基础数理统计 （研究生公共课）

肖柳青 教授 博士主讲

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

第 2 章 参数估计（Parametric Estimation）

主要内容

1.点估计的基本概念
2.两种基本的点估计方法
3.点估计的优良标准（无偏性及有效估计和C-R下界）
4.最佳点估计（充分统计量）
5.置信区间估计
6.贝叶斯估计

参数估计的原理

设总体X 的分布函数的形式已知，但含有一个或多个未知参数：$\theta_1, \theta_2, \ldots, \theta_k$

参数是刻画总体某方面概率特性的数量.

设 $X_1, X_2, \ldots, X_n$为总体的一组样本，如何构造 k 个统计量：

$$\left.\begin{array}{l} \theta_1(X_1, X_2, \cdots, X_n) \\ \theta_2(X_1, X_2, \cdots, X_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \theta_k(X_1, X_2, \cdots, X_n) \end{array}\right\} \text{随机变量！}$$

分布中参数的重要性

参数估计就是当参数未知时，从总体抽出一个子样，用某种方法对这个未知参数进行估计.

例如，$X \sim N(\mu, \sigma^2)$,

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—— 即求置信区间：

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.

1. 点估计的基本概念（Point Estimator）

点估计： 由样本$X_1, X_2, \ldots, X_n$确定一个统计量

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

用它估计总体的未知参数，称为总体参数的估计量。

当具体的样本抽出后，可求得出样本统计量的值。用它作为总体参数的估计值，称作总体参数的点估计值。

2. 几种基本的点估计方法

- 矩估计(Moment Estimate)
- 极大似然估计 (Maximum Likelihood Estimate)
  - 多项分布的极大似然估计
  - 极大似然估计的渐进分布
  - 极大似然估计的置信区间解法
- 最小二乘法
- 贝叶斯方法
- ……

（1） 矩估计法

它是基于一种简单的“替换”思想建立起来的一种估计方法 .

是英国统计学家*K.*皮尔逊最早提出的 .

其基本思想是用样本矩估计总体矩 .

理论依据：辛钦大数定律

矩估计法具体做法：

设X 是一随机变量， $X_1, X_2, \ldots\ldots X_n$ 是它的一个样本。

若 $E\left(X^k\right)$ 存在，则称之为 X 的 k 阶原点矩。记作V_k

若$E\left[\left(X-E\left(X\right)\right)^k\right]$存在，则称之为 X 的 k 阶中心矩。记作 U_k

称 $A_k=\dfrac{1}{n}\sum_{i=1}^{n}X_i^k$ 为样本的 k 阶原点矩。

称 $B_k=\dfrac{1}{n}\sum_{i=1}^{n}\left(X_i-\overline{X}\right)^k$ 为样本的 k 阶中心矩。

$$\boxed{\text{矩法估计：}\ \widehat{V_k}=A_k,\quad \widehat{U_k}=B_k}$$

这是包含k个未知参数$\theta_1, \cdots, \theta_k$的联立方程组，

$$\begin{cases} A_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_k) \\ A_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_k) \\ \cdots\ \cdots\ \cdots\ \cdots \\ A_k = \mu_k(\theta_1, \theta_2, \cdots, \theta_k) \end{cases}$$

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k,$$

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

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ \hat{\theta}_2 = \hat{\theta}_2(X_1, X_2, \cdots, X_n) \\ \cdots\ \cdots\ \cdots\ \cdots \\ \hat{\theta}_k = \hat{\theta}_k(X_1, X_2, \cdots, X_n) \end{cases}$$

矩估计法

例1 设总体X的概率密度为

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

其中 $\alpha>-1$ 是未知参数,

$X_1,X_2,\ldots,X_n$是取自X的样本,求参数α 的矩估计.

解: $\mu_1=E(X)=\int_0^1 x(\alpha+1)x^{\alpha}dx$

数学期望是一阶原点矩

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

由矩法,

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

样本矩

总体矩

从中解得 $\hat{\alpha}=\dfrac{2\bar{X}-1}{1-\bar{X}}$, 即为$\alpha$的矩估计.

矩法的优点是简单易行,并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．

其主要原因在于建立矩法方程时,选取那些总体矩用相应样本矩代替带有一定的随意性．

2. 极大似然法 MLE

是在总体类型已知条件下使用的一种参数估计方法.

