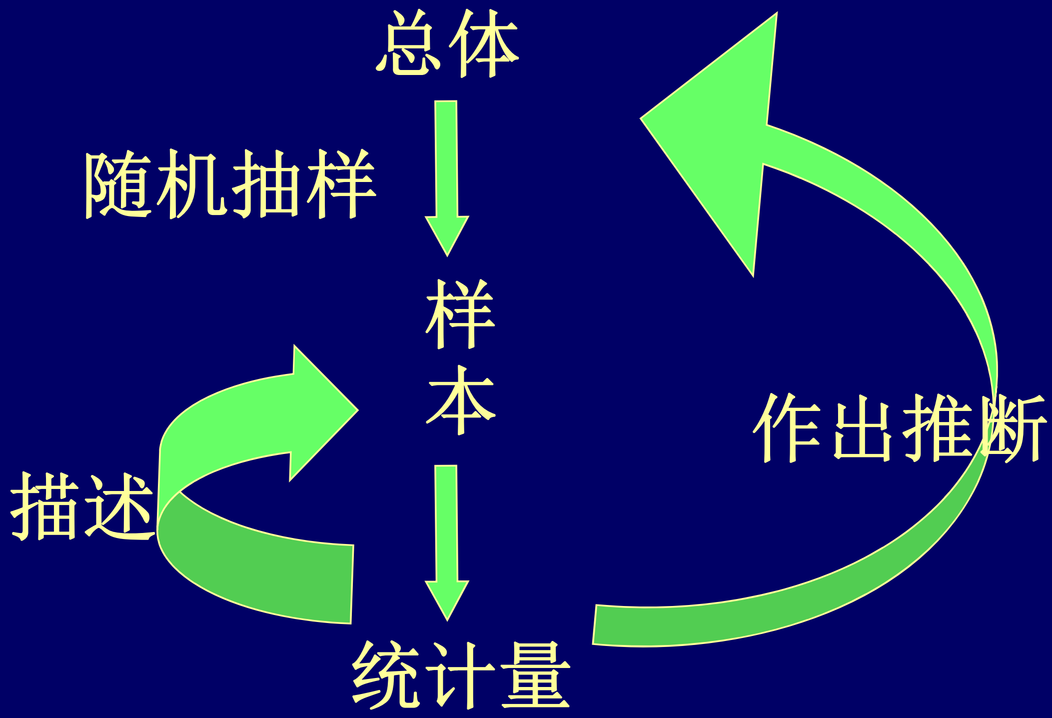
第7章 参数估计

引言

上一章，我们介绍了总体、样本、简单随机样本、统计量和抽样分布的概念，介绍了统计中常用的三大分布，给出了几个重要的抽样分布定理．它们是进一步学习统计推断的基础．



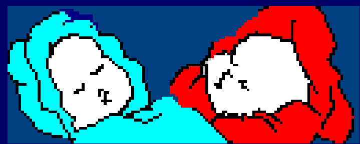
研究统计量的性质和评价一个统计推断的优良性，完全取决于其抽样分布的性质.

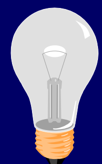
- 统计推断的基本问题可分为参数估计、假设检验、预测预报等等
  - 参数估计（§7）：用样本统计量去估计总体的未知参数。
  - 假设检验（§8）：先对总体参数提出一个假设值，然后利用样本信息判断这一假设是否成立。

参数估计

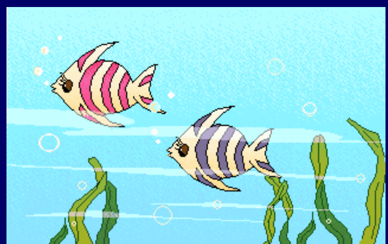
现在我们来介绍一类重要的统计推断问题

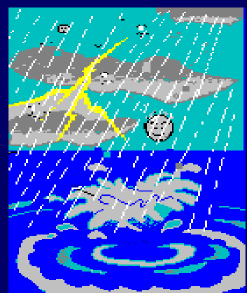
参数估计问题是利用从总体抽样得到的信息来估计总体的某些参数或者参数的某些函数.

估计新生儿的体重 

估计废品率 

估计湖中鱼数　估计降雨量





…

…

在参数估计问题中，假定总体分布形式已知，未知的仅仅是一个或几个参数.

