概率论与数理统计

第七章　参数估计

练习：

1、从正态总体 $X \sim N(\mu,\sigma^2)$ 中，抽取了

$n = 20$的样本 $(X_1, X_2, \cdots, X_{20})$

(1) 求 $P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.76\sigma^2\right)$

(2) 求 $P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \mu\right)^2 \le 1.76\sigma^2\right)$

解 (1) $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$

即 $\frac{19S^2}{\sigma^2} = \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \sim \chi^2(19)$

故 $$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 1.76\sigma^2\right)$$

$$= P\left(7.4 \le \frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \le 35.2\right)$$

$$= P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 7.4\right) - P\left(\frac{1}{\sigma^2}\sum_{i=1}^{20}\left(X_i - \overline{X}\right)^2 \ge 35.2\right)$$

$$\overset{\text{查表}}{=} 0.99 - 0.01 = 0.98$$

(2) $$\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi^2(20)$$

故 $$P\left(0.37\sigma^2 \le \frac{1}{20}\sum_{i=1}^{20}(X_i-\mu)^2 \le 1.76\sigma^2\right)$$

$$= P\left(7.4 \le \sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \le 35.2\right)$$

$$= P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 7.4\right) - P\left(\sum_{i=1}^{20}\left(\frac{X_i-\mu}{\sigma}\right)^2 \ge 35.2\right)$$

$$= 0.995 - 0.025 = 0.97$$

2、设随机变量X与Y相互独立，$X \sim N(0,16)$, $Y \sim N(0,9)$, $X_1, X_2, \ldots, X_9$ 与$Y_1, Y_2, \ldots, Y_{16}$ 分别是取自 X 与 Y 的简单随机样本，求

统计量 $Z = \dfrac{X_1 + X_2 + \cdots + X_9}{\sqrt{Y_1^2 + Y_2^2 + \cdots + Y_{16}^2}}$ 所服从的分布.

解 $X_1 + X_2 + \cdots + X_9 \sim N(0, 9 \times 16)$

$$\frac{1}{3 \times 4}(X_1 + X_2 + \cdots + X_9) \sim N(0,1)$$

$$\frac{1}{3}Y_i \sim N(0,1) \ , i = 1, 2, \cdots, 16 \quad \sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2 \sim \chi^2(16)$$