它首先是由德国数学家**高斯**在1821年提出的，然而，这个方法常归功于英国统计学家**费歇**.

费歇在1922年重新发现了这一方法，并首先研究了这种方法的一些性质.

极大似然法的原理

先看一个简单例子：

某位同学与一位猎人一起外出打猎．

一只野兔从前方窜过．

只听一声枪响，野兔应声倒下．

如果要你推测，

是谁打中的呢？

你会如何想呢?

你就会想，只发一枪便打中，猎人命中的概率一般大于这位同学命中的概率. 看来这一枪是猎人射中的 .

这个例子所作的推断已经体现了极大似然法的基本思想 .

下面我们再看一个具体例子，进一步体会极大似然法的基本思想 .

例2 设$X\sim B(1,p)$, p未知．假设我们事先知道p只有两种可能:

p=0.7 或 p=0.3

如今重复试验3次,得结果: 0， 0， 0

问:应如何估计p?

由概率论的知识, 3次试验中出现“1”的次数

$$Y \sim B(3, p)$$

$$P(Y=k)=\binom{3}{k}p^k(1-p)^{n-k}$$ k=0,1,2,3

将…结果…如下

估计 估计 出现 出现 出现 出现

p值	$P(Y=0)$	$P(Y=1)$	$P(Y=2)$	$P(Y=3)$
0.7	0.027	0.189	0.441	0.343
0.3	0.343	0.441	0.189	0.027
	0.343	0.441	0.441	0.343

估计 估计

应如何估计p? p=0.7 或 p=0.3

$$P(Y=k)=\binom{3}{k}p^k(1-p)^{3-k}\quad k=0,1,2,3$$

P=0.7，p=0.3
假设如何确认？
多种情况？
不是常数？

定义：设总体 X 的密度为 $f(x;\theta)$ （当 X 为离散型时 $f(x;\theta)$ 为分布列），$\theta=(\theta_1,\theta_2,\cdots,\theta_l)$ 是总体的未知参数，$x_1,x_2,\cdots x_n$ 是样本 $X_1,X_2,\cdots X_n$ 的一组样本值，则称

①

$$L(\theta)=L(x_1x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n} f(x_i;\theta)$$

求导麻烦!!

转向 $\ln L(\theta)$!!

为样本的**似然函数**. 若存在某个 $\hat{\theta}=(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_l)$，使得 ②

$$L(x_1x_2,\cdots,x_n;\hat{\theta})=\max_{\theta\in\Theta} L(x_1x_2,\cdots,x_n;\theta)$$

即选取的估计量 $\hat{\theta}$ 应满足

成立(其中 Θ 为参数空间)，则称

$$\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)=(\hat{\theta}_1(X_1,\cdots,X_n),\hat{\theta}_2(X_1,\cdots,X_n),\cdots,\hat{\theta}_l(X_1,\cdots,X_n)),$$

为 θ 的**极大似然估计量**. 称

$$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)=(\hat{\theta}_1(x_1,\cdots,x_n),\hat{\theta}_2(x_1,\cdots,x_n),\cdots,\hat{\theta}_l(x_1,\cdots,x_n)),$$

为 θ 的**极大似然估计值**. ③

如 何 求 ?

$L(\theta)$ 是 θ 的函数，可用求导的方法找到使 $L(\theta)$ 达到最大值的 θ.

注意到 $\ln L(\theta)$ 为单增函数，故 $L(\theta)$ 与 $\ln L(\theta)$ 达到最大值的自变量相同，故问题可转化为求 $\ln L(\theta)$ 的最大值点.

求极大似然估计的一般步骤是:

① 由总体分布建立似然函数 $L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)$ ——把自变量 x 看成常数，把未知参数 $\theta=(\theta_1,\theta_2,\cdots,\theta_l)$ 看成自变量;