参数估计问题的一般提法

设有一个统计总体，总体的分布函数为 F(x,θ)，其中θ为未知参数 (θ可以是向量) . 现从该总体抽样，得样本

$$X_1,X_2,\ldots,X_n$$

要依据该样本对参数 θ 作出估计, 或估计 θ 的某个已知函数 $g(\theta)$.

这类问题称为参数估计.

参数估计 $\begin{cases} \text{点估计} \\ \\ \text{区间估计} \end{cases}$

例如我们要估计某队男生的平均身高.

（假定身高服从正态分布 $N(\mu,0.1^2)$）

现从该总体选取容量为5的样本，我们的任务是要根据选出的样本（5个数）求出总体均值 μ 的估计. 而全部信息就由这5个数组成 .

设这5个数是:

1.65 1.67 1.68 1.78 1.69

估计 μ 为1.68， 这是点估计.

估计 μ 在区间 [1.57, 1.84] 内，这是区间估计.

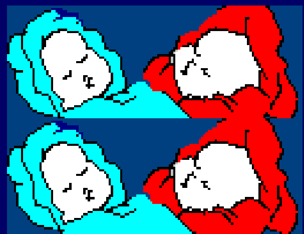
第一节 参数的点估计

- 点估计概念
- 求估计量的方法
- 课堂练习
- 小结 布置作业

一、点估计概念

例1 已知某地区新生婴儿的体重 $X \sim N(\mu,\sigma^2)$,

$(\mu,\sigma$ 未知 $)$

 … 

随机抽查100个婴儿 ,得100个体重数据

10,7,6,6.5,5,5.2, …

而全部信息就由这100个数组成 .

据此, 我们应如何估计 μ 和 σ 呢 ?

为估计 μ:

我们需要构造出适当的样本的函数 $T(X_1,X_2,\ldots X_n)$ ，每当有了样本，就代入该函数中算出一个值，用来作为 μ 的估计值 .

$T(X_1,X_2,\ldots X_n)$ 称为参数 μ 的点估计量，

记做 $\hat{\mathrm{T}}(\mathrm{X}_1,\mathrm{X}_2,\cdots\mathrm{X}_n)$

把样本值代入 $T(X_1,X_2,\ldots X_n)$ 中，得到 μ 的一个点估计值 . 记做 $\hat{\mathrm{T}}(\mathrm{x}_1,\mathrm{x}_2,\cdots\mathrm{x}_n)$，简记做 $\hat{\mathrm{T}}$。

我们知道, 若 $X \sim N(\mu, \sigma^2)$, 则 $E(X) = \mu$.

由大数定律,

样本体重的平均值

$$\lim_{n\to\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} X_i - \mu| < \varepsilon\} = 1$$

自然想到把样本体重的平均值作为总体平均体重的一个估计.

用样本体重的均值 $\overline{X}$ 估计 μ.

类似地，用样本体重的方差 S^2 估计 σ^2.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

问题是:

使用什么样的统计量去估计 μ ?

可以用样本均值;

也可以用样本中位数;

还可以用别的统计量 .

二、寻求估计量的方法

1. 矩估计法

2. 极大似然法

3. 最小二乘法

4. 贝叶斯方法 ……

这里我们主要介绍前面两种方法 .

【复习】原点矩　　中心矩

定义　设X是随机变量，若 $E(X^k), k=1,2,\cdots$ 存在，称它为X的k阶原点矩，简称 k阶矩

若 $E\{[X-E(X)]^k\}, k=2,3,\cdots$ 存在，称它为X的k阶中心矩

- 均值 E(X) 是 X 的一阶原点矩
- 方差 D(X) 是 X 的二阶中心矩

设X是总体，$X_1, X_2, \ldots. X_n$是X的一个样本.

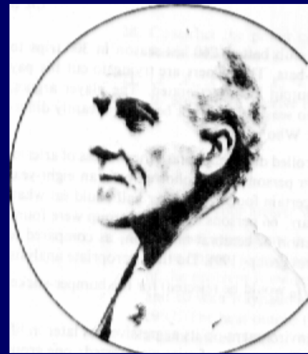
$E(X^k)$ 称为总体X的k阶原点矩μ_k;

$\frac{1}{n}\sum_{i=1}^{n} X_i^k$ 称为样本的k阶原点矩A_k.

