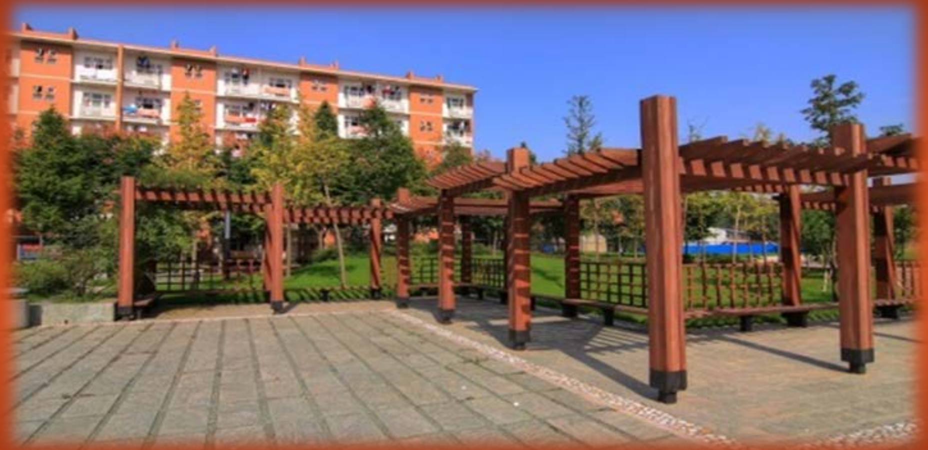
西安电子科技大学
XIDIAN UNIVERSITY

第7章 参数估计



概率论与数理统计课程组

CHAPTER 7

参数估计

CHAPTER 7

参数估计

7.1 点估计

问题的提出

参数估计是统计推断的基本问题之一，实际工作中碰到的总体，其分布类型往往是知道的，只是不知道其中的某些参数

因此，要求估计该参数的值，或是以一定的可靠性估计该参数在某个范围内或者不低于某数。参数估计问题就是要求通过样本估计总体分布所包含未知参数的值

参数估计的分类
- 点估计
  - 矩估计法
  - 最大似然估计法
- 区间估计

点估计

设总体X的分布函数$F(x;\theta_1,\theta_2,\dots,\theta_k)$形式已知，$\theta_1,\theta_2,\dots,\theta_k$是未知参数，$X_1,X_2,\dots,X_n$是来自总体$X$的一个样本，$x_1,x_2,\dots,x_n$是相应的一个样本值。

点估计问题就是要构造一个适当的统计量

$$\hat{\theta}_i=\hat{\theta}_i(X_1,X_2,\dots,X_n)\qquad i=1,2,\dots$$

用它的观察值作为未知参数θ_i的近似值

$\hat{\theta}_i=\hat{\theta}_i(X_1,X_2,\dots,X_n)$ 称为θ_i的点估计量

本质是一个随机变量

$\hat{\theta}_i=\hat{\theta}_i(x_1,x_2,\dots,x_n)$ 称为θ_i的点估计值

矩估计法

设X是连续型随机变量，其概率密度为

$f(x;\theta_1,\theta_2,\dots,\theta_k)$

或X为离散型随机变量，其分布律为

$P(X=x)=p(x;\theta_1,\theta_2,\dots,\theta_k)$

其中$\theta_1,\theta_2,\dots,\theta_k$为待估参数，$X_1,X_2,\dots,X_n$是来自总体$X$的样本，假设总体$X$的前$k$阶矩

$$\mu_l=E(X^l)=\int_{-\infty}^{\infty}x^l f(x;\theta_1,\theta_2,\dots,\theta_k)dx \quad X\text{连续型}$$

$$\mu_l=E(X^l)=\sum_{x\in R_X}x^l p(x;\theta_1,\theta_2,\dots,\theta_k) \quad X\text{离散型}$$

