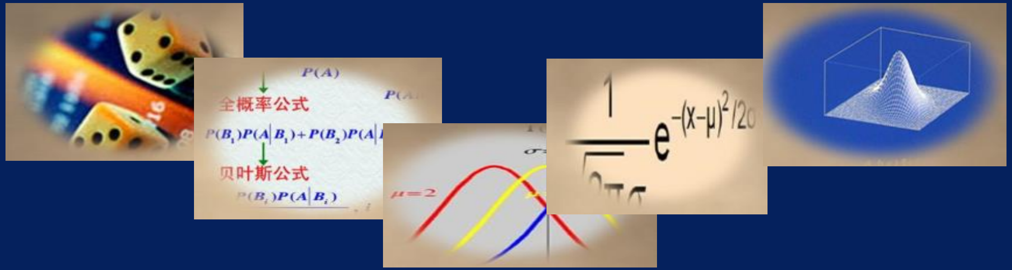
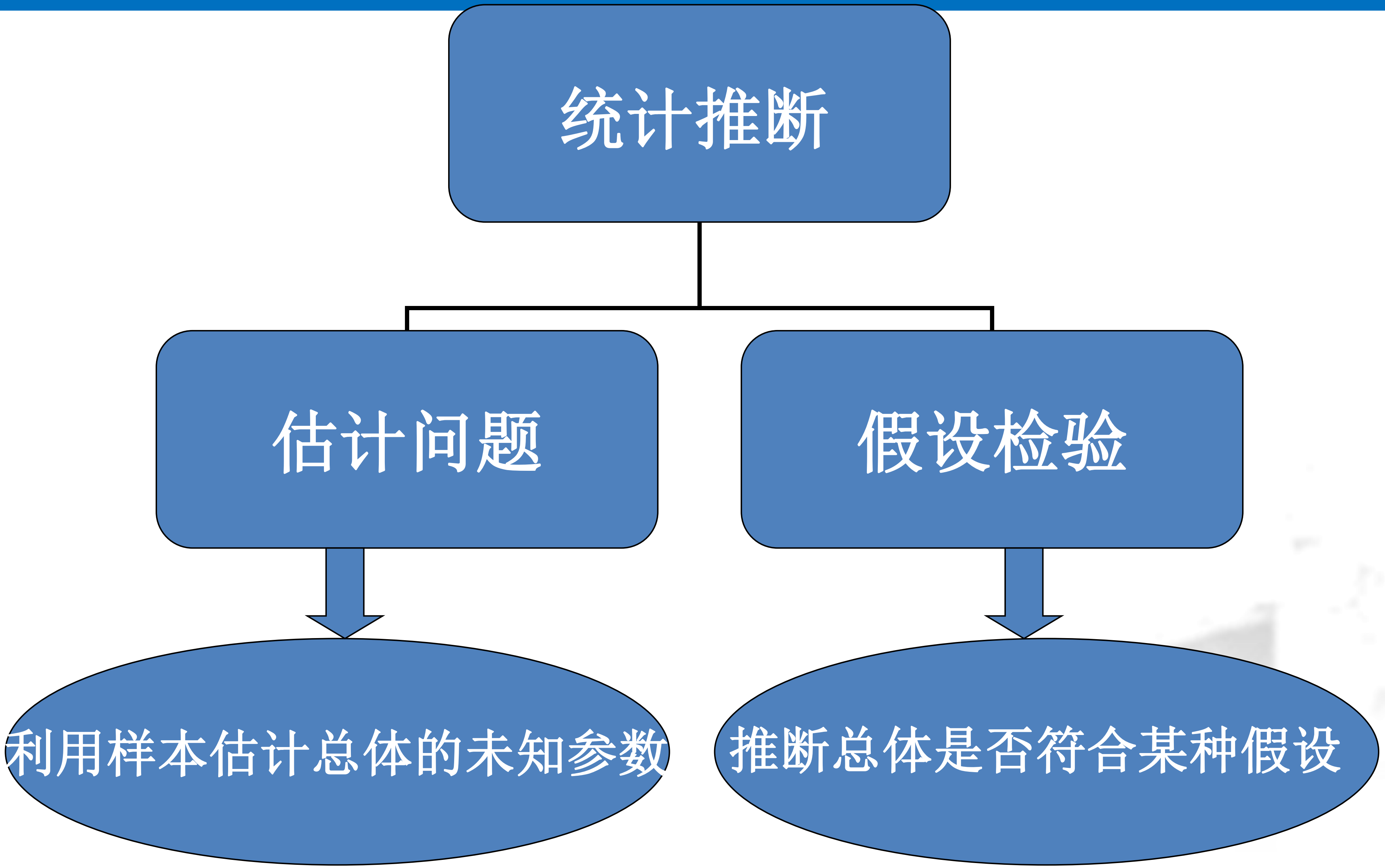