$E(X-\mu)^k$ 称为总体X的k阶中心矩;

$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^k$ 称为样本的k阶中心矩B_k;

1. 矩估计法



矩估计法是英国统计学家*K.*皮尔逊最早提出来的．由辛钦定理，

若总体 X 的数学期望 $E(X)=\mu$ 有限，

则有 $$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} E(X)=\mu$$

弱大数定理（辛钦大数定理）　设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立，服从同一分布且具有数学期望 $E(X_k)=\mu\ (k=1,2,\cdots)$，则序列 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于 μ，即 $\overline{X}\xrightarrow{P}\mu$.

$X_1^k,X_2^k,\cdots,X_n^k$独立且与X^k同分布，故有$E(X_1^k)=E(X_2^k)=\cdots=E(X_n^k)=\mu_k$

$\Downarrow$

样本k阶矩　　　　　　　　　　　　总体k阶矩

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X^k)=\mu_k\ (k=1,2,\cdots)$$

1. 矩估计法

若总体 X 的数学期望 $E(X)=\mu$ 有限，则有

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i \xrightarrow{P} E(X)=\mu$$

$$\Downarrow$$

$$A_k=\frac{1}{n}\sum_{i=1}^{n}X_i^k \xrightarrow{P} E(X^k)=\mu_k \ (k=1,2,\cdots)$$

再由依概率收敛性质知，可将上述性质推广为

$$g(A_1,A_2,\cdots,A_k)\xrightarrow{p} g(\mu_1,\mu_2,\cdots,\mu_k)$$

其中g为连续函数.

设$X_n \xrightarrow{P} a, Y_n \xrightarrow{P} b$，又设函数$g(x,y)$在点$(a,b)$连续，

则$g(X_n,Y_n)\xrightarrow{P} g(a,b)$

【定义】用样本原点矩估计相应的总体原点矩，又用样本原点矩的连续函数估计相应的总体原点矩的连续函数，这种参数点估计法称为矩估计法．

理论依据：大数定律

特点：不需要事先知道总体是什么分布．

【例】设总体X的概率密度为

$$f(x;\theta)=\begin{cases}(\theta+1)x^{\theta} & 0<x<1 \\ 0 & \text{其它}\end{cases} \qquad (\theta>-1)$$

$X_1,X_2,\dots,X_n$是来自总体X的样本。0.1,0.2,0.9,0.8,0.7,0.7为一个样本观察值，试求θ的矩估计值。

解：$$E(X)=\int_{-\infty}^{+\infty}xf(x;\theta)dx=\int_0^1 x(\theta+1)x^{\theta}dx=(\theta+1)\int_0^1 x^{\theta+1}dx$$

$$=(\theta+1)\frac{1}{\theta+2}x^{\theta+2}\Big|_0^1=\frac{\theta+1}{\theta+2}=\mu_1$$

总体矩

令 $\dfrac{\theta+1}{\theta+2}=\overline{X}$ 　解之得 θ的矩估计 $\hat{\theta}=\dfrac{2\overline{X}-1}{1-\overline{X}}$

样本矩

由样本值 0.1,0.2,0.9,0.8,0.7,0.7计算得 $\overline{x}=0.5667$

故θ的矩估计值为 $\hat{\theta}=\dfrac{2\overline{x}-1}{1-\overline{x}}=0.3079$

矩估计法的具体做法

设总体的分布函数中含有k个未知参数 $\theta_1,\theta_2,\cdots,\theta_k$,那么它的前$k$阶矩 $\mu_1,\mu_2,\cdots,\mu_k$,一般都是这 k 个参数的函数,记为:

$$\mu_i=\mu_i(\theta_1,\theta_2,\cdots,\theta_k) \qquad i=1,2,\ldots,k$$

从这 k 个方程中解出

$$\theta_j=\theta_j(\mu_1,\mu_2,\cdots,\mu_k) \qquad j=1,2,\ldots,k$$

那么用诸 μ_i 的估计量 A_i 分别代替上式中的诸 μ_i ,即可得诸 θ_j 的矩估计量 :

$$\hat{\theta}_j=\theta_j(A_1,A_2,\cdots,A_k) \qquad j=1,2,\ldots,k$$

矩估计量的观察值称为矩估计值 .