存在，一般它们是$\theta_1,\theta_2,\dots,\theta_k$的函数

基于样本矩 $A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l$

依概率收敛于相应的总体矩μ_l

样本矩的连续函数依概率收敛于相应的总体矩的连续函数

定义

由*Khinchine*大数定律，若总体X的期望$E(X)$有限，则样本均值$\overline{X}$收敛于$E(X)$

因此，可以用样本矩作为相应的总体矩的估计量，而以样本矩的连续函数作为相应总体矩连续函数的估计量

该方法称为**矩估计法**

❖ 样本矩的连续函数收敛于总体矩的连续函数

统计量 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k,\ k=1,\ 2,\cdots$　　样本值 $a_k=\frac{1}{n}\sum_{i=1}^{n}x_i^k,\ k=1,\ 2,\cdots$

设总体X的k阶矩存在，记为$E(X^k)\xrightarrow{\text{记成}}\mu_k$

当$n\to\infty$时，$A_k\xrightarrow{P}\mu_k,\ k=1,2,\ldots$

因为$X_1,X_2,\ldots,X_n$互相独立且与X同分布，故 $X_1^k,X_2^k,\ \ldots,X_n^k$ 独立且与X^k同分布

$$E(X_1^k)=E(X_2^k)=\ldots=E(X_n^k)=\mu_k$$

由*Khinchine*大数定律 $A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k\xrightarrow{P}\mu_k,\ k=1,2,\ldots$

由依概率收敛的性质 $g(A_1,A_2,\ldots,A_k)\xrightarrow{P}g(\mu_1,\mu_2,\ldots,\mu_k)$　　g是连续函数

具体流程

设 $\begin{cases}\mu_1=\mu_1(\theta_1,\theta_2,\ \dots\ ,\theta_k)\\ \mu_2=\mu_2(\theta_1,\theta_2,\ \dots\ ,\theta_k)\\ \quad\vdots\\ \mu_k=\mu_k(\theta_1,\theta_2,\ \dots\ ,\theta_k)\end{cases}\Rightarrow\begin{cases}\theta_1=\theta_1(\mu_1,\mu_2,\ \dots\ ,\mu_k)\\ \theta_2=\theta_2(\mu_1,\mu_2,\ \dots\ ,\mu_k)\\ \quad\vdots\\ \theta_k=\theta_k(\mu_1,\mu_2,\ \dots\ ,\mu_k)\end{cases}$

其中 $\mu_l=E(X^l)$

这是一个包含k个参数$\theta_1, \theta_2, \dots, \theta_k$的联立方程组，可以从中解出$\theta_1, \theta_2, \dots, \theta_k$

以样本矩A_l分别代替上式中总体矩μ_l, $l = 1, 2, \dots, k$

而 $A_l=\frac{1}{n}\sum_{i=1}^{n}X_i^l$

以 $\hat{\theta}_i=\theta_i(A_1, A_2,\ \dots\ ,A_k),\ i=1,\ 2,\ \dots,\ k$

分别作为θ_i, $i = 1, 2, \dots, k$的估计量

这种估计量称为参数θ_i的矩估计量，其样本值称为参数θ_i的矩估计值

流程概要

2、3顺序可互换

1. 写出总体矩的k个方程，是k个未知参数的表达式
2. 解方程组，将每个未知参数用总体矩表示
3. 用样本l阶矩代替总体l阶矩

例 设在总体X的均值μ和方差σ^2均未知，$(X_1, X_2,\ldots, X_n)$是取自X的一个样本，试求μ和σ^2的矩估计。

解 先求总体矩 Step 1 求总体矩，有几个参数待估就写几个

$$\mu_1 = E(X) = \mu$$

$$\mu_2 = E(X^2) = D(X) + [E(X)]^2 = \mu^2 + \sigma^2$$

再求样本矩 Step 3 求样本矩，用来替代总体矩

$$A_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X},\quad A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

令 $\begin{cases}\mu_1 = A_1 \\ \mu_2 = A_2\end{cases}$ 得到 $\begin{cases}\hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i - \overline{X})^2\end{cases}$ Step 2