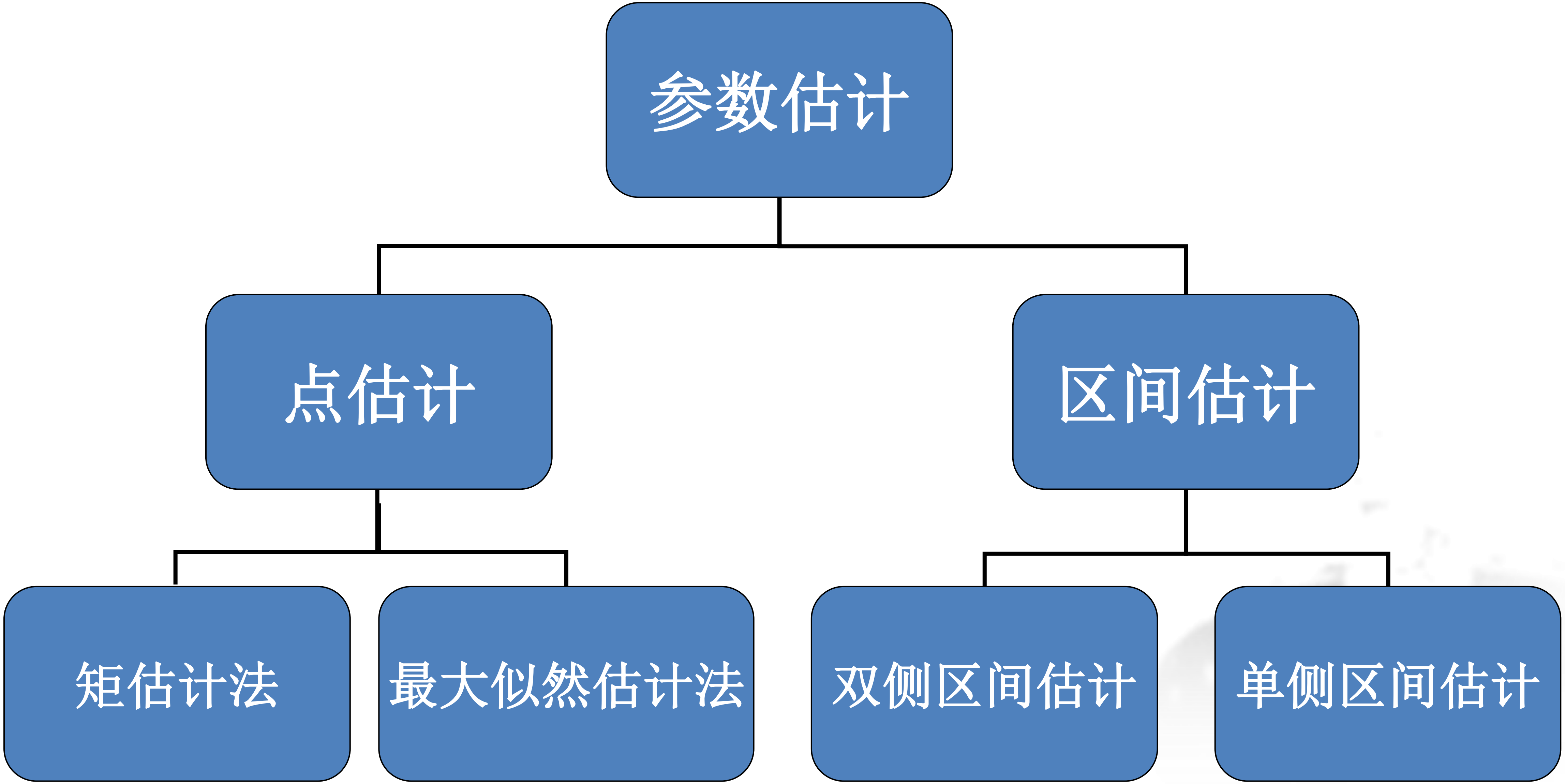
# 第七章 参数估计

