第六章　　参数估计

点估计的几种方法：

矩估计，极大似然估计

点估计的优良准则

区间估计的“枢轴量”方法

数理统计的基本问题是如何根据样本所提供的信息，对于总体的分布以及分布的某些数字特征进行统计推断

根据样本, 对于分布中的未知参数进行估计，这类问题常称为参数估计问题

$$\text{参数估计}\begin{cases} \text{点估计——估计参数的值} \\ \text{区间估计——估计参数的范围} \end{cases}$$

设θ为未知参数，找统计量 $T=T(x_1,x_2\cdots,x_n)$

将$T=T(x_1,x_2\cdots,x_n)$作为θ的估计值，记为

$$\hat{\theta}=T(x_1,x_2\cdots,x_n).$$

矩法估计

用样本矩替换总体矩

假定总体的 k 阶原点矩 $\mu_k = E\xi^k$ 存在，它们都是

$\theta_1, \theta_2, \cdots, \theta_k$ 的函数。构造 “矩方程组”：

$$\overline{X^j} = E\xi^j, \quad j = 1,2,\cdots,k$$

求解方程组得解：

$$\hat{\theta}_j = \hat{\theta}_j(X_1, X_2, \cdots, X_n), \quad j = 1,2,\cdots,k$$

例1.设$x_1, x_2, \cdots, x_n$为来自总体 X的样本，

已知总体 X的期望和方差均存在， 但未知

令$EX = \mu, DX = \sigma^2$。求μ、σ^2与σ的矩估计。

解：解方程组：$\begin{cases} EX = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i \\ EX^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n} x_i^2 \end{cases}$ 即 $\begin{cases} \mu = \overline{X} \\ \sigma^2 + \mu^2 = \overline{X^2} \end{cases}$

所求矩估计为 $\hat{\mu} = \overline{X}, \hat{\sigma}^2 = \overline{X^2} - (\overline{X})^2 = \dfrac{n-1}{n} s^2,$

$$\hat{\sigma} = \sqrt{\overline{X^2} - (\overline{X})^2} = \sqrt{\frac{n-1}{n}} s$$

注意：期望、方差的矩估计没有涉及总体的分布。

已知分布类型，参数未知。求参数的矩估计。

例2.设$x_1, x_2, \cdots, x_n$为来自总体 X的样本，已知总体 $X \sim E(\lambda)$,求λ的矩估计。

解： $$EX = \frac{1}{\lambda} = \overline{X},$$

$$\Rightarrow \hat{\lambda} = \frac{1}{\overline{X}}$$

即λ的矩估计为 $\frac{1}{\overline{X}}$。

例3.设$x_1, x_2, \cdots, x_n$为来自总体X的样本，

已知总体X服从$[0, \theta]$上的均匀分布，$\theta > 0$

未知，求θ的矩估计。

解：$EX = \frac{\theta}{2} = \overline{X},$

$$\Rightarrow \hat{\theta} = 2\overline{X}$$

即θ的矩估计为$2\overline{X}$。

若$(x_1, x_2, x_3, x_4, x_5) = (1,2,3,5,9)$,

$$\overline{X} = \frac{1+2+3+5+9}{5} = 4 \quad \Rightarrow \hat{\theta} = 2\overline{X} = 8$$

矩估计的优缺点

优点：计算简单。

缺点：（1）总体的矩不一定存在。故矩估计不一定可行。

（2）可能会有不同的矩估计。

已知总体 X服从 $P(\lambda)$，则 λ的矩估计：

解1：$\hat{\lambda} = EX = \overline{X}$

解2：$\hat{\lambda} = DX = EX^2 - (EX)^2 = \overline{X^2} - (\overline{X})^2$

规定：尽量使用低阶矩。

（3）可能会得到不合理的解。

例3中，$x_5 = 9 > \hat{\theta} = 8, \quad \therefore \hat{\theta} = 8$ 不合理。

极大似然估计

一个随机试验如果有若干个可能的结果 $A, B, C, \cdots$，而在一次试验中，结果 A 出现了，则一般认为试验的机理最有利于结果 A 出现，即使得 A 出现的可能性最大。

