- 统计推断
  - 统计估计
    - 参数估计
      - 点估计（第7章1-2节）
      - 区间估计（第7章3-5节）
    - 非参数估计（不讲）
  - 假设检验
    - 参数假设检验（第8章1-2节）
    - 非参数（分布）假设检验（第8章3节*）
  - 回归分析（第九章*）
  - 方差分析（第十章*）

第七章　参数估计

“参数”---刻画总体某些概率特征的数量

- 分布中所含的未知参数：如二点分布$B(1, p)$中的概率p
- 分布中所含的未知参数的函数：如$X \sim N(\mu,\sigma^2)$,则X不超过

  某定值a的概率$P\{X \le a\} = \Phi(\frac{a-\mu}{\sigma})$是未知参数$\mu,\sigma$的函数.
- 分布的各种数字特征也都是未知参数：如$E(X)$，$D(X)$等.

一般场合, 常用θ表示参数, 参数θ所有可能取值组成的集合称为参数空间, 常用Θ表示.

参数估计问题：讨论如何根据样本对上述各种未知参数作出估计.

- 如何给出估计
- 如何评判不同的估计优劣

参数估计的形式: 点估计与区间估计

第七章　参数估计

一、参数的点估计

二、估计量的评选标准

三、参数的区间估计

四、正态总体参数的区间估计

五、单侧置信区间

第一节　参数的点估计

一、　矩估计法

二、　最大似然估计法

点估计的概念

设总体 X 的分布函数为已知的$F(x,\theta)$,其中x是自变量，θ为未知参数（一个或多个），借助于总体 X 的一个样本来估计未知参数θ的值的问题称为**点估计问题**.

点估计的具体思想

设总体 X 的分布函数为 $F(x;\theta_1,\theta_2,\cdots,\theta_m)$, $\theta_1,\theta_2,\cdots,\theta_m$是未知参数,

$X_1,X_2,\cdots,X_n$ 是 X 的一个样本, $x_1,x_2,\cdots,x_n$ 为相应的一个样本值.

构造统计量：(称为θ的点估计量)　用其观测值(称为θ的点估计值)

$$\hat{\theta}(X_1,X_2,\cdots,X_n)=\begin{pmatrix}\hat{\theta}_1(X_1,X_2,\cdots,X_n)\\ \hat{\theta}_2(X_1,X_2,\cdots,X_n)\\ \vdots \\ \hat{\theta}_m(X_1,X_2,\cdots,X_n)\end{pmatrix} \qquad \hat{\theta}(x_1,x_2,\cdots,x_n)=\begin{pmatrix}\hat{\theta}_1(x_1,x_2,\cdots,x_n)\\ \hat{\theta}_2(x_1,x_2,\cdots,x_n)\\ \vdots \\ \hat{\theta}_m(x_1,x_2,\cdots,x_n)\end{pmatrix}$$

来估计未知参数$\theta=(\theta_1,\theta_2,\ldots,\theta_m)$.

问题:

(1) 如何构造估计量?

(2) 如何评价估计量?

常用构造估计量的方法: (两种)

矩估计法和最大似然估计法.

1. 矩估计法

基本原理：“替换原理”

用样本矩代替相应的总体矩,如：令$E(X)=\bar{X}$

用样本矩的连续函数代替相应的总体矩的连续函数，

建立含有待估参数的方程，从而解出待估参数.

矩估计法

这是由英国统计学家K. 皮尔逊最早提出的 .

矩估计的思想方法：

设总体 $X \sim F(x;\theta_1,\theta_2,\cdots,\theta_r)$，参数 $\theta_1,\theta_2,\cdots,\theta_r$ 未知，

总体 X 的 $1,2,\cdots,r$ 阶矩存在：$\mu_k(\theta_1,\theta_2,\cdots,\theta_r)=E(X^k)\quad(k=1,2,\cdots,r).$

