第七章 参数估计

§1 点估计

§2 估计量的评选标准

§1 点估计

参数估计与非参数估计

- 参数估计——总体分布的类型是已知的，通过样本估计其中的未知参数
- 非参数估计——总体分布的类型是未知的，通过样本估计总体的分布

点估计问题

设总体X的分布函数$F(x,\theta)$的形式已知(例如：正态分布)，θ是待估计参数(例如：期望μ)。

$X_1, X_2, \dots, X_n$是X的一个样本， $x_1, x_2, \dots, x_n$是相应的样本值。

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

这种对未知参数进行定值估计的问题就是点估计问题

概念区分：

- 估计量与估计值
  - **估计量是统计量**，因而是随机变量
  - **估计值**则是一个**标量或向量**
- 在不会导致混淆的情况下，统称估计量与估计值为未知变量θ的估计
- 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, 如何构造和求解估计量是关键问题

常用构造估计量的方法

- 矩估计法
- 最大似然估计法
- 贝叶斯估计

矩估计法

若X为连续型随机变量，其概率密度函数为

$$f(x;\theta_1,\theta_2,\dots,\theta_k)$$

若X为离散型随机变量，其分布律为

$$P\{X=x\}=p(x;\theta_1,\theta_2,\dots,\theta_k)$$

其中，$\theta_1,\theta_2,\dots,\theta_k$是待估计参数，$X_1,X_2,\dots,X_n$为来自X的样本。

设相应的l阶矩$E[X^l]=\mu_l$存在，$l=1,2,\dots,k$

则$\mu_l=\mu_l(\theta_1,\theta_2,\dots,\theta_k),l=1,2,\dots,k$

从样本获得的相应的l阶矩为

$$A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l\ ,\ l=1,2,\dots,k$$

于是有包含$\theta_1, \theta_2, \ldots, \theta_k$的联立方程组，

$$\begin{cases} A_1 = \mu_1(\theta_1, \theta_2, \ldots, \theta_k) \\ A_2 = \mu_2(\theta_1, \theta_2, \ldots, \theta_k) \\ \qquad \ldots \ldots \\ A_n = \mu_n(\theta_1, \theta_2, \ldots, \theta_k) \end{cases}$$

从中解出方程组的解，记为$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ ，即

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \ldots, X_k) \\ \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \ldots, X_k) \\ \qquad \ldots \ldots \\ \hat{\theta}_k = \hat{\theta}_k(X_1, X_2, \ldots, X_k) \end{cases}$$

用$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$分别作为$\theta_1, \theta_2, \ldots, \theta_k$的估计量，这种利用矩求估计量的方法称为矩估计法，这种估计量称为**矩估计量**；相应的观察值称为**矩估计值**

矩估计法的理论分析

由于$X_1, X_2, \dots, X_n$相互独立，且与总体X同分布，

于是有$X_1^l, X_2^l, \dots, X_n^l$相互独立，且与总体$X^l$同分布，

故有

$$E\left(X_1^l\right) = E\left(X_2^l\right) = \cdots = E\left(X_n^l\right) = \mu_l, l = 1, 2, \dots, k$$

由辛钦大数定律知

$$A_l = \frac{1}{n}\sum_{i=1}^{n} X_i^l \xrightarrow{P} \mu_l \ , \ l = 1, 2, \dots, k$$

因此可以假设

$$A_l = \mu_l \ , \ l = 1, 2, \dots, k$$

用A_l估计μ_l

独立同分布的随机变量序列$\{X_n\}$，且具有数学期望$E(X_k) = \mu,\ (k = 1,2, \dots)$，则有
$\frac{1}{n}\sum_{k=1}^{n} X_k \xrightarrow{P} \mu$

矩法求估计量的步骤：

1. 求$\mu_1 = E[X]$ $(\mu_2 = E[X^2], \dots)$
2. 设$A_1 = \mu_1$ $(A_2 = \mu_2, \dots)$
3. 解上面的方程(组)，得$\hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \dots, X_k)$，$\left(\hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \dots, X_k), \dots\right)$

