第七章　参数估计

参数估计是统计推断的基本问题之一．在实际问题中，总体 X 的分布类型可能已知，也可能未知．但不论如何，都需要依据样本所提供的信息，估计总体 X 中如数字特征等未知参数 θ 的取值，这就是参数估计问题．其主要内容包含点估计、估计量的评价标准和区间估计．

第一节 点估计

点估计的概念

所谓点估计就是要构造一个合适的统计量

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

作为未知参数θ的估计．统计学上称$\hat{\theta}$为θ**的估计量**.

对应于样本$(X_1,X_2,\cdots,X_n)$的每个观察值$(x_1,x_2,\cdots,x_n)$，

估计量$\hat{\theta}$的值$\hat{\theta}(x_1,x_2,\cdots,x_n)$称为$\theta$**的估计值**.

一、矩估计法

矩估计法是由英国统计学家$K\cdot$皮尔逊$(K\cdot Pearson)$在1894年提出的方法. 矩估计法的原理是来自第五章的大数定律.

设$(X_1, X_2, \cdots, X_n)$为来自总体X的一个样本，且$E(X^r)=\mu_r$，则

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}X_i^r \overset{P}{=} \mu_r = E(X^r).$$

因此，当n充分大时，$\frac{1}{n}\sum_{i=1}^{n}X_i^r \approx E(X^r)$.

定义 1 用来自总体 X 的样本 $(X_1, X_2, \cdots, X_n)$ 的 r 阶原点矩

$A_r = \frac{1}{n}\sum_{i=1}^{n} x_i^r$ 作为总体 X 的 r 阶原点矩 $E(X^r)$ $(r = 1, 2, \cdots, k)$ 的估计量，所产生的参数估计方法称为**矩估计法**，由矩估计法得到的估计量叫做**矩估计量**.

思想与方法： 用样本矩代替理论矩，建立 k 个方程，从中解出 k 个未知参数的矩估计量.

当$k=1$时，方程$\overline{X}=EX$最为常用.

但有时EX中不含有未知参数θ，因此从$\overline{X}=EX$中不能求得$\hat{\theta}$，故此时根据低阶矩优先的原则，如改用二阶原点矩建立方程

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

当$k=2$时，最常用的二个方程为

$$\begin{cases} \overline{X} = EX, \\ \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2). \end{cases}$$

由于此方程组与下列方程组

$$\begin{cases} \overline{X} = EX, \\ \dfrac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 = DX \end{cases}$$

等价，因此后者的使用更为方便.

例 1 设 $(X_1, X_2, \cdots, X_n)$ 为来自总体 $X \sim P(\lambda)$ 的样本，其中 λ 为未知参数．试求 λ 的矩估计量 $\hat{\lambda}$．

解 由于只有一个未知参数 λ，故只需建立一个方程．由 $\overline{X} = EX = \lambda$，解得 $\hat{\lambda} = \overline{X}$．

【直观意义】 设某电话在一定时间段内的被呼唤次数 $X \sim P(\lambda)$，$\lambda = EX$ 为平均呼唤次数．X_i 为第 i 次观察时的呼唤次数，则 $\overline{X}$ 为此 n 次观察时的平均次数．故 $\hat{\lambda} = \overline{X}$ 意味着用观察的平均次数，近似代替一般情况下该电话在此时间段内的平均呼唤次数．

例 2 设总体 $X \sim N(\mu,\sigma^2)$，$(X_1,X_2,\cdots,X_n)$ 为来自总体 X 的样本.

(1) 如果 σ^2 已知， μ 未知，求 μ 的矩估计量 μ；

(2) 如果 μ 已知， σ^2 未知，求 σ^2 的矩估计量 σ^2；

(3) 如果 μ,σ^2 均未知，求 μ 和 σ^2 的矩估计量 μ 和 σ^2 .

解 (1) 由 $\overline{X} = EX = \mu$，解得 $\mu=\overline{X}$.

（续解）（2）由于 EX 中不含有 σ^2，故根据低阶矩优先的原则，改用二阶原点矩建立方程

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) = \sigma^2 + \mu^2,$$

解得 $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \mu^2$.