## 7.1 点估计（1）
## 矩估计法




杨建芳


统计推断
估计问题
假设检验
利用样本估计总体的未知参数
推断总体是否符合某种假设


参数估计
点估计
区间估计
矩估计法
最大似然估计法
双侧区间估计
单侧区间估计

- 参数估计
  - 点估计法
    - 矩估计
    - 最大似然估计
  - 估计量评选标准
    - 无偏性
    - 有效性
    - 相合性
  - 区间估计
    - 双侧置信区间
    - 单侧置信区间

# 参数的点估计（Point Estimation）

设总体 X 的分布函数$F(x;\theta)$的形式已知，其中未知参数为 $\theta$，我们要构造样本$(X_1,X_2,\cdots,X_n)$的一个适当的函数——即统计量$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$，每当有了样本值$(x_1,x_2,\cdots,x_n)$，代入后得 $\hat{\theta}$ 的一个值，作为未知参数 $\theta$ 的估计值。为着这样的目的而构造的统计量 $\hat{\theta}$ 称为( $\theta$ 的)**估计量**。由于未知参数 $\theta$ 是数轴上一个点，用 $\hat{\theta}$ 去估计 $\theta$，等于用一个点去估计另一个点，所以这样的估计叫做**点估计**。

# 矩估计法(The Method of Moments)

它是基于一种简单的“替换”思想建立起来的一种估计方法.

是英国统计学家*K.* 皮尔逊最早提出的.

其基本思想是用样本原点矩估计总体原点矩.

**矩法估计的理论基础是: 辛钦大数定律.**

# 矩估计法(The Method of Moments)

记总体 $l$ 阶矩为 $\mu_l = E(X^l)$

样本 $l$ 阶矩为 $A_l = \dfrac{1}{n}\sum_{i=1}^{n} X_i^l$

大数定律：$\overline{X} \xrightarrow{P} \mu, A_l \xrightarrow{P} \mu_l$

基本原理：总体矩=样本矩，令 $\mu_l = A_l$

# 矩估计法(The Method of Moments)

**基本步骤：**

步骤1：求出总体原点矩 $u_l = E(X^l)$

（含有待估未知参数）

注意：

选择原点矩；

几个未知参数选择几个矩；

从低阶到高阶选择矩。

步骤2：列方程或者方程组，原点矩

$$u_l = A_l, l = 1, 2, \ldots, k$$

$k$为未知参数的个数

只有一个未知参数时，则令

$$E(X) = \overline{X}$$

步骤3：解方程或者方程组，求出矩估

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

**例1** 设总体 $X$ 服从均匀分布 $U(0,\theta)$，$(X_1,\cdots,X_n)$ 是取自 $X$ 的样本，求未知参数 $\theta$ 的矩估计量。

**解** 步骤1 求出总体均值 $E(X)=\theta/2$，

步骤2 列方程

$$令\ \frac{\theta}{2}=\overline{X}$$

步骤3 解方程

得 $\theta$ 的矩估计量为 $\hat{\theta}=2\overline{X}$ .

一个未知数，
选择一个原点
即均值

估计值 $\hat{\theta}=2\overline{x}$ .

区别于真实值，记得给估计量戴上帽子哦！估计量实质是个随机变量

**例2** 设总体 **X** 服从正态分布 $N(\mu,\sigma^2)$，$(X_1,\cdots,X_n)$ 是取自 **X** 的样本，试求 $\mu,\sigma^2$ 矩估计量。

**解** (1) $E(X)=\mu;\ E(X^2)=\mu^2+\sigma^2$

(2) 令 $\begin{cases} E(X)=\overline{X}=A_1; \\ E(X^2)=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2=A_2. \end{cases}$

(3) 得 $\begin{cases} \hat{\mu}=\overline{X}\ ; \\ \hat{\sigma}^2=A_2-A_1^2=\dfrac{1}{n}\sum_{i=1}^{n}X_i^2-\overline{X}^2. \end{cases}$

两个未知数，选择两个原点矩，即一阶原点矩和二阶原点矩

# 练习题

1、设 $X \sim f(x;\theta) = \begin{cases} \theta c^{\theta} x^{-(\theta+1)}, & x > c, \\ 0, & x \le c. \end{cases}$

其中 $c > 0$ 为已知常数，$\theta > 1$，$\theta$ 为未知常数，

求未知参数 $\theta$ 估计量和估计值.

2、考虑泊松分布、指数分布等其他常见分布的矩估计量.