从而 $\dfrac{X_1+X_2+\cdots+X_9}{\sqrt{Y_1^2+Y_2^2+\cdots+Y_{16}^2}}$

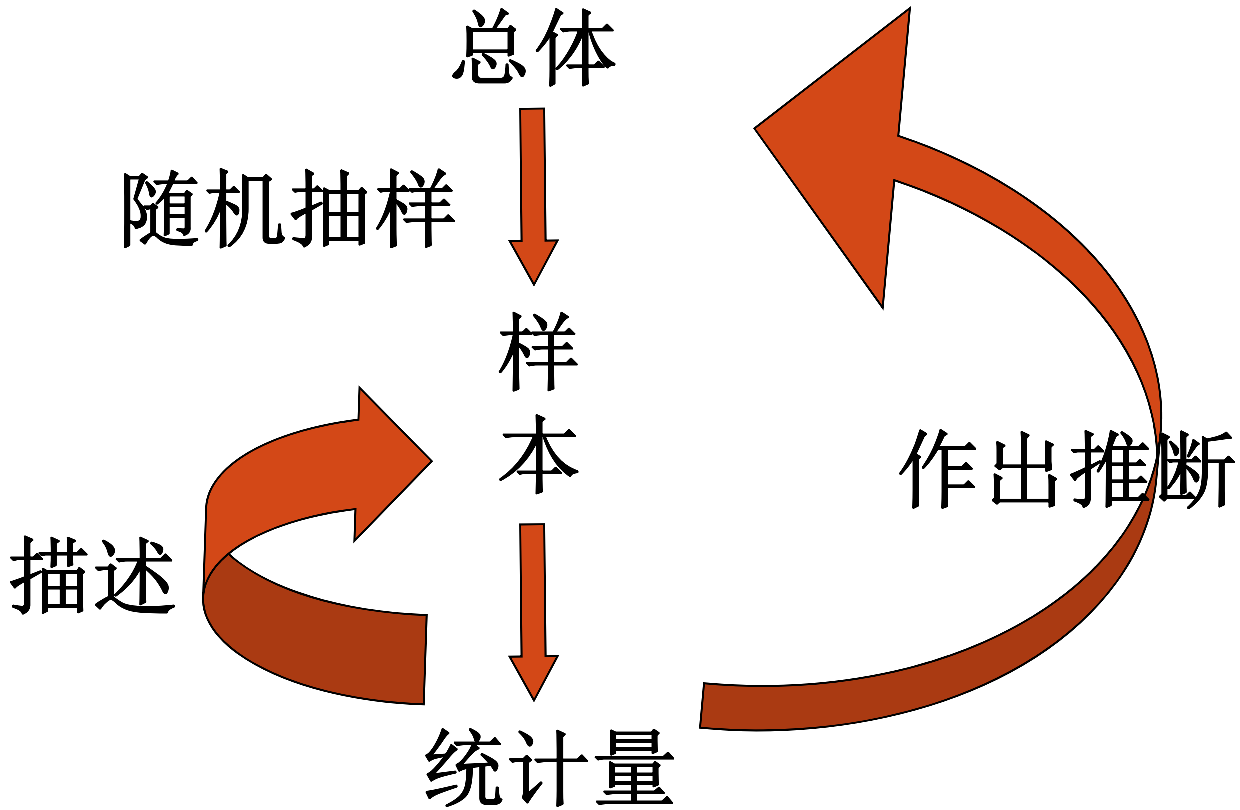
$$=\frac{\frac{1}{3\times 4}\left(X_1+X_2+\cdots+X_9\right)}{\sqrt{\dfrac{\sum_{i=1}^{16}\left(\frac{1}{3}Y_i\right)^2}{16}}}\sim t(16)$$

第七章　参数估计

- 点估计
- 基于截尾样本的最大似然估计
- 估计量的评选标准
- 区间估计
- 正态总体均值与方差的区间估计
- (0-1)分布参数的区间估计
- 单侧置信区间

引言

上一讲，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．



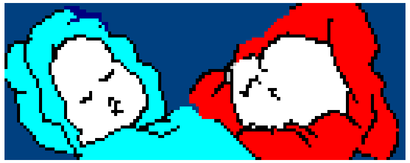
研究统计量的性质和评价一个统计推断的优良性，完全取决于其抽样分布的性质.

参数估计

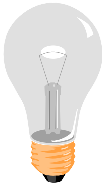
现在我们来介绍一类重要的统计推断问题

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.

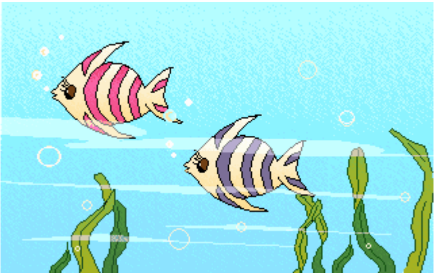
估计新生儿的体重

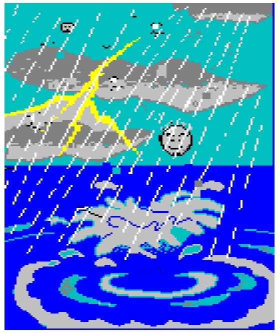


估计废品率



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.

什么是参数估计?

参数是刻画总体某方面概率特性的数量.

当此数量未知时,从总体抽出一个样本，用某种方法对这个未知参数进行估计就是参数估计.

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 $F(x,\theta)$，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.

参数估计的类型

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

点估计 —— 估计未知参数的值.

区间估计 —— 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.

