第四章：参数估计

中国科学技术大学

第四章：参数估计

参数估计问题:

- 总体: $X \sim f_\theta(x)$,f 形式已知, $\theta = (\theta_1, \dots, \theta_k)$ 为未知参数
- 样本: $X_1, \dots, X_n$

利用样本对参数 θ 的作出估计或估计它们的某个已知函数 $g(\theta)$.

- **点估计**: 用样本的一个函数 $T(X_1, \dots, X_n)$ 去估计 $g(\theta)$
- **区间估计**: 用一个区间 (区域) 去估计 $g(\theta)$

4.4 点估计

根据样本 $X_1, \cdots, X_n$ 来估计参数 θ, 就是要构造适当的统计量 $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$. 当有了样本 $X_1, \cdots, X_n$ 的值后, 就代入 $\hat{\theta} = \hat{\theta}(X_1, \cdots, X_n)$ 中算出一个值, 用来作为 θ 的估计值. 为这样特定目的而构造的统计量 $\hat{\theta}$ 叫做 θ 的**估计量**. 由于参数 θ 是数轴上的一个点, 用 $\hat{\theta}$ 估计 θ, 等于用一个点去估计另一个点, 所以这样的估计叫做**点估计**.

求点估计的方法有多种, 下面介绍两种点估计方法:

4.4.1 矩估计方法

矩方法追溯到 19 世纪的**Karl Pearson**. 矩方法是基于一种简单的 "替换" 思想建立起来的一种估计方法. 其基本思想是用样本矩估计总体矩. 由大数律, 如果未知参数和总体的某个 (些) 矩有关系, 我们很自然的来构造未知参数的估计。

回忆一下以前关于矩的记法:

$$\text{样本}k\text{阶矩:}\quad a_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \qquad m_k = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^k$$

$$\text{总体}k\text{阶矩:}\quad \alpha_k = EX^k \qquad \mu_k = E(X - EX)^k$$

因此在 k 阶矩存在的情况下，根据大数律有

$$a_k \xrightarrow{p} \alpha_k, \quad m_k \xrightarrow{p} \mu_k$$

从而我们可以使用 a_k, m_k 分别估计 α_k, μ_k, 进而得到 θ 的估计. 介绍如下: 假设总体 X 包含 k 个未知参数 $\theta = (\theta_1, \cdots, \theta_k)$, 由方程组

$$\begin{cases} \alpha_1 = f_1(\theta_1, \cdots, \theta_k) \\ \qquad\vdots \\ \alpha_k = f_k(\theta_1, \cdots, \theta_k) \end{cases}$$

反解得到

$$\begin{cases} \theta_1 = g_1(\alpha_1, \cdots, \alpha_k) \\ \quad\quad\quad \vdots \\ \theta_k = g_k(\alpha_1, \cdots, \alpha_k) \end{cases}$$

将其中的总体矩用相应的样本矩代替,则我们可以得到参数 $\theta_1, \cdots, \theta_k$ 的一个估计:

$$\begin{cases} \hat{\theta}_1 = g_1(a_1, \cdots, a_k) \\ \quad\quad\quad \vdots \\ \hat{\theta}_k = g_k(a_1, \cdots, a_k) \end{cases}$$

若要估计参数 $\theta_1, \cdots, \theta_k$ 的某函数 $g(\theta_1, \cdots, \theta_k)$, 则用 $g(\hat{\theta}_1, \cdots, \hat{\theta}_k)$ 去估计它.

这里我们用的都是原点矩 α_k，当然也可以使用中心矩 μ_k，或者两个都使用。在这种情况下，只需要把相应的总体矩换成样本矩。我们称这种估计方法为矩估计法，得到的估计量称为矩估计量。**矩估计方法应用的原则是：能用低阶矩处理的就不用高阶矩。**

矩估计法的优点是简单易行, 有些情况下不需要事先知道总体是什么分布. 缺点是，当总体类型已知时，没有充分利用分布提供的信息. 一般场合下, 矩估计量不具有唯一性.

投掷一枚硬币, 为了解正面出现的概率, 现独立重复的投掷 n 次, 用 $X_1, \cdots, X_n$ 表示投掷结果. 显然此时总体 X 的分布为 $B(1, p)$, p 为感兴趣的量. 而 $X_1, \cdots, X_n$ 为样本, 则求参数 p 的矩估计量。

↑Example
↓Example

解: 由于 $EX = p$, 而样本均值 $\bar{X}$ 收敛到总体均值 EX, 因此 p 的一个矩估计量为 $\hat{p} = \bar{X}$.