例2 设总体 X 在 $[a, b]$ 上服从均匀分布，a, b 未知．$X_1,\ldots,X_n$ 是来自 X 的样本，试求 a, b 的矩估计量．

解 $\mu_1 = E(X) = \dfrac{a+b}{2}$

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

$$= \frac{(b-a)^2}{12} + \frac{(a+b)^2}{4} \quad \mu_1^2$$

即 $\begin{cases} a+b = 2\mu_1 \\ b-a = \sqrt{12(\mu_2 - \mu_1^2)} \end{cases}$ 解得 $a = \mu_1 - \sqrt{3(\mu_2 - \mu_1^2)}$

$b = \mu_1 + \sqrt{3(\mu_2 - \mu_1^2)}$

总体矩

$$a=\mu_1-\sqrt{3(\mu_2-\mu_1^2)}\qquad b=\mu_1+\sqrt{3(\mu_2-\mu_1^2)}$$

分别以A_1, A_2代替μ_1, μ_2,

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

$$a=A_1-\sqrt{3(A_2-A_1^2)}\qquad b=A_1+\sqrt{3(A_2-A_1^2)}$$

$$=\overline{X}-\sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2\right)}\qquad =\overline{X}+\sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2\right)}$$

又因为$\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2=\frac{1}{n}\sum_{i=1}^{n}[X_i^2-2X_i\overline{X}+(\overline{X})^2]$$

$$=\frac{1}{n}\{\sum_{i=1}^{n}X_i^2-2\sum_{i=1}^{n}X_i\overline{X}+\sum_{i=1}^{n}(\overline{X})^2\}=\frac{1}{n}\{\sum_{i=1}^{n}X_i^2-2n(\overline{X})^2+n(\overline{X})^2\}$$

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

$$a=\mu_1-\sqrt{3(\mu_2-\mu_1^2)} \qquad b=\mu_1+\sqrt{3(\mu_2-\mu_1^2)}$$

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

分别以A_1, A_2代替μ_1, μ_2,

$$a=A_1-\sqrt{3(A_2-A_1^2)} \qquad b=A_1+\sqrt{3(A_2-A_1^2)}$$

$$=\overline{X}-\sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2\right)} \qquad =\overline{X}+\sqrt{3\left(\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2\right)}$$

又因为$\frac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$

于是 a, b 的矩估计量为

$$\hat{a}=\overline{X}-\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}, \quad \hat{b}=\overline{X}+\sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2}$$

样本矩

样本2阶中心矩

注意：

1　定义中选用的是原点矩，也可以用中心矩，只要给定总体矩，采用相应的样本矩就可以。

2　若$\hat{\theta}$为θ的矩估计量，$g(\theta)$为θ的连续函数，亦称$g(\hat{\theta})$为$g(\theta)$的矩估计量。

例如，B_2为总体方差σ^2的矩估计量，

则$\sqrt{B_2}$为标准差σ的矩估计量。

3　矩估计的关键是计算总体矩，因此使用矩估计法其前提是总体矩必须存在。

例 设$X_1,X_2,\dots X_n$是取自总体 X 的一个样本

$$X \sim f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu \\ 0, & 其它\end{cases}$$ θ,μ为未知参数

其中$\theta>0$，求 θ,μ 的矩估计.

解 由密度函数可知，$X-\mu$具有均值为 θ 的指数分布

故 $\begin{cases}E(X-\mu)=\theta \\ D(X-\mu)=\theta^2\end{cases}$ 即 $\begin{cases}E(X)=\mu+\theta \\ D(X)=\theta^2\end{cases}$

解得 $\theta=\sqrt{D(X)}$ $\mu=E(X)-\sqrt{D(X)}$

于是 θ,μ 的矩估计量为 $\begin{cases}\hat{\theta}=\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2} \\ \hat{\mu}=\overline{X}-\sqrt{\dfrac{1}{n}\sum\limits_{i=1}^{n}(X_i-\overline{X})^2}\end{cases}$

例3 设总体 X 的均值 μ 和方差 $\sigma^2(>0)$ 都存在，μ,σ^2 未知．$X_1,\ldots,X_n$ 是来自 X 的样本，试求 μ,σ^2 的矩估计量．

总体均值与方差的矩估计量的表达式
不因不同的总体分布而异

解 $\mu_1 = E(X) = \mu$

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

解得 $\mu = \mu_1$ $\sigma^2 = \mu_2 - \mu_1^2$

总体矩

于是 μ,σ^2 的矩估计量为

$$\hat{\mu} = A_1 = \bar{X}$$

样本矩

$$\widehat{\sigma^2} = A_2 - A_1^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

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，选取哪些总体矩用相应样本矩代替带有一定的随意性 .

