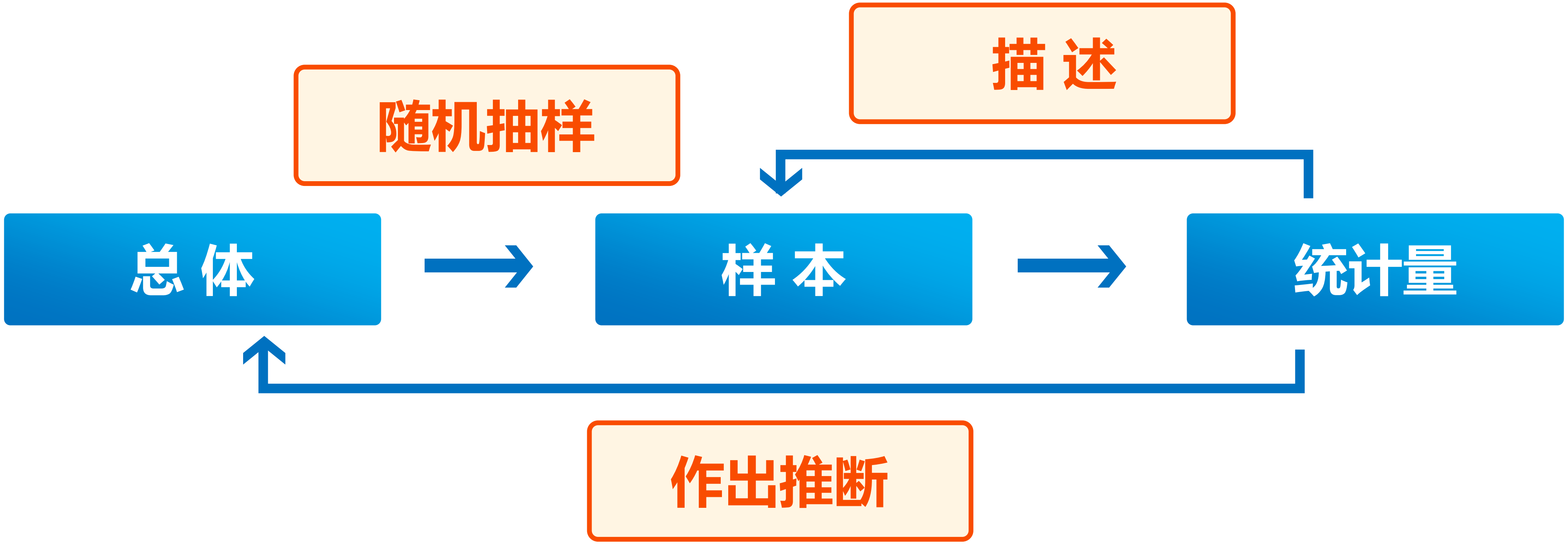
概率论与数理统计

第18讲

点估计的概念、矩估计和最大似然估计方法



统计推断：参数估计和假设检验

一. 参数估计

数理统计的任务是根据样本所提供的信息，对总体的分布或分布的某些数字特征作推断。

最常见的统计推断问题是总体分布的类型已知，而它的某些参数未知。

例如，已知总体X服从正态分布$N(0,\sigma^2)$，其中$\sigma^2>0$未知，只要对未知参数σ^2作出了估计，也就对整个总体分布作出了推断。

这类问题称为参数估计问题

参数估计问题是利用样本$X_1, X_2, \ldots, X_n$提供的信息，对总体中的未知参数或参数的函数作出估计。

一. 参数估计

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——根据样本构造出适当的区间，使它以一定的概率包含未知参数或未知参数函数的真值。

一. 参数估计

假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，分别为

1.65　1.67　1.68　1.71　1.69

求出总体均值μ的估计。

全部信息就由这5个数组成 。 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。

二. 点估计

设总体X的分布函数$F(x,\theta)$形式已知，θ是待估参数，X_1，X_2，…，X_n为抽自总体X的样本，x_1，x_2,…，x_n是相应的一个样本值。据此，应如何估计未知参数θ呢?

