第八章 参数估计

参数是刻画总体某方面概率特性的数量指标.

实际问题中，往往是总体分布类型已知，但所含参数未知，从总体抽出一个样本，用某种方法对这个未知参数进行估计就是参数估计.

例如，总体$X \sim N(\mu, \sigma^2)$,

若μ, σ^2**未知, 通过构造样本的函数, 给出它们的估计值或取值范围就是参数估计的内容.**

点估计　　**区间估计**

参数估计的类型

点估计 —— 估计未知参数的值

区间估计 ——

估计未知参数的取值范围，
并使此范围包含未知参数
真值的概率为给定的值.

§8.1　点估计

设X 是总体，其分布函数为F(x, θ),θ是未知参数，

$X_1, X_2, \cdots, X_n$ 是X 的样本。现构造统计量 $\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$ 来估计未知参数θ 。 将抽样后的样本值 $x_1, x_2, \cdots, x_n$ 代入上述统计量而得一具体数值 $\hat{\theta}=\hat{\theta}(x_1, x_2, \cdots, x_n)$ 这个数值就称为θ 的一个估计值，而 $\hat{\theta}=\hat{\theta}(X_1, X_2, \cdots, X_n)$ 称为 θ 的一个估计量（一个随机变量）。这样的估计量以及估计值均称为 θ 的一个**点估计**，相应的估计值（量）称为**点估计值**（点估计量），均简记 $\hat{\theta}$ 。

从上述点估计的定义，似乎样本的任何一个统计量都可以作为未知参数θ的**点估**计量？实际上不行！**统计量** $\hat{\theta} = \hat{\theta}(X_1, X_2, \cdots, X_n)$ **能作为**θ的点估计量，必须要合理，需要满足一定的理论基础。**下面，我们介绍两种常用的合理的点估计法：矩估计法和极大似然估计法。**

一、 矩估计法（用样本矩作为总体矩的估计）

设总体X 含有未知参数θ，则总体的分布函数可记为 $F(x,\theta)$，其中未知参数 $\theta=(\theta_1,\theta_2,\cdots,\theta_k)$ 是k 维未知参数向量，即总体X 含有k 个未知参数。我们假设总体X 的前k 阶原点矩都存在.

（1）建立总体原点矩与未知参数的关系：

$$m_r=E(X^r)=h_r(\theta_1,\theta_2,\cdots,\theta_k),\quad r=1,2,\cdots,k.$$

即

$$\begin{cases} m_1=h_1(\theta_1,\theta_2,\cdots,\theta_k), \\ m_2=h_2(\theta_1,\theta_2,\cdots,\theta_k), \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ m_k=h_k(\theta_1,\theta_2,\cdots,\theta_k), \end{cases}$$

（2）由上述方程组可以解出未知参数 $\theta_1,\theta_2,\cdots,\theta_k$ 即有

$$\begin{cases}\theta_1=g_1(m_1,m_2,\cdots,m_k),\\ \theta_2=g_2(m_1,m_2,\cdots,m_k),\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \theta_k=g_k(m_1,m_2,\cdots,m_k),\end{cases}$$

（3）用样本原点矩**替换**相应的总体原点矩，从而得到点估计量

$$\begin{cases}\hat{\theta}_1=g_1(A_1,A_2,\cdots,A_k)=\hat{\theta}_1(X_1,X_2,\cdots,X_n),\\ \hat{\theta}_2=g_2(A_1,A_2,\cdots,A_k)=\hat{\theta}_2(X_1,X_2,\cdots,X_n),\\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\ \hat{\theta}_k=g_k(A_1,A_2,\cdots,A_k)=\hat{\theta}_k(X_1,X_2,\cdots,X_n),\end{cases}$$

则称 $\hat{\theta}_i$ 为 θ_i 的矩估计量，这种参数估计法就是**矩估计法**。

矩估计理论基础：

矩估计法是合理的, 是因为：

（1）由大数定理知：

$$A_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r \xrightarrow{P} m_r = E(A_r)(r = 1, 2, \cdots, k).$$