常见于选择填空中

# 练习1

**解**　步骤1 求出总体均值 $E(X)=\int_c^{+\infty}\theta c^{\theta}x^{-(\theta+1)}dx=\dfrac{c\theta}{\theta-1}$

步骤2　列方程 $\dfrac{c\theta}{\theta-1}=\overline{X}$

步骤3　解方程　估计量　$\hat{\theta}=\dfrac{\overline{X}}{\overline{X}-c}$

估计值　$\hat{\theta}=\dfrac{\overline{x}}{\overline{x}-c}$


1、矩估计法的基本原理？解题步骤？

2、矩估计法采用的是什么矩？

3、估计量和估计值的区别？

# 最大似然估计法（极大似然估计法）
# (The Method of Maximum Likelihood)

最大似然法, 也叫极大似然法, 它最早是由**高斯**所提出的, 后来由英国统计学家**费歇**(R·A·Fisher)于1912年在其一篇文章中重新提出, 并且证明了这个方法的一些性质. 极大似然估计这一名称也是费歇给的. 它是建立在极大似然原理的基础上的一个统计方法. 为了对极大似然原理有一个直观的认识, 我们先来看个例子.

# 最大似然估计法（极大似然估计法）

引例：设一个盒子中装有黑球、白球，两种球的数量之比为**3:1**，但不知哪种球多，$p$表示从盒子中任取一球是黑球的概率，那么$p$=1/4或者$p$=3/4。现在有放回的从盒中取**3**个球，试根据样本中的黑球数$X$来估计参数$p$。

# 最大似然估计法（极大似然估计法）

解：样本中黑球数 $X \sim b(3,p)$，即：

$$P\{X=x\}=C_3^x p^x(1-p)^{3-x} \qquad (x=0,1,2,3)$$

对$p$的两个可能值，$X$的分布律如下：

| $X$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $p$=1/4 ,P{$X$=$x$} | 27/64 | 27/64 | 9/64 | 1/64 |
| $p$=3/4 ,P{$X$=$x$} | 1/64 | 9/64 | 27/64 | 27/64 |

若$X$=0，$p$=1/4 P{$X$=0}=27/64>P{$X$=0}=1/64 ($p$=3/4)

根据这个样本可以推断$p$=1/4，以此类推。

# 最大似然估计法（极大似然估计法）

极大似然估计法的基本思想：

根据样本值来选择参数，使该样本发生的概率最大。

离散型随机变量：

联合分布律

连续型随机变量：

联合密度函数

似然函数

# 最大似然估计法（极大似然估计法）

（1）设总体为离

离散型随机变量的联合分布律

为未知参数, $(x_1,\cdots,x_n)$

$$L(\theta)=\prod_{i=1}^{n}p(x_i;\theta)$$ 为 **“似然函数”**。

是$x_i$,不是$x$

(注意：小写的$x_i$，$x_i$代入的$n$个分布律的连乘)

求 $\hat{\theta}$ , 使 $L(\theta)$ 达到最大, 则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 $\theta$ 的**最大似然估计量**

求最大似然估计量实际上就是求似然函数最大值点问题。

# 最大似然估计法（极大似然估计法）

（2）设总体为连续型，其概率密度函数为 $f(x;\theta)$，其中 $\theta$ 为未知参数，$(x_1,\cdots$

连续型随机变量的联合密度函数

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)$$ 为"**似然函数**"。

是$x_i$，不是$x$

求 $\hat{\theta}$，使 $L(\theta)$ 达到最大，则称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\cdots,X_n)$ 为 $\theta$ 的**最大似然估计量**。

求最大似然估计量实际上就是求似然函数最大值点问题。

# 最大似然估计法（极大似然估计法）

求最大似然估计(MLE)的一般步骤是：

把样本联合概率分布律(或联合密度)中自变量看成已知常数，而把参数 $\theta$看作自变量

(1) 由总体分布导出样本的联合分布 律 (或

似然函数 $L(\theta) = L(x_1, x_2, \ldots, x_n; \theta)$

对数似然函数

(2) 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$\theta$的MLE：

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

一个未知参数：求导；

多个未知参数：求偏导.