设 $X_1,X_2,\cdots,X_n$ 是 X 的一个样本，由辛钦大数定律，当 $n\to\infty$ 时有

$$\text{样本矩} A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{\text{依概率收敛}} \text{总体矩}\mu_k(\theta_1,\theta_2,\cdots,\theta_r), k=1,2,\ldots r.$$

$$\text{样本矩的连续函数} \xrightarrow{\text{依概率收敛}} \text{总体矩的连续函数.}$$

因此当 n 较大时有

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \approx \mu_k(\theta_1,\theta_2,\cdots,\theta_r),(k=1,2,\ldots.,r).$$

$$g(A_1,A_2,\cdots,A_r)\approx g(\mu_1,\mu_2,\cdots,\mu_r).$$

令 $\begin{cases} \mu_1 = \mu_1(\theta_1, \theta_2, \cdots, \theta_r) = A_1, \\ \mu_2 = \mu_2(\theta_1, \theta_2, \cdots, \theta_r) = A_2, \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ \mu_r = \mu_r(\theta_1, \theta_2, \cdots, \theta_r) = A_r \cdot \end{cases}$ 解得 $\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(A_1, A_2, \cdots, A_r), \\ \hat{\theta}_2 = \hat{\theta}_2(A_1, A_2, \cdots, A_r), \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ \hat{\theta}_r = \hat{\theta}_r(A_1, A_2, \cdots, A_r) \cdot \end{cases}$

以此作为未知参数 $\theta_1, \theta_2, \ldots, \theta_r$的估计量, 称为矩估计量.

如果样本观测值为$x_1, x_2, \ldots, x_n$， 则得未知参数$\theta_1, \theta_2, \ldots, \theta_r$ 的矩估计值为

$$\begin{cases} \hat{\theta}_1 = \hat{\theta}_1(a_1, a_2, \cdots, a_r), \\ \hat{\theta}_2 = \hat{\theta}_2(a_1, a_2, \cdots, a_r), \\ \cdots \quad \cdots \quad \cdots \quad \cdots \\ \hat{\theta}_r = \hat{\theta}_r(a_1, a_2, \cdots, a_r) \cdot \end{cases}$$

其中$a_k = \dfrac{1}{n}\sum_{i=1}^{n} x_i^{\,k}, k = 1, 2, \ldots., r.$

矩估计法的一般步骤(解题方法，重要！！):

(1)根据未知参数的个数求需要的总体矩，

如 $\mu_1 = E(X),\ \mu_2 = E(X^2)$ 等，

(2)用样本矩替换相应的总体矩得到方程组:

如：令 $\mu_1 = E(X) = A_1 = \bar{X}, \mu_2 = A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

(3) 解方程组得到参数的矩估计量（值）.

注：一般先找低阶的矩.

例7.1.1 设总体 $X \sim B(m, p)$,其中p $(0<p<1)$未知，又设X_1, $X_2, \ldots, X_n$为X 的样本，求 p 的矩估计量和矩估计值.

因为仅有一个未知参数，仅需要一个方程

解：第一步:求总体矩

——求期望(总体一阶原点矩):

因$X \sim B(m, p)$, 则$E(X) = mp$,

第二步：用样本矩替换相应的总体矩：

令$\mu_1 = A_1$,即$mp = \overline{X}$,

解得未知参数 p 的矩估计量 $\hat{p} = \dfrac{\overline{X}}{m} = \dfrac{1}{mn}\sum_{i=1}^{n} X_i$,

如果将X_i替换为x_i，则相应的估计称为“矩估计值”．

故未知参数 p 的矩估计值为 $\hat{p} = \dfrac{\overline{x}}{m} = \dfrac{1}{mn}\sum_{i=1}^{n} x_i$.

例7.1.2 设总体X服从指数分布，其概率密度为：

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0, \quad\quad 其它\end{cases}，\quad \lambda>0未知，$$

因为仅有一个未知参数，仅需要一个方程

设$X_1, X_2, \dots, X_n$是总体X的一个样本，求λ的矩估计量.

解：第一步：求总体矩

——求期望（总体一阶原点矩）：

$$\mu_1 = \qquad \frac{1}{\lambda},$$

第二步：用样本矩替换相应的总体矩：

$$令\mu_1 = A_1，\ 即\frac{1}{\lambda}=\bar{X},$$

如果将X_i替换为x_i，则相应的估计称为"矩估计值".

第三步：求解矩估计量：

$$\Rightarrow \hat{\lambda}=\frac{1}{\bar{X}}=\frac{1}{\frac{1}{n}\sum_{i=1}^{n}X_i}.$$

故$\hat{\lambda}$为真值λ的矩估计量.

例7.1.3 总体X的均值μ及方差σ^2都存在，且有$\sigma^2>0$，但μ和σ^2未知，$X_1,X_2,\dots,X_n$是X的一个样本，试求μ和σ^2的矩估计量.