例 1 设某炸药厂一天中发生着火现象的次数X服从参数为λ的泊松分布，λ未知，有以下样本：

着火的次数 k	0	1	2	3	4	5	6	
发生 k次着火天数 n_k	75	90	54	22	6	2	1	$\sum = 250$

试用矩估计估计参数λ。

解：因为泊松分布有：$\mu_1 = E[X] = \lambda$

样本的一阶矩为：

$$A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

则$\hat{\lambda} = \bar{X} = \frac{1}{250}(0 \times 75 + 1 \times 90 + \cdots + 6 \times 1) = 1.22$

于是λ的估计值$\hat{\lambda}$=1.22

例2. 设总体$X \sim U[a, b]$，a, b未知， $X_1, X_2, \ldots, X_n$是一个样本，求a, b的矩估计量。

解：$\mu_1 = E[X] = \frac{a+b}{2}$

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

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

导出：$\begin{cases} a + b = 2A_1 \\ b - a = \sqrt{12(A_2 - A_1^2)} \end{cases}$

解得

$$\begin{cases}\hat{a} = A_1 - \sqrt{3(A_2 - A_1^2)} = \bar{X} - \sqrt{\dfrac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \\ \hat{b} = A_1 + \sqrt{3(A_2 - A_1^2)} = \bar{X} + \sqrt{\dfrac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}\end{cases}$$

注意对样本方差的有偏估计

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

例3. 设总体X的均值μ、方差σ^2均存在，且$\sigma^2 > 0$，但μ、σ^2未知，有$X_1, X_2, \dots, X_n$是一个样本。求μ、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，$\mu_1 = A_1,\ \mu_2 = A_2$

即有 $\begin{cases} \mu = A_1 \\ \sigma^2 + \mu^2 = A_2 \end{cases}$

于是解得

$$\begin{cases} \hat{\mu} = A_1 = \bar{X} \\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 \end{cases}$$

例4. 设总体X服从参数为λ的指数分布，其中$\lambda > 0$未知，$X_1, X_2, \dots, X_n$是一个样本，试求参数λ的矩估计。

解：总体X的密度函数为

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \le 0 \end{cases}$$

于是

$$E[X] = \int_{-\infty}^{+\infty} x f(x) dx = \int_{0}^{+\infty} x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

令$\bar{X} = 1/\lambda$

于是得到λ的矩估计量为

$$\hat{\lambda} = \frac{1}{\bar{X}}$$

例5. 设总体X的密度函数为

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

其中$\alpha>0$为未知参数，试求参数α的矩估计量。

解：

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

于是令，

$$\bar{X}=\frac{\alpha+1}{\alpha+2}$$

由此解得α的矩估计量为：

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

极大似然估计(MLE)

问题引出

实例1 已知某盒中装有一些黑球和白球，不知道哪种球多，但知道它们的数目比是1:2. 从中有放回地抽取5个球，发现黑球有2只，白球3只. 问盒中哪个球多？

设X表示抽取的5个球中黑球的数目，则$X \sim B(5, p)$

由于数目比为1:2，因此$p = 1/3$或$p = 2/3$。

若$p = 1/3$，则$P(X = 2) = C_5^2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^3 = \frac{80}{243}$

若$p = 2/3$，则$P(X = 2) = C_5^2 \left(\frac{2}{3}\right)^2 \left(\frac{1}{3}\right)^3 = \frac{40}{243}$

据此，认为$p = 1/3$，则黑球数目少。

极大似然估计

基本假设：

设某一随机试验共有$A_1, A_2, \dots, A_k, \dots$若干可能结果，若在一次试验中$A_m$出现，则认为$A_m$出现概率最大。

离散型分布的极大似然估计

若总体X属离散型，其分布律

$$P\{X=x\}=p(x;\theta),\theta\in\Theta$$

的形式已知， θ为待估计参数， Θ是θ的取值范围。

设$X_1,X_2,\dots,X_n$是来自X的一个样本，则$X_1,X_2,\dots,X_n$的联合分布率为

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

又设$x_1,x_2,\dots,x_n$ 是$X_1,X_2,\dots,X_n$的一个样本值，事件$\{X_1=x_1,X_2=x_2,\dots,X_n=x_n\}$发生的概率为：