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. 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.

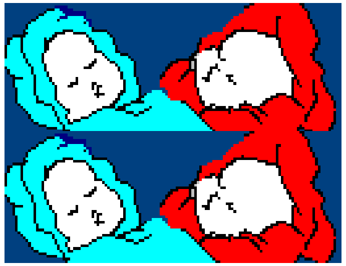
§1　点估计

- 点估计概念
- 求估计量的方法

一、点估计概念

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

(μ,σ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$f(X_1,X_2,\ldots X_n)$，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 .

$f(X_1,X_2,\ldots X_n)$称为参数μ的点估计量，

把样本值代入$f(X_1,X_2,\ldots X_n)$中，得到μ的一个点估计值 .

点估计的思想方法

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：$\theta_1, \theta_2, \ldots, \theta_k$

设 $X_1, X_2, \ldots, X_n$为总体的一个样本

构造 k 个统计量：

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

并建立k个方程。

当测得样本值$(x_1, x_2,\ldots, x_n)$时,代入上述方程组，即可得到 k 个数：

$$\left.\begin{aligned}&\hat{\theta}_1(x_1, x_2, \cdots, x_n)\\&\hat{\theta}_2(x_1, x_2, \cdots, x_n)\\&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\\&\hat{\theta}_k(x_1, x_2, \cdots, x_n)\end{aligned}\right\}\text{数 值}$$

称数 $\hat{\theta}_1 \cdots, \hat{\theta}_k$ 为未知参数 $\theta_1, \cdots, \theta_k$的估计值

对应统计量 为未知参数 $\theta_1, \cdots, \theta_k$ 的估计量

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

由大数定律，

样本的平均值

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

自然想到把样本的平均值作为总体平均值的一个估计.

用样本的均值 $\overline{X}$ 估计 μ .

类似地，用样本的方差 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 .

注：中位数：将一组数据按大小顺序排列，把处在最中间位置的一个数据（或最中间两个数据的平均数）叫做这组数据的中位数.

二、估计量的方法

1. 矩估计法

2. 最大似然估计法

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 .

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

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

$$\Downarrow$$

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

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

其中 g 为连续函数．

这表明，当样本容量很大时，在统计上，可以用用样本矩去估计总体矩．这一事实导出矩估计法.

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

矩估计法的具体做法如下

矩估计法

方法 用样本 k 阶矩作为总体 k 阶矩的估计量, 建立含有待估参数的方程, 从而解出待估参数

一般, 不论总体服从什么分布, 总体期望 μ 与方差 σ^2 存在, 则它们的矩估计量分别为

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

事实上，按矩估计法原理，令

$$\begin{cases} \overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i = \mu \\ A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2 = E(X^2) \end{cases}$$

$$\begin{cases} \hat{\mu} = \overline{X} \\ \hat{\sigma}^2 = \overset{\wedge}{E(X^2)} - \overset{\wedge}{E^2(X)} = A_2 - \hat{\mu}^2 \end{cases}$$

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

设待估计的参数为 $\theta_1, \theta_2, \cdots, \theta_k$

设总体的 r 阶矩存在，记为

$$E(X^r) = \mu_r(\theta_1, \theta_2, \cdots, \theta_k)$$

样本 $X_1, X_2, \ldots, X_n$ 的 r 阶矩为 $A_r = \dfrac{1}{n}\sum_{i=1}^{n} X_i^r$

令

$$\mu_r(\theta_1, \theta_2, \cdots, \theta_k) = \frac{1}{n}\sum_{i=1}^{n} X_i^r \quad r = 1, 2, \cdots, k$$

—— 含未知参数 $\theta_1, \theta_2, \ldots, \theta_k$ 的方程组

解方程组，得 $\boldsymbol{k}$ 个统计量：

$$\left.\begin{array}{l}\hat{\theta}_1(X_1, X_2, \cdots, X_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_k(X_1, X_2, \cdots, X_n)\end{array}\right\}$$

