frac{1}{4})^2=\frac{9}{64}$$

认为 p=1/4.

□似然函数

(1) **离散型总体** 设总体X分布律 $P\{X=x\}=p(x;\theta)$, $\theta\in\Theta$为待估参数, Θ是θ可能取值的范围. 设$X_1, X_2, \ldots, X_n$是来自X的样本，样本观察值为$x_1, x_2, \ldots, x_n$，则 $(X_1, X_2, \ldots, X_n)$的联合分布律为

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

对固定的样本观察值$x_1, x_2, \ldots, x_n$，它是未知参数的函数. 记为

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

称其为<u>**样本的似然函数**</u>.

(2) **连续型总体** 设总体X的概率密度为$f(x;\theta)$, $\theta\in\Theta$为未知参数，

<u>此时定义**样本的似然函数为:**</u>

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

定义 若存在 $\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots, x_n)$，使得

$$L(\hat{\theta}) = \max_{\theta \in \Theta} L(\theta)$$

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

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

如何求 $L(\theta)$ 的最大值?

由于 $L(\theta)$ 与 $\ln L(\theta)$ 在 Θ 上有相同的最大值点, 所以求 $L(\theta)$ 的最大值点可以改为求 $\ln L(\theta)$ 的最大值点.

当 $\ln L(\theta)$ 关于 θ 可微时, 必满足方程:

$$\frac{\partial \ln L(\theta_i)}{\partial \theta_i} = 0, (i = 1,2,\ldots, k)$$ ----对数似然方程(组)

当 $\ln L(\theta)$ 关于 θ 不可微时，回到原式定义。

例5 设$X\sim b(1,p)$，$X_1, X_2, \ldots, X_n$是来自X的一个样本，求参数 p 的最大似然估计量.

解 设$x_1, x_2, \ldots, x_n$是相应于$X_1, X_2, \ldots, X_n$一个样本值，X的分布律为 $P\{X=x\}=p^x(1-p)^{1-x}$， $x=0,1$，

故似然函数为

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

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

令 $\frac{d}{dp}\ln L(p)=0 \Rightarrow \hat{p}=\frac{1}{n}\sum_{i=1}^{n} x_i=\bar{x}$， $\hat{p}=\frac{1}{n}\sum_{i=1}^{n} X_i=\bar{X}$

似然估计值　　似然估计量

这一估计量与矩估计量是相同的.

例6 设总体$X\sim N(\mu,\sigma^2)$, μ,σ^2均未知，又设$X_1, X_2,\ldots,X_n$为总体X的样本，$x_1, x_2,\ldots, x_n$为X的一组样本观测值，试求μ,σ^2的最大似然估计值量.

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

似然函数为 $L(\mu,\sigma^2)==\dfrac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left\{-\dfrac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$

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

似然方程
$$\begin{cases}\dfrac{\partial}{\partial\mu}\ln L=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L=-\dfrac{n}{2\sigma^2}+\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2=0\end{cases}\Rightarrow\begin{cases}\hat{\mu}=\overline{X},\\[2ex]\hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2\end{cases}$$

它们与相应的矩估计量相同.

例7 设总体$X\sim U(a, b)$，a, b 均未知，又设$X_1, X_2,\ldots,X_n$ 为总体X 的样本，$x_1, x_2 ,\ldots, x_n$为X的一组样本观测值，试求a, b 的极大似然估计值量.（用定义）

$$\hat{a} = \min_{1\le i\le n} X_i , \hat{b} = \max_{1\le i\le n} X_i$$

例8 设总体 X 服从参数为 θ 指数分布，求θ 的极大似然估计值量.

$$\hat{\theta} = \overline{X}$$

例9 已知一批灯泡的使用寿命T服从参数为θ 的指数分布, 现随机抽取18只, 测得使用寿命(小时)如下:

16, 29, 50, 68, 100, 130, 140, 270, 280, 340, 410, 450, 520,620, 190, 210,800, 1100

求参数θ 的极大似然估计值与 $P\{T \geq 1000\}$.

解: 因为T 服从指数分布，故参数θ的极大似然估计为

$$\hat{\theta} = \overline{X} \quad \text{计算得} \ \overline{x} \approx 318 \quad \text{所以} \ \hat{\theta} \approx 318.$$

$$P\{T \geq 1000\} = \int_{1000}^{+\infty} 0.003 e^{-0.003t} dt = \frac{1}{e^3}$$

最大似然估计的性质

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 .

证明 因为$\hat{\theta}$是θ 的最大似然估计值，所以

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Theta} L(x_1, x_2, \cdots, x_n; \theta),$$

其中$x_1, x_2, \cdots, x_n$是来自总体 X的一个样本值，

由于

$$\hat{u} = u(\hat{\theta}), \qquad \hat{\theta} = \theta(\hat{u}),$$

故

$$L(x_1, x_2, \cdots, x_n; \theta(\hat{u})) = \max_{u \in U} L(x_1, x_2, \cdots, x_n; \theta(u)),$$

于是 $\hat{u} = u(\hat{\theta})$是$u(\theta)$的最大似然估计.