例如：

$$X \sim P(\lambda),\ E(X)=D(X)=\lambda$$

故 $\hat{\lambda}=\overline{X}$ 或 $\hat{\lambda}=B_2$

【例】设总体X的概率密度为

$$f(x,\theta)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}} \qquad -\infty<x<+\infty,\theta>0$$

试求 θ的矩估计量 $\hat{\theta}$。

解：虽然$f(x;\theta)$中仅含一个未知参数 θ，但因

$$E(X)=\int_{-\infty}^{+\infty}x\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=0$$

不含θ，不能由此解出，故需 继续求出总体二阶原点 矩：

$$E(X^2)=\int_{-\infty}^{+\infty}x^2\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\int_{0}^{+\infty}x^2\frac{1}{\theta}e^{-\frac{x}{\theta}}dx$$

$$=\theta^2+\theta^2=2\theta^2$$

故令$2\theta^2=A_2=\frac{1}{n}\sum_{i=1}^{n}X_i^2$， 得$\theta$的矩估计量 $\hat{\theta}=\sqrt{\frac{1}{2n}\sum_{i=1}^{n}X_i^2}$

【例】设总体X的概率密度为

$$f(x,\theta)=\frac{1}{2\theta}e^{-\frac{|x|}{\theta}} \qquad -\infty<x<+\infty,\theta>0$$

试求 θ的矩估计量 $\hat{\theta}$。

【解2】

考虑$|X|$的数学期望，

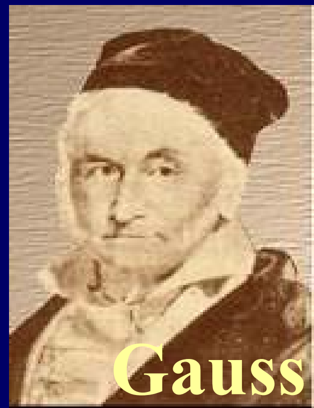
$$E(|X|)=\int_{-\infty}^{+\infty}|x|\frac{1}{2\theta}e^{-\frac{|x|}{\theta}}dx=\int_0^{+\infty}x\frac{1}{\theta}e^{-\frac{x}{\theta}}dx=\theta$$

可得出θ的另一矩估计量$\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}|X_i|$

2. 最大似然法

它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．

最大似然法的基本思想

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.

一般说，事件A发生的概率与参数$\theta\in\Theta$有关，θ取值不同，则P(A)也不同。因而应记事件A发生的概率为$P(A|\theta)$.若A发生了，则认为此时的θ值应是在Θ中使$P(A|\theta)$达到最大的那一个。这就是**极大似然思想**

例1 袋中放有白球和黑球共4个，今进行3次有放回抽样，每次抽取1个，结果抽得2次白球1次黑球，试估计袋中白球个数。

解 设袋中白球个数为m，

X为3次抽样中抽得的白球数，则

$$X \sim b(3,p), \quad p = m/4$$

当袋中白球数m分别为1, 2, 3时，

p对应的值分别为1/4, 2/4, 3/4，

X对应的分布律见下表

袋中白球数m	p	抽到白球数x			
		x=0	x=1	x=2	x=3
1	1/4	27/64	27/64	9/64	1/64
2	2/4	8/64	24/64	24/64	8/64
3	3/4	1/64	9/64	27/64	27/64