（2）更一般地，样本矩的连续函数依概率收敛到对应的总体矩的连续函数。设g是连续函数，则

$$g(A_1, A_2, \cdots, A_r) \xrightarrow{P} g(m_1, m_2, \cdots, m_r).$$

$$g(B_1, B_2, \cdots, B_r) \xrightarrow{P} g(\mu_1, \mu_2, \cdots, \mu_r).$$

例、已知总体X 的密度函数为

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

其中 c 为已知正常数，θ>1 是未知参数。设$X_1, X_2, \cdots, X_n$
总体X 的一个样本。

（1） 求θ的矩估计量 $\hat{\theta}$

（2）若有样本$8.9, 9.2, 7.5, 8.0, 8.8, 7.9, 8.4, 8.6, 9.0, 8.7$
且$c=3.5$, 求θ 的矩估计值 $\hat{\theta}$

解: (1)因总体仅含有一个未知参数，故考虑用总体的1阶原点矩

$$m_1 = E(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{c}^{+\infty} x\theta c^{\theta} x^{-\theta-1}dx = \frac{c\theta}{\theta-1},$$

解出θ得到 $$\theta = \frac{m_1}{m_1 - c}$$

用样本的1阶原点矩A_1 替换总体的1阶原点矩 m_1，得到θ的矩估计量：

$$\hat{\theta} = \frac{A_1}{A_1 - c} = \frac{\overline{X}}{\overline{X} - c}$$

(2) 由样本观测值得样本均值为 $\bar{x} = \frac{1}{10}\sum_{i=1}^{10} x_i = 8.5,$

则$\hat{\theta} = \frac{\bar{x}}{\bar{x} - c} = \frac{8.5}{8.5 - 3.5} = 1.7,$**即** $\hat{\theta} = 1.7$ **为θ的矩估计值**

定理(例8.2) 设X 是任意总体，数学期望 $E(X)=u$ 和方差 $D(X)=\sigma^2$ 均存在，其中 u,σ^2 为未知参数，则 u,σ^2 的矩估计量分别为 $\hat{u}=\overline{X}$，$\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=B_2$.

证明： 由于总体含有两个未知参数，故考虑总体的1，2阶原点矩

$$m_1=E(X)=u,\quad m_2=E(X^2)=D(X)+(E(X))^2=\sigma^2+u^2,$$

解出未知参数 $\begin{cases} u=m_1 \\ \sigma^2=m_2-u^2 \end{cases}$ 样本原点矩替换相应的总体矩得到矩估计量

$$\begin{cases} \hat{u}=\overline{X}, \\ \hat{\sigma}^2=A_2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2=B_2. \end{cases}$$

例：设总体 $X \sim \Gamma(\alpha,\beta)$，其中 α,β 未知，$X_1, X_2, \cdots, X_n$ 为X 的样本，求 α,β 的矩估计量。

解：因 $m_1 = E(X) = \dfrac{\alpha}{\beta}, \mu_2 = D(X) = \dfrac{\alpha}{\beta^2}$，解得

$$\begin{cases} \alpha = \dfrac{[E(X)]^2}{D(X)} = \dfrac{m_1^{\ 2}}{\mu_2} \\ \beta = \dfrac{E(X)}{D(X)} = \dfrac{m_1}{\mu_2} \end{cases} \quad \text{替换} \quad \begin{cases} \hat{\alpha} = \dfrac{\bar{X}^2}{B_2} \\ \hat{\beta} = \dfrac{\bar{X}}{B_2} \end{cases}$$

结论：若 $\hat{\theta}$ 为 θ 的矩估计量，$g(x)$ 为连续函数，则 $g(\hat{\theta})$ 为 $g(\theta)$ 的矩估计。

结论： 若$\hat{\theta}$为θ 的矩估计量，$g(x)$为连续函数，则$g(\hat{\theta})$为$g(\theta)$的矩估计。

练习： 设总体$X \sim P(\lambda)$，λ未知，用矩估计法估计$h(\lambda)=3\lambda^2+1$。

解：E(X)= λ 由前面定理知，$\hat{\lambda}=\bar{X}$，而$h(\lambda)$是λ的连续函数，则$h(\lambda)$的矩估计为 $\hat{h}(\lambda)=h(\hat{\lambda})=3\hat{\lambda}^2+1=3\bar{X}^2+1.$

