第七章
参数估计

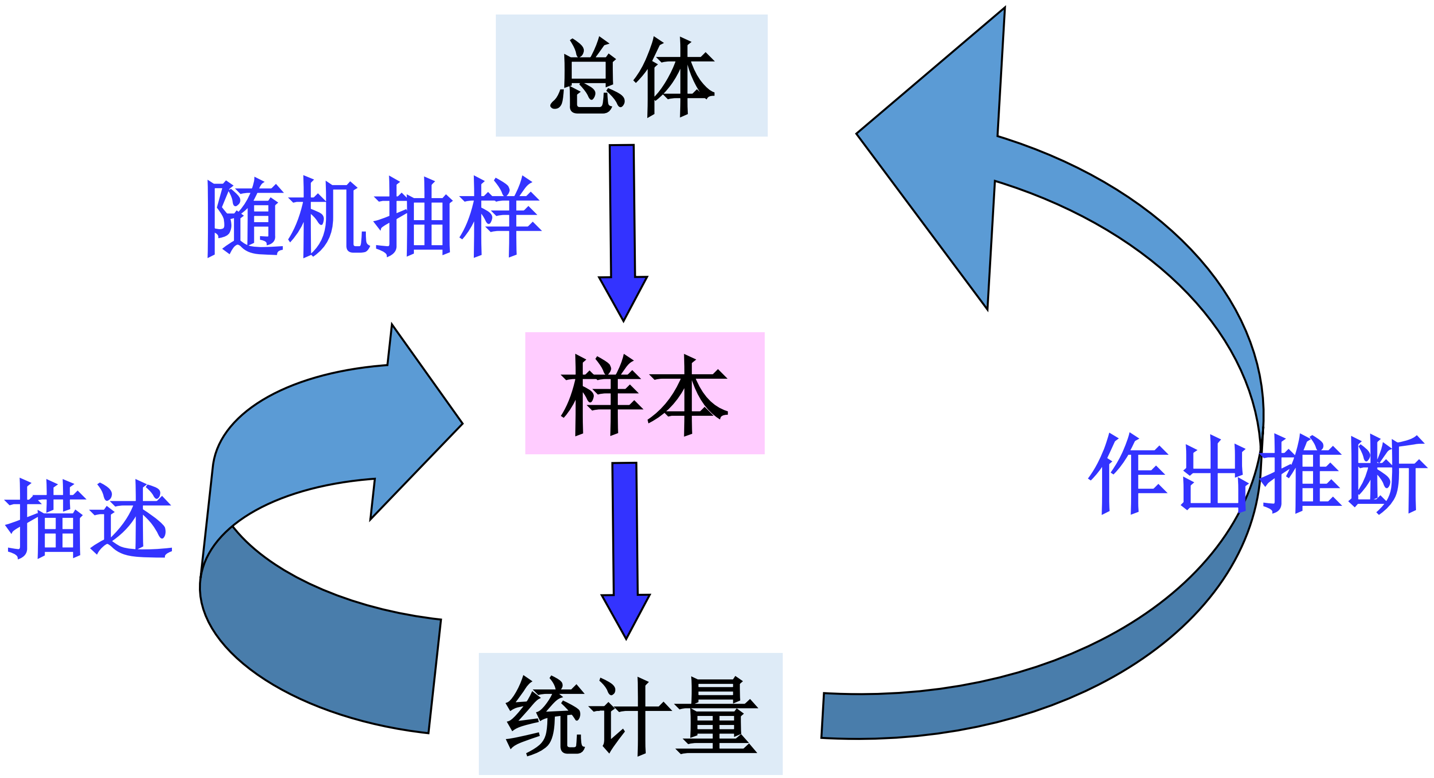
王笑尘

北京邮电大学网络空间安全学院

wxiaochen@bupt.edu.cn

第七章 参数估计

- 数理统计方法具有“部分推断整体”的特征



研究统计量的性质和评价一个统计推断的优良性，
完全取决于其抽样分布的性质。

- **参数估计问题:** 利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数。

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数。

■ 参数估计问题的一般提法

设有一个总体 X， 总体的分布函数为 $F(x;\theta)$ 其中 θ 为未知参数 (θ 可以是向量) 。现从该总体抽样，得样本： $X_1, X_2, \ldots, X_n$