当p=3/4时，P{X=2}的概率最大，

⇒估计袋中白球个数为3比较合理。

【最大似然估计的基本思想】当从模型总体随机抽取n组样本观测值后，最合理的参数估计量应该使得从模型中抽取该n组样本观测值的概率最大。

【例】假如有一个罐子，里面有黑白两种颜色的球，数目未知，两种颜色的比例未知，且不能把罐中的球全部拿出来数，该如何求罐中白球和黑球的比例？

[方法] 每次任意从已经摇匀的罐中拿一个球出来，记录球的颜色，然后把拿出来的球再放回罐中。重复此过程，用记录的球的颜色来估计罐中黑白球的比例。

【问题】 若在一百次记录中，有七十次是白球，请问罐中白球所占的比例最有可能是多少？ **【答案】70%**

【分析】假设罐中白球的比例是p，那么黑球的比例就是1-p。

- 把一次抽出来球的颜色称为一次抽样，如果第一抽样的结果记为x_1，第二抽样的结果记为x_2... 那么Data = $(x_1,x_2,\ldots,x_{100})$（独立同分布）。
- 在一百次抽样中，七十次是白球的概率是P(Data | M)，其中M就是这个罐子（即题目所给出的模型）。

$$
\begin{aligned}
&P(\text{Data} \mid M) \\
&= P(x_1, x_2, \ldots, x_{100} \mid M) \\
&= P(x_1 \mid M) P(x_2 \mid M) \ldots P(x_{100} \mid M) \\
&= p^{70}(1-p)^{30}.
\end{aligned}
$$

- p取何值时，$P(\text{Data} \mid M)$的值最大？
- 将$p^{70}(1-p)^{30}$对p求导，并令其等于零。

$$70p^{69}(1-p)^{30}-p^{70}\times 30(1-p)^{29}=0$$

$$p^{69}(1-p)^{29}(7-10p)=0$$

解方程可以得到$p=0.7$或$p=0$或$p=1$。

- 在边界点$p=0$、1，$P(\text{Data}|M)=0$。
- 所以，当$p=0.7$时，$P(\text{Data}|M)$的值最大。

最大似然估计原理：

设$X_1,X_2,\ldots X_n$是取自总体X的一个样本，样本的联合密度(连续型）或联合分布律 (离散型)为$f(x_1,x_2,\ldots ,x_n;\theta)$.

当给定样本$X_1,X_2,\ldots X_n$时，定义似然函数为：

$$\boxed{L(\theta)=f(x_1,x_2,\ldots,x_n;\theta)}$$

这里 $x_1,x_2,\ldots,x_n$ 是样本的观察值 .

似然函数：

$$\boxed{L(\theta)=f(x_1,x_2,\ldots,x_n;\theta)}$$

$L(\theta)$看作参数θ的函数，它可作为θ将以多大可能产生样本值$x_1,x_2,\ldots,x_n$的一种度量.

最大似然估计法就是用使$L(\theta)$达到最大值的$\hat{\theta}$去估计θ.

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

称$\hat{\theta}$为θ的最大似然估计值．而相应的统计量$\hat{\theta}(X_1,\ldots,X_n)$称为$\theta$的最大似然估计量.

两点说明：

1、求似然函数$L(\theta)$的最大值点，可以应用微积分中的技巧。由于ln(x)是 x 的增函数，ln$L(\theta)$与$L(\theta)$在θ的同一值处达到它的最大值，假定θ是一实数，且ln$L(\theta)$是θ的一个可微函数。通过求解方程：

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

可以得到θ的MLE .

若 θ 是向量，上述方程必须用方程组代替 .

2、用上述求导方法求参数的MLE有时行不通，这时要用最大似然原则来求 .

下面举例说明如何求最大似然估计

例5 设$X_1,X_2,\ldots X_n$是取自总体 $X\sim B(1,p)$ 的一个样本，求参数p的最大似然估计量.

解：似然函数为:

$$X_i \sim \begin{Bmatrix} 0 & 1 \\ 1-p & p \end{Bmatrix}$$

$$L(p)=f(x_1,x_2,\ldots,x_n;p)$$

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

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

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

对数似然函数为：

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

对p求导并令其为0，

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

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

从而 p 的最大似然估计量为

$$\hat{p}(X_1,\ldots,X_n)=\frac{1}{n}\sum_{i=1}^{n}X_i=\overline{X}$$

求最大似然估计(MLE)的一般步骤是:

(1) 由总体分布导出样本的联合分布率(或联合密度);

(2) 把样本联合分布率 (或联合密度) 中自变量看成已知常数,而把参数θ看作自变量,得到似然函数$L(\theta)$;

(3) 求似然函数$L(\theta)$的最大值点(常常转化为求$\ln L(\theta)$的最大值点) ，即θ的MLE;

(4) 在最大值点的表达式中, 用样本值代入就得参数的最大似然估计值 .