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

$L(\theta)$称为样本的似然函数

连续型分布的极大似然估计

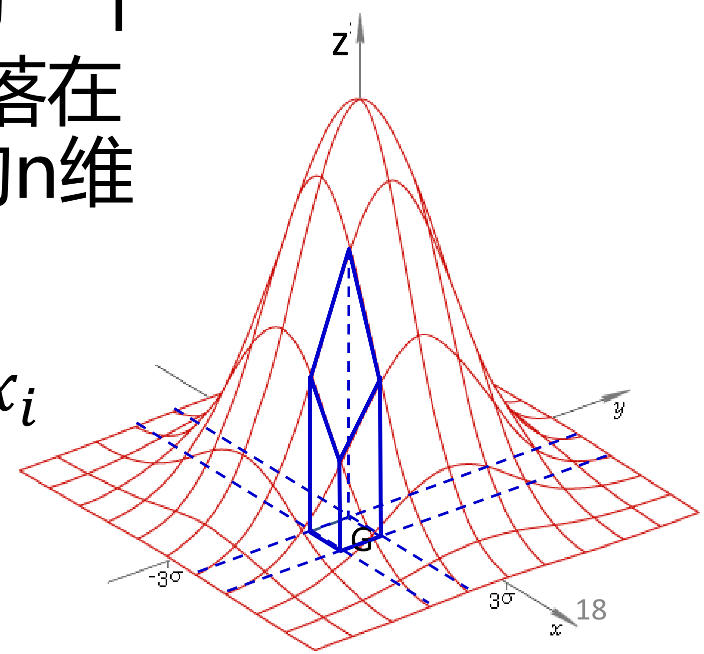
若总体X属连续型，其概率密度$f(x;\theta),\theta\in\Theta$的形式已知， θ为待估计参数， Θ是θ的取值范围。

设$X_1,X_2,\dots,X_n$是来自X的一个样本，则$X_1,X_2,\dots,X_n$的联合概率密度为

$$\prod_{i=1}^{n} f(x_i;\theta)$$

又设$x_1,x_2,\dots,x_n$ 是$X_1,X_2,\dots,X_n$的一个样本值，则随机点$(X_1,X_2,\dots,X_n)$落在$(x_1,x_2,\dots,x_n)$的邻域(dx_i为边长的n维立方体)内的概率近似为：

$$\prod_{i=1}^{n} f(x_i;\theta)dx_i$$



连续型分布的极大似然估计

选取得到观察值$(x_1, x_2, \dots, x_n)$最大化上式情况下的θ作为其估计值$\hat{\theta}$

注意到，$\prod_{i=1}^{n} dx_i$不随θ改变，故只需考虑

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

$L(\theta)$同样称为样本的似然函数

极大似然估计原理：

固定$x_1, x_2, \dots, x_n$，选择使概率$L(x_1, x_2, \dots, x_n; \theta)$达到最大的参数$\hat{\theta}$作为对$\theta$的估计值，即取$\hat{\theta}$使得：

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

$\hat{\theta}$ 与$x_1, x_2, \dots, x_n$ 有关，记为$\hat{\theta}(x_1, x_2, \dots, x_n)$，称其为参数$\theta$的极大似然估计值。

这种求未知参数θ的方法称为极大似然法

若$L(x_1, x_2, \dots, x_n; \hat{\theta}) = \max\limits_{\theta \in \Theta} L(x_1, x_2, \dots, x_n; \theta)$