所研究的问题是：要依据该样本对参数 θ 作出估计，或估计 θ 的某个已知的函数 $g(\theta)$

这类问题称为参数估计。

■ 参数估计问题的分类

参数估计 $\begin{cases} \text{点估计} \\ \text{区间估计} \end{cases}$

例如

- 现要估计某班男生的平均身高。假定身高服从正态分布 $N(\mu, 0.1^2)$
- 现从该总体选取容量为 5 的样本，所研究的问题是要根据选出的样本（5个数）求出总体均值 μ 的估计。
- 而全部信息就由这 5 个数组成 。设这 5 个数是：1.65，1.67，1.68，1.78，1.69
- 则估计 μ 为1.68，这是点估计问题（估计未知参数的值）。
- 估计 μ 在区间 $[1.57, 1.84]$ 内，这是区间估计问题。

第一节 点估计

解决问题：

总体 X 的分布函数的形式已知，但它的一个或多个参数未知，根据总体 X 的一个样本来估计总体未知参数或对总体未知参数作出一个估计。

一. 估计量的定义

定义：设 θ 为总体X的分布函数 $F(x;\theta)$ 中的待估计的参数， $X_1, X_2, \cdots X_n$ 是总体 X 的一个样本，用 $X_1, X_2, \cdots X_n$ 构成的一个统计量：

$\boxed{\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots X_n)\text{ 称为 }\theta\text{ 的估计量}}$。

如果 $X_1, X_2, \cdots X_n$ 的一组样本值为：$x_1, x_2 \cdots x_n$

$\boxed{\text{则 }\hat{\theta} = \hat{\theta}(x_1, x_2, \cdots x_n)\text{ 为 }\theta\text{ 的估计值}}$。

二. 构造统计量的方法

1. 矩估计法 (数字特征法)

- 矩估计法是由统计学家卡. 皮尔逊（**K. Pearson**）在**19**世纪末引入的。

- 矩是描写随机变量最简单的数字特征，由大数定律可知，在一定条件下可以用样本的矩作为总体矩的估计，从而得**矩估计法的基本思想**为：

用**样本的各阶矩**来估计**总体的各阶矩**

▲ 若总体 **X** 的 k 阶原点矩 $E(X^k)=\mu_k$ 存在，

则当 $n\to\infty$ 时有：$A_k \xrightarrow{p} \mu_k$，　$k=1,2,\cdots$

矩估计法的具体步骤

设总体 X 的分布函数 $F(x;\theta_1,\theta_2,\cdots,\theta_k)$ 中含有 k 个未知参数 $\theta_1,\theta_2,\cdots\theta_k$，假定总体 X 的前 k 阶矩存在，则可通过下列步骤求未知参数的矩估计量：

(1)

若总体 X 是离散型随机变量，其分布律为：

$P(x;\theta_1,\theta_2,\cdots,\theta_k)$ 则：求总体 X 的前 k 阶矩

$$\boxed{E(X^i)=\sum_{l=1}^{n} x_l^i\cdot P(x,\theta_1,\theta_2,\cdots\theta_k)\quad i=1,2,\cdots k}$$

若总体 X 是连续型随机变量，其密度函数为：

$f(x;\theta_1,\theta_2,\cdots,\theta_k)$ 则：

$$\boxed{E(X^i)=\int_{-\infty}^{+\infty} x^i\cdot f(x;\theta_1,\theta_2,\cdots,\theta_k)\,dx}\quad i=1,2,\cdots,k$$