此性质可以推广到总体分布中含有多个未知参数的情况.

在例6中，σ^2 的最大似然估计值为

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

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 的最大似然估计值为

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

极大似然法克服了矩法的一些缺点，它利用总体的样本和分布函数表达形式所提供的信息建立未知参数的估计量，同时它也不要求总体原点矩存在，因此极大似然估计量有比较良好的性质。但求极大似然估计量一般要解似然方程，而有时解似然方程很困难，只能用数值方法求似然方程的近似解。

练习：第7章习题2、3

小结

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

在统计问题中往往先使用最大似然估计法,

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

§7.3 估计量的评选标准

问题的提出

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.

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?

本节介绍几个常用标准.

1. 无偏性　　2. 有效性　　3. 相合性

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.

定义1 若估计量 $\hat{\theta}=\theta(X_1,X_2,\cdots,X_n)$ 的数学期望 $E(\hat{\theta})$ 存在, 且对于任意 $\theta\in\Theta$ 有 $E(\hat{\theta})=\theta$, 则称 $\hat{\theta}$ 是 θ 的无偏估计量．这里 Θ 是 θ 的取值范围 .

在科学技术中 $E(\hat{\theta})-\theta$ 称为以 $\hat{\theta}$ 作为 θ 的估计的系统误差. 无偏估计的实际意义就是无系统误差.

例1 设总体X的数学期望 μ 与方差σ^2存在, $X_1, X_2,\ldots,X_n$ 为总体X 的样本，证明：

(1) $\hat{\mu}_1 = \overline{X} = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i$ 是 μ 的无偏估计量;

(2) $\hat{\mu}_2 = \sum\limits_{i=1}^{n} c_i X_i$, 其中 $\sum\limits_{i=1}^{n} c_i = 1, c_i > 0, i = 1,2,\cdots,n$

也是 μ 的无偏估计量;

(3) $\hat{\sigma}_1^2 = S^2 = \dfrac{1}{n-1}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ 是 σ^2 的无偏估计量;

(4) $\hat{\sigma}_2^2 = B_2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2$ 不是 σ^2 的无偏估计量.

证明(4)：

$$\hat{\sigma}_2^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = A_2 - \overline{X}^2,$$

因为 $E(A_2) = \mu_2 = \sigma^2 + \mu^2,$

又因为 $E(\overline{X}^2) = D(\overline{X}) + [E(\overline{X})]^2 = \dfrac{\sigma^2}{n} + \mu^2,$

所以 $E(\hat{\sigma}_2^2) = E(A_2 - \overline{X}^2) = E(A_2) - E(\overline{X}^2)$

$= \dfrac{n-1}{n}\sigma^2 \neq \sigma^2,$ 所以 $\hat{\sigma}_2^2$ 是有偏的.

证明(3)：

若以 $\dfrac{n}{n-1}$ 乘 $\hat{\sigma}_2^2$, 所得到的估计量就是无 偏的.

(这种方法称为无偏化).

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

因为 $\dfrac{n}{n-1}\hat{\sigma}_2^2=S^2=\dfrac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X}^2)$,

即 S^2是σ^2 的无偏估计，故通常取S^2作σ^2的估计量.

例 设总体 X 的k 阶矩 $\mu_k = E(X^k)\ (k \geq 1)$存在，又设 $X_1, X_2, \cdots, X_n$ 是 X 的一个样本，试证明不论总体服从什么分布，k阶样本矩 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$ 是 k阶总体矩 μ_k 的无偏估计.

证 因为$X_1, X_2, \cdots, X_n$与X同分布，

故有 $$E(X_i^k) = E(X^k) = \mu_k, \quad i = 1, 2, \cdots, n.$$

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

故 k 阶样本矩 A_k 是k阶总体矩 μ_k 的无偏估计.

例2 设总体 X 服从指数分布，其概率密度

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

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 $\overline{X}$ 和 $nZ=n[\min(X_1, X_2, \cdots, X_n)]$ 都是 θ 的无偏估计.

证 因为 $E(\overline{X})=E(X)=\theta$,

所以 $\overline{X}$ 是 θ 的无偏估计量.

而$Z=\min(X_1,\dots,X_n)$具有概率密度

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

$F_{\min}(x)=1-[1-F_X(x)]^n$

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

所以 nZ 也是 θ 的无偏估计量 .

由以上两例可知, 一个参数可以有不同的无偏估计量.

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, $\hat{\theta}_1$ 的观测值较 $\hat{\theta}_2$ 更密集, 则认为 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ 为理想 .

二. 有效性

用 $\hat{\theta}$ 估计 θ 时, 仅具有无偏性是不够的. 我们希望 $\hat{\theta}$ 的取值能集中于 θ 附近, 而且密集的程度越高越好. 方差是描述随机变量取值的集中程度的，所以无偏估计以方差小者为好，因此提出所谓“**有效性**”标准.