此外，$3B_2^{\ 2}+1$ 也是 $h(\lambda)=h(\lambda)=3\lambda^2+1$ 矩估计量。

这是因为 $D(X)=\lambda$，而 B_2 是方差的矩估计量，从而 $3B_2^{\ 2}+1$ 也是 $h(\lambda)=h(\lambda)=3\lambda^2+1$ 矩估计量。

说明: （1）矩估计法的基本原理：用样本矩作为总体矩的估计；（2）矩估计（值或量）不一定唯一。例如 Poission 分布的参数λ，既是总体均值，又是总体方差。按照矩估计方法，既可以用 $\bar{x}$ 来估计，也可以用 $b_2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$ 来估计，遇到这种情形，总是选低阶矩的估计。

二、 极大似然估计法

一个随机试验有多种可能结果，一次抽样中，结果A发生了，则我们认为A是全部可能结果中发生概率最大的那一个。

引例：一袋小球分为黑白两种，形状一样，现要估计白球所占的百分比p. 为此有放回的抽取10次，每次一个小球。 引入随机变量 X_i,

$$X_i=\begin{cases}1,\text{第i次摸得白球},\\0,\text{第i次摸得黑球}。\end{cases}\quad \text{则}X_i\sim B(1,p), i=1,2,\cdots,10.$$

试验的结果是(1, 0, 1, 0, 0, 0,1, 0, 0, 0)

这表明此次抽样中事件

$$A=\begin{Bmatrix} X_1=1, X_2=0, X_3=1, X_4=0, X_5=0, \\ X_6=0, X_7=1, X_8=0, X_9=0, X_{10}=0 \end{Bmatrix}$$

发生了。考察A发生的概率 $P(A)=p^3(1-p)^7.$

A事件发生的概率与p有关，可看作p的函数：

$$L(p)=P(A)=p^3(1-p)^7.$$

抽样的结果有很多种可能，为何在一次抽样中A就出现了呢？我们认为是试验的条件（对应于p的值）使得结果A出现的概率最大！利用微积分的知识，可解得当p=0.3时，P(A)达到最大概率。从而可估计白球所占比例为30%.

总结上面的想法就是： 一个随机事件有若干个可能的结果，若在一次抽样中某一结果出现了 ，我们自然地认为， 该结果是若干个可能结果中发生概率最大的一个， 因此参数的取值应使得已发生的事件概率最大。这种想法就是<u>极大似然估计法</u>的基本思想。

1、当总体X离散且分布律为p(X=x) =p(x,θ) 时，其中θ是未知参数，且样本 $X_1, X_2, \ldots, X_n$ 的观察值为$x_1, x_2, \ldots, x_n$时，则事件$\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$发生的概率为

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

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

由极大似然估计法的思想，θ的极大似然估计值 $\hat\theta$ 应使得L(θ)（即已发生事件$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$ 发生的概率）达到最大，即 $\hat\theta$ 应是

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

的解。L(θ) 称为似然函数。

2、当总体X为连续型且密度函数为f(x, θ)时，其中θ是未知参数，且样本$X_1, X_2, \dots, X_n$的观察值为x_1，x_2，…，x_n 时，因X_i 落在x_i的邻域(设其长度为 Δx_i)内的近似概率为 $f(x_i,\theta)\Delta x_i, i=1,\dots,n$，则每个Xi 都落在xi的邻域(设其长度为 Δx_i)内发生的概率近似为

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

按照极大似然估计法的思想，θ的估计值 $\hat{\theta}$ 应使得

$$\prod_{i=1}^{n} f(x_i,\hat{\theta})\Delta x_i = \max_{\theta}\prod_{i=1}^{n} f(x_i,\theta)\Delta x_i \qquad (**)$$

由于 Δx_i 与θ无关，故

$$\prod_{i=1}^{n} f(x_i,\hat{\theta})\Delta x_i = \left(\prod_{i=1}^{n} f(x_i,\hat{\theta})\right)\left(\prod_{i=1}^{n}\Delta x_i\right)$$