未知参数 $\boldsymbol{\theta_1, \ldots, \theta_k}$ 的矩估计量

代入一组样本值得 $\boldsymbol{k}$ 个数：

$$\left.\begin{array}{l}\hat{\theta}_1 = \hat{\theta}_1(x_1, x_2, \cdots, x_n) \\ \cdots\cdots\cdots\cdots\cdots\cdots \\ \hat{\theta}_k = \hat{\theta}_k(x_1, x_2, \cdots, x_n)\end{array}\right\}$$

未知参数 $\boldsymbol{\theta_1, \ldots, \theta_k}$ 的矩估计值

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　$$\mu_1 = E(X) = \frac{a+b}{2}$$

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

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

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

于是 a , b 的矩估计量为

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

样本矩

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．

解　$\mu_1 = E(X) = \mu$

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

总体矩

解得 $\mu = \mu_1$

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

于是 μ, σ^2 的矩估计量为

样本矩

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

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

矩估计法的优点是简单易行,并不需要事先知道总体是什么分布.

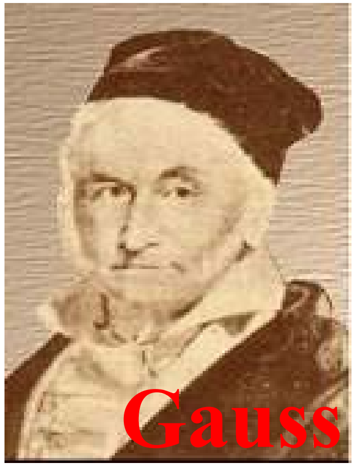
缺点是，当总体类型已知时，没有充分利用分布提供的信息.一般场合下,矩估计量不具有唯一性.

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

2. 最大似然估计法

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 尔.



费歇尔在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 .

最大似然估计法的基本思想

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最大似然估计法的基本思想．

思想方法：一次试验就出现的

事件有较大的概率.

例如: 有两外形相同的箱子,各装100个球

一箱　99个白球　　1 个红球

一箱　　1 个白球　　99个红球

现从两箱中任取一箱, 并从箱中任取一球,

结果所取得的球是白球.

问: 所取的球来自哪一箱?　　　　答: 第一箱.

最大似然估计原理:

一般, 设 X 为离散型随机变量, 其分布律为

$$P(X=x)=p(x,\theta),\quad x=u_1,u_2,\cdots,\theta\in\Theta$$

则样本 $X_1, X_2,\ldots, X_n$的概率分布为

$$P(X_1=x_1,X_2=x_2,\cdots,X_n=x_n)$$

$$=p(x_1,\theta)p(x_2,\theta)\cdots p(x_n,\theta)$$

记为

$$=L(x_1,x_2,\cdots,x_n,\theta)\overset{\text{或}}{=}L(\theta)\qquad x_i=u_1,u_2,\cdots,$$

$$i=1,2,\cdots,n,\theta\in\Theta$$

称$L(\theta)$为样本的似然函数.

最大似然估计法的思想

选择适当的$\boldsymbol{\theta}=\hat{\theta}$,使 $L(\theta)$取最大值, 即

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

$$=\max_{\theta\in\Theta}\{\, p(x_1,\theta)p(x_2,\theta)\cdots p(x_n,\theta)\}$$

称这样得到的 $\hat{\theta}=g(x_1,x_2,\cdots,x_n)$

为参数 $\boldsymbol{\theta}$ 的**最大似然估计值.**

称统计量 $\widehat{\theta}=g(X_1,X_2,\cdots,X_n)$

为参数 $\boldsymbol{\theta}$的**最大似然估计量.**

注1 若 X 连续, 取 $f(x_i,\theta)$ 为 X_i 的密度函数

似然函数为 $L(\theta)=\prod_{i=1}^{n} f(x_i,\theta)$

注2 未知参数可以不止一个, 如 $\theta_1,\ldots,\theta_k$

设 X 的密度(或分布)为 $f(x,\theta_1,\cdots,\theta_k)$

则定义似然函数为

$$L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$$

$$=L(\theta_1,\cdots,\theta_k)=\prod_{i=1}^{n} f(x_i,\theta_1,\cdots,\theta_k)$$

$$-\infty<x_i<+\infty,\ i=1,2,\cdots,n \qquad (\theta_1,\cdots,\theta_k)\in\Theta$$