总之，$\mu_i=E(X^i)$ 是参数 $\theta_1,\theta_2,\cdots,\theta_k$ 的函数，

记为：

$$\begin{cases}\mu_1=\mu_1(\theta_1,\theta_2,\cdots,\theta_k)\\ \mu_2=\mu_2(\theta_1,\theta_2,\cdots,\theta_k)\\ \cdots\cdots\cdots\cdots\cdots\cdots\\ \mu_k=\mu_k(\theta_1,\theta_2,\cdots,\theta_k)\end{cases}\qquad (*)$$

(2) 解（*）式解出 $\theta_1, \theta_2, \cdots, \theta_k$ 得到：

$$\begin{cases} \theta_1 = \theta_1(\mu_1, \mu_2, \cdots, \mu_k) \\ \theta_2 = \theta_2(\mu_1, \mu_2, \cdots, \mu_k) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \theta_k = \theta_2(\mu_1, \mu_2, \cdots, \mu_k) \end{cases} \quad (**)$$

(3) 用 μ_i 的估计量$A_i = \frac{1}{n}\sum_{j=1}^{n} X_j^i$ 分别代替（**）中的 μ_i，则得 θ_i 的矩估计量 $\hat{\theta}_i$：

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

例 1. 设总体X的均值为 μ, 方差为 σ^2都存在，且 $\sigma^2 > 0$, $X_1, X_2, \cdots X_n$ 是总体 X 的一个样本

(1). μ，σ^2 均未知, 求: μ, σ^2 的矩估计量

(2). 当总体(某种灯泡寿命) $X \sim N(\mu, \sigma^2)$ ，μ, σ^2 未知，今取 4 只灯泡， 测得其寿命（小时）如下: 1502, 1453, 1367, 1650（小时）

求：μ, σ^2的矩估计值

解：(1). $\because$ 总体 X 的数学期望是 X 的一阶原点矩；

总体 X 的方差是 X 的二阶中心矩。

$$\therefore \begin{cases} \mu = E(X) \\ \sigma^2 = E(X^2) - [E(X)]^2 \end{cases}$$

现令：
$$\begin{cases} E(X) = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i \\ E(X^2) = \dfrac{1}{n}\sum\limits_{i=1}^{n} X_i^2 \end{cases}$$

估计：样本一阶原点矩代替总体一阶原点矩

估计：样本二阶原点矩

解之得: $\begin{cases} \hat{\mu} = \boxed{\dfrac{1}{n}\sum_{i=1}^{n} X_i} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 - (\hat{\mu})^2 \end{cases}$ 用样本进行估计

从而得 μ, σ^2 的矩估计量为:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

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

结论：不论总体服从什么分布，总体均值与方差的矩估计量的表达式是不变的。

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

$$\overline{X} = \frac{1}{4}(1502 + 1453 + 1367 + 1650) = 1493$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{4}[(1502-1493)^2 + (1453-1493)^2$$

$$+ (1367-1493)^2 + (1650-1493)^2]$$

$$= 10551$$

$\therefore$ 某种灯泡寿命的均值与方差的矩估计值分布为:

$$\hat{\mu} = 1493, \quad \hat{\sigma}^2 = 10551$$

例 2. 设 $X_1, X_2, \ldots X_n$ 是取自总体 X 的一个样本，其概率密度为：

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

其中 θ, μ 为未知参数，$\theta > 0$

求：$\theta,\ \mu$ 的矩估计量

1、参数和总体矩的关系；
2、用样本矩代替总体矩进行估计。

例 2. 设 $X_1, X_2, \ldots X_n$ 是取自总体 X 的一个样本，其概率密度为：

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

其中 θ,μ 为未知参数，$\theta>0$

求：θ, μ 的矩估计量

1、参数和总体矩的关系；
2、用样本矩代替总体矩进行估计。

解: 由密度函数可知：

$X-\mu$ 为均值为 θ 的指数分布，故有：

$$E(X-\mu)=\theta, \quad D(X-\mu)=\theta^2$$

即： $E(X)=\mu+\theta,\quad D(X)=\theta^2$

1、参数和总体矩的关系；

令：$\begin{cases} \mu+\theta=\bar{X} \\ \theta^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2 \end{cases}$