② 求似然函数 $L(\theta)$ 的最大值点—— 转化为求 $\ln L(\theta)$ 的最大值点，即

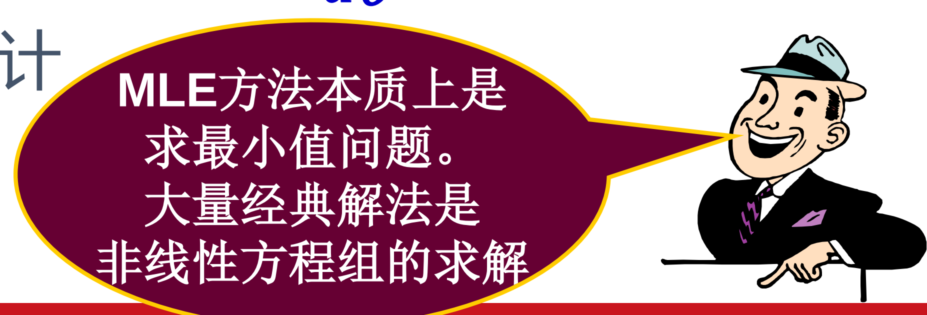
1^0 建立似然方程组: $\dfrac{\partial \ln L(\theta)}{\partial \theta_i}=0 \quad (i=1,2,\cdots,l)$,

2^0 解似然方程组得到 $L(\theta)$ 的最大值点;

③ 将样本 $X_1,X_2,\cdots X_n$ 代入最大值点的表达式中，就得未知参数的极大似然估计量 $\hat{\theta}$,将样本值 $x_1,x_2,\cdots x_n$ 代入最大值点的表达式中，就得未知参数的极大似然估计值 $\hat{\theta}$.

- θ 是实数时，似然方程组就是方程 $\dfrac{d\ln L(\theta)}{d\theta}=0$.

下面举例说明如何求极大似然估计



例4 设总体 $X \sim B(1, p)$，其分布列为

$$f(x;\ p) = p^x(1-p)^{1-x},\quad x = 0,\ 1$$

1, 0, 0, 1, 0, 0 是取自总体的一组样本值求参数 p 的极大似然估计.

解 样本的似然函数为: $L(p) = \prod_{i=1}^{n} P(x_i;\ p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

对数似然函数为: $\ln L(p) = \sum_{i=1}^{n} x_i \ln p + \sum_{i=1}^{n}(1-x_i)\ln(1-p)$

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

对 p 求导并令其为 0 得似然方程: $\frac{d\ln L(p)}{dp} = \frac{1}{p}\sum_{i=1}^{n} x_i - \frac{1}{1-p}(n - \sum_{i=1}^{n} x_i) = 0$，

解之得 p 的极大似然估计量: $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$,

代入样本值即得极大似然估计值为: $\hat{p} = \frac{1}{6}(1+0+0+1+0+0) = \frac{1}{3}$.

例3 设总体 X 的密度为 $f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0;\\ 0, & 其它,\end{cases}$

其中 $\lambda>0$ 为未知参数, $X_1, X_2, \cdots, X_n$ 是取自总体 X 的一组样本, 求 λ 的极大似然估计量与矩估计量.

解（1） 样本的似然函数为

$$L(\lambda)=\prod_{i=1}^{n} f(x_i;\lambda)=\begin{cases}\lambda^n \prod_{i=1}^{n} e^{-\lambda x_i}, & x_i>0,\ i=1,2,\cdots,n;\\ 0, & 其他.\end{cases}$$

当 $x_i>0$ 时, $L(\lambda)>0$, $1\le i\le n$,

故有对数似然函数: $\ln L(\lambda)=n\ln\lambda-\lambda\sum_{i=1}^{n}x_i$,

对 λ 求导并令其为0 可得似然方程: $\dfrac{d\ln L(\lambda)}{d\lambda}=\dfrac{n}{\lambda}-\sum_{i=1}^{n}x_i=0$,

解得 极大似然估计量: $\hat{\lambda}=n\Big/\sum_{i=1}^{n}X_i=\dfrac{1}{\overline{X}}$;

（2） $EX=\int_{-\infty}^{+\infty}xf(x;\lambda)dx=1/\lambda$, 令 $\dfrac{1}{\lambda}=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$,

解得矩估计量: $\hat{\lambda}=\dfrac{1}{\overline{X}}$.

例5 设总体 X 服从均匀分布 $U[0,\theta]$，为 θ 未知参数，$X_1, X_2, \cdots, X_n$ 是总体 X 的一组样本，求 θ 的极大似然估计量.

解 样本的似然函数为

不必建立
对数似然方程

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_i\le\theta,\ i=1,2,\cdots,n;\\ 0, & 其他.\end{cases}$$

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

无法用求导建立似然方程的方法确定其极大似然估计量!!

用极大似然原则来求：即用其他方法求似然函数 $L(\theta)$ 的最大值点.