则称$\hat{\theta}(x_1, x_2, \dots, x_n)$为$\theta$的极大似然估计值

称$\hat{\theta}(X_1, X_2, \dots, X_n)$为$\theta$的极大似然估计量

一般情况下$p(x_i; \theta)$，$f(x_i; \theta)$关于θ可导，故θ可由下式求得：

$$\frac{dL(\theta)}{d\theta} = 0 \quad \text{– 似然方程}$$

又因为$L(\theta)$与$\ln L(\theta)$在同一θ处取得极值，因此有等价的对数似然方程

$$\frac{d}{d\theta} \ln L(\theta) = 0$$

样本$X_1, X_2, \ldots, X_n$的联合密度函数与似然函数的异同

相同点——形式上一致，都是

$$\prod_{i=1}^{n} f(x_i;\theta)$$

不同点——

在联合密度函数中θ 看作是固定数

在似然函数中将θ看作是变量

从单一参数到多参数

若总体的分布中包含多个参数，则可令

$$\frac{d\boldsymbol{L}}{d\theta_i} = 0, \qquad i = 1,2,\dots,k$$

或

$$\frac{d\ln\boldsymbol{L}}{d\theta_i} = 0, \qquad i = 1,2,\dots,k$$

解k个方程组求得$\theta_1, \theta_2, \dots, \theta_k$的极大似然估计值

一般而言，极大似然估计优于矩估计

注意：若似然方程（组）无解，或似然函数不可导，此法失效，改用其它方法

例6. 设$X\sim B(1,p)$， $X_1,X_2,\ldots,X_n$来自X的一个样本，试求参数p的MLE。

解：设$x_1,x_2,\ldots,x_n$是一个样本值，X的分布律为：

$$P\{X=x\}=p^x(1-p)^{1-x},x=0,1$$

故似然函数为：

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

而

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

令

$$\frac{d\ln \boldsymbol{L}(p)}{dp} = 0$$

得到

$$\frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p} = 0$$

解得p的MLE值为

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

p的MLE量为

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

注意：与矩估计量是相同的

例7. 设总体$X\sim N(\mu,\sigma^2)$ ，但μ、σ^2未知，$x_1,x_2,\ldots,x_n$是一个样本。求μ、σ^2的极大似然估计量

解：X的概率密度为

$$f(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

于是似然函数为：

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

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

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

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

令$\begin{cases}\dfrac{d}{d\mu}\ln L(\mu,\sigma^2)=0\\ \dfrac{d}{d\sigma^2}\ln L(\mu,\sigma^2)=0\end{cases}$，

得$\begin{cases}\dfrac{\sum_{i=1}^{n}x_i-n\mu}{\sigma^2}=0\\ -\dfrac{n}{2\sigma^2}+\dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0\end{cases}$

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

相应的极大似然估计量为$\begin{cases}\mu=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}\\ \sigma^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2\end{cases}$

与相应的矩估计量是相同的

例8. 设总体X服从参数为$\lambda(>0)$的泊松分布，$X_1, X_2, \dots, X_n$是来自X的一个样本，试求参数λ的最大似然估计量。

解：X的分布律为

$$P\{X=x\}=\frac{\lambda^x}{x!}e^{-\lambda}, \qquad (x=0,1,\dots,n)$$

于是λ的似然函数为

$$L(\lambda)=\prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right)=e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n}x_i}}{\prod_{i=1}^{n}x_i!}$$

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

令

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

解得λ的MLE值为

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

λ的最大似然估计量为

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

与相应的矩估计量是相同的

例9. 设总体X的密度函数为

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

其中θ未知，$\theta>1$，$X_1,X_2,\ldots,X_n$是来自X的一个样本，试求参数θ的最大似然估计

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

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

令$\frac{\mathrm{d}}{\mathrm{d}\theta}\ln L=0$，有似然方程$\frac{n}{\theta}+\sum_{i=1}^n \ln x_i=0$

解得$\hat{\theta}=-\frac{n}{\sum_{i=1}^n \ln x_i}$，

因此θ的极大似然估计量为$\hat{\theta}=-\frac{n}{\sum_{i=1}^n \ln X_i}$

例10. 设总体$X \sim U[a,b]$，a,b未知，$X_1, X_2, \ldots, X_n$是一个样本，求a,b的极大似然估计量。

注意：X的概率密度为$f(x;a,b)=\begin{cases}\frac{1}{b-a}, & a \le x \le b\\ 0, & otherwise\end{cases}$

似然函数为

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

$$\ln L(a,b)=-n\ln(b-a)$$

于是$\frac{\partial \ln L}{\partial a}=\frac{n}{b-a}=0, \frac{\partial \ln L}{\partial b}=-\frac{n}{b-a}=0,$

显然，似然方程组无解，但这不能说明不存在极大似然估计量，只是不能由似然方程组求解。

解：将$x_1, x_2, \ldots, x_n$排序$x_{(1)} \le x_{(2)} \le \cdots \le x_{(n)}$

则

$$L(a,b) = \begin{cases} \dfrac{1}{(b-a)^n}, a \le x_{(1)} \le \cdots \le x_{(n)} \le b \\ 0, \qquad otherwise \end{cases}$$