(3) 解方程求出驻点，即为$\theta$ 的最大似然估计值$\theta = \theta(x_1, x_2, \ldots x_n)$ ，

称 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots X_n)$ 为 $\theta$ 的最大似然估计量。

# 最大似然估计法（极大似然估计法）

## 对数函数的基本性质：

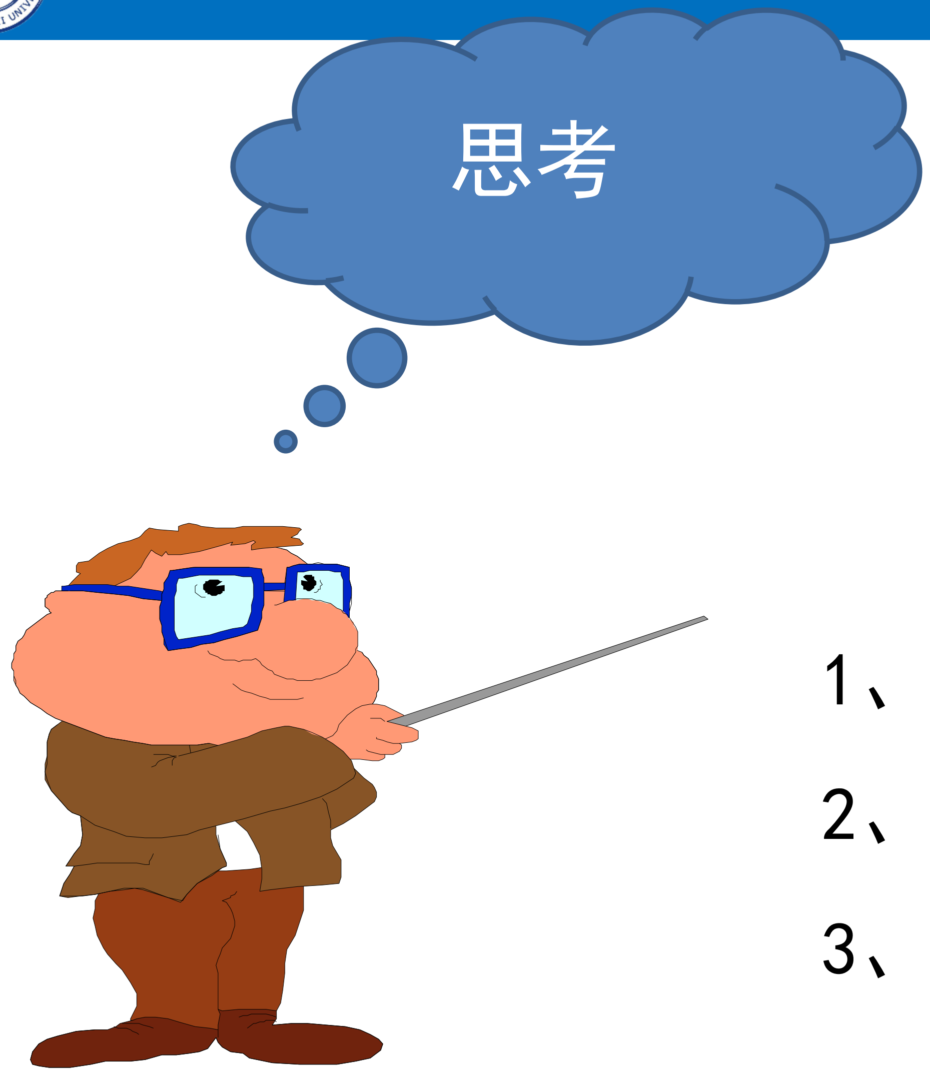
若$a>0, b>0$,则有：

$$(1)\quad \ln ab = \ln a + \ln b$$

$$(2)\quad \ln \frac{a}{b} = \ln a - \ln b$$

$$(3)\quad \ln a^b = b \ln a$$

把似然函数转换成对数似然函数，不改变其最值点，但是可以简化计算，乘积转换成求和，商转换成差，乘方转换成乘积




1、如何理解最大似然估计法的基本思想？

2、如何确定似然函数？

3、似然估计法的基本步骤？

# 第七章 参数估计

## 7.1 点估计（3）

## 最大似然估计法的计算

杨建芳

# 最大似然估计法

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 律 (或联合概率密度)， 即：

似然函数 $L(\theta)=L(x_1,x_2,\ldots,x_n;\theta)$

(2) 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$\theta$的MLE：

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

(3) 解方程求出驻点，即为$\theta$ 的最大似然估计值$\hat{\theta}=\hat{\theta}(x_1,x_2,\ldots x_n)$ ，

称 $\hat{\theta}=\hat{\theta}(X_1,X_2,\ldots X_n)$ 为 $\theta$ 的最大似然估计量。

# 最大似然估计法的计算

例1　设总体X服从指数分布，其密度函数为

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\geq 0\\ 0, & x<0\end{cases},$$