解方程组，采用规范的统计量写法

例 设总体概率密度为

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

$\theta > 0$是未知参数，$(X_1, X_2,\ldots, X_n)$是取自X的一个样本，试求θ的矩估计。

解

$$E(X) = \int_{-\infty}^{+\infty} x f(x)dx = \int_0^1 \sqrt{\theta}\, x^{\sqrt{\theta}} dx$$

$$= \frac{\sqrt{\theta}}{\sqrt{\theta}+1}$$ Step 1

令 $E(X) = \overline{X}$ Step 3

$$\Rightarrow \frac{\sqrt{\theta}}{\sqrt{\theta}+1} = \overline{X} \Rightarrow \hat{\theta} = \left(\frac{\overline{X}}{1-\overline{X}}\right)^2$$ Step 2

最大似然估计法

引例

考察下例，假设在一个罐中放着许多白球和黑球，并假定已经知道两种球的数目之比是1:3，但不知道哪种颜色球多。如果用放回抽样方法从罐中任取 n 个球，其中黑球的个数为 x 的概率为

$$P(x;p)=C_n^x p^x(1-p)^{n-x}$$

由假设知，$p=\dfrac{1}{4}$ 或 $p=\dfrac{3}{4}$

若取 n=3，如何通过x来估计p值

先计算抽样的可能结果 x 在这两种 p 值之下的概率，列出分布律

x	0	1	2	3
$P(x, 3/4)$	1/64	9/64	27/64	27/64
$P(x, 1/4)$	27/64	27/64	9/64	1/64

从上表可知

$x=0,\ P(0,1/4)=27/64>P(0,\ 3/4)=1/64$

$\hat{p}=1/4$ 更合理；x=1类似；

$x=2,\ P(2,1/4)=9/64<P(2,\ 3/4)=27/64$

$\hat{p}=3/4$ 更合理；x=3类似；

因此 $\hat{p}(x)=\begin{cases}1/4 & x=0,\ 1\\ 3/4 & x=2,\ 3\end{cases}$

对每个x，取 $\hat{p}(x)$ 使得 $P[x;\hat{p}(x)]\geq P(x;p')$

p'是不同于p的另一个值

最大似然估计的原理

抽样中，样本$X_1, X_2, \dots, X_n$取值为$x_1, x_2, \dots, x_n$这一事件发生的可能性应该是较大的（因为结果已然出现），即参数的取值应使得事件$\{X_1 = x_1, X_2 = x_2, \dots, X_n = x_n\}$的概率最大

具体流程

若X为离散型随机变量，其分布律为 $P(X=x) = p(x; \theta)$，或若X是连续型随机变量，其概率密度为$f(x; \theta)$

$\theta = (\theta_1, \theta_2, \dots, \theta_k)$是待估参数，$\theta \in \Theta$，$\Theta$为参数空间，也就是$\theta$的取值范围

设$(x_1, x_2, \dots, x_n)$是样本$(X_1, X_2, \dots, X_n)$的样本值

(i) 做似然函数

似然函数通常简写为$L(\theta)$

若X为离散型随机变量，则$(X_1, X_2, \dots, X_n)$的联合分布律为

$$P(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \prod_{i=1}^{n} p(x_i; \theta)$$

似然函数为

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

若X为连续型随机变量，则$(X_1, X_2, \dots, X_n)$的联合概率密度为

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

似然函数为

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

(ii) 使似然函数取最大值

求使 $L(x_1,x_2,...,x_n;\theta)$ 达到最大，即 $L(x_1,x_2,\dots,x_n;\hat{\theta})=\max\limits_{\theta\in\Theta}L(x_1,x_2,\dots,x_n;\theta)$

$\hat{\theta}(X_1,X_2,\dots,X_n)$ 称为 θ 的**最大似然估计量**　　$\hat{\theta}(x_1,x_2,\dots,x_n)$ 称为θ的**最大似然估计值**

注意： 求$L(\theta)$的最大值通常转为求 $\ln L(x_1,x_2,\dots,x_n;\theta)$ 的最大值，称**对数似然函数**

一般是从方程 $\frac{d}{d\theta}L(\theta)=0$ 或是 $\frac{d}{d\theta}\ln L(\theta)=0$ 解得 $\hat{\theta}$