2、用样本矩代替总体矩进行估计。

解得：$\begin{cases} \hat{\mu}=\bar{X}-\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2} \\ \hat{\theta}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\bar{X})^2} \end{cases}$

用样本矩估计总体矩

$\hat{\mu},\hat{\theta}$ 即为总体参数 μ,θ 的矩估计量。

2. 极大似然法

- 极大似然法是在总体类型已知条件下使用的一种参数估计方法。

- 它首先是由德国数学家高斯（Gauss）在 1821 年提出的。

- 然而，这个方法常归功于英国统计学家费歇（Fisher），费歇在 1922 年重新发现了这一方法，并首先研究了这种方法的一些性质。

Gauss

Fisher

极大似然法的基本思想

引例　设$X \sim B(1, p)$, p未知，若事先知道 p 只有两种可能：$p=0.7$　或　$p=0.3$

如今重复试验 3 次,得结果：0，0,　0

试问：应如何估计 p ?

分析：由概率论的知识，可知：

3 次试验中出现 “1” 的次数 $Y \sim B(3, p)$且：

$$P(Y=k)=\binom{3}{k}p^k(1-p)^{n-k} \qquad k=0, 1, 2, 3$$

现将这计算结果列出如下：

将计算结果列表如下：

估计

估计

p值	$P(Y=0)$	$P(Y=1)$	$P(Y=2)$	$P(Y=3)$
0.7	0.027	0.189	0.441	0.343
0.3	0.343	0.441	0.189	0.027

注：引例体现了极大似然法的基本思想：
当试验中得到一个结果时，应选择使得这个试验结果出现的概率达到最大的这个值作为参数的估计值。

(1). 极大似然估计量的定义

定义： 设总体X的概率密度函数为 $f(x,\theta_1,\theta_2,\cdots\theta_l)$ 或分布律为$P(x,\theta_1,\theta_2,\cdots\theta_l)$，其中 $\theta_1,\theta_2,\cdots\theta_l$ 为未知参数。又设 $x_1,x_2,\cdots x_n$ 是相应于样本 $X_1,X_2,\cdots X_n$ 的一组样本值。

作**似然函数：**

$$L=\prod_{k=1}^{n} f(\boxed{x_k},\theta_1,\theta_2,\cdots\theta_l)$$

或

$$=\prod_{k=1}^{n} P(\boxed{x_k},\theta_1,\theta_2,\cdots\theta_l)$$

使得似然函数 L 达到极大值的 $\bar{\theta}_1, \bar{\theta}_2, \cdots \bar{\theta}_l$

称为参数 $\theta_1, \theta_2, \cdots \theta_l$ 的极大似然估计值，记为：

$\hat{\theta}_i(x_1, x_2, \cdots x_n)$ (它与样本值有关)，记统计量：

$\hat{\theta}_i(X_1, X_2, \cdots X_n)$ 为参数 θ_i 的极大似然估计量

似然函数代表了取观测值的概率：

注：▲ 若总体X属于连续型，随机点 $(X_1, X_2, \cdots X_n)$ 落在点 $(x_1, x_2, \cdots x_n)$ 的邻域 (边长分别为 $dx_1, dx_2, \cdots dx_n$ 的 n 维立方体) 内的概率为