$(X_1,\cdots,X_n)$是取自 $X$ 的样本，$(x_1,x_2,\cdots,x_n)$为样本值，求未知参数 $\lambda$ 极大似然估计量。

# 最大似然估计法的计算

**解：**（1）似然函数：$L(\lambda)=\prod_{i=1}^{n}(\lambda e^{-\lambda x_i})=\lambda^n e^{-\lambda\sum_{i=1}^{n}x_i}$

是$x_i$,不是$x$

（2）取对数:$\ln L(\lambda)=n\ln\lambda-\lambda\sum_{i=1}^{n}x_i$，

取对数，简化计算

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

一个未知参数，求导，对于实际问题，唯一的驻点为最值

(3) 得：$\hat{\lambda}=\dfrac{n}{\sum_{i=1}^{n}x_i}$　　即 $\hat{\lambda}=\dfrac{1}{\overline{X}}$ 为$\lambda$的极大似然估计量。

小写的为估计值，大写的为估计量.

# 最大似然估计法的计算

**例2** 设总体 $X\sim N(\mu,\sigma^2)$，样本 $(X_1,\cdots,X_n)$，其中和 $\sigma^2$ 未知，试求 $\mu$ 和 $\sigma^2$ 的极大似然估计量。

**解：**(1) 似然函数 $L(\mu,\sigma^2)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$ … $\sum\frac{(x_i-\mu)^2}{2\sigma^2}$ ,

是 $x_i$ ,不是 $x$

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

取对数， 简化计算

# 最大似然估计法的计算

令 $\begin{cases} \dfrac{\partial \ln L}{\partial \mu} = \dfrac{1}{\sigma^2}\sum(x_i-\mu)=0 \\ \dfrac{\partial \ln L}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2}+\dfrac{1}{2(\sigma^2)^2}\sum(x_i-\mu)^2=0 \end{cases}$

两个未知参数，求偏导，对于实际问题，唯一的驻点为最值点

(3) 解得 $\begin{cases} \hat{\mu} = \dfrac{1}{n}\sum X_i = \bar{X}, \\ \hat{\sigma}^2 = \dfrac{1}{n}\sum(X_i-\bar{X})^2. \end{cases}$

整理后，正态分布最大似然估计量与相应的矩估计量相同

即为 $\mu$ 和 $\sigma^2$ 的最大似然估计量。

如果似然函数不可导（或驻点不存在），如何求最大值点？

利用高数中的知识即可！

# 最大似然估计法的计算

**例3 （了解）** 设总体 $X$ 服从均匀分布 $U(0,\theta)$，$(X_1,\cdots,X_n)$ 是取自 $X$ 的样本，求未知参数 $\theta$ 的最大似然估计量。

**解** $U(0,\theta)$的密度函数为

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

# 最大似然估计法的计算

似然函数为

$$L(\theta)=\prod_{i=1}^{n} f(x_i;\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0\le x_1,\cdots,x_n\le\theta\\ 0, & \text{其它}\end{cases}=\begin{cases}\dfrac{1}{\theta^n}, & \theta\ge\max\{x_1,\cdots,x_n\}\\ 0, & \text{其它}\end{cases}$$

显然驻点不存在

显然当 $\hat{\theta}=\max\limits_{1\le i\le n} x_i$ 时，$L$ 最大，

故 $\theta$ 的极大似然估计量为 $\hat{\theta}=\max\limits_{1\le i\le n} X_i$ .

比较：$\theta$ 的矩法估计量为 $\hat{\theta}=2\overline{X}$ .

# 练习题

设总体$X$的概率密度为

$$f(x)=\begin{cases}(\theta+1)x^{\theta}, & 0<x<1\\ 0, & \text{else}\end{cases}$$

$(X_1,\cdots,X_n)$ 是取自 $\boldsymbol{X}$ 的样本，$\theta>-1$ 是未知参数，试分别用矩估计法和最大似然估计法给出 $\theta$ 的估计量。

# 练习题

**解 矩估计法**

**(1) 总体X的数学期望为**

$$E(X)=\int_{-\infty}^{+\infty} xf(x)\,\mathrm{d}x=\int_0^1(\theta+1)x^{\theta+1}\,\mathrm{d}x=\frac{\theta+1}{\theta+2},$$

(2) 令 $\dfrac{\theta+1}{\theta+2}=\overline{X}$

(3) 得$\theta$的矩估计量为 $\hat{\theta}=\dfrac{2\overline{X}-1}{1-\overline{X}}.$

# 最大似然估计法

(1) $L(\theta)=\begin{cases}(\theta+1)^n(\prod_{i=1}^{n}x_i)^\theta, & 0<x_i<1(i=1,\cdots,n)\\ 0, & \text{else}\end{cases}$