为考察某种考试成绩分布情况, 使用正态分布 $N(a,\sigma^2)$ 来作为总体 X 的分布. 现在从中随机调查 n 个人, 即样本为 $X_1,\cdots,X_n$. 试求参数 a,σ^2 的矩估计量。

↑Example
↓Example

解: 由于

$$EX=a,\quad Var(X)=\sigma^2$$

所以 a,σ^2 的一个矩估计量为

$$\hat{a}=\bar{X},\quad \hat{\sigma}^2=m_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$$

我们知道 $ES^2=\sigma^2$，因此，σ^2 的另一个矩估计量为 $\hat{\sigma}^2=S^2$.

设 $(X_1, X_2, \cdots, X_n)$ 是从均匀分布总体 $X \sim U(\theta_1, \theta_2)$ 中抽取的一个样本, 求 $\hat{\theta}_1, \hat{\theta}_2$

↑Example

↓Example

解: 由于 $EX = (\theta_1 + \theta_2)/2, \mu_2 = \mathrm{Var}(X) = (\theta_2 - \theta_1)^2/12$, 用 $\bar{X}$ 代替 EX, S^2 代替 μ_2, 解方程组, 得

$$\hat{\theta}_1 = \bar{X} - \sqrt{3}S, \quad \hat{\theta}_2 = \bar{X} + \sqrt{3}S$$

设 $(X_1, X_2, \cdots, X_n)$ 是从总体 $X \sim F$ 中抽取的一个样本, 求偏度系数 $\beta_1 = \mu_3/\mu_2^{3/2}$, 峰度系数 $\beta_2 = \mu_4/\mu_2^2$ 的矩估计。

↑Example
↓Example

解: 由于偏度系数和峰度系数都是总体中心矩的函数, 所以可以用样本中心矩代替总体中心矩得到, 即

$$\hat{\beta}_1 = m_3/m_2^{3/2}, \quad \hat{\beta}_2 = m_4/m_2^2$$

设 $(X_1, X_2, \cdots, X_n)$ 是从正态总体 $X \sim N(\mu, \sigma^2)$ 中抽取的一个样本, 求概率 $P(X > 3)$ 的矩估计。

↑Example
↓Example

解: 由于

$$P(X > 3) = 1 - \Phi\left(\frac{3-\mu}{\sigma}\right)$$

这是 μ, σ^2 的函数, 用 $\bar{X}$ 代替 μ, S 代替 σ, 得

$$\hat{P}(X > 3) = 1 - \Phi\left(\frac{3-\bar{X}}{S}\right)$$

4.4.2 最大 (极大) 似然估计方法

最大似然方法到目前为止应用最广的的点估计方法. 这种方法是基于如下的看法:

设样本 $X = (X_1, \ldots, X_n)$ 有概率函数

$$f(x;\theta) = f(x;\theta_1, \cdots, \theta_k)$$

这里参数 $\theta = (\theta_1, \cdots, \theta_k) \in \Theta$，$x = (x_1, \ldots, x_n)$ 为样本 X 的观察值. 当固定 x 时把 $f(x;\theta)$ 看成为 θ 的函数, 称为似然函数，常记为 $L(x;\theta)$ 或 $L(\theta)$.

Definition

当固定参数 θ 时，$f(x;\theta)$ 可以看成是得到样本观察值 x 的可能性，这样，当把参数 θ 看成变动时，也就得到 "**在不同的 θ 值下能观察到 x 的可能性大小, 即 $L(x;\theta)$**"; **由于我们已经观察到了 x，所以**

使得能观察到 x 的可能性 $L(x;\theta)$ 最大的 θ 值, 看起来应该最像未知的 θ。这个 θ 的值即称为 θ **最大似然估计值**(看上去最有可能的)。我们先看一个例子:

从鱼池里随机捕捞 500 条鱼, 做好记号后重新放入鱼池中, 待充分混合后再捕捞 1000 条鱼, 结果发现其中有 72 条带有记号. 试问鱼池中可能有多少条鱼.

↑Example
↓Example

解: 先将问题一般化. 设池中有 N 条鱼, 其中 r 条做好记号. 鱼在鱼池里均匀. 随机捕捞 s 条, 发现 x 条有记号. 用上述信息来估计 N.