$$\prod_{k=1}^{n} f(x_k, \theta_1, \theta_2, \cdots \theta_l) dx_k$$

由于因子 $\prod_{k=1}^{n} dx_k$ 不随 θ_i 而变，故似然函数表示为

$$L = \prod_{k=1}^{n} f(x_k, \theta_1, \theta_2, \cdots \theta_l)$$

注：▲ 若总体X属于离散型，随机点 $(X_1, X_2, \cdots X_n)$

取到 $x_1, x_2, \cdots x_n$ 的概率：

$$L = \prod_{k=1}^{n} P(x_k, \theta_1, \theta_2, \cdots \theta_l)$$

▲ 似然函数 L 是 θ_k 的函数。

(2). 极大似然法的具体步骤

思路: 现要求 $\prod_{k=1}^{n} f(x_k,\theta_1,\theta_2,\cdots\theta_l)$ 或 $\prod_{k=1}^{n} P(x_k,\theta_1,\theta_2,\cdots\theta_l)$

的最大值，即求 $\theta_1,\theta_2\cdots\theta_l$ 取什么值时函数 L

达到最大。即其随机点 $(X_1,X_2,\cdots X_n)$ 落在

$(x_1,x_2,\cdots x_n)$ 的邻域内的概率或 随机点

$(X_1,X_2,\cdots X_n)$ 取到 $(x_1,x_2,\cdots x_n)$ 的概率最大。

从而此问题就转化为一般的求函数的最大值问题。

具体步骤

(1) 作似然函数 $L(\theta)=\prod_{k=1}^{n} f(x_k,\theta)$ 或

$$L(\theta)=\prod_{k=1}^{n} P(x_k,\theta)$$

(2) 当似然函数可微且 $L(\theta)$ 的最大值能在参数空间取得时，求方程组: $\dfrac{\partial \ln L(\theta)}{\partial \theta}=0$ 的解，解得一解为 $\hat{\theta}$，则 $\hat{\theta}$ 为极大似然估计量（值）。

注: ▲ 因为 $\ln L(\theta)$ 与 $L(\theta)$ 有相同的最大值点，有时求 $\max \ln L(\theta)$ 比求 $\max L(\theta)$ 方便，所以常取前者作为似然函数。

▲ 按照求函数极值的方法，在求方程组：

$$\frac{\partial \ln L(\theta)}{\partial \theta}=0$$ 的解后还应该用极值的充分条件

对解做进一步的判断。但又由最值原理，如果最值存在，此方程组求得的驻点即为所求的最值点。极大似然估计法一般属于这种情况，所以可直接按步骤(2)求得其值。

▲ 当似然函数不可微或方程组无解时，则应根据定义直接寻求能使 $L(\theta)$ 达到最大值的解作为极大似然估计值。

▲ 极大似然估计法也适用于多个未知参数的情形。

例3. 设 $X \sim N(\mu,\sigma^2)$, $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$ 为未知参数，

$x_1, x_2 \cdots x_n$ 是 X 的一个样本值

求: μ, σ^2 的极大似然估计量

例3. 设 $X \sim N(\mu,\sigma^2)$, μ, σ^2 为未知参数,

$x_1, x_2 \cdots x_n$ 是 X 的一个样本值

求: μ, σ^2 的极大似然估计量

解: $\because$ X 的密度函数为: $f(x;\mu,\sigma^2) = \dfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

作似然函数:

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

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

为计算方便对 L 两边取对数得:

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

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

解得所求为:

$$\begin{cases} \hat{\mu} = \dfrac{1}{n}\sum_{i=1}^{n} x_i = \bar{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\hat{\mu})^2 = \dfrac{1}{n}\sum_{i=1}^{n}(x_i-\bar{X})^2 \end{cases}$$

与矩估计法所得的结论是一致的

例　设总体服从指数分布 $X \sim EXP(\theta)$,其密度为

$$f(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0,\\ \quad 0 \quad , & x\le 0,\end{cases} \qquad (\theta>0)$$