解：第一步：求总体矩：两个未知数，需要两个方程，即二阶矩

$$\begin{cases}\mu_1=E(X)=\mu,\\ \mu_2=E(X^2)=D(X)+\left(EX\right)^2=\sigma^2+\mu^2.\end{cases}$$

第二步：用样本矩替换相应的总体矩：

令$\mu_1=A_1,\mu_2=A_2$. 即$\begin{cases}\mu=\overline{X},\\ \sigma^2+\mu^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2,\end{cases}$

第三步：求矩估计量：解得μ和σ^2的矩估计量为

$$\begin{cases}\hat{\mu}=\overline{X},\\ \hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\overline{X}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\overline{X}\right)^2.\end{cases}$$

故$\hat{\mu}$和$\hat{\sigma}^2$分别为μ和σ^2的矩估计量。

例7.1.3 总体X的均值 μ及方差 σ^2都存在，且有$\sigma^2>0$，但μ和σ^2未知，$X_1, X_2, \ldots, X_n$是X的一个样本，试求μ和σ^2的矩估计量？

前面得到
$$\begin{cases} \hat{\mu}=\bar{X}, & \text{样本均值} \\ \hat{\sigma}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\bar{X}^2=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2. \end{cases}$$

样本二阶中心矩

故：总体X的均值μ和方差σ^2的矩估计量分别为样本均值$\bar{X}$和样本二阶中心矩B_2，即

$$\begin{cases} \hat{\mu}=\bar{X}=\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i, \\ \widehat{\sigma^2}=\dfrac{1}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2 \end{cases}.$$

与总体X服从的分布无关

例7.1.4 设总体$X \sim U(a,b)$，a,b均未知，设$X_1,X_2,\ldots,X_n$是总体X的一个样本，求a,b的矩估计量.

解： 因为$X \sim U(a,b)$，则有$E(X)=\dfrac{a+b}{2}$，$D(X)=\dfrac{(b-a)^2}{12}$.

$$\text{于是}\begin{cases}\mu_1=E(X)=\dfrac{a+b}{2},\\[2ex] \mu_2=E(X^2)=D(X)+(EX)^2=\dfrac{(b-a)^2}{12}+\left(\dfrac{a+b}{2}\right)^2\end{cases}$$

令$\mu_k=A_k$，$k=1,2$，即$\begin{cases}\frac{a+b}{2}=A_1,\\ \frac{(b-a)^2}{12}+\frac{(a+b)^2}{4}=A_2,\end{cases}$ 得$\begin{cases}a+b=2A_1,\\ b-a=2\sqrt{3(A_2-A_1^2)},\end{cases}$

解得a和b的矩估计量分别为

$$\begin{cases}\hat{a}=A_1-\sqrt{3\left(A_2-A_1^2\right)}=\bar{X}-\sqrt{3\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\bar{X}^2\right)}=\bar{X}-\sqrt{\dfrac{3}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2}=\bar{X}-\sqrt{3B_2},\\[3ex] \hat{b}=A_1+\sqrt{3(A_2-A_1^2)}=\bar{X}+\sqrt{3\left(\dfrac{1}{n}\sum\limits_{i=1}^{n}X_i^2-\bar{X}^2\right)}=\bar{X}+\sqrt{\dfrac{3}{n}\sum\limits_{i=1}^{n}\left(X_i-\bar{X}\right)^2}=\bar{X}+\sqrt{3B_2}.\end{cases}$$

例7.1.4 设总体$X \sim U(a,b)$，a,b均未知，设$X_1, X_2, \ldots, X_n$是总体X的一个样本，求a,b的矩估计量.

解： 若从均匀总体$U(a,b)$中获得一个容量为5的样本：

4.5, 5.0, 4.7, 4.0, 4.2,

经计算$\overline{x} = 4.48, b_2 = 0.1256,$

于是可得a和b的矩估计值分别为

$$\begin{cases} \hat{a} = \overline{x} - \sqrt{3b_2} = 4.48 - \sqrt{3 \times 0.1256} = 3.8662, \\ \hat{b} = \overline{x} + \sqrt{3b_2} = 4.48 + \sqrt{3 \times 0.1256} = 5.0938. \end{cases}$$

例7.1.5 设总体X具有以下概率密度:

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, 0<x<1, \\ 0, \quad 其它\end{cases} \quad \theta>0 未知,$$

因为仅有一个未知参数，仅需要一个方程

设$X_1, X_2, \ldots, X_n$是总体X的一个样本，求θ的矩估计量.