用 X 表示捕捞的 s 条鱼中带记号鱼的数目, 则

$$P(X=x)=\frac{C_{N-r}^{s-x}C_r^x}{C_N^s}.$$

目前发现在捕捞的 s 条鱼中有记号的鱼 x 条, 要寻求 N 取何值时, 使得观察到这个事件 $\{X = x\}$ 的可能性最大. 即 x 是固定的, N 是变化的, 记 $p(x; N) = P(X = x)$. 因为

$$g(N) := \frac{p(x;N)}{p(x;N-1)} = \frac{(N-s)(N-r)}{N(N-r-s+x)} = \frac{N^2 - N(s+r) + rs}{N^2 - N(r+s) + Nx},$$

当 $rs > Nx$ 时, $g(N) > 1$; $rs < Nx$ 时, $g(N) < 1$. 所以 $P(X = x)$ 在 $N = \frac{rs}{x}$ 附近达到最大, 注意到 N 只能取正整数, 故 N 的最可能的估计即最大似然估计为

$$\hat{N} = \left\lceil \frac{rs}{x} \right\rceil.$$

其中 $\lceil \ \rceil$ 表示下取整, 即小于该值的最大整数. 将题目中的数字代入,

$$\hat{N} = \left\lceil \frac{500 \times 1000}{72} \right\rceil = 6944.$$

即鱼池中的总的鱼数为 6694 条.

现给出最大似然估计的一般性定义:

设 $X = (X_1, \cdots, X_n)$ 为从具有概率函数 $f_\theta(x)$ 的总体中抽取的样本，θ 为未知参数或者参数向量. $x = (x_1, \cdots, x_n)$ 为样本的观察值。若在给定 x 时，值 $\hat{\theta} = \hat{\theta}(x)$ 满足下式

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

Definition

则称 $\hat{\theta}$ 为参数 θ 的最大似然估计值，而 $\hat{\theta}(X)$ 称为参数 θ 的最大似然估计量。若待估参数为 θ 的函数 $g(\theta)$, 则 $g(\theta)$ 的最大似然估计量为 $g(\hat{\theta})$。

求最大似然估计值相当于求似然函数的最大值。在简单样本的情况下,

$$L(x;\theta)=\prod_{i=1}^{n} f_\theta(x_i)$$

而把似然函数的对数 $l(\theta)=\log L(\theta)$ 称为对数似然函数 (这是由于在一些情况下，处理对数似然函数更方便)

当似然函数对变量 θ 单调时, 我们可以容易得到其最大值点. 反之当似然函数为非单调函数且对变量 θ 可微分时, 我们可以求其驻点: 令

$$\frac{dl(\theta)}{d\theta}=0 \quad (\text{或者}\frac{dL(\theta)}{d\theta}=0)$$

当 θ 为多维时, 比如 $\theta=(\theta_1,\cdots,\theta_k)$ 时令

$$\frac{\partial l(\theta)}{\partial \theta_i}=0 \quad (\text{或者}\frac{\partial L(\theta)}{\partial \theta_i}=0)\ i=1,\cdots,k$$

然后判断此驻点是否是最大值点。

设 $X_1, \cdots, X_n$ 为从总体 $X \sim N(a, \sigma^2)$ 中抽取的样本，求参数 a, σ^2 的最大似然估计量。

↑Example
↓Example

解: 易得对数似然函数为

$$l(a, \sigma^2) = c - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - a)^2 - \frac{n}{2} log(\sigma^2)$$

其中 c 是与参数无关的常数. 令

$$\begin{cases} \frac{\partial l(a, \sigma^2)}{\partial a} = 0 \\ \frac{\partial l(a, \sigma^2)}{\partial \sigma^2} = 0 \end{cases} \Rightarrow \begin{cases} a = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \\ \sigma^2 = \frac{1}{n} \sum\limits_{i=1}^{n} (x_i - a)^2 \end{cases}$$

容易验证此驻点是唯一的最大值点，因此得到 a, σ^2 的最大似然估计量:

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

设总体 X 服从 $[a,b]$ 上的均匀分布, $a<b$, 求参数 a,b 的最大似然估计.

↑Example
↓Example

解: 易得似然函数为

$$L(a,b)=\frac{1}{(b-a)^n}\prod_{j=1}^{n}I(a\le x_j\le b)=\frac{1}{(b-a)^n}I(a\le x_{(1)}\le x_{(n)}\le b).$$

于是对任何满足条件 $a\le x_j\le b$ 的 a,b 都有

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

即似然函数 $L(a,b)$ 在 $a=x_{(1)},b=x_{(n)}$ 时取到最大值. 于是 a,b 的最大似然估计量为 $\hat{a}=X_{(1)},\hat{b}=X_{(n)}$.