定义2 设 $\hat{\theta}_1$ 与 $\hat{\theta}_2$ 都是参数 θ 的无偏估计, 若 $D(\hat{\theta}_1) \le D(\hat{\theta}_2)$ 则称 $\hat{\theta}_1$ 较 $\hat{\theta}_2$ **有效**. （存在 θ 使不等号成立）

有效性的意义是：用 $\hat{\theta}$ 估计 θ 时, 除无系统偏差外，还有估计精度高的意义.

例3 设总体 X 的数学期望 μ, 方差 σ^2 存在，X_1, X_2 是 X 的样本，证明估计 μ 时,

$$\hat{\mu}_1 = \frac{1}{2}X_1 + \frac{1}{2}X_2 \quad \text{较} \quad \hat{\mu}_2 = \frac{1}{4}X_1 + \frac{3}{4}X_2 \text{ 有效.}$$

证明 因为 $\hat{\mu}_1, \hat{\mu}_2$ 均为 μ 的无偏估计，又因为

$$D(\hat{\mu}_1) = D(\bar{X}) = D(\frac{1}{2}X_1 + \frac{1}{2}X_2) = \frac{1}{4}D(X_1) + \frac{1}{4}D(X_2) = \frac{1}{2}\sigma^2$$

$$D(\hat{\mu}_2) = D(\frac{1}{4}X_1 + \frac{3}{4}X_2) = \frac{1}{16}D(X_1) + \frac{9}{16}D(X_2) = \frac{5}{8}\sigma^2$$

所以 $D(\hat{\mu}_1) \le D(\hat{\mu}_2)$

由定义知 $\hat{\mu}_1$ 较 $\hat{\mu}_2$ 有效.

一般地 设总体X的数学期望μ,方差 σ^2存在,$X_1, X_2,..,X_n$ 是X的样本，证明估计 μ 时,

$$\hat{\mu}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \text{ 较 } \hat{\mu}_2 = \sum_{i=1}^{n} c_i X_i \text{ 有效.}$$

其中 $\sum_{i=1}^{n} c_i = 1, c_i > 0, i = 1,2,\cdots,n$.

例2补充： 试证当$n>1$时，θ的无偏估计量

$\overline{X}$ 较θ的无偏估计量 nZ 有效.

证 由于 $D(X)=\theta^2$, 故有 $D(\overline{X})=\dfrac{\theta^2}{n}$,

又因为 $D(Z)=\dfrac{\theta^2}{n^2}$，故有 $D(nZ)=\theta^2$,

当$n>1$时，$D(nZ)>D(\overline{X})$,

故θ的无偏估计量 $\overline{X}$ 较 nZ 有效.

三. 相合性(一致性)

无偏性和有效性是在样本容量n一定的情况下对估计量提出的要求，一个好的估计量 $\hat{\theta}$, 当样本容量增大时，$\hat{\theta}$ 的取值与参数 θ 的真值任意接近的可能性应该更大. 因此，还有所谓“一致性”标准.

定义3 设 $\hat{\theta}_n$是未知参数θ 的估计量，如果对于任意 ε>0, 有

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

则称 $\hat{\theta}_n$为θ 的**相合估计量**. 即 $\hat{\theta}_n \xrightarrow{P} \theta$

结论:**若$\hat{\theta}_n$是θ 的无偏估计量，且当n→∞时,** $D(\hat{\theta}_n)\to 0$

则 $\hat{\theta}_n$是θ 的相合估计量。

（可利用切比雪夫不等式证明）

例4 设总体X的数学期望 μ 与方差 σ^2存在，$X_1, X_2,\ldots, X_n$是X的样本，证明 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ 是μ相合估计量.

证明 由大数定理可知，对于任意的 $\varepsilon > 0$ ，有

$$\lim_{n\to\infty} P\{|\bar{X} - E(X_i)| < \varepsilon\} = 1$$

一般地，由辛钦定理，若总体 X 的k阶矩$E\left(X^k\right) = \mu_k$ 存在，则有 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} E(X^k) = \mu_k \ (k = 1, 2, \cdots)$

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

矩法得到的估计量一般为相合估计量.

例5 试证：样本方差 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$ 及样本的二阶中心矩 $B_2=\frac{1}{n}\sum_{i=1}^{n}\left(X_i-\bar{X}\right)^2$都是总体方差 σ^2 的相合估计量.

证

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

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

(A_2 是样本二阶原点矩)

由大数定律知,

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

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

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

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

小结

估计量的评选的三个标准

- 无偏性
- 有效性
- 相合性

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.

由极大似然法得到的估计量，在一定条件下也具有一致性，这里就不再讨论了．（可利用大数定律和Jensen不等式证明）

作业

- **第七章习题1, 4(1), 5, 7, 10, 14,**