$X_1, X_2, \cdots, X_n$ 是来自总体的样本，

试求 θ 的最大似然估计（MLE）？

例 设总体服从指数分布 $X \sim EXP(\theta)$，其密度为

$$f(x;\theta) = \begin{cases} \frac{1}{\theta} e^{-\frac{x}{\theta}}, & x > 0, \\ 0, & x \le 0, \end{cases} \quad (\theta > 0)$$

$X_1, X_2, \cdots, X_n$ 是来自总体的样本，

试求 θ 的最大似然估计（MLE）？

似然函数为

$$L(\theta) = \prod_{i=1}^{n} \frac{1}{\theta} e^{-\frac{X_i}{\theta}} = \theta^{-n} e^{-\frac{1}{\theta}\sum_{i=1}^{n} X_i} = \theta^{-n} e^{-\frac{n\overline{X}}{\theta}}$$

因为 $L(\theta)$ 与 $\ln L(\theta)$ 有相同的极值点，故令

$$\frac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta} = -\frac{n}{\theta} + \frac{n\overline{X}}{\theta^2} = 0$$

称为对数似然方程

解似然方程，求得 θ 的 MLE 为 $\hat{\theta} = \overline{X}$.

(3). 性质

$u=u(\theta)$ 是 θ 的函数，$\theta\in H$

H 是 θ 的取值范围

且具有单值反函数 $\theta=\theta(u)$,又设 $\hat{\theta}$ 是 X 的概率密度函数 $f(x,\theta)$ 中参数 θ 的极大似然估计，

则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的极大似然估计。

证: $\because\hat{\theta}$ 是 θ 的极大似然估计

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

又 $\because\hat{u}=u(\hat{\theta})$ 且有 $\hat{\theta}=\theta(\hat{u})$

$\therefore$ 上式可写为:

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

即表明:

$\hat{u} = \hat{u}(\theta)$ 是 $u(\theta)$ 的极大似然估计

注:

此性质对总体 X 中含有多个未知参数时也成立。

例4. 设 $X_1, X_2 \cdots X_n$ 为参数都是未知的正态总体的一个样本。

求: $P(\overline{X} < t)$ 的极大似然估计

解:

例4. 设 $X_1, X_2 \cdots X_n$ 为参数都是未知的正态总体的一个样本。

求: $P(\overline{X} < t)$ 的极大似然估计

解: $\because \quad X_i \sim N(\mu, \sigma^2), \quad \mu, \sigma^2$ 未知

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

$$P(\overline{X} < t) = P\left(\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} < \frac{t - \mu}{\sigma / \sqrt{n}}\right) = \Phi\left(\frac{t - \mu}{\sigma / \sqrt{n}}\right)$$

由例 2可知：μ的极大似然估计为 $\bar{X}$

σ^2的极大似然估计为 $\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$$P(\bar{X}<t)=P\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}<\frac{t-\mu}{\sigma/\sqrt{n}}\right)=\Phi\left(\frac{t-\mu}{\sigma/\sqrt{n}}\right)$$

$\therefore\ P(\bar{X}<t)$ 的极大似然估计为：$\Phi\left(\frac{t-\bar{X}}{\hat{\sigma}}\sqrt{n}\right)$

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

例5. 设 $X_1, X_2, \ldots X_n$ 是取自总体 X 的一个样本，其密度函数为：

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

求 θ 的极大似然估计.

解：作似然函数：

$$L(\theta)=\prod_{i=1}^{n}\theta\, x_i^{\theta-1}=\theta^{n}\Big(\prod_{i=1}^{n}x_i\Big)^{\theta-1}\qquad \begin{matrix}(0<x_i<1)\\ 1\le i\le n\end{matrix}$$

则对数似然函数为：

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

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

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

第二节 估计量的评选标准

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。如果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。这就引出无偏性这个评选标准 。

定义：设 $\hat{\theta}$ 是 θ 的估计量，若对任意的 $\theta \in H$ 有：

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

则称 $\hat{\theta}$ 是 θ 的无偏估计量。

注：无偏性的实际意义是指没有系统性的偏差 。

它是用数学期望衡量其靠近真值的程度。

在科学技术中称 $E(\hat{\theta})-\theta$ 为以 $\hat{\theta}$ 作为 θ 的估计的系统误差。则无偏估计即无系统误差。

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则不会产生系统偏差 。

例1. 设总体 X 的均值 μ, 方差 $\sigma^2>0$ 都存在，若 μ, σ^2 均未知.