于是 (**) 等价于 $\prod_{i=1}^{n} f(x_i,\hat{\theta}) = \max_{\theta}\prod_{i=1}^{n} f(x_i,\theta)$

让 $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$

则θ 的估计值 $\hat{\theta}$ 应满足 $L(\hat{\theta}) = \max_{\theta} L(\theta)$ 我们

称 $L(\theta)$ 为似然函数。

综上所述，我们有

定义：设X 为总体，θ为参数，且总体的分布律或密度已知，$x_1,x_2,\cdots,x_n$为一组样本观测值。若存在 θ 的一个值$\hat{\theta}(x_1,x_2,\cdots,x_n)$使得$L(\hat{\theta})=\max_{\theta} L(\theta)$时，则称$\hat{\theta}(x_1,x_2,\cdots,x_n)$为 θ 的极大似然估计值，$\hat{\theta}(X_1,X_2,\cdots,X_n)$为$\theta$的极大似然估计量。

按定义，极大似然估计值可以由方程 $\frac{dL(\theta)}{d\theta}=0$解得 $\theta=\hat{\theta}$，由于$\ln L(\theta)$和$L(\theta)$在同一处取得极值点，故上式的求解可通过求解

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

综上所述，求θ 的极大似然估计值的一般步骤：

（1）写出似然函数

$$L(\theta)=\prod_{i=1}^{n} p(x_i,\theta) \qquad (or \qquad \prod_{i=1}^{n} f(x_i,\theta))$$

（2）求出 $\ln L(\theta)$

（3）令 $\dfrac{d\ln L(\theta)}{d\theta}=0$**，求** $\hat{\theta}$ **。**

例8.6 设总体 $X \sim B(m,p)$，$x_1, x_2, \cdots, x_n$为样本观测值，其中m 已知而 p未知。求p 的极大似然估计值和极大似然估计量。

解：总体X的分布律为

$$p(x,p) = C_m^x p^x (1-p)^{m-x}, \quad x = 0,1,2,\cdots,m$$

从而，似然函数为

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

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

则 $\ln L(p)=\sum_{i=1}^{n}x_i\ln p+(mn-\sum_{i=1}^{n}x_i)\ln(1-p)+\ln(\prod_{i=1}^{n}C_m^{x_i})$

令 $\frac{d\ln L(p)}{dp}=\frac{1}{p}\sum_{i=1}^{n}x_i-\frac{1}{1-p}(mn-\sum_{i=1}^{n}x_i)=0$，解得

p 的极大似然估计值为 $\hat{p}=\frac{\sum_{i=1}^{n}x_i}{mn}=\frac{\bar{x}}{m}$，相应的，$p$的极大似然估计量为 $\hat{p}=\frac{\bar{X}}{m}$。

例题：设总体 $X \sim e(\theta)$, $x_1, x_2, \cdots, x_n$ 为样本观测值

求 θ 的极大似然估计。

解：总体X 密度函数为 $f(x,\theta)=\begin{cases}\theta e^{-\theta x}, & x>0,\\ 0, & x\le 0,\end{cases}$

似然函数为 $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\prod_{i=1}^{n}\theta e^{-\theta x_i}=\theta^n e^{-\theta\sum_{i=1}^{n}x_i}$

于是，$\ln L(\theta)=n\ln\theta-\theta\sum_{i=1}^{n}x_i$, 令 $\dfrac{d\ln L(\theta)}{d\theta}=\dfrac{n}{\theta}-\sum_{i=1}^{n}x_i=0$,

解得 θ 的极大似然估计值为 $\hat{\theta}=\dfrac{n}{\sum_{i=1}^{n}x_i}=\dfrac{1}{\bar{x}}$，相应的，$\theta$ 的极大似然估计量为 $\hat{\theta}=\dfrac{1}{\bar{X}}$。

思考题：设总体 $X \sim e(\lambda)$， $x_1, x_2, \ldots, x_n$ 为样本观测值，求 λ 的矩估计。

结论： $\eta = g(x)$ 具有单值反函数，$\hat{\theta}$ 为 θ 的极大似然估计，则 $\hat{\eta} = g(\hat{\theta})$ 为 $\eta = g(\theta)$ 的极大似然估计。