设 $X_1, \cdots, X_n$ 为从具有如下形式密度的总体中抽取的样本:

$$f(x;a,b) = \begin{cases} \frac{1}{b}\exp\{-\frac{x-a}{b}\} & ,x > a \\ 0 & ,x \le a \end{cases}$$

求参数 a, b 的最大似然估计量.

↑Example
↓Example

解: 易得似然函数为

$$L(a,b)=\prod_{i=1}^{n}f(x_i;a,b)=\frac{1}{b^n}\exp\{-\frac{1}{b}\sum_{i=1}^{n}(x_i-a)\}I(x_{(1)}>a)$$

在固定 b 时，显然似然函数为 a 的单调增函数，因此 $L(a)$ 的驻点为 $\hat{a}=x_{(1)}$。再令 $\frac{\partial L(a,b)}{\partial b}=0$，得到 $b=\frac{1}{n}\sum\limits_{i=1}^{n}(x_i-x_{(1)})$，容易验证此解是最大值点。从而得到 a,b 的最大似然估计量:

$$\begin{cases}\hat{a}=X_{(1)}\\ \hat{b}=\frac{1}{n}\sum\limits_{i=1}^{n}(X_i-X_{(1)}).\end{cases}$$

设总体 $X_1, \ldots, X_n$ 服从 0-1 分布 $B(1,p)$, $0 < p < 1$, 求参数 p 的最大似然估计.

↑Example
↓Example

解: 似然函数为

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

从而令 $\frac{\partial log L(p)}{\partial p} = 0$ 得到

$$\frac{\sum x_i}{p} = \frac{n - \sum x_i}{1-p}$$

因此 p 的似然估计为

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

设总体 $X_1, \ldots, X_n$ 服从柯西分布 $f(x) = \frac{1}{\pi}\frac{1}{1+(x-\theta)^2}, x \in R, \theta \in R$, 求参数 θ 的最大似然估计.

↑Example
↓Example

解: 因为柯西分布不存在矩, 因此矩方法不适用. 其对数似然函数为

$$l(\theta) = logL(\theta) = \sum_{i=1}^{n} logf(x_i) = \sum_{i=1}^{n}[-log\pi - log(1 + (x_i - \theta)^2)$$

从而令 $\frac{\partial l(\theta)}{\partial \theta} = 0$ 得到

$$\sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2} = 0$$

此方程没有显式解, 可以使用数值方法求解. 使用起来不太方便, 因此在应用中, 考虑到柯西分布的对称性, 使用样本中位数来估计 θ.

↑Example

设 $X_1, \cdots, X_n$ 为从如下分布中抽取的简单样本，求 θ 的最大似然估计.

$$f(x) = \frac{1}{x!(2-x)!}[\theta^x(1-\theta)^{2-x}+\theta^{2-x}(1-\theta)^x], \quad x = 0, 1, 2; \quad \theta \in (0, \frac{1}{2})$$

↓Example

解: 由题设知 $f(x)$ 为离散型，其分布律为

X	0	1	2
P	$\frac{1}{2}[(1-\theta)^2+\theta^2]$	$2\theta(1-\theta)$	$\frac{1}{2}[(1-\theta)^2+\theta^2]$

若直接从此分布出发，则不能得到 θ 的最大似然估计的显式表达。为此，我们重新参数化，记 $\eta = 2\theta(1-\theta)$. 则由题设知 $\eta < 1/2$。则

X	0	1	2
P	$\frac{1}{2}(1-\eta)$	η	$\frac{1}{2}(1-\eta)$

再记 $n_i = \#\{X_1, \cdots, X_n$中等于$i$的个数$\}, \quad i = 0, 1, 2$, 则得到似然函数为

$$L(\eta) = (\frac{1}{2}(1-\eta))^{n_0}\eta^{n_1}(\frac{1}{2}(1-\eta))^{n_2} = (\frac{1}{2}(1-\eta))^{n-n_1}\eta^{n_1}$$

求解并注意 η 的上界即得到 η 的最大似然估计为

$$\hat{\eta} = \min\{\frac{n_1}{n}, \frac{1}{2}\}$$

再由 $\theta = \frac{1-\sqrt{1-2\eta}}{2}$ 得到 θ 的最大似然估计为

$$\hat{\theta} = \frac{1-\sqrt{1-2\hat{\eta}}}{2}$$