例：从一批产品中抽出5件检查，发现有前2件是次品，问这批产品的次品率是多少？

设产品的次品率为 p，$p \in (0,1)$

任取一产品，次品数（ 总体）$\xi \sim b(1, p)$

设第i次抽到的次品数为 $\xi_i (i = 1,2,3,4,5)$,

样本$\xi_1, \xi_2, \xi_3, \xi_4, \xi_5$独立同分布， $\xi_i \sim b(1, p)$

$$P(\xi_1 = 1, \xi_2 = 1, \xi_3 = 0, \xi_4 = 0, \xi_5 = 0) = p^2(1-p)^3$$

p	0.2	0.4	0.6	0.8
$p^2(1-p)^3$	0.02048	0.03456	0.02304	0.00512

以样本值出现可能性最 大的p值作为其估计，这是极大似然估计的思想方 法。

这里， p的极大似然估计为： $\hat{p} = 0.4$

设$X_1, X_2, \cdots, X_n$为来自总体X的一组样本，

样本值出现可能性

(1)若总体 X是离散型随机变量，

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

(2)若总体X是连续型随机变量，概 率密度为$p(x)$，

与$p^*(x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} p(x_i)$成正比。

（1）、（2）中的分布概率或密度函数均含有未知参数。

问题转化：

当参数取何值时，样本分布概率或概率密度函数值最大？

称以上样本的联合概率分布或联合密度为极大似然函数，

记为L。

似然函数:

离散型

$$L(\theta) = P(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)$$

$$= \prod_{i=1}^{n} P(X_i = x_i),$$

连续型

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

极大似然估计：寻找统计量 $\hat{\theta}$，使其满足

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

为了方便乘积求导，常常改求

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

一般步骤：

1：　写出似然函数 $L(\theta)$ 。

2：将似然函数取对数，求得对数似然函数 $\ln L(\theta)$ 。

3：$\ln L(\theta)$ 关于 θ 求导（若 θ 为多个参数，则分别关于每个参数求偏导），并令其为 **0**，得对数似然方程组。

4：解上述似然方程组，求得 θ 的极大似然估计 $\hat{\theta}$ 。

例 1 设总体$\xi \sim P(\lambda)$，具有分布列

$$P(\xi = k) = \frac{e^{-\lambda} \cdot \lambda^{k}}{k!}, \ k = 0,1,2,\cdots$$

$\lambda > 0$ 为一未知参数。求λ的极大似然估计。

解 设$(X_1, X_2, \cdots, X_n)$为一个样本, $(x_1, x_2, \cdots, x_n)$

为一组观测值。似然函数为

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

取对数，得 $\ln L(\lambda) = \sum_{i=1}^{n}(-\lambda + x_i \ln\lambda - \ln x_i!)$

求导，得 $\dfrac{d\ln L(\lambda)}{d\lambda} = -n + \sum_{i=1}^{n}\dfrac{x_i}{\lambda} = 0$

解得 $\hat{\lambda} = \overline{x}$

由于 $\left.\dfrac{d^2\ln L}{d\lambda^2}\right|_{\lambda=\overline{x}} < 0$，可知 $\hat{\lambda}$ 使 L 达到最大，从而得出 λ

的极大似然估计为 $\hat{\lambda} = \overline{X}$ 。

例 2 设 $(X_1, X_2, \cdots, X_n)$ 为取自正态总体 $N(\mu, \sigma^2)$ 的一个样本，其中 μ, σ^2 为未知参数，参数空间 $\Theta = \{-\infty < \mu < \infty, \sigma^2 > 0\}$。试求参数 μ 与 σ^2 的极大似然估计。

解 设 $(x_1, x_2, \cdots, x_n)$ 为样本的一组观测值，似然函数为

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

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

关于 μ,σ^2 求偏导，并令其为 **0**，得似然方程组

$$\begin{cases}\dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0\\ \dfrac{\partial\ln L(\mu,\sigma^2)}{\partial\sigma^2}=\dfrac{1}{2\sigma^4}\sum\limits_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2\sigma^2}=0\end{cases}$$

解此方程组，可得 μ,σ^2 的极大似然估计为

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

一般步骤：

1： 写出似然函数 $L(\theta)$ 。。

2：将似然函数取对数，求得对数似然函数 $\ln L(\theta)$ 。

3：求得 θ 的极大似然估计 $\hat{\theta}$， 使得

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

例 3 设总体$\xi \sim U[0,\theta]$， $\theta > 0$ 为未知参数。试

求θ的极大似然估计。

解 设$(X_1, X_2, \cdots, X_n)$为一个样本，$(x_1, x_2, \cdots, x_n)$

为一组观测值。似然函数为

$$L(\theta) = \prod_{i=1}^{n} (\frac{1}{\theta}) = \frac{1}{\theta^n}。$$

两边取对数，得对数似然函数

$$\ln L(\theta) = -n \ln \theta$$

关于θ求导，可得$\dfrac{d\ln L(\theta)}{d\theta}=\dfrac{-n}{\theta}<0, \quad (\theta>0)$

$\ln L(\theta)$关于θ严格单调递减，似然函数的最大值应在左面的边界点上达到。注意到

$$0\le x_1\le\theta, 0\le x_2\le\theta, \ldots, 0\le x_n\le\theta$$

$$\theta\ge\max_{1\le i\le n} x_i,$$

当$\theta=\max\limits_{1\le i\le n} x_i$时，$L(\theta)$取到最大值。因此$\theta$的极大似然估计为：$\hat{\theta}=\max\limits_{1\le i\le n} x_i$