例8 设总体X的分布律为

X	1	2	3
p_k	θ	θ	$1-2\theta$

其中参数$\theta>0$未知，现有一组样本值**1,1,1,3,2,1,3,2,2,1,2,2,3,1,1,2**

试求$\boldsymbol{\theta}$的矩估计值和最大似然估计值。

解：（**1**）$\mu_1 = E(X) = \theta + 2\theta + 3(1-2\theta) = 3 - 3\theta$

解得$\theta = 1 - \frac{\mu_1}{3}$，

故$\boldsymbol{\theta}$的矩估计量为$\hat{\theta} = 1 - \frac{\bar{X}}{3}$

上述样本值的$\bar{x} = 7/4$，故$\boldsymbol{\theta}$的矩估计值为**5/12.**

（**2**）$L(\theta) = \prod_{i=1}^{n} p(x_i,\theta) = (P\{X=1\})^7 (P\{X=2\})^6 (P\{X=3\})^3 = \theta^7\theta^6(1-2\theta)^3$

$$\ln L(\theta) = 13\ln\theta + 3\ln(1-2\theta)$$

$$\frac{\partial}{\partial\theta}\ln L(\theta) = \frac{13}{\theta} - \frac{6}{1-2\theta} = 0$$

$$\theta_{MLE} = \frac{13}{32}$$

例3 设 $X_1, X_2, \ldots, X_n$ 是取自总体 X 的一个样本，

$X \sim \pi(\lambda)$，求参数λ的最大似然估计量。

解 $P\{X = x_i\} = \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!}$ $, x_i = 0,1,\cdots$

似然函数为：$L(\lambda) = \prod_{i=1}^{n} \dfrac{\lambda^{x_i} e^{-\lambda}}{x_i!}$

$$\ln L(\lambda) = \sum_{k=1}^{n} (x_k \ln \lambda + \ln e^{-\lambda} - \ln x_k!)$$

$$= \sum_{k=1}^{n} (x_k \ln \lambda - \lambda - \ln x_k!)$$

$$\frac{d \ln L(\lambda)}{d\lambda} = \sum_{k=1}^{n} (\frac{x_k}{\lambda} - 1) = 0 \quad \Rightarrow \hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}$$

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

例6 设总体 $X\sim N(\mu,\sigma^2)$，μ,σ^2 未知．$x_1,\dots,x_n$ 是来自 X 的样本值，试求 μ,σ^2 的最大似然估计量．

解 X 的概率密度为

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

似然函数为

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

$$=(2\pi)^{-n/2}(\sigma^2)^{-n/2}\exp[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2]$$

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

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

令 $$\frac{\partial}{\partial\mu}LnL = \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}x_i - n\mu\right) = 0$$

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

解得 $$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n}x_i = \bar{x}$$

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

μ,σ^2 的最大似然估计量为

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

例： 设总体 $X \sim U[a,b]$，$(x_1, x_2, \cdots, x_n)$ 为一样本值，求 a, b 的极大似然估计.

解： X 的概率密度 $f(x;a,b)=\begin{cases} 1/(b-a), & a \le x \le b, \\ 0, & 其它. \end{cases}$

似然函数 $L(a,b)=\prod_{i=1}^{n} f(x_i;a,b)$

$$=\begin{cases} 1/(b-a)^n, & a \le x_1, x_2, \cdots, x_n \le b, \\ 0, & 其它. \end{cases}$$

利用求导方法无法确定未知参数的极大似然估计，

由 $L(a,b)$ 的表达式知：若 $b-a$ 取最小，则 $L(a,b)$ 达到最大，

故得 $\hat{a}=\min\limits_{1\le i\le n}\{x_i\}$，$\hat{b}=\max\limits_{1\le i\le n}\{x_i\}$

例7 设$X_1,X_2,\dots X_n$是取自总体X的一个样本

$$X \sim f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu\\ 0, & \text{其它}\end{cases}$$ θ,μ为未知参数

其中$\theta>0$,求θ,μ的最大似然估计.