对于满足$a \le x_{(1)} \le \cdots \le x_{(n)} \le b$的任意$a, b$有

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

即$L(a,b)$在$b = x_{(n)}, a = x_{(1)}$时，取最大值

MLE值为：$\hat{b} = x_{(n)} = \max x_i, \hat{a} = x_{(1)} = \min x_i$

MLE量为：$\hat{b} = \max X_i, \hat{a} = \min X_i$

例11　为了估计湖中有多少条鱼，从湖中捞出1000条鱼，标上记号后又放回湖中，过一段时间后，再捞出150条鱼，发现其中有10条鱼带有标记，估计湖中鱼的总数为多少时使上述事件的概率为最大？

解： 设湖中鱼的总数为N，则带记号的鱼所占比例为$1000/N$．令X为从湖中捞起一条鱼所带有的记号数．则

$$X\sim B(1,p)$$

$$P\{X=x\}=\begin{cases}\dfrac{1000}{N}, & x=1\\ 1-\dfrac{1000}{N}, & x=0\end{cases}$$

设$X_i = \begin{cases} 1, 第i条鱼有标记 \\ 0, 否则 \end{cases}$，$i = 1,2,\ldots,150$

由于湖中鱼的数量众多，第二次捞鱼可近似为有放回抽样

于是令

$$Y = \sum_{i=1}^{150} X_i$$

则$Y = \sum_{i=1}^{150} X_i \sim B\left(150, \frac{1000}{N}\right)$，则题中的事件为$\{Y = 10\}$，取似然函数为：

$$L(N) = P\{Y = 10\} = C_{150}^{10}\left(\frac{1000}{N}\right)^{10}\left(\frac{N-1000}{N}\right)^{140}$$

$$L(N)=C_{150}^{10}\frac{1000^{10}(N-1000)^{140}}{N^{150}}$$

于是，

$\ln L(N)=\ln C_{150}^{10}+10\ln 1000+140\ln(N-1000)-150\ln N$

令，

$$\frac{d}{dN}\ln L(N)=\frac{140}{N-1000}-\frac{150}{N}=0$$

解得N=15000

即当湖中的鱼数为15000条时，概率$P\{Y=10\}$为最大。

极大似然估计量的不变性

设θ的函数$u = u(\theta), \theta \in \Theta$具有单值反函数，$\hat{\theta}$是$\theta$的极大似然估计，则

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

例：$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$是$\sigma^2$的极大似然估计，

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

故$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$是$\sigma$的极大似然估计

例12. 设$X\sim N(\mu,\sigma^2)$, μ,σ^2未知，求使$P\{X>A\}=0.05$的点A的极大似然估计量。

解：

$$P\{X>A\}=1-\Phi\left(\frac{A-\mu}{\sigma}\right)=0.05$$

查表有$\frac{A-\mu}{\sigma}=1.645$，故有$A=\mu+1.645\sigma$

由前面知μ,σ^2的极大似然估计量分别为：

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

故A的极大似然估计量为

$$\hat{A}=\hat{\mu}+1.645\hat{\sigma}=\bar{X}+1.645\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2}$$

§2 估计量的评选标准

缘起：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。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?

考虑估计的性质

(1) 无偏性

(2) 有效性

(3) 相合性(一致性)

一、无偏性

若$X_1, X_2, \ldots, X_n$为总体X的一个样本，$\theta \in \Theta$是包含在总体X的分布中的待估计参数，

若估计量$\hat{\theta} = \theta(X_1, X_2, \ldots, X_n)$的数学期望$E(\hat{\theta})$存在且对于任意$\theta \in \Theta$有$E\left(\hat{\theta}\right) = \theta$，则称$\hat{\theta}$为$\theta$的无偏估计量

无偏估计意味着无系统性误差

例13. 设总体X的k阶矩$\mu_k = E\left(X^k\right)\ (k \geq 1)$存在，又设$X_1, X_2, \ldots, X_n$为总体X的一个样本，试证明不论总体服从何种分布，k阶样本矩$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$是k阶总体矩$\mu_k$的无偏估计。

证明：因为$X_1, X_2, \ldots, X_n$与X同分布，故有

$$E\left(X_i^k\right) = E\left(X^k\right) = \mu_k$$

即

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

故k阶样本矩A_k是k阶总体矩μ_k的无偏估计

注意：不论总体 X 服从什么分布, 只要它的数学期望存在,$\overline{X}$总是总体X的数学期望$\mu_1 = E(X)$的无偏估计

例14. 对于均值μ，方差σ^2都存在的总体，若μ和σ^2均为未知，则σ^2的估计量$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$是有偏的。