为估计θ，需要构造一个适当的统计量 $\hat{\theta}(X_1,X_2,\cdots,X_n)$，每当有了样本观测值$x_1, x_2,\ldots, x_n$，就代入该统计量中算出一个值 $\hat{\theta}(x_1,x_2,\cdots,x_n)$ 作为未知参数θ的近似值。

由于未知参数θ 和 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 的取值都是实数轴上的一点，用 $\hat{\theta}(X_1,X_2,\cdots,X_n)$ 去估计θ 等于用一个点去估计另外一个点，所以这样的估计称为点估计。

定义2 设总体X的分布为$f(x,\theta)$，其中θ是未知参数，$X_1, X_2, \dots, X_n$是来自总体X的样本，构造统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$，对于样本观察值$x_1, x_2, \dots, x_n$，若将统计量的观察值$\hat{\theta}(x_1,x_2,\cdots,x_n)$作为未知参数$\theta$的值，则称$\hat{\theta}(x_1,x_2,\cdots,x_n)$为参数$\theta$的估计值，而统计量$\hat{\theta}(X_1,X_2,\cdots,X_n)$称为参数$\theta$的估计量。$\theta$的估计量和估计值常简记为$\hat{\theta}$，在不引起混淆的情况下统称为$\theta$的估计，这种估计称为参数的点估计。

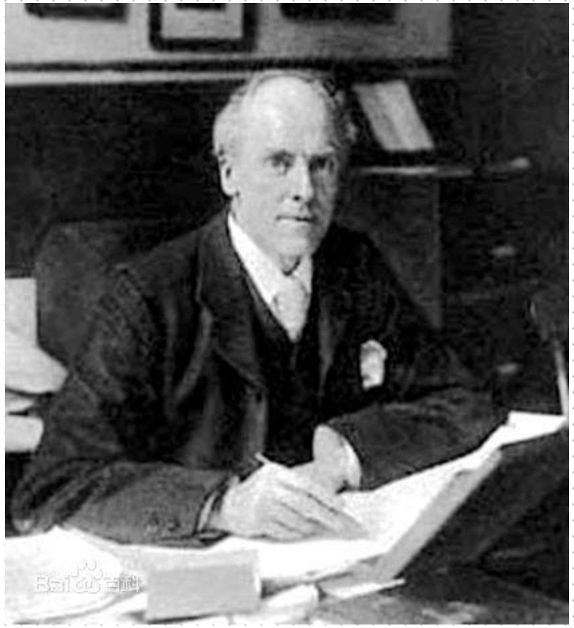
若总体分布中含有k个未知参数$\theta_1, \theta_2, \dots, \theta_k$，则由样本建立$k$个不带任何未知参数的统计量$\hat{\theta}_i(X_1,X_2,\cdots,X_n)$，$i=1,2,\cdots,k$

将它们分别作为k个未知参数$\theta_1, \theta_2, \dots, \theta_k$的点估计量。

矩估计法由英国统计学家卡尔·皮尔逊(Karl Pearson)在20世纪初提出。

1.矩估计方法的基本思想

用样本矩估计总体矩



1857-1936

三. 矩估计

设$X_1, X_2, \ldots, X_n$为来自总体X中抽出的一个样本，则

总体k阶原点矩为 $\mu_k = E(X^k)$　样本k阶原点矩为 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$

由大数定律 $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k \xrightarrow{P} \mu_k$　若g为连续函数，则 $g(A_k) \xrightarrow{P} g(\mu_k)$

因此，可以用A_k估计μ_k，用$g(A_k)$估计$g(\mu_k)$

用样本矩作为相应总体矩的估计量，样本矩的连续函数作为相应总体矩的连续函数的估计量, 这种估计方法称为**矩估计法**

2.矩估计的步骤

设总体$X \sim F(x, \theta_1, \theta_2, \ldots, \theta_k)$, $X_1, X_2, \ldots, X_n$为样本。

（1）根据未知参数的个数，求出总体的各阶矩 $\boxed{\mu_l = \mu_l(\theta_1, \cdots, \theta_k), l = 1, 2, \cdots, k}$