解：似然函数为

$$L(\theta,\mu)=\begin{cases}\prod_{i=1}^{n}\dfrac{1}{\theta}e^{-(x_i-\mu)/\theta}, & x_i\ge\mu\\ 0, & \text{其它}\end{cases}$$ $i=1,2,\dots,n$

$$=\begin{cases}\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum_{i=1}^{n}(x_i-\mu)}, & \min x_i\ge\mu\\ 0, & \text{其它}\end{cases}$$

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

对数似然函数为

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

对 θ, μ 分别求偏导并令其为0,

$$\frac{\partial \ln L(\theta,\mu)}{\partial\theta} = -\frac{n}{\theta} + \frac{1}{\theta^2}\sum_{i=1}^{n}(x_i-\mu) = 0 \qquad (1)$$

$$\frac{\partial \ln L(\theta,\mu)}{\partial\mu} = \frac{n}{\theta} = 0 \qquad (2)$$

由(1)得

$$\theta = \frac{1}{n}\sum_{i=1}^{n}x_i - \mu$$

用求导方法无法最终确定 θ、μ,
用最大似然原则来求.

$$L(\theta,\mu)=\begin{cases}\dfrac{1}{\theta^n}e^{-\frac{1}{\theta}\sum\limits_{i=1}^{n}(x_i-\mu)}, & \min x_i \geq \mu\\ 0, & 其它\end{cases}$$

对 $\mu \leq \min x_i, L(\theta,\mu)>0$, 且是$\mu$的增函数

μ 取其它值时，$L(\theta,\mu)=0.$

故使$L(\theta,\mu)$达到最大的μ, 即 μ的MLE是

$$\mu^*=\min_{1\leq i\leq n} x_i$$

于是 $\theta^*=\dfrac{1}{n}\sum\limits_{i=1}^{n}x_i-\mu^*$

即 θ^*,μ^*为 θ,μ的MLE .

最后，我们用最大似然法估计湖中的鱼数

为了估计湖中的鱼数N，第一次捕上 r 条鱼，做上记号后放回. 隔一段时间后，再捕出 S 条鱼，结果发现这S条鱼中有k条标有记号.根据这个信息，如何估计湖中的鱼数呢？

第二次捕出的有记号的鱼数X是$r.v$, X具有超几何分布：

$$P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S}, 0 \le k \le \min(S, r)$$

把上式右端看作 N 的函数，记作 $L(N;k)$.

$$P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S},\ 0\le k\le \min(S,r)$$

应取使$L(N;k)$达到最大的N，作为N的极大似然估计.
但用对N求导的方法相当困难, 我们考虑比值：

$$\frac{P(X=k;N)}{P(X=k;N-1)}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S}\times\frac{C_{N-1}^S}{C_r^k C_{N-1-r}^{S-k}}=\frac{C_{N-r}^{S-k}C_{N-1}^S}{C_N^S C_{N-1-r}^{S-k}}$$

$$=\frac{\dfrac{(N-r)\cancel{!}}{\cancel{(S-k)!}(N-r-S+k)\cancel{!}}\times\dfrac{\cancel{(N-1)!}}{\cancel{S!}\cancel{(N-1-S)!}}}{\dfrac{N\cancel{!}}{\cancel{S!}(N-S)\cancel{!}}\times\dfrac{\cancel{(N-1-r)!}}{\cancel{(S-k)!}\cancel{(N-1-r-S+k)!}}}$$

$$=\frac{(N-S)(N-r)}{N(N-r-S+k)}=\frac{N^2-(S+r)N-Sr}{N^2-(S+r)N+kN}=1+\frac{Sr-kN}{N^2-(S+r)N+kN}$$

$$\frac{P(X=k;N)}{P(X=k;N-1)}=1+\frac{Sr-kN}{N^2-(S+r)N+kN}$$

经过简单的计算知，这个比值大于或小于 1，

由 $N<\frac{Sr}{k}$ 或 $N>\frac{Sr}{k}$ 而定 .

这就是说，当N增大时，序列$P(X=k;N)$先是上升而后下降；当N为小于 $\frac{Sr}{k}$ 的最大整数时，达到最大值 . 故N的极大似然估计为 $\hat{N}=[\frac{Sr}{k}]$.

三、小结

这一讲，我们介绍了参数点估计, 给出了寻求估计量最常用的矩估计法和极大似然法 .

通常，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.

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 .