解:

$$\mu_1=E(X)=\int_{-\infty}^{+\infty} x f(x)dx=\int_0^1 x\theta x^{\theta-1}dx=\int_0^1 \theta x^{\theta}dx=\frac{\theta}{\theta+1}.$$

令$\mu_1=A_1$,

即 $\dfrac{\theta}{\theta+1}=\overline{X}$,

解得矩估计量:

$$\Rightarrow \hat{\theta}=\frac{\overline{X}}{1-\overline{X}}=\frac{\frac{1}{n}\sum_{i=1}^{n}X_i}{1-\frac{1}{n}\sum_{i=1}^{n}X_i}.$$

故$\hat{\theta}$为真值θ的矩估计量.

如果将X_i替换为x_i，则相应的估计称为“矩估计值”.

矩估计法小结

求解步骤:

1. 根据未知数个数求解总体矩, 并用未知参数表示总体矩;

2. 替换原理: 用样本矩 A_k 代替总体矩 μ_k;

3. 求未知参数的矩估计量（值）.

优点: 原理直观，简单易行;

缺点: 总体矩要存在;

矩估计基于大数定律，在大样本条件下才有较好的效果。

一般情况下, 矩估计量不具有唯一性.

例如:若总体$X \sim P(\lambda)$,则$E(X)=D(X)=\lambda$,因此λ有两个估计量:

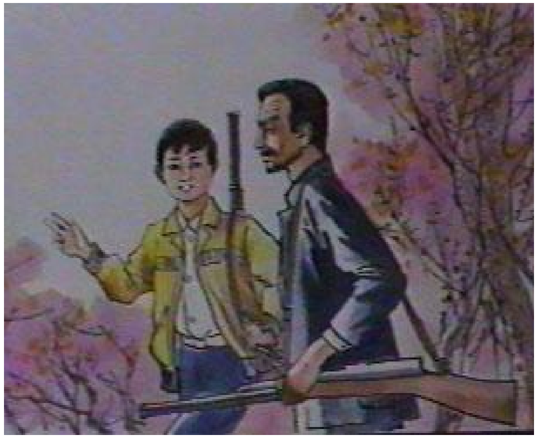
$$\hat{\lambda}=\overline{X}, \qquad \hat{\lambda}=B_2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

2. 最大似然估计法（Maximum Likelihood Estimate）

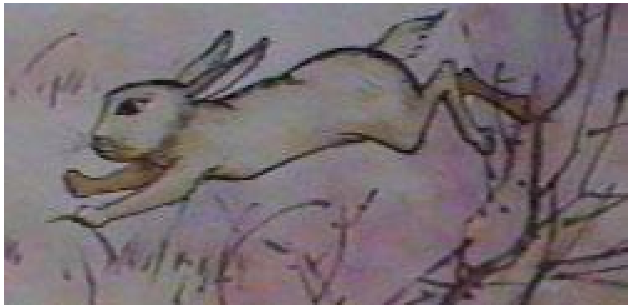
1) 最大似然法的基本思想

先看一个简单例子：

某位新手与一位老猎人一起外出打猎．一只野兔从前方窜过．只听枪响，野兔应声倒下，仅有一个弹痕.



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，老猎人命中的概率一般大于这位新手命中的概率。 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最大似然法的基本思想：

概率最大的事件在一次试验中最可能发生。

在已知试验结果的情况下，认为试验条件对试验结果最有利。

最大似然估计法是由英国统计学家Fisher引进的.

Fisher的最大似然思想:

概率最大的事件在一次试验中最可能发生。

设某试验的可能结果为: $A_1, A_2, \ldots, A_n, \ldots,$
发生概率为$P(A_n;\theta)$, $n=1,2,\ldots$

若在一次试验中，某结果A_i发生，则应选择参数θ使得A_i出现的概率最大。

$$\text{即若} P(A_i;\hat{\theta}) = \max_{\theta\in\Theta} P(A_i;\theta),$$

则应选择$\hat{\theta}$作为未知参数θ的估计值。

利用已知总体的概率密度(或概率分布)及样本, 根据概率较大的事件在一次试验中最有可能出现的原理, 求总体的概率密度(或概率分布)中所含未知参数的点估计的方法叫做最大似然估计法.