（3）属 $k=2$ 的情形，故需要建立二个方程．由

$$\begin{cases} \overline{X} = EX = \mu, \\ \frac{1}{n}\sum_{i=1}^{n} (X_i - \overline{X})^2 = DX = \sigma^2, \end{cases}$$

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

例 3 设总体 $X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & \theta^2 & 1-2\theta \end{pmatrix}$，其中 θ 是未知参数，利用总体 X 的样本值

$$(3,\ 1,\ 3,\ 0,\ 3,\ 1,\ 2,\ 3),$$

求 θ 的矩估计值 $\hat{\theta}$.

解 $EX = 0\times\theta^2 + 1\times 2\theta(1-\theta) + 2\times\theta^2 + 3\times(1-2\theta) = 3-4\theta$,

$$\overline{x} = \frac{3+1+3+0+3+1+2+3}{8} = 2,$$

由 $\overline{x} = EX$，即 $2 = 3-4\theta$，故解得 θ 的矩估计值为 $\hat{\theta} = \frac{1}{4}$.

例 4 设总体 X 的密度函数为

$$f(x;\theta)=\begin{cases} e^{-(x-\theta)}, & x\ge\theta, \\ 0, & x<\theta, \end{cases}$$ 其中 θ 为未知参数.

从总体 X 中取得样本 $(X_1,X_2,\cdots,X_n)$，求 θ 的矩估计量 $\hat{\theta}$.

解 由于

$$EX=\int_{-\infty}^{+\infty}xf(x;\theta)dx=\int_{\theta}^{+\infty}xe^{-(x-\theta)}dx=1+\theta,$$

故由 $\overline{X}=EX=1+\theta$，解得 $\hat{\theta}=\overline{X}-1$.

二、极大似然估计法

极大似然估计法是由英国统计学家 $R \cdot A$ 费歇尔（$R \cdot A\ Fisher$）于1912年提出，并在1921年的工作中又加以发展的一种重要且普遍使用的点估计法.

极大似然估计法是依据“概率最大的事件最有可能出现”的“实际推断”原理产生的估计法．其基本思想是：如果在一次试验中事件 A 已出现，则一般说来，当时的试验条件应更有利于事件 A 的出现.

定义 2 如果总体 X 为离散型随机变量，其分布律为

$$P\{X=x\}=p(x;\theta),\quad x=a_1,a_2,\cdots,a_n,\cdots,$$

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

如果总体 X 为连续型随机变量，其密度函数为

$$f(x;\theta),\quad -\infty<x<+\infty,$$

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

称 $L(x_1,x_2,\cdots,x_n;\theta)$ 为**似然函数**，简记为 $L(\theta)$.

定义 3 如果θ满足

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

就称θ为未知参数θ的**极大似然估计量**.

【思想】求θ的极大似然估计量，即是求似然函数

$$L(\theta) = L(x_1, x_2, \cdots, x_n; \theta)$$

的**最大值点**θ.

极大似然估计量的求解步骤：

常规方法

第一步：写出似然函数 $L(\theta)$，取对数 $\ln L(\theta)$；

第二步：令 $\dfrac{d\ln L(\theta)}{d\theta}=0$（或 $\dfrac{\partial \ln L(\theta_1,\theta_2)}{\partial\theta_i}=0$， $i=1,2$），

如果从中解得惟一驻点 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$，（或

$\hat{\theta}_1=\hat{\theta}_1(X_1,X_2,\cdots,X_n)$，$\hat{\theta}_2=\hat{\theta}_2(X_1,X_2,\cdots,X_n)$），则 $\hat{\theta}$ 即为 θ 的极大似然估计；

非常规方法

第三步：如果上述方程无解，则通过 $L(\theta)$ 单调性的讨论，在某边界点处求出 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 的极大似然估计量.

例 5 设总体 $X \sim P(\lambda)$，$(X_1, X_2, \cdots, X_n)$ 为来自总体 X 的样本，试求未知参数 λ 的极大似然估计量 $\hat{\lambda}$.

解 似然函数为

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

所以

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

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

令 $\dfrac{d \ln L(\lambda)}{d\lambda} = 0$，解得 $\hat{\lambda} = \dfrac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}$.