若 $L(x_1,\cdots,x_n;\theta_1,\cdots,\theta_k)$ 关于$\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_k}$可微,则称

$$\frac{\partial}{\partial\theta_r}L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)=0 \quad r=1,2,\cdots,k$$

为似然方程组.

若对于某组给定的样本值 $\boldsymbol{x_1, x_2, \ldots, x_n}$,

参数 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$ 使似然函数取得最大值, 即

$$L(x_1,\cdots,x_n;\widehat{\theta}_1,\cdots,\widehat{\theta}_k)$$
$$=\max_{(\theta_1,\theta_2,\cdots,\theta_k)\in\Theta}\{L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)\}$$

则称 $\widehat{\theta}_1,\cdots,\widehat{\theta}_k$ 为$\boldsymbol{\theta_1}, \ldots, \boldsymbol{\theta_k}$ 的最大似然估计值

若：

$$\hat{\theta}_r = g(x_1, x_2, \cdots, x_n) \qquad r = 1,2,\cdots,k$$

称统计量

$$\hat{\theta}_r = g(X_1, X_2, \cdots, X_n) \qquad r = 1,2,\cdots,k$$

为$\boldsymbol{\theta}_1$， $\boldsymbol{\theta}_2$, …， $\boldsymbol{\theta}_k$ 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的最大似然估计值.

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最大似然估计值有时行不通，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4 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：设 $x_1,x_2,\cdots,x_n$为相应于样本 $X_1,X_2,\cdots,X_n$的一个样本值,

X的分布律为 $P\{X=x\}=p^x(1-p)^{1-x}, x=0,1,$

似然函数为:

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

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

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

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

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 $\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i=\overline{x}$

即为 p 的最大似然估计值.

从而 p 的最大似然估计量为

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

例5 设总体 $X \sim N(\mu,\sigma^2)$, $x_1, x_2,\ldots, x_n$ 是 X 的样本值, 求 μ, σ^2 的最大似然估计.

解　X的概率密度为 $f(x;\mu,\sigma^2)=\dfrac{1}{\sqrt{2\pi}\sigma}\mathrm{e}^{-\frac{(x-\mu)^2}{2\sigma^2}}$,

X 的似然函数为

$$L(x_1,x_2,\cdots,x_n;\mu,\sigma^2)$$

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

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

$\boxed{\text{似然方程组为}}\begin{cases} \left(\dfrac{\partial}{\partial \mu}\ln L\right) = \dfrac{1}{\sigma^2}\sum\limits_{i=1}^{n}(x_i-\mu)=0 \\ \left(\dfrac{\partial}{\partial(\sigma^2)}\ln L\right) = \dfrac{1}{2(\sigma^2)^2}\sum\limits_{i=1}^{n}(x_i-\mu)^2-\dfrac{n}{2(\sigma^2)}=0 \end{cases}$

$$\Rightarrow \begin{cases} \hat{\mu} = \dfrac{1}{n}\sum\limits_{i=1}^{n}x_i = \bar{x} \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum\limits_{i=1}^{n}(x_i-\bar{x})^2 \end{cases}$$

$\boldsymbol{\mu}, \boldsymbol{\sigma}^2$ 的最大似然估计量分别为

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

最大似然估计方法

1） 写出似然函数 L

2）求出 $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$，使得

$$L(x_1, x_2, \cdots, x_n; \hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$$

$$= \max_{(\theta_1, \theta_2, \cdots, \theta_k) \in \Theta} \{L(x_1, x_2, \cdots, x_n; \theta_1, \theta_2, \cdots, \theta_k)\}$$

若　L是$\theta_1,\cdots,\theta_k$的可微函数,解似然方程组

$$\frac{\partial}{\partial\theta_r}L(x_1,x_2,\cdots,x_n;\theta_1,\theta_2,\cdots,\theta_k)=0$$

$$r=1,2,\cdots,k$$

可得未知参数的最大似然估计值 $\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$

然后,再求得最大似然估计量.