流程概要

1. 写出似然函数（联合概率密度或联合分布律）
2. 求似然函数或对数似然函数取最大值的条件
3. 写出以统计量表示未知参数的结果

例 设总体概率密度为 $f(x)=\begin{cases}\sqrt{\theta}x^{\sqrt{\theta}-1} & 0\le x\le 1\\ 0 & \text{其他}\end{cases}$

$\theta>0$是未知参数，$(X_1, X_2,\ldots, X_n)$是取自X的一个样本，试求θ的最大似然估计。

解 似然函数 $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)=\prod_{i=1}^{n}\sqrt{\theta}x_i^{\sqrt{\theta}-1}=\theta^{n/2}\left(\prod_{i=1}^{n}x_i\right)^{\sqrt{\theta}-1}$

Step 1 写出似然函数，即联合概率密度或联合分布律

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

Step 2 求使似然函数最大值的条件

令 $\frac{d}{d\theta}\ln L(\theta)=\frac{n}{2}\cdot\frac{1}{\theta}+\frac{1}{2\sqrt{\theta}}\sum_{i=1}^{n}\ln x_i=0$ 即 $\frac{n}{\sqrt{\theta}}=-\sum_{i=1}^{n}\ln x_i$

最大似然估计量为 $\hat{\theta}=\frac{n^2}{\left(\sum_{i=1}^{n}\ln X_i\right)^2}$

Step 3 以规范的统计量写法表示

例 设总体X概率密度为 $f(x)=\begin{cases}\frac{1}{\theta}e^{-(x-\mu)/\theta} & x\geq\mu\\ 0 & \text{其他}\end{cases}$

其中$\theta>0$，θ、μ是未知参数，$(X_1, X_2, \ldots, X_n)$是取自X的样本，求θ、μ的**矩估计**与**最大似然估计**。

解 (i) 矩估计

$$E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{\mu}^{+\infty}x\frac{1}{\theta}e^{-(x-\mu)/\theta}dx=\mu+\theta$$

$$E(X^2)=\int_{\mu}^{+\infty}x^2\frac{1}{\theta}e^{-(x-\mu)/\theta}dx=\mu^2+2\theta\int_{\mu}^{+\infty}\frac{x}{\theta}e^{-(x-\mu)/\theta}dx=\mu^2+2\mu\theta+2\theta^2$$

$$D(X)=E(X^2)-\left[E(X)\right]^2=\theta^2$$

令 $E(X)=\overline{X}$ $\quad D(X)=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 样本二阶中心矩B_2

$$\Rightarrow \hat{\theta}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2} \qquad \hat{\mu}=\overline{X}-\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

例 设总体X概率密度为 $f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta} & x\geq\mu\\ 0 & \text{其他}\end{cases}$

其中$\theta>0$，θ、μ是未知参数，$(X_1, X_2,\ldots, X_n)$是取自X的样本，求θ、μ的矩估计与最大似然估计。

解 (ii) 最大似然估计

$$L(\theta,\mu)=\prod_{i=1}^{n}\frac{1}{\theta}e^{-(x_i-\mu)/\theta}=\frac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)} \qquad x_i\geq\mu$$

此时不能通过求偏导数获得μ的最大似然估计量

因为$x_i\geq\mu$，故μ的取值范围最大不超过 $x=\min\{x_1,x_2,\cdots,x_n\}$

而 $L(\theta,\mu)=\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i+\frac{n\mu}{\theta}}$ 是μ的增函数，μ取最大值时L达到最大，故 $\hat{\mu}=X_{(1)}=\min\{X_1,X_2,\cdots,X_n\}$

又 $\ln L(\theta)=-n\ln\theta-\dfrac{1}{\theta}\sum_{i=1}^{n}(X_i-\hat{\mu})$ 令 $\dfrac{d}{d\theta}\ln L(\theta)=-\dfrac{n}{\theta}+\dfrac{1}{\theta^2}\sum_{i=1}^{n}\left[X_i-X_{(1)}\right]=0$ 故 $\hat{\theta}=\overline{X}-X_{(1)}$