最后一步改变

例 6 设总体 $X\sim N(\mu,\sigma^2)$，$(X_1,X_2,\cdots,X_n)$ 为来自总体 X 的样本.

(1) 如果 σ^2 已知，μ 未知，求 μ 的极大似然估计量 μ；

(2) 如果 μ 已知，σ^2 未知，求 σ^2 的极大似然估计量 σ^2；

(3) 如果 μ,σ^2 均未知，求 μ 和 σ^2 的极大似然估计量 μ 和 σ^2.

【提示】 正态分布 $N(\mu,\sigma^2)$ 的密度函数为

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}},\ -\infty<x<+\infty.$$

解 （1）由于σ^2已知，μ未知，似然函数为

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

所以

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

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

令$\dfrac{d\ln L(\mu)}{d\mu}=0$，解得$\mu=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$．

（续解）（2）由于μ已知，σ^2未知，似然函数为

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

所以

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

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

令$\dfrac{d\ln L(\sigma^2)}{d(\sigma^2)}=0$，解得$\sigma^2=\dfrac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$.

（续解）(3) 由于μ，σ^2均未知，属$k=2$情形，故似然函数为

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

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

令 $$\frac{\partial\ln L(\mu,\sigma^2)}{\partial\mu}=\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu)=\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i-n\mu\right)=0,$$

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

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

例 7 设总体 X 的分布律为

$$X \sim \begin{pmatrix} 0 & 1 & 2 & 3 \\ \theta^2 & 2\theta(1-\theta) & \theta^2 & 1-2\theta \end{pmatrix},$$

其中 $\theta\ (0<\theta<\dfrac{1}{2})$ 是未知参数，利用总体 X 的样本值

(3, 1, 3, 0, 3, 1, 2, 3)，求 θ 的极大似然估计值 $\hat{\theta}$.

解 由于样本值为 (3, 1, 3, 0, 3, 1, 2, 3)，故似然函数为

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

$$=P\{X=0\}(P\{X=1\})^2 P\{X=2\}(P\{X=3\})^4$$

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

（续解） 所以

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

$$\frac{d\ln L(\theta)}{d\theta}=\frac{6}{\theta}-\frac{2}{1-\theta}-\frac{8}{1-2\theta},$$

令 $\frac{d\ln L(\theta)}{d\theta}=0$，解得 $\theta_1=\frac{7-\sqrt{13}}{12},\theta_2=\frac{7+\sqrt{13}}{12}$.

因为 $\theta_2=\frac{7+\sqrt{13}}{12}>\frac{1}{2}$ 不合题意，所以 θ 的极大似然估计

值为 $\hat{\theta}=\frac{7-\sqrt{13}}{12}$.

例 8 设总体 X 的密度函数为 $f(x;\theta)=\begin{cases} e^{-(x-\theta)}, & x\geq\theta, \\ 0, & x<\theta, \end{cases}$ 其中 θ 为未知参数. 从总体 X 中取得样本 $(X_1,X_2,\cdots,X_n)$，求 θ 的极大似然估计量 $\hat{\theta}$.

解 由于求 θ 的极大似然估计量 $\hat{\theta}$ 就是求似然函数

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

的最大值点. 而当 $L(\theta)=0$ 时，$L(\theta)$ 取得最小值，不合题意. 因此， $L(\theta)=0$ 的情况可以不予考虑，故似然函数为

$$L(\theta)=\prod^{n} e^{-(x_i-\theta)}=e^{n\theta-\sum\limits_{i=1}^{n}x_i}.$$

（续解） 因为 $\dfrac{d\ln L(\theta)}{d\theta}=n>0$，所以 $L(\theta)$ 为 θ 单调增加函数，此时 θ 的极大似然估计量必在 θ 取值范围的右端点处取得. 又由于 $\theta\le x_i, i=1,2,\cdots,n$，故 θ 的取值范围为

$$\theta\le\min\left\{x_1,x_2,\cdots,x_n\right\}.$$

故当 $\theta=\min\left\{x_1,x_2,\cdots,x_n\right\}$ 时，$L(\theta)$ 取得最大值，所以 θ 的极大似然估计量为

$$\hat{\theta}=\min\left\{X_1,X_2,\cdots,X_n\right\}=\min_{1\le i\le n}X_i=X_1^*.$$