1. 离散型总体的概率分布中只含一个未知参数情形

设离散型总体X分布律 $P\{X=x\}=p(x;\theta), \theta$ 为待估参数, $\theta\in\Theta$,

(其中 Θ 是 θ 可能的取值范围)

设$X_1, X_2, \cdots, X_n$是来自总体 X 的样本，相应的样本观测值为 $x_1, x_2, \cdots, x_n$,

样本观测值为 $x_1, x_2, \cdots, x_n$, 说明事件$\{X_1=x_1\}, \cdots, \{X_n=x_n\}$同时发生。

由于 $X_1, X_2, \cdots, X_n$ 相互独立且与X同分布，则有

$$P\{X_1=x_1, X_2=x_2, \cdots, X_n=x_n\}=P\{X_1=x_1\}P\{X_2=x_2\}\cdots P\{X_n=x_n\}$$

$$=\prod_{i=1}^{n}p(x_i;\theta), \quad \theta\in\Theta,$$

似然函数的定义

记$L(\theta)=L(x_1,x_2,\cdots,x_n;\theta)=\prod_{i=1}^{n}p(x_i;\theta),\quad \theta\in\Theta,$

称$L(\theta)$为样本的似然函数.

最大似然估计值（量）

设Θ是θ可能的取值范围，

若有$\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$使得对一切$\theta\in\Theta$,有

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})\geq L(x_1,x_2,\cdots,x_n;\theta).$$

则称$\boxed{\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)}$为 参数$\theta$的最大似然估计值，

$\boxed{\hat{\theta}(X_1,X_2,\cdots,X_n)}$ 参数θ的最大似然估计量.

2. 连续型总体的概率密度中只含一个未知参数情形

似然函数的定义

设连续型总体X概率密度为 $f(x;\theta)$, θ 为待估参数， $\theta \in \Theta$,

(其中 Θ 是 θ 可能的取值范围)

设$X_1, X_2, \cdots, X_n$是来自总体 X 的样本,

相应的样本观测值为 $x_1, x_2, \cdots, x_n$,

随机点 X_i落在点x_i的长度为Δx_i的邻域内的概率近似等于

$f(x_i;\theta)\Delta x_i \quad (i = 1,2,\cdots,n)$,

则随机点 $(X_1, X_2, \cdots, X_n)$落在点$(x_1, x_2, \cdots, x_n)$的邻域

(边长分别为$\Delta x_1, \Delta x_2, \cdots, \Delta x_n$的$n$维立方体)内的概率

近似地为$\prod_{i=1}^{n} f(x_i;\theta)\Delta x_i$.

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

称$L(\theta)$为样本的似然函数.

最大似然估计值（量）

设Θ 是 θ 可能的取值范围,

若有 $\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)$使得对一切$\theta\in\Theta$,有

$$L(x_1,x_2,\cdots,x_n;\hat{\theta})\geq L(x_1,x_2,\cdots,x_n;\theta).$$

则称$\boxed{\hat{\theta}=\hat{\theta}(x_1,x_2,\cdots,x_n)}$为参数 θ 的最大似然估计值,

$\boxed{\hat{\theta}(X_1,X_2,\cdots,X_n)}$ 参数 θ 的最大似然估计量.

求最大似然估计值（量）---求似然函数$L(\theta)$的最大值

求似然函数的最大值的步骤:

(一) 写出似然函数

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

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

(二) 取对数

$$\ln L(\theta)=\sum_{i=1}^{n}\ln p(x_i;\theta) \quad 或 \quad \ln L(\theta)=\sum_{i=1}^{n}\ln f(x_i;\theta);$$

对数似然函数

(三) 对 θ 求导 $\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}$, 并令 $\boxed{\dfrac{\mathrm{d}\ln L(\theta)}{\mathrm{d}\theta}=0}$, 对数似然方程

解方程即得未知参数 θ 的最大似然估计值 $\hat{\theta}$.

最大似然估计法也适用于分布中含有多个未知参数的情况. 此时只需令

$$\boxed{\frac{\partial}{\partial\theta_i}\ln L=0,\qquad i=1,2,\cdots,k.}$$

对数似然方程组

解出由 k 个方程组成的方程组, 即可得各未知参数 $\theta_i\ (i=1,2,\cdots,k)$ 的最大似然估计值 $\hat{\theta}_i$.

例7.1.6 设$X_1, X_2, \ldots, X_n$是取自总体$X \sim B(m, p)$的样本,求参数p $(0 < p < 1)$的最大似然估计量.