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

总体X的密度函数

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

总体X的分布律

三. 矩估计

(2) 解方程(组)，得

$$\theta_l = \theta_l(\mu_1, \cdots, \mu_k), l = 1, 2, \cdots, k$$

(3) 用样本矩估计相应的总体矩，即：用A_l替代相应的μ_l，得到θ_l的矩估计

$$\hat{\theta}_l = \theta_l(A_1, A_2, \cdots, A_k), l = 1, 2, \cdots, k$$

(3) $g(\theta_1, \ldots, \theta_k)$的矩估计为

$$g(\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k)$$

例1.设总体X的概率密度为 $f(x)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1\\ 0, & 其它\end{cases}$

其中$\alpha>-1$是未知参数，$X_1, X_2, \ldots, X_n$是取自X的样本，求(1)参数α的矩估计；(2) $g(\alpha)=(\alpha+1)/\alpha$的矩估计。

解:(1) 1^0.求总体的1阶矩

$$\mu_1=E(X)=\int_{-\infty}^{+\infty} xf(x)dx=(\alpha+1)\int_0^1 x^{\alpha+1}dx=\frac{\alpha+1}{\alpha+2}$$

三. 矩估计

(1)参数α的矩估计　(2) $g(\alpha)=(\alpha+1)/\alpha$的矩估计。　$\mu_1=\dfrac{\alpha+1}{\alpha+2}$

2^0. 解方程 $\alpha=\dfrac{2\mu_1-1}{1-\mu_1}$

3^0. 用 $A_1=\bar{X}=\dfrac{1}{n}\sum_{i=1}^{n}X_i$ 代替μ_1，得α的矩估计为 $\hat{\alpha}=\dfrac{2\bar{X}-1}{1-\bar{X}}$

(2) $g(\alpha)=(\alpha+1)/\alpha$的矩估计为

$$g(\alpha)=\frac{\hat{\alpha}+1}{\hat{\alpha}}=\frac{\bar{X}}{2\bar{X}-1}$$

例2.设总体X的均值μ和方差σ^2都存在，且$\sigma^2>0$，但μ和σ^2均未知，设$X_1, X_2, \dots, X_n$是来自总体X样本，求μ，σ^2的矩估计量。

解：1^0. 求总体的1阶矩和2阶矩

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

2^0. 解方程组

$$\begin{cases} \mu = \mu_1 \\ \sigma^2 = \mu_2 - \mu_1^2 \end{cases}$$

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

3^0.分别以A_1, A_2代替μ_1, μ_2得到μ , σ^2的矩估计量分别为

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

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

特别，若$X\sim N(\mu,\sigma^2)$, μ, σ^2未知,则μ, σ^2的矩估计量分别为

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

例3. 设总体X的概率密度函数为 $f(x)=\dfrac{1}{2\lambda}e^{-\frac{|x|}{\lambda}},\quad x\in R$

其中$\lambda>0$为未知参数，$X_1, X_2,\ldots, X_n$是来自X的简单随机样本。求参数θ的矩估计量。

解：易知

$$\mu_1=E(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-\infty}^{+\infty}x\frac{1}{2\lambda}e^{-\frac{|x|}{\lambda}}dx=0$$

此结果与未知参数λ无关。

三. 矩估计

$$\mu_2 = EX^2 = \int_{-\infty}^{+\infty} x^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}} dx = \int_{0}^{+\infty} x^2 \frac{1}{\lambda} e^{-\frac{x}{\lambda}} dx = 2\lambda^2$$

解方程得：

$$\lambda = \sqrt{\frac{\mu_2}{2}}$$

用 $A_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$ **代替**μ_2**得参数**λ **的矩估计为**

$$\hat{\lambda} = \sqrt{\frac{A_2}{2}} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2}{2n}}$$

优点：直观、简单　　缺点：（1）不唯一，如例1

例1.设总体X的概率密度为 $f(x)=\begin{cases}(\alpha+1)x^{\alpha}, & 0<x<1\\ 0, & 其它\end{cases}$

其中$\alpha>-1$是未知参数，X_1，X_2，…，X_n是取自X的样本，求(1)参数α的矩估计；

可以求总体的二阶矩μ_2，用A_2代替μ_2得到矩估计。

$$\mu_2=E(X^2)=\int_{-\infty}^{+\infty}x^2f(x)dx=(\alpha+1)\int_0^1 x^{\alpha+2}dx=\frac{\alpha+1}{\alpha+3}$$