证明: $\hat{\sigma}^2$ 的两个估计量 $\hat{\sigma}_1^{\,2}=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

$$\hat{\sigma}_2^{\,2}=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

前者是有偏的，后者是无偏的。

证明: $\because E(\hat{\sigma}_1^{\,2})=E(\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2)$

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

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

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

数学期望的性质

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

$$= (\sigma^2 + \mu^2) - (\frac{\sigma^2}{n} + \mu^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$$

$\therefore\ \hat{\sigma}_1^{\,2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ 是有偏的。

$$\because\ E(\hat{\sigma}_2^{\,2}) = E\left(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right)$$

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

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

$$\therefore\quad \hat{\sigma}_2^{\,2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

$= S^2$（样本方差）是无偏的。

结论:

- 样本均值、样本方差作为总体均值、总体方差的估计量是无偏的；
- 它要比矩估计法，极大似然估计法求出来的方差估计量更接近于真值。
- 所以：一般可取样本均值与方差作为总体均值与方差的估计量。

二. 有效性

注意到，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, 则可通过比较$E[(\hat{\theta}_1-\theta)^2]$和$E[(\hat{\theta}_2-\theta)^2]$的大小来决定二者谁更优 。又由于$D(\hat{\theta})=E[(\hat{\theta}-\theta)^2]$，这就引出有效性这个评选标准。

定义： 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 θ 的无偏的估计。若：

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

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

注： 有效性指的是在同是 θ 的无偏估计量的前提下，希望估计值与真值的偏离程度越小越好。一般称方差越小的估计量越有效。

例 2. 若总体 X 的均值为 μ,方差 σ^2 ，但均为未知，现有两个 μ 的无偏估计量：

$$\hat{\mu}_1 = \overline{X}, \quad \hat{\mu}_2 = X_1$$

求: μ_1 与 μ_2 哪个作为 μ 的无偏估计更有效?

例 2. 若总体 X 的均值为 μ,方差 σ^2 ，但均为未知，现有两个μ 的无偏估计量：

$$\hat{\mu}_1 = \overline{X}, \quad \hat{\mu}_2 = X_1$$

求: μ_1 与 μ_2 哪个作为 μ 的无偏估计更有效?

解: $\because$ $D(\hat{\mu}_1) = D(\overline{X}) = D(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sigma^2$

$$D(\hat{\mu}_2) = D(X_1) = \sigma^2$$

且显然，$\frac{\sigma^2}{n} \le \sigma^2$

$\therefore$ 用 $\hat{\mu}_1 = \overline{X}$ 作为 μ 的估计量更有效。

三. 一致性 (相合性)

- 注意到，前面介绍的无偏性和有效性都是在样本容量n固定的前提下提出的。当样本容量n增大时自然希望估计量对未知参数的估计更精确；
- 再注意到，在无偏估计类中所讨论的是以估计量的方差的大小（有效性）作为衡量估计量为“最优”的准则。
- 但是无偏估计类中方差为最小或较小的估计量不一定比某个有偏的估计量的方差来的小；

- 另外，有偏与无偏是反映估计量的数学期望是否等于被估计的参数的真值；
- 方差的大小是反映估计量的观测值与被估计的参数的真值的离散程度。
- 如果希望在偏差性与离散性两者兼顾的原则下建立估计量为“最优”的准则，这就引出相合性的概念。

定义： 设 $\hat{\theta}$ 为参数 θ 的估计量，若对于任意的

$\theta \in H$ 当 $n \to \infty$ 时 $\hat{\theta}$ 依概率收敛于 θ,

即： $\forall \varepsilon > 0$, $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$