解: 设样本取值为 $x_1, x_2, \ldots, x_n$. 因为$X_i \sim B(m, p)$,

X_i各自的分布律为$P\{X_i = x_i\} = \binom{m}{x_i} p^{x_i}(1-p)^{m-x_i}, x_i = 0,1,\cdots,m.$

$$L(p) = \prod_{i=1}^{n}\binom{m}{x_i} p^{x_i}(1-p)^{m-x_i} = [\prod_{i=1}^{n}\binom{m}{x_i}]p^{\sum_{i=1}^{n}x_i}(1-p)^{mn-\sum_{i=1}^{n}x_i},$$

$$\ln[L(p)] = \sum_{i=1}^{n}\ln\binom{m}{x_i} + \left(\sum_{i=1}^{n}x_i\right)\ln p + \left(mn - \sum_{i=1}^{n}x_i\right)\ln(1-p),$$

令

$$\frac{\mathrm{d}\left\{\ln\left[L(p)\right]\right\}}{\mathrm{d}p}=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{mn-\sum_{i=1}^{n}x_i}{1-p}=0,$$

从而解出真值 p 的最大似然估计值为 $\hat{p}=\frac{1}{mn}\sum_{i=1}^{n}x_i$,

p的最大似然估计量为$\hat{p}=\frac{1}{mn}\sum_{i=1}^{n}X_i=\frac{\bar{X}}{m}$.

例7.1.7 设总体X具有分布律

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中$\theta(0<\theta<1/2)$为未知参数. 已知取得了样本值3, 1, 3, 0, 3, 1, 2, 3, 试求θ的矩估计值和最大似然估计值.

解：矩估计法：

由于$\mu_1 = E(X) = 0\times\theta^2 + 1\times 2\theta(1-\theta) + 2\times\theta^2 + 3\times(1-2\theta) = 3-4\theta$，

令$\mu_1 = A_1$, 即$3-4\theta = \overline{X}$,

解得$\hat{\theta} = \dfrac{1}{4}(3-\overline{X})$为真值$\theta$的矩估计量.

又因为样本值分别为3, 1, 3, 0, 3, 1, 2, 3,

所以 $\overline{x} = \dfrac{1}{8}(3+1+3+0+3+1+2+3) = 2$,

代替A_1，得到真值θ的矩估计值为$\hat{\theta} = \dfrac{1}{4}(3-\overline{x}) = 0.25$.

例7.1.7 设总体X具有分布律

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

其中$\theta(0<\theta<1/2)$为未知参数. 已知取得了样本值3, 1, 3, 0, 3, 1, 2, 3, 试求θ的矩估计值和最大似然估计值.

解：最大似然估计法：

对于给定的样本值，似然函数为

$$L(\theta)=\theta^2\times[2\theta(1-\theta)]^2\times\theta^2\times(1-2\theta)^4=4\theta^6(1-\theta)^2(1-2\theta)^4$$

取对数，得到

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

对未知参数θ求导得到

$$\frac{\mathrm{d}[\ln L(\theta)]}{\mathrm{d}\theta}=\frac{6}{\theta}-\frac{2}{1-\theta}-\frac{8}{1-2\theta}=\frac{6-28\theta+24\theta^2}{\theta(1-\theta)(1-2\theta)}\overset{\text{令}}{=}0,\ \text{解得}\hat{\theta}=\frac{7-\sqrt{13}}{12},$$

故θ的最大似然估计值为$\hat{\theta}=\dfrac{7-\sqrt{13}}{12}\approx 0.2829.$

例7.1.8 从一大批产品中随机抽取n件, 发现其中有k件次品, 求这批产品的次品率的最大似然估计.

解: 设该批产品的次品率为p, 从这批产品中随机抽取一件产品时, 对这件产品的检验结果可用随机变量

$$X=\begin{cases}0, & \text{产品是合格品,}\\ 1, & \text{产品是次品}\end{cases}$$

表示, 则X服从(0-1)分布, 概率分布为

$$P\{X=x\}=p(x;p)=p^{x}(1-p)^{1-x},\ x=0,1.$$

以X为总体, 从中抽取样本观测值 $x_1,\ x_2,\ \ldots\ x_n$, 则似然函数为

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

取对数得,

例7.1.8 从一大批产品中随机抽取n件，发现其中有k件次品，求这批产品的次品率的最大似然估计.