显然，似然函数 $L(\theta)$ 的值随 θ 的减小而增大，故应取 θ 的值尽量地小；

另一方面，θ 必须满足条件 $0\le x_i\le\theta$ $(i=1,2,\cdots,n)$，

而事件 $\{0\le X_i\le\theta,\ i=1,2,\cdots,n\}=\{\max\limits_{1\le i\le n}\{X_i\}\le\theta\}$

故可取极大似然估计量为 $\hat{\theta}=\max\limits_{1\le i\le n}\{X_i\}\le\theta$.

可证明极大似然估计具有下面的单调函数不变性:

若$\hat{\theta}$为未知参数θ的极大似然估计量，而 $g(\theta)$为θ的单调函数，则 $g(\hat{\theta})$ 也是 $g(\theta)$的极大似然估计量.

[例] 一罐中装有白球和黑球，有放回地抽取一个容量为n的样本，其中有k个白球， 求罐中黑球与白球之比R的极大似然估计.

解 设 $X_1, X_2, \cdots, X_n$ 为所取样本，$X_i = \begin{cases} 1, & \text{取到白球} \\ 0, & \text{取到黑球} \end{cases}$ $i = 1, \cdots, n$

则 $X_1, \cdots, X_n \sim B(1, p)$，$p$是每次抽取时取到白球的概率，且 p未知.

容易求得 p 的极大似然估计为：$\hat{p} = \dfrac{k}{n}$， $\because\ R = \dfrac{1-p}{p}$，

由极大似然估计的不变性知 R 的极大似然估计是 $\hat{R} = \dfrac{1-\hat{p}}{\hat{p}} = \dfrac{n}{k} - 1$.

- 上述解法是应用微积分中的技巧求似然函数$L(\theta)$的最大值点.但当似然函数$L(\theta)$不可微或偏导数不为零时，就不能用上述求导方法求未知参数的极大似然估计了. 这时要用极大似然原则来求.

[例] 设总体 X 服从均匀分布 $U[0, \theta]$，为 θ 未知参数，$X_1, X_2, \cdots, X_n$ 是总体X的一组样本，求 θ 的极大似然估计量.

解 样本的似然函数为

不必建立
对数似然方程

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_i\le\theta,\ i=1,2,\cdots,n;\\ 0, & 其他.\end{cases}$$

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

无法用求导建立似然方程的方法确定其极大似然估计量!!

用极大似然原则来求： 即用其他方法求似然函数 $L(\theta)$ 的最大值点.

显然，似然函数 $L(\theta)$ 的值随 θ 的减小而增大，故应取 θ 的值尽量地小；

另一方面，θ 必须满足条件 $0\le x_i\le\theta$ $(i=1,2,\cdots,n)$，

而事件 $\{0\le X_i\le\theta,\ i=1,2,\cdots,n\}=\{\max\limits_{1\le i\le n}\{X_i\}\le\theta\}$

故可取极大似然估计量为 $\hat{\theta}=\max\limits_{1\le i\le n}\{X_i\}\le\theta$.

[例] 设X的分布函数为

$$F(x;\alpha,\beta)=\begin{cases}1-(\frac{\alpha}{x})^{\beta}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$$

其中 $\alpha>0$, $\beta>1$, $X_1,\cdots,X_n$ 是总体 X 的样本， 求 $\alpha=1$ 时, 未知参数 β 的矩估计量.

(1) 求 $\alpha=1$ 时, 未知参数 β 的极大似然估计量.

(2) 求 $\beta=2$ 时, 未知参数 α 的极大似然估计量.

β 的矩估计量为

$$\hat{\beta}=\frac{\overline{X}}{\overline{X}-1}.$$

解 (1) $\alpha=1$ 时,

$$f(x;\beta)=\begin{cases}\frac{\beta}{x^{\beta+1}}, & x>1,\\ 0, & x\le1,\end{cases}$$

$$L(\beta)=\prod_{i=1}^{n}f(x_i;\beta)=\begin{cases}\prod_{i=1}^{n}\frac{\beta}{x_i^{\beta+1}}, & x_i>1,\ i=1,\cdots,n;\\ 0, & \text{其他}.\end{cases}$$

当 $x_i>1$时, $L(\beta)>0$, $1\le i\le n$,

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

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

解得 β 的极大似然估计量: $\boxed{\hat{\beta}=n\Big/\sum_{i=1}^{n}\ln X_i}$.