若　L不是$\theta_1,\cdots,\theta_k$ 的可微函数,需用其它方法求最大似然估计值. 请看下例:

例6 设 $X \sim U(a,b)$, $x_1, x_2, \ldots, x_n$ 是 X 的一个样本值, 求 a, b 的最大似然估计值与最大似然估计量.

解 X 的密度函数为

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

似然函数为

$$L(x_1,x_2,\cdots,x_n;a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a<x_i<b,\ i=1,2,\cdots,n\\ 0, & \text{其它}\end{cases}$$

似然函数只有当 $a < x_i < b, i = 1,2,\ldots, n$ 时

才能获得最大值, 且 a 越大, b 越小, L 越大.

令
$$x_{\min} = \min\{x_1, x_2,\ldots, x_n\}$$
$$x_{\max} = \max\{x_1, x_2,\ldots, x_n\}$$

取
$$\hat{a} = x_{\min}, \quad \hat{b} = x_{\max}$$

则对满足 $a \le x_{\min} \le x_{\max} \le b$ 的一切 $a < b$,

都有
$$\frac{1}{(b-a)^n} \le \frac{1}{(x_{\max} - x_{\min})^n}$$

故 $\hat{a} = x_{\min}$, $\hat{b} = x_{\max}$

是 $\boldsymbol{a}$, $\boldsymbol{b}$ 的最大似然估计值.

$$X_{\min} = \min\{X_1, X_2, \cdots, X_n\}$$

$$X_{\max} = \max\{X_1, X_2, \cdots, X_n\}$$

分别是 $\boldsymbol{a}$, $\boldsymbol{b}$ 的最大似然估计量.

最大似然估计的不变性

设$\hat{\theta}$是θ的最大似然估计值, $u(\theta)$

$(\theta \in \Theta)$是θ的函数, 且有单值反函数

$$\theta = \theta(u), \quad u \in U$$

则 $\hat{u} = u(\hat{\theta})$ 是 $u(\theta)$ 的最大似然估计值.

如 在正态总体$N(\mu,\sigma^2)$中, σ^2的最大似然估计值为

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

$\sigma=\sqrt{\sigma^2}$ 是σ^2的单值函数, 且具有单值反函数，故σ的最大似然估计值为

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

$\lg\sigma$的最大似然估计值为

$$\lg\hat{\sigma}=\lg\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2}$$

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最大似然估计法．

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．

作业：课后习题 2、3、5

练习：

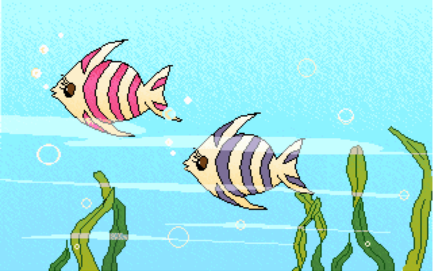
用最大似然估计法估计湖中的鱼数

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是随机变量，X 具有超几何分布：

$$P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S},$$

$$0 \le k \le \min(S,r)$$



$$P\{X=k\}=\frac{C_r^k C_{N-r}^{S-k}}{C_N^S},$$

把上式右端看作 N 的函数，记作 $L(N;k)$.

应取使$L(N;k)$达到最大的N，作为N的最大似然估计.

但用对N求导的方法相当困难, 我们考虑比值:

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

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的最大似然估计为 $\hat{N}=[\frac{Sr}{k}]$.