证明：$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\bar{X}^2=A_2-\bar{X}^2$

因为$E(A_2)=\mu_2=\sigma^2+\mu^2$

而$E(\bar{X}^2)=D(\bar{X})+[E(\bar{X})]^2=\frac{\sigma^2}{n}+\mu^2$

于是有

$$E(\hat{\sigma}^2)=E(A_2-\bar{X}^2)=E(A_2)-E(\bar{X}^2)$$
$$=\sigma^2+\mu^2-\frac{\sigma^2}{n}-\mu^2=\frac{n-1}{n}\sigma^2\neq\sigma^2$$

故是有偏的。

若以$\frac{n}{n-1}$乘以$\hat{\sigma}^2$，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

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

即S^2是σ^2的无偏估计，通常取S^2作为σ^2的估计量

例16. 设$X \sim N(\mu, \sigma^2)$，μ已知，σ^2未知，$x_1, x_2, \dots, x_n$是来自X的一个样本值，求σ^2的极大似然估计量。

解：X的概率密度函数为：

$$f(x;\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

似然函数为：

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

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

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

令

$$\frac{d}{d\sigma^2}\ln L=-\frac{n}{2\sigma^2}-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4}=0$$

解得：

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

注意到

$$E(\hat{\sigma}^2)=E\left(\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2\right)$$

$$=\frac{1}{n}\sum_{i=1}^{n}E[(x_i-\mu)^2]=\frac{1}{n}\times n\sigma^2=\sigma^2$$

故，$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2$是总体的无偏估计

例17. 设总体X服从参数为θ的指数分布，概率密度

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

参数$\theta>0$，又设$X_1,X_2,\dots,X_n$是来自总体X的样本，试证$\bar{X}$和$nZ=n[\min(X_1,X_2,\dots,X_n)]$都是$\theta$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。

而$Z=\min(X_1,X_2,\dots,X_n)$服从参数为$\frac{\theta}{n}$的指数分布，

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

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

所以nZ也是θ的无偏估计量。

由此可知，一个参数可以有不同的无偏估计量

注意：如果未知参数θ有两个不同的无偏估计量$\hat{\theta}_1$和$\hat{\theta}_2$，则一定有无穷多个无偏估计

$$\alpha\hat{\theta}_1+(1-\alpha)\hat{\theta}_2$$

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$\hat{\theta}_2$更为密集，则认为$\hat{\theta}_1$较$\hat{\theta}_2$有效

由于方差是随机变量取值与其数学期望的偏离程度,所以无偏估计以方差小者为好

形式化：

设$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\ldots,X_n)$和$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\ldots,X_n)$都是$\theta$的无偏估计量，若有$D\left(\hat{\theta}_1\right)\le D\left(\hat{\theta}_2\right)$，则称$\hat{\theta}_1$较$\hat{\theta}_2$有效

例18. 设总体X服从参数为θ的指数分布，概率密度

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

参数$\theta>0$，又设$X_1,X_2,\dots,X_n$是来自总体X的样本，试证θ的无偏估计中，当$n>1$时$\bar{X}$较nZ更为有效。

证明：由于$D(X)=\theta^2$，故有$D(\bar{X})=\frac{\theta^2}{n}$

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

当n>1时，$D(nZ)>D(\bar{X})$,

所以θ的无偏估计中，当$n>1$时$\bar{X}$较nZ更为有效

三、相合性(一致性)

若$\hat{\theta}=\hat{\theta}(X_1,X_2,\ldots,X_n)$为参数$\theta$的估计量，如果对于任意$\theta\in\Theta$，当$n\to\infty$时$\hat{\theta}_n\xrightarrow{P}\theta$，

则称$\hat{\theta}$是θ的一致估计。

例19. 若$X_1,X_2,\ldots,X_n$是总体X的样本，$E[X]=\mu$，$E[X^k]=\mu_k$，有辛钦大数定理，知对任意给定的$\epsilon>0$，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{i=1}^{n}X_i-\mu\right|\geq\epsilon\right\}=0$$

因此，样本均值$\bar{X}$是总体期望μ的相合(一致)估计量

作业

概率论与数理统计

pp. 173-175, #2，#5，#10，#12