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

由 $$\frac{d\ln L(p)}{dp}=\frac{1}{p}\sum_{i=1}^{n}x_i-\frac{1}{1-p}\left(n-\sum_{i=1}^{n}x_i\right)$$

$$=\frac{1}{p(1-p)}\left(\sum_{i=1}^{n}x_i-p\sum_{i=1}^{n}x_i-np+p\sum_{i=1}^{n}x_i\right)$$

$$=\frac{1}{p(1-p)}\left(\sum_{i=1}^{n}x_i-np\right)=0,$$

解得p的最大似然估计值为 $\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}.$

由于$\sum_{i=1}^{n}x_i=k$，所以这批产品的次品率的最大似然估计值为$\hat{p}=\frac{k}{n}.$

例7.1.9 设 X 服从参数为 $\lambda\ (\lambda>0)$ 的泊松分布，$X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本，求 λ 的最大似然估计量.

解 设样本值为$x_1, x_2, \ldots, x_n$.

因为X的分布律为

$$P\{X=x\}=\frac{\lambda^x}{x!}\mathrm{e}^{-\lambda},\quad (x=0,1,2,\cdots)$$

所以 λ 的似然函数为

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

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

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

解得 λ 的最大似然估计值 $$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

λ 的最大似然估计量为 $$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

Q: 与矩估计法的结果一样么？

例7.1.10 设总体X服从指数分布，其概率密度为：

$$f(x;\lambda)=\begin{cases}\lambda e^{-\lambda x}, x>0\\ 0, \quad\quad 其它\end{cases}，\ \lambda>0未知，$$

设$X_1,X_2,\ldots,X_n$是总体X的一个样本，求λ的最大似然估计量.

解：(1) 设样本值为$x_1,x_2,\ldots,x_n$.

作似然函数：

$$f_{X_i}(x_i)=f(x_i;\lambda)=\begin{cases}\lambda e^{-\lambda x_i}, x_i>0\\ 0, \quad\quad 其它\end{cases}.$$

$$L(\lambda)=f(x_1,x_2,\ldots,x_n)=\prod_{i=1}^{n}f_{X_i}(x_i)=\prod_{i=1}^{n}f(x_i;\lambda)$$

$$=\begin{cases}\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}x_i}, x_i>0, i=1,2,\cdots,n.\\ 0, \quad\quad\quad 其它\end{cases}$$

$$L(\lambda)=\begin{cases}\lambda^n e^{-\lambda\sum\limits_{i=1}^{n}x_i}, & x_i>0, i=1,2,\cdots,n.\\ 0, & \text{其它}\end{cases}$$

(2) 求对数：

当$x_i > 0$时，$L(\lambda) > 0$,取对数似然函数得到

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

(3) 求导求驻点：

$$令\frac{\mathrm{d}\left[\ln L(\lambda)\right]}{\mathrm{d}\lambda}=n\times\frac{1}{\lambda}-\sum_{i=1}^{n}x_i=0$$

(4) 求解：

解得λ的最大似然估计值为 $\hat{\lambda}=\dfrac{n}{\sum\limits_{i=1}^{n}x_i}=\dfrac{1}{\overline{x}}$,

最大似然估计量为$\hat{\lambda}=\dfrac{n}{\sum\limits_{i=1}^{n}X_i}=\dfrac{1}{\overline{X}}$.

例7.1.11 设总体X具有以下概率密度:

$$f(x;\theta)=\begin{cases}\theta x^{\theta-1}, 0<x<1\\ 0, \quad\quad 其它\end{cases}, \quad \theta>0$$

设$X_1,X_2,\ldots,X_n$是总体X的一个样本,求θ的最大似然估计量.

解:(1) 设样本值为$x_1,x_2,\ldots,x_n$.

作似然函数:

$$f_{X_i}(x_i)=f(x_i;\theta)=\begin{cases}\theta x_i^{\theta-1}, 0<x_i<1\\ 0, \quad 其它\end{cases}.$$

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

$$=\begin{cases}\theta^n\left(x_1\cdots x_n\right)^{\theta-1}, 0<x_i<1\\ 0, 其它\end{cases}, i=1,2,\ldots,n.$$

$$L(\theta)=\begin{cases}\theta^n\left(x_1\cdots x_n\right)^{\theta-1}, 0<x_i<1\\ 0, \quad 其它\end{cases}, i=1,2,\ldots,n..$$

(2) 求对数：

当$0<x_i<1$时，$L(\theta)>0$,取对数似然函数得到

$$\ln L(\theta)=n\ln\theta+(\theta-1)\left[\sum_{i=1}^{n}\ln x_i\right];$$

(3) 求导求驻点：

$$令\frac{\mathrm{d}\left[\ln L(\theta)\right]}{\mathrm{d}\theta}=n\times\frac{1}{\theta}+\sum_{i=1}^{n}\ln x_i=0$$

(4) 求解：

解得θ的最大似然估计值为 $\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln x_i}$,

最大似然估计量为$\hat{\theta}=-\dfrac{n}{\sum_{i=1}^{n}\ln X_i}$.