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

注:
- **样本k阶矩是总体X的k阶矩的相合估计量;**
- **一般**，一致性（相合性）是要在样本容量相当大时才能显示出其优越性；但这在实际中很难做到。

定义： 设 $\hat{\theta}$ 为参数 θ 的估计量，若对于任意的 $\theta \in H$ 当 $n \to \infty$ 时 $\hat{\theta}$ 依概率收敛于 θ，

即： $\forall \varepsilon > 0$， $\lim\limits_{n\to\infty} P(|\hat{\theta}_n - \theta| < \varepsilon) = 1$

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

注:
- 因此**在工程中经常使用的是无偏性和有效性**这两条衡量估计量为“最优”的准则。
- 但**值得指出**的是：若估计量不具有相合性，那么不论将样本容量 n 取的多大，都不能使得待估计量估计的足够准确。

例 2. 设 $X_1, X_2, \cdots, X_n$ 为总体 $X \sim b(m, p)$ $(0 < p < 1)$ 的样本,

求未知参数 p 的MLE $\hat{p}$, 试证 $\hat{p}$ 是 p 的无偏估计与相合估计。

似然函数为 $$L(p) = \prod_{i=1}^{n} C_m^{X_i} p^{X_i} (1-p)^{m-X_i}$$

$$= (\prod_{i=1}^{n} C_m^{X_i}) \cdot p^{\sum_{i=1}^{n} X_i} \cdot (1-p)^{\sum_{i=1}^{n}(m-X_i)}$$

$$= (\prod_{i=1}^{n} C_m^{X_i}) \cdot p^{n\bar{X}} \cdot (1-p)^{mn-n\bar{X}}$$

令 $\dfrac{\partial \ln L}{\partial p} = \dfrac{n\bar{X}}{p} - \dfrac{mn-n\bar{X}}{1-p} = 0$, 求得 p 的MLE为 $\hat{p} = \dfrac{\bar{X}}{m}$

例 2. 设 $X_1, X_2, \cdots, X_n$ 为总体 $X \sim b(m, p)$ $(0 < p < 1)$的样本，求未知参数 p 的MLE $\hat{p}$，试证 $\hat{p}$ 是 p 的无偏估计与相合估计。

似然函数为 $$L(p) = \prod_{i=1}^{n} C_m^{X_i} p^{X_i} (1-p)^{m-X_i}$$

$$= (\prod_{i=1}^{n} C_m^{X_i}) \cdot p^{\sum_{i=1}^{n} X_i} \cdot (1-p)^{\sum_{i=1}^{n}(m-X_i)}$$

$$= (\prod_{i=1}^{n} C_m^{X_i}) \cdot p^{n\bar{X}} \cdot (1-p)^{mn-n\bar{X}}$$

令 $\dfrac{\partial \ln L}{\partial p} = \dfrac{n\bar{X}}{p} - \dfrac{mn - n\bar{X}}{1-p} = 0$，求得 p 的MLE为 $\hat{p} = \dfrac{\bar{X}}{m}$

$$\because\ E(\hat{p}) = \frac{E(\bar{X})}{m} = \frac{mp}{m} = p$$

无偏估计

$$\hat{p} = \frac{\bar{X}}{m} \xrightarrow{P} \frac{mp}{m} = p \quad (n \to \infty)$$

相合估计

$\therefore$ $\hat{p} = \dfrac{\bar{X}}{m}$ 是 p 的无偏估计与相合估计。

第七章部分作业（教材第五版）：

P176：1、2、3、5

P177：8、10

P178：12、13、15

注：作业不得抄袭；写上姓名、班级、学号和页码（如1/5），待第七章讲授结束，第五章、第六章、第七章作业一起提交至教学云平台，标明题目属于第五章/第六章/第七章。