(2) $\ln L=n\ln(\theta+1)+\theta\sum_{i=1}^{n}\ln x_i$，令 $\dfrac{\mathrm{d}\ln L}{\mathrm{d}\theta}=\dfrac{n}{\theta+1}+\sum_{i=1}^{n}\ln x_i=0$，

(3) 解得$\theta$的最大似然估计值为：$\hat{\theta}=-1-\dfrac{n}{\sum_{i=1}^{n}\ln x_i}$.

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


1、如何通过求导或者偏导求最大似然估计值？

2、对于驻点不存在的似然函数如何求其最值点？

# 一、无偏性 (Unbiasedness)

估计量是随机变量，对于不同的样本值就会得到不同的估计值，希望估计值在未知参数真值左右徘徊，最好它的数学期望等于未知参数的真值，这就导致了无偏性这个标准。

**定义** 设 $(X_1,\cdots,X_n)$ 是总体 $\boldsymbol{X}$ 的一个样本，$\hat{\theta}=\hat{\theta}(X_1,\cdots,X_n)$ 是未知参数 $\theta$ 的估计量，如果有

$$E(\hat{\theta})=\theta,$$

则称 $\hat{\theta}$ 为 $\theta$ 的**无偏估计量**。

# 一、无偏性 (Unbiasedness)

$$E(\hat{\theta}) = \theta$$

无偏估计量的含义是：$\hat{\theta}$ 作为样本的函数是一个随机变量，它在 $\theta$ 的真值附近波动，但其平均值恰好是 $\theta$ 的真值。比如用一台秤去称物品，误差有两个来源：一是秤本身制作结构上的问题，这属于系统误差；另一种是操作上或其它随机因素的干扰，这属于随机误差。

无偏性即要求没有系统误差。

## 一、无偏性 (Unbiasedness)

设 $(X_1,\cdots,X_n)$ 为取自总体 **X** 的样本，

$$E(\overline{X})=E(X),$$

说明 $\overline{X}$ 是总体均值 $E(X)$ 的无偏估计；

**样本方差** $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$, $\quad E(S^2)=D(X)$,

说明 $S^2$ 是总体方差 $D(X)$ 的无偏估计.

# 一、无偏性 (Unbiasedness)

**例1** 设总体 $X$ 服从均匀分布 $U(0,\theta)$，试证 $\theta$ 的矩法估计量 $\hat{\theta}=2\bar{X}$ 是 $\theta$ 的无偏估计量。

**证** $E(\hat{\theta})=E(2\bar{X})=2E(\bar{X})=2E(X)=2\cdot\dfrac{\theta}{2}=\theta.$

所以 $\hat{\theta}$ 是 $\theta$ 的无偏估计量。

# 一、无偏性 (Unbiasedness)

**例2** 设 $D(X) \neq 0$, $E(\bar{X}) = \mu$, 试问 $\bar{X}^2$ 是否为 $\mu^2$ 的无偏估计?

**证** $$E(\bar{X}^2) = D(\bar{X}) + [E(\bar{X})]^2$$

$$= \frac{1}{n}D(X) + \mu^2 \neq \mu^2,$$

故 $\bar{X}^2$ 不是 $\mu^2$ 的无偏估计。

# 二、有效性（Efficiency）

一般来说,一个参数往往有多个无偏估计量.

若 $\theta$有两个无偏估计量： $\hat{\theta}_1, \hat{\theta}_2$ ，则$\hat{\theta} = a\hat{\theta}_1 + b\hat{\theta}_2$

当$a+b=1$时也是 $\theta$的无偏估计量。

**定义** 设总体有一未知参数 $\theta$，样本$(X_1,\cdots,X_n)$，$\hat{\theta}_1, \hat{\theta}_2$ 均为$\theta$的无偏估计，如果 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$, 则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ **有效**。

# 二、有效性（Efficiency）

**例3** 设$(X_1, X_2, X_3)$为取自总体 $X$ 的样本，试证明下列三个统计量均为$E(X)$的无偏估计量，并比较有效性.

$$\hat{\theta}_1 = \frac{1}{5}X_1 + \frac{3}{10}X_2 + \frac{1}{2}X_3,$$

$$\hat{\theta}_2 = \frac{2}{3}X_1 + \frac{1}{2}X_2 - \frac{1}{6}X_3,$$

$$\hat{\theta}_3 = \frac{1}{3}X_1 + \frac{1}{3}X_2 + \frac{1}{3}X_3.$$

$$\hat{\theta}_1=\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3,\ \hat{\theta}_2=\frac{2}{3}X_1+\frac{1}{2}X_2-\frac{1}{6}X_3,\ \hat{\theta}_3=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3.$$