[例]（续）设 X 的分布函数为 $F(x;\alpha,\beta)=\begin{cases}1-(\frac{\alpha}{x})^{\beta}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$

其中 $\alpha>0$, $\beta>1$, $X_1,\cdots,X_n$ 是总体 X 的样本,

（2）求 $\beta=2$ 时, 未知参数 α 的极大似然估计量.

解（2） $\beta=2$ 时， $f(x;\alpha)=\begin{cases}\frac{2\alpha^2}{x^3}, & x>\alpha,\\ 0, & x\le\alpha,\end{cases}$

$$L(\alpha)=\prod_{i=1}^{n}f(x_i;\alpha)=\begin{cases}\prod_{i=1}^{n}\frac{2^n\alpha^{2n}}{(x_1x_2\cdots x_n)^3}, & x_i>\alpha,\ i=1,2,\cdots,n;\\ 0, & \text{其他}.\end{cases}$$

因为 $L(\alpha)$ 随 α 的增大而增大, 故应取 α 的值尽量地大；

另一方面, 由于 $L(\alpha)=0$ 不可能是最大值, 故 α 必须满足 $x_i>\alpha$,
$(1\le i\le n)$,

而事件 $\{X_i>\alpha,\ i=1,2,\cdots,n\}$ $=\{\min\limits_{1\le i\le n}\{X_i\}\ge\alpha\}$

故 α 的极大似然估计量为 $\hat{\alpha}=\min\limits_{1\le i\le n}\{X_i\}\ge\alpha$.

[例]多项分布参数的极大似然估计

很多情况下, 假定一个变量X可能取m个状态,$m>2$,每个状态假定可能性为$p_1,\ldots,p_m$, $\sum_{i=1}^{m} p_i = 1$, 独立进行n次试验, 用X_i 表示第 i种状态出现的频数, $X_1,\cdots,X_m$会有多项分布,

$$p(x_1,\ldots,x_m \mid p_1,\ldots,p_m) = \frac{n!}{\prod_{i=1}^{m} x_i!} \prod_{i=1}^{m} p_i^{x_i}$$

$$l(p_1,\ldots,p_m) = \log(n!) - \sum_{i=1}^{m} \log x_i! + \sum_{i=1}^{m} x_i \log p_i$$

$$\sum_{i=1}^{m} p_i = 1$$

$$l(p_1,\ldots,p_m,\lambda)=\log(n!)-\sum_{i=1}^{m}\log x_i!+\sum_{i=1}^{m}x_i\log p_i+\lambda(\sum_{i=1}^{m}p_i-1)$$

$$\frac{\partial l(p_1,\ldots,p_m,\lambda)}{\partial p_i}=\frac{x_i}{p_i}+\lambda=0\Rightarrow p_i=-\frac{x_i}{\lambda},i=1,\ldots,m$$

$$\sum_{i=1}^{m}p_i=1\Rightarrow\sum_{i=1}^{m}-\frac{x_i}{\lambda}=1\Rightarrow\lambda=n$$

$$\hat{p}_i=\frac{x_i}{n}$$

极大似然估计的理论结果

$$\sqrt{nI(\theta)}(\hat{\theta}_{mle}-\theta)\xrightarrow{n\to\infty}N(0,1)$$

$$I(\theta)=E\left(\frac{\partial\log(p(x,\theta))}{\partial\theta}\right)^2=-E\left(\frac{\partial^2\log(p(x,\theta))}{\partial\theta^2}\right)$$

极大似然估计的分布有渐近的正态分布！

我们介绍了参数点估计，给出了寻求估计量最常用的方法——　矩法和极大似然法 .

计算简单
无需总体分布
不唯一、不够稳定

使用了总体分布，质量好
计算困难

- 参数点估计是用一个确定的值去估计未知的参数. 看来似乎精确，实际上把握不大.

自然要问:

样本均值、样本方差是否是一个好的估计量?
多个估计量时，哪一个估计量更好?

这就需要讨论以下几个问题:

估计量的评选标准

（1）我们希望一个“好的”估计量具有什么特性?
（2）怎样决定一个估计量是否比另一个估计量“好”?
（3）应该如何寻求一个合理的估计量?

第1次作业:

- 电子书上p.109 习题二
- 2（1） （3） （5）、
- 4（2） （4） （6）、
- 5、6、

谢 谢！