三. 矩估计

解方程 $\alpha = \dfrac{3\mu_2 - 1}{1 - \mu_2}$

用 $A_2 = \dfrac{1}{n}\sum_{i=1}^{n} X_i^2$ 代替μ_2，得α的矩估计为$\hat{\alpha} = \dfrac{3A_2 - 1}{1 - A_2}$

显然，与$\hat{\alpha} = \dfrac{2\bar{X} - 1}{1 - \bar{X}}$ 不同

规定：用尽量低阶的矩求相应的参数估计

缺点：（2）损失信息，如例2

例2.设总体X的均值μ和方差σ^2都存在，且$\sigma^2>0$，但μ和σ^2均未知，设$X_1, X_2, \dots, X_n$是来自总体X样本，求μ, σ^2的矩估计量。

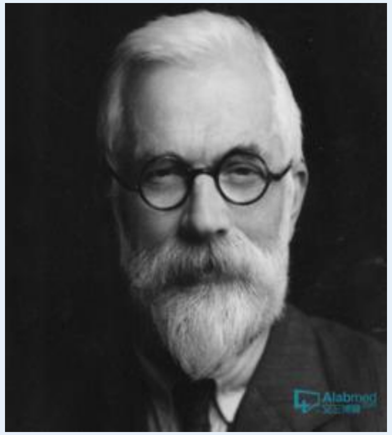
若已知总体X的服从正态分布，则该分布形式已知的信息没有用到，从而造成信息的损失。

（3）有不适用的情况存在。矩估计方法要求总体的原点矩存在，而有些随机变量(如柯西分布)的原点矩不存在，因此就不能用矩估计方法求点估计。

四. 最（极）大似然估计（MLE）

Gauss(1777-1855)

首先是由德国数学家高斯在1821年提出。



Fisher(1890-1962)

然而，这个方法常归功于英国统计学家费歇(Fisher)。

费歇在1922年重新发现了这一方法，并首先研究了该方法的一些性质。

一.最大似然估计的基本思想

例4.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?

因为只发一枪便打中，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。

其数学模型为

令X为打一枪的中弹数，则$X \sim b(1,p)$，p未知。依题意p有两种可能：

$\{0.9, \quad 0.1\}$

p的取值范围

两人中有一人打枪，估计这一枪是谁打的，即估计总体X的参数p的值是0.9还是0.1？

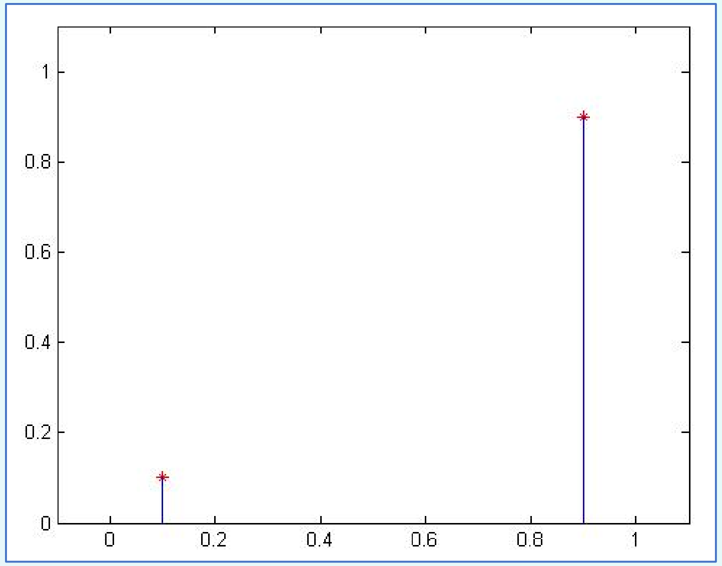
四. 最（极）大似然估计（MLE）

打了1枪，相当于得到一个样本，记为X_1

兔子中弹，相当于样本观测值为1， 即$\{X_1=1\}$发生了

考虑此样本观测值出现的概率，有

$L(p)=P\{X_1=1\}=p=0.9$或0.1



选择是猎人打的，相当于选择p的值，使得样本观测值1出现的可能性最大。

最大似然估计法的基本思想：根据样本观测值，选择参数p的值，使得该样本值出现的可能性最大。

四. 最（极）大似然估计（MLE）

说明： 虽然参数p可取参数空间Θ中的所有值，但在给定样本观察值x_1，x_2，…，x_n后，不同的p值对应样本X_1，X_2，…，X_n落入x_1，x_2，…，x_n的邻域的概率大小也不同，既然在一次试验中观察到X_1，X_2，…，X_n取值x_1，x_2，…，x_n，因此有理由认为X_1，X_2，…，X_n落入x_1，x_2, …, x_n的邻域中的概率较其它地方大。