例 设总体X服从$(0,\theta)$上的均匀分布，$\theta>0$未知，$x_1, x_2, \ldots, x_n$是取自X的一个样本，试求θ的最大似然估计和矩估计。

解 (i) 最大似然估计　　X概率密度为 $f(x;\theta)=\begin{cases}\dfrac{1}{\theta} & 0\le x\le\theta\\ 0 & \text{其他}\end{cases}$

似然函数为 $L(\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_1,x_2,\cdots,x_n\le\theta\\ 0 & \text{其他}\end{cases}$　　由于$\dfrac{d}{d\theta}\ln(\theta)=-\dfrac{n}{\theta}\ne 0$　不能用微分法求$\hat\theta$

只能从定义上去求$\hat\theta$　　因为$0\le x_i\le\theta$　故θ的最小取值为 $x_{(n)}=\max\{x_1,x_2,\cdots,x_n\}$

又 $\ln L(\theta)=\dfrac{1}{\theta^n}$　　对$\theta>x_{(n)}$的θ是减函数　　θ越小，L越大，故 $\hat\theta=x_{(n)}$ 时，L最大

故θ的最大似然估计值是　$\hat\theta=x_{(n)}=\max\{x_1,x_2,\cdots,x_n\}$

(ii) 矩估计　　由 $E(X)=\int_0^\theta \dfrac{1}{\theta}x\,dx=\dfrac{\theta}{2}=\overline{X}$　　$\Rightarrow \hat\theta=2\overline{X}=2\dfrac{x_1+x_2+\cdots+x_n}{n}$

本题给出条件是样本值，因此答案用样本估计值（估计量的观测值），一般是用样本的估计量表示

例 设总体X分布律为

x_k	1	2	3
p_k	θ	$\theta/2$	$1-3\theta/2$

θ是未知参数， 现得到X的一组样本观测值(2, 3, 2, 1, 3)，试求θ的矩估计和最大似然估计。

解 (i) 矩估计

$$E(X)=\sum_{k=1}^{3}x_k p_k=\theta+2\times\theta/2+3\times(1-3\theta/2)=3-5\theta/2$$

$$\overline{X}=2.2 \qquad 令\ E(X)=\overline{X} \qquad \Rightarrow \hat{\theta}=0.32$$

(ii) 最大似然估计

$$L(\theta)=(\theta/2)(1-3\theta/2)(\theta/2)\theta(1-3\theta/2)=\frac{1}{16}\theta^3(2-3\theta)^2$$

$$\ln L(\theta)=-\ln 16+3\ln\theta+2\ln(2-3\theta) \qquad \frac{d}{d\theta}\ln L(\theta)=\frac{3}{\theta}-\frac{6}{2-3\theta}=0 \qquad \Rightarrow \hat{\theta}=0.4$$

上述例题矩估计和最大似然估计的结果

分布	矩估计	最大似然估计
例 2&4	$\hat{\theta}=\left(\dfrac{\overline{X}}{1-\overline{X}}\right)^2$	$\hat{\theta}=\dfrac{n^2}{\left(\sum\limits_{i=1}^{n}\ln X_i\right)^2}$
例 5	$\hat{\mu}=\overline{X}-\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$ $\hat{\theta}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}$	$\hat{\mu}=X_{(1)}=\min\{X_1,X_2,\cdots,X_n\}$ $\hat{\theta}=\overline{X}-X_{(1)}$
例 6	$\hat{\theta}=2\overline{X}$	$\hat{\theta}=X_{(n)}=\max\{X_1,X_2,\cdots,X_n\}$

本节回顾

矩估计法

1. 写出总体矩的k个方程，是k个未知参数的表达式
2. 解方程组，将每个未知参数用总体矩表示
3. 用样本l阶矩代替总体l阶矩

最大似然估计法

1. 写出似然函数（联合概率密度或联合分布律）
2. 求似然函数或对数似然函数取最大值的条件
3. 写出以统计量表示未知参数的结果