如在上例中，$\theta > 0$，因而 $\eta = \frac{1}{\theta}$ 的极大似然估计量为 $\hat{\eta} = \frac{1}{\hat{\theta}} = \bar{X}$。

在前面的讨论中，所 涉及未知参数均只有一个，如果未知参数为多个，即 $\theta=(\theta_1,\theta_2,\cdots,\theta_n)$时，似然函数为 $L(\theta_1,\theta_2,\cdots,\theta_n)$，则$\theta=(\theta_1,\theta_2,\cdots,\theta_n)$ 的极大似然估计值 $\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_n)$ 应由如下方程组求得。

$$\begin{cases} \dfrac{\partial \ln L(\theta_1,\theta_2,\cdots,\theta_n)}{\partial\theta_1}=0 \\ \dfrac{\partial \ln L(\theta_1,\theta_2,\cdots,\theta_n)}{\partial\theta_2}=0 \\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots\ \cdots \\ \dfrac{\partial \ln L(\theta_1,\theta_2,\cdots,\theta_n)}{\partial\theta_n}=0 \end{cases}$$

例8.7： 设总体$X \sim N(u,\sigma^2)$，其中u,σ^2未知，$x_1, x_2, \cdots, x_n$为X 的一组样本观测值，求u,σ^2的极大似然估计量。

解： 似然函数为

$$L(u,\sigma^2) = \prod_{i=1}^{n} f(x_i; u,\sigma^2)$$

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

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

解方程组

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

得 u,σ^2 的

极大似然估计值为

$$\hat{u} = \bar{x},\quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2;$$

相应的，u,σ^2 的极大似然估计量为

$$\hat{u} = \bar{X},\quad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = B_2.$$

由此可见，当总体 $X \sim N(\mu,\sigma^2)$ 时，μ,σ^2 的矩估计量和极大似然估计量是相同的，均为

$$\hat{\mu} = \overline{X},\ \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 = B_2.$$

一般来说，未知参数的矩估计和极大似然估计可以不一样。

需要注意的是，当 $\frac{dL(\theta)}{d\theta}$ 无解时，极大似然估计值应该怎样取呢?

例8.8：设总体 $X \sim U(0,\theta)$，未知 $\theta > 0$，$x_1, x_2, \cdots, x_n$ 为样本观测值，求 θ 的极大似然估计。

解：似然函数为 $L(\theta) = \prod_{i=1}^{n} f(x_i, \theta) = \begin{cases} \frac{1}{\theta^n}, & 0 < x_i < \theta, \\ 0, & else. \end{cases}$

由于 $\frac{dL(\theta)}{d\theta} = -n\theta^{-(n-1)} = 0$ 在 $\theta > 0$ 时无解，所以无法由该方程解出 θ的极大似然估计值 $\hat{\theta}$。

但另一方面，当 $\theta > 0$ 时

$$\frac{dL(\theta)}{d\theta} = -n\theta^{-(n-1)} < 0,$$

故在$\theta > 0$ 时$L(\theta)$单调递减；同时每个 x_i应满足$0 \le x_i \le \theta$,

故 $0 \le \max_i x_i \le \theta$, 所以 时 $\hat{\theta} = \max_i x_i$

$$L(\hat{\theta}) = \max_{\theta} L(\theta),$$

即θ 的极大似然估计值为 $\hat{\theta} = \max_i x_i$;

θ 的极大似然估计量为 $\hat{\theta} = \max_i X_i$。

可以看到，和极大似然估计法相比，矩估计的好处在于：无需知道总体的分布，就可以求出估计：不论总体是什么，对 k 阶原点矩的估计量总是 $\frac{1}{n}\sum_{i=1}^{n} x_i^{\,k}$，总是用样本均值 $\bar{x}$ 估计总体均值，用样本的二阶中心矩 b_2 估计总体方差。但这也是它的缺点，即当知道总体分布的类型时，矩估计法和极大似然估计法相比，矩估计法没能利用分布类型提供的信息，其估计方法可能会不如极大似然估计法好。