哪一个参数使得X_1, X_2, …, X_n落入x_1, x_2, …, x_n的邻域中的概率最大，这个参数就是最可能的参数，我们用它作为参数的估计值，这就是**最大似然原理。**

四. 最（极）大似然估计（MLE）

设总体X为离散型，其分布列为

$$P\{X=x\}=f(x;\theta)$$

其中$\theta\in\Theta$，θ为待估参数，Θ是参数空间即参数θ的取值范围。

设$X_1, X_2, \ldots, X_n$是来自总体X的样本，则样本$X_1, X_2, \ldots, X_n$的分布列为

$$\prod_{i=1}^{n} f(x_i;\theta)$$

四. 最（极）大似然估计（MLE）

又设$x_1, x_2, \dots, x_n$是样本$X_1, X_2, \dots, X_n$的一个观察值，易知样本$X_1, X_2, \dots, X_n$取$x_1, x_2, \dots, x_n$的概率，即事件$\{X_1=x_1，X_2=x_2, \dots, X_n=x_n\}$发生的概率为

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

这一概率随θ的取值而变化，它是θ的函数，$L(\theta)$称为样本的**似然函数**。

四. 最（极）大似然估计（MLE）

由最大似然原理，固定样本观察值$x_1, x_2, \ldots, x_n$，在θ的取值范围Θ内挑选使似然函数$L(\theta)$达到最大的值 $\hat{\theta}$ 作为参数θ的估计值，即取使得

$$L(\hat{\theta}; x_1, \cdots, x_n) = \max_{\theta \in \Theta} L(\theta; x_1, \cdots, x_n)$$

$\hat{\theta}$ 与$x_1, x_2, \ldots, x_n$有关，记做 $\hat{\theta}(x_1, \cdots, x_n)$，称其为参数$\theta$的最大似然估计值；称 $\hat{\theta}(X_1, X_2, \cdots, X_n)$ 为参数θ的最大似然估计量。

四. 最（极）大似然估计（MLE）

设总体X为连续型，其密度函数为$f(x;\theta)$，其中$\theta\in\Theta$，θ为待估参数，Θ是参数空间即参数θ的取值范围。

设$X_1, X_2, \dots, X_n$是来自总体X的样本，则样本$X_1, X_2, \dots, X_n$的联合密度函数为

$$\prod_{i=1}^{n} f(x_i;\theta)$$

又设$x_1, x_2, \dots, x_n$是样本$X_1, X_2, \dots, X_n$的一个观察值，易知样本$X_1, X_2, \dots, X_n$落在$x_1, x_2, \dots, x_n$的邻域（边长分别为$dx_1, dx_2, \dots, dx_n$的维立方体）的概率近似为

四. 最（极）大似然估计（MLE）

$$\prod_{i=1}^{n} f(x_i;\theta)dx_i$$

这一概率随θ的取值而变化，它是θ的函数，取θ的估计值$\hat{\theta}$使得上式达到最大，但$\prod_{i=1}^{n} dx_i$不随θ取值的变化而改变，故只需考虑

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

的最大值，$L(\theta)$称为样本的似然函数。

四. 最（极）大似然估计（MLE）

由最大似然原理，固定样本观察值$x_1, x_2, \ldots, x_n$，在θ的取值范围Θ内挑选使似然函数$L(\theta)$达到最大的值$\hat{\theta}$作为参数θ的估计值，即取使得

$$L(\hat{\theta}; x_1, \cdots, x_n) = \max_{\theta \in \Theta} L(\theta; x_1, \cdots, x_n)$$

$\hat{\theta}$与$x_1, x_2, \ldots, x_n$有关，记做$\hat{\theta}(x_1, \cdots, x_n)$，称其为参数$\theta$的最大似然估计值；称$\hat{\theta}(X_1, X_2, \cdots, X_n)$为参数$\theta$的最大似然估计量。