**证** $$E(\hat{\theta}_1)=E(\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3)$$

$$=(\frac{1}{5}+\frac{3}{10}+\frac{1}{2})E(X)=E(X),$$

$$E(\hat{\theta}_2)=(\frac{2}{3}+\frac{1}{2}-\frac{1}{6})E(X)=E(X),$$

$$E(\hat{\theta}_3)=(\frac{1}{3}+\frac{1}{3}+\frac{1}{3})E(X)=E(X),$$

所 以 $\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3$ 均为 $E(X)$的无偏估计量。

$$\hat{\theta}_1=\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3,\ \hat{\theta}_2=\frac{2}{3}X_1+\frac{1}{2}X_2-\frac{1}{6}X_3,\ \hat{\theta}_3=\frac{1}{3}X_1+\frac{1}{3}X_2+\frac{1}{3}X_3.$$

**证** $D(\hat{\theta}_1)=D(\frac{1}{5}X_1+\frac{3}{10}X_2+\frac{1}{2}X_3)$

$$=(\frac{1}{25}+\frac{9}{100}+\frac{1}{4})D(X)=0.38D(X),$$

$$D(\hat{\theta}_2)=(\frac{4}{9}+\frac{1}{4}+\frac{1}{36})D(X)=0.72D(X),$$

$$D(\hat{\theta}_3)=(\frac{1}{9}+\frac{1}{9}+\frac{1}{9})D(X)=0.33D(X).$$

所以 $\hat{\theta}_3$ 最为有效。

# 三、一致性（Consistency）（不要求）

**定义** 如果对 $\forall \varepsilon > 0$，有 $\lim\limits_{n\to\infty} P\{|\hat\theta_n - \theta| \geq \varepsilon\} = 0$，

称 $\hat\theta_n$ 是 $\theta$ 的**一致估计量**或**相合估计量**。

即当 $n \to \infty$ 时，$\hat\theta(X_1, X_2, \cdots, X_n)$ 依概率收敛于 $\theta$.

直观上看，当$n$增大时，样本信息增多，当然希望估计量越来越靠近真值的概率也越来越大，这种想法就引出了上面的一致性概念。一致估计量一般地是当样本容量很大时，才能显示其优点。

大数定律：$\bar X \xrightarrow{P} \mu, A_l \xrightarrow{P} \mu_l$

可知样本原点矩 $\bar X$ 和 $A_l$ 分别是 $E(X)$ 和 $\mu_l$ 的一致估计量。

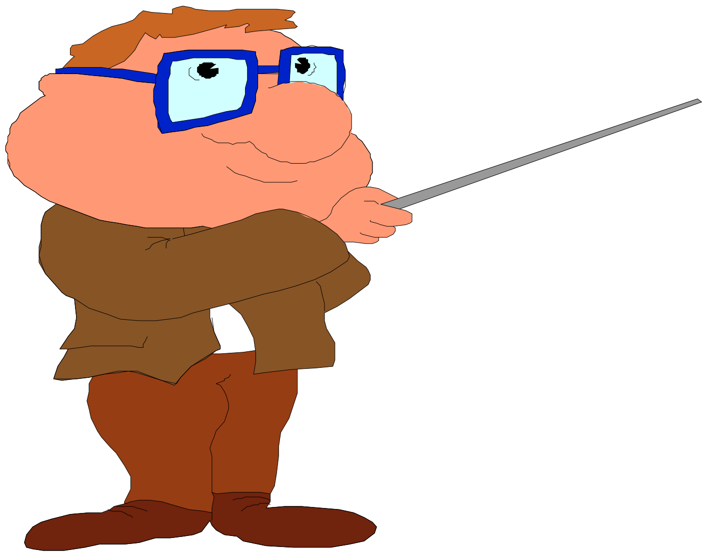
# 估计量的评选标准

一、无偏性 $E(\hat{\theta}) = \theta$

二、有效性 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$

三、一致性 $\hat{\theta}_n \xrightarrow[n\to\infty]{P} \theta$

## 思考



如何理解无偏性、有效性、一致性？

谢谢大家！