3. 总体的分布中含有多个未知参数的情形

设总体 X 的分布中含有 r 个参数$\theta_1, \theta_2, \ldots \theta_r$，则似然函数仍然是这些未知参数的函数

$$L = L(\theta_1, \theta_2, \cdots, \theta_r),$$

求L 或$\ln L$关于θ_i 的偏导数, 并令它们等于0, 得（或对数）似然方程组:

$$\frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, r \quad \text{或} \quad \frac{\partial \ln L}{\partial \theta_i} = 0, \quad i = 1, 2, \cdots, r$$

解此方程组得θ_i 的最大似然估计值和最大似然估计量.

例 7.1.12 设总体 $X \sim N(\mu, \sigma^2)$, μ, σ^2为未知参数， $x_1, x_2, \cdots, x_n$ 是来自X的一个样本值，求μ和σ^2的最大似然估计量.

解 X 的概率密度为

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

似然函数为

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

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

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

$$令\begin{cases}\dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2)=0,\\[2ex]\dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2)=0,\end{cases}$$

$$\begin{cases}\dfrac{1}{\sigma^2}\left[\displaystyle\sum_{i=1}^{n}x_i-n\mu\right]=0\ ,\\[2ex]-\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2=0\ ,\end{cases}$$

由 $\frac{1}{\sigma^2}\left[\sum_{i=1}^{n} x_i - n\mu\right] = 0$ 解得 $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$,

由 $-\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_{i=1}^{n}(x_i - \mu)^2 = 0$ 解得

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

故 μ 和 σ^2 的最大似然估计量分别 为

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

例 7.1.13 设总体 X 在 $[a,b]$ 上服从均匀分布，其中 a, b 未知，$x_1,x_2,\cdots,x_n$ 是来自总体 X 的一个样本值，求 a,b 的最大似然估计量.

解 X 的概率密度为

$$f(x;a,b)=\begin{cases}\dfrac{1}{b-a}, & a\le x\le b,\\ 0, & \text{其他}.\end{cases}$$

作为 a,b 的函数的似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_i\le b, i=1,\cdots,n.\\ 0, & \text{其它}\end{cases}$$

$L(a,b)$ 关于 a 单调增，关于 b 单调减。

记 $x_{(1)}=\min(x_1,x_2,\cdots,x_n)$,

$$x_{(n)}=\max(x_1,x_2,\cdots,x_n),$$

因为$a\le x_1,x_2,\cdots,x_n\le b$等价于$a\le x_{(1)},x_{(n)}\le b$,

所以似然函数可写为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a\le x_{(1)},\ b\ge x_{(n)},\\ 0, & \text{其他}\end{cases}$$

于是对于满足条件 $a\le x_{(1)},\ b\ge x_{(n)}$的任意$a,b$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 $(x_{(n)} - x_{(1)})^{-n}$,

a, b 的最大似然估计值

$$\hat{a} = x_{(1)} = \min_{1 \le i \le n} x_i, \qquad \hat{b} = x_{(n)} = \max_{1 \le i \le n} x_i,$$

a, b 的最大似然估计量

$$\hat{a} = \min_{1 \le i \le n} X_i, \qquad \hat{b} = \max_{1 \le i \le n} X_i .$$

4. 最大似然估计的性质

设 θ 的函数 $u=u(\theta),\theta\in\Theta$ 具有单值反函数 $\theta=\theta(u),u\in U$. 又设 $\hat{\theta}$ 是 X 的概率密度函数 $f(x;\theta)$ (f 形式已知) 中的参数 θ 的最大似然估计, 则 $\hat{u}=u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计.

此性质可以推广到总体分布中含有多个未知参数的情况.

如例7.1.12中, σ^2 的最大似然估计值为 $\hat{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$,

函数 $u=u(\sigma^2)=\sqrt{\sigma^2}$ 有单值反函数 $\sigma^2=u^2(u\geq 0)$,

故标准差 σ 的最大似然估计值为

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

三、小结

两种求点估计的方法: $\begin{cases} \text{矩估计法} \\ \text{最大似然估计法} \end{cases}$

在统计问题中往往先使用最大似然估计法,
在最大似然估计法使用不方便时, 再用矩估计法.

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

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