四. 最（极）大似然估计（MLE）

1.似然函数： 设总体X的概率密度(或分布律)为$f(x;\theta)$，$\theta\in\Theta$，X_1, …,X_n为来自总体的样本，则$(X_1,\dots,X_n)$的密度函数(或分布律)为

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

若已知样本观测值$(x_1,\dots,x_n)$，则令θ的函数为

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

称其为基于样本$(x_1,\dots,x_n)$的似然函数。

称 $\ln L(\theta)=\ln L(\theta;x_1,\cdots,x_n)=\ln\prod_{i=1}^{n}f(x_i;\theta)$ 为样本的对数似然函数。

2.最大似然估计： 如果似然函数$L(\theta;x_1,\ldots,x_n)$，在 $\hat{\theta}$ 达到最大值，即

$$L(\hat{\theta};x_1,\cdots,x_n)=\max_{\theta\in\Theta}L(\theta;x_1,\cdots,x_n)$$

则称 $\hat{\theta}$ 为 θ 的最大似然估计值。

它一般是$x_1,\ldots,x_n$的函数，也常记为 $\hat{\theta}(x_1,\ldots,x_n)$ $\hat{\theta}(X_1,\ldots,X_n)$ 称为最大似然估计量。

在不引起混淆的情况下，统称为最大似然估计(*Maximum Likelihood Estimator*简记为*MLE*)。

最大似然估计的求法

求似然函数$L(\theta; x_1, \ldots, x_n)$在$\theta \in \Theta$内关于$\theta$的最大值点。

若$f(x, \theta)$关于θ可微，则θ的MLE可由下式得到

$$\frac{dL}{d\theta} = 0$$

似然方程

又因为$L(\theta)$和$\ln L(\theta)$在同一θ处取得极值，因此MLE也可由下述方程得到

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

对数似然方程

当似然函数$L(\theta)$不是θ的可微函数时，此时需要用定义的方法求最大似然估计。

四. 最大似然估计

最大似然估计也适用于分布中含多个未知参数$\theta=(\theta_1, \theta_2,\dots,\theta_k)$的情况，此时，似然函数$L(\theta)$是这些未知参数的函数。若似然函数$L(\theta; x_1,\dots,x_n)$关于$\theta$各分量的偏导数都存在，则$\theta=(\theta_1, \theta_2,\dots,\theta_k)$的最大似然估计可由下述方程组求得

$$\frac{\partial L(\theta)}{\partial \theta_i}=0, \quad i=1,2,\cdots,k$$

又因为$L(\theta)$与$\ln L(\theta)$在同一θ处取得极值，因此$\theta=(\theta_1, \theta_2,\dots,\theta_k)$的最大似然估计也可由下述方程组解得

$$\frac{\partial \ln L(\theta)}{\partial \theta_i}=0, \quad i=1,2,\cdots,k$$

该方程组称为对数似然方程组。利用对数似然方程组求解更方便。

最大似然估计的不变性

设总体X的分布类型已知，其概率密度函数(或分布列)为$f(x, \theta)$，$\theta=(\theta_1, \theta_2,\ldots,\theta_k)\in\Theta\subset R^k$为未知参数，参数$\theta=(\theta_1, \theta_2,\ldots,\theta_k)$的已知函数为$g(\theta_1, \theta_2,\ldots,\theta_k)$，函数$g$具有单值反函数。

若$\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k$分别为$\theta_1, \theta_2,\ldots,\theta_k$的最大似然估计，则$g(\hat{\theta}_1,\hat{\theta}_2,\cdots,\hat{\theta}_k)$是$g(\theta_1, \theta_2,\ldots,\theta_k)$的最大似然估计。

求最大似然估计(MLE)的一般步骤是：

1. 由总体X的分布写出似然函数$L(\theta)$；
2. 求对数似然函数$\ln L(\theta)$；
3. 对 $\ln L(\theta)$关于θ求(偏)导数，并令(偏)导函数为0；
4. 解方程(组)，得到未知参数的最大似然估计。

当以上方法不适用时，需要从定义出发，直接求似然函数的极值点。

作业：1(矩估计),2,5(矩估计)

第 18 讲

谢谢观看