如果$\hat{\theta}$是θ的极大似然估计，则对任何函数$g(\theta)$，其

极大似然估计为$g(\hat{\theta})$ -- “不变性”

例 1 设$(X_1, X_2, \cdots X_n)$是来自正态总体$N(\mu, \sigma^2)$的样本，求标准差σ以及变异系数$\gamma \overset{\Delta}{=} \sigma/\mu$的极大似然估计。

解 根据例 2 中给出的参数μ与σ^2的极大似然估计

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

标准差σ以及变异系数$\gamma \overset{\Delta}{=} \sigma/\mu$的极大似然估计为

$$\hat{\sigma} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2} = \sqrt{\frac{n-1}{n}}s, \quad \hat{\gamma} = \frac{\hat{\sigma}}{\hat{\mu}} = \sqrt{\frac{n-1}{n}}\frac{s}{\overline{x}}。$$

例 2. 设总体 X 具有密度函数

$$p(x)=\begin{cases}\beta\alpha^{\beta}x^{-(\beta+1)} & x>\alpha\\ 0 & x\le\alpha\end{cases}$$

其中 $\alpha>0$ 且 $\beta>0$，从中抽得一子样 $X_1, X_2, \cdots X_n$，求

（1）β 已知时，α 的最大似然估计量；

（2）α 已知时，β 的最大似然估计量；

（3）α 为已知时，β 之矩法估计量存在否？

解：（1） β 已知时，

$$L(\alpha)=\begin{cases}\prod_{i=1}^{n}(\beta\alpha^{\beta}x_i^{-(\beta+1)}) & x_i>\alpha,\quad i=1,2,\cdots n\\ 0 & \text{其它}\end{cases}$$

当 $L(\alpha)\neq 0$ 时，

$$\ln L(\alpha)=n\ln\beta+n\beta\ln\alpha-(\beta+1)\sum_{i=1}^{n}\ln x_i \quad ,$$

$$\frac{d\ln L(\alpha)}{d\alpha}=\frac{n\beta}{\alpha}>0$$

故有最大似然估计量 $\hat{\alpha}=\min\limits_{1\leq i\leq n}\{x_i\}$。

（2） α 已知时，

$$L(\beta)=\begin{cases}\prod_{i=1}^{n}(\beta\alpha^{\beta}x_i^{-(\beta+1)}) & x_i>\alpha,\quad i=1,2,\cdots n\\ 0 & \text{其它}\end{cases}$$

$$\ln L(\beta)=n\ln\beta+n\beta\ln\alpha-(\beta+1)\sum_{i=1}^{n}\ln x_i\quad,$$

$$\frac{d\ln L(\beta)}{d\beta}=\frac{n}{\beta}+n\ln\alpha-\sum_{i=1}^{n}\ln x_i$$

令 $\dfrac{d\ln L(\beta)}{d\beta}=0$，故有最大似然估计量 $\hat{\beta}=\dfrac{n}{\sum_{i=1}^{n}\ln x_i-n\ln\alpha}$。

(3) α 为已知时，母体的一阶矩为

$$EX=\int_{\alpha}^{+\infty} x\beta\alpha^{\beta}x^{-(\beta+1)}dx=\frac{\beta}{1-\beta}\alpha^{\beta}x^{1-\beta}\Big|_{\alpha}^{+\infty},$$

显然当 $\mathbf{0}<\beta<\mathbf{1}$ 时， β 的矩法估计量不存在。

当 $\beta>1$ 时， $EX=\dfrac{\beta}{\beta-1}\alpha$，得方程： $\dfrac{\alpha\beta}{1-\beta}=\overline{X}$，

从中解出矩法估计量 $\hat{\beta}=\dfrac{\overline{X}}{\overline{X}+\alpha}$。

矩法估计

假定总体的k阶原点矩$\mu_k = E\xi^k$存在，它们都是

$\theta_1,\theta_2,\cdots,\theta_k$的函数。构造“矩方程组”：

$$\overline{X^j} = E\xi^j, \quad j = 1,2,\cdots,k$$

求解方程组得解：

$$\hat{\theta}_j = \hat{\theta}_j(X_1, X_2, \cdots, X_n), \quad j = 1,2,\cdots,k$$

极大似然估计

一般步骤：

1：　写出似然函数 $L(\theta)$。。

2：将似然函数取对数，求得对数似然函数 $\ln L(\theta)$。

3：求得 θ 的极大似然估计 $\hat{\theta}$，　使得

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