本次课的主要内容：第六章 三大统计分布及抽样分布

下次课内容：

第七章 参数估计-点估计

统计推断三个方面：

1. 抽样分布；(精确分布)
2. 参数估计；(已知分布类型)
3. 假设检验。

- 参数估计
  - 点估计
    - 矩估计
    - 极大似然估计
  - 区间估计

第七章　参 数 估 计

- 参数估计是对已知分布类型的总体，利用样本对其未知参数作出估计.

  例如: 总体为正态分布$N(\mu,\sigma^2)$,
  对两个参数进行估计.

- 参数估计可作如下划分:

参数估计
- 点估计
  - 矩估计
  - 极大似然估计
- 区间估计

样本矩与总体矩 (第四章中定义的随机变量矩)

$$A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^{\,k}$$ (随机变量) $\neq$ $E(X^K)$ (数)

样本矩：是随机变量, 包含了部分个体的信息,
取值是变化的, 由具体的实验结果决定.

总体矩：包含了所有个体信息, 取值是确定的数.

思想来源：依大数定律有

$$\lim_{n\to\infty} P\{|\overline{X}_n - E(\overline{X}_n)| < \varepsilon\} = 1 \quad \text{或} \quad \overline{X}_n - E(\overline{X}_n) \xrightarrow{p} 0$$

即样本的一阶原点矩依概率收敛于总体的一阶原点矩.

实际中用 $\overline{X}$ 估计$E(X)$是很有说服力的.

§7.1 参数的点估计

一、矩估计法

(总体)矩是总体随机变量最简单的数字特征,

设$X_1, X_2, \ldots, X_n$是总体X的样本, 有

$E(A_K)=$ (随机变量) $\frac{1}{n}\sum_{i=1}$ $\frac{1}{n}$ $\neq$ $)=E(X^K)$ (数)

$$E(X_i^K)=E(X^k)$$

矩估计法的基本思想:

* 用样本矩去替换相应的总体矩;

* 或用样本矩的函数替换相应的总体矩的同一函数,

称之为替换原则.

例如: 期望$E(X)$可用样本均值(样本一阶原点矩)估计.

定义：设总体X的分布中含有m个未知参数$\theta_1,\theta_2,\ldots,\theta_m$

X 的m阶原点矩存在, 即$E(X^m)$, 构造方程组

$$\frac{1}{n}\sum_{i=1}^{n}X_i^K = E(X^K), \qquad K=1, 2, \ldots, m.$$

解得 $\hat{\theta}_K = \theta_K(X_1, X_2, \cdots, X_n)$, $K=1, 2, \ldots, m.$

称 $\hat{\theta}_K$为θ_K 的矩估计量.

注意:样本矩是随机变量,而总体矩是数值

对样本进行一次观测就可以得到参数的一个估计值.

- 参数估计是对已知分布类型的总体，利用样本对其未知参数作出估计.

矩估计法前提: 总体X 的m阶原点矩存在, 即$E(X^m)$,

替换原则 令 $\frac{1}{n}\sum_{i=1}^{n} X_i^K = E(X^K)$, K=1，2，… ,m.

解得 $\hat{\theta}_K = \theta_K(\mathrm{X}_1, \mathrm{X}_2, \cdots, \mathrm{X}_n)$ ，K=1，2，… ,m.

称 $\hat{\theta}_K$为θ_K 的矩估计量.

常用情况: $m=1$, 令 $\overline{X} = E(X)$ $m=2$, 令 $\begin{cases} \overline{X} = E(X) \\ M_2 = A_2 - A_1^2 = D(X) \end{cases}$

注意: 一般选用低阶矩(若均值和方差相等).

缺陷: 总体X 的m阶原点矩存在, 例Cauchy分布(例4.1.4)?

例7.1: 不合格品率的矩法估计

设某车间生产一批产品，为估计该批产品不合格品率，抽取一组样本$X_1, X_2 \ldots X_n$, 总体X的分布? 即抽一件产品的不合格产品数.

记 $p= P\{X=1\}=P\{$产品不合格$\}$.

解：有 $X_i=\begin{cases}1, & \text{第}i\text{次取到不合格品；}\\ 0, & \text{第}i\text{次取到合格品。}\end{cases}$

由替换原则: $\frac{1}{n}\sum_{i=1}^{n} X_i^K = E(X^K)$, $K=1，2，\ldots,m.$

因 $p=E(X)=1\times p+0\times(1-p) \Rightarrow \frac{1}{n}\sum_{i=1}^{n} X_i = E(X)$, $k=1$

故 p 的矩估计量: $\hat{p}=\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = f_n(A)$

(即出现不合格产品的频率)思考:总体能用二项分布吗?

例7.2:设总体X的概率密度函数: 两参数的指数分布的矩估计

$$f(x)=\begin{cases}\dfrac{1}{\theta}e^{-(x-\mu)/\theta}, & x\ge\mu;\\ 0, & x<\mu.\end{cases}$$

$$\frac{1}{n}\sum_{i=1}^{n}X_i^K=E(X^K),$$

$K=1$, 2, … ,m.

其中$\theta>0$，μ与θ是未知参数，X_1，X_2，…, X_n，是X 的一组样本，求μ与 θ的矩估计量.

解:
$$E(X)=\int_{\mu}^{+\infty}\frac{x}{\theta}e^{-\frac{x-\mu}{\theta}}dx \quad 令 y=x-\mu$$

$$=\int_0^{+\infty}(y+\mu)\frac{1}{\theta}e^{-\frac{y}{\theta}}dy=E(Y+\mu)$$

$$=\int_0^{+\infty}y\frac{1}{\theta}e^{-\frac{y}{\theta}}dy+\int_0^{+\infty}\mu\frac{1}{\theta}e^{-\frac{y}{\theta}}dy=\theta+\mu$$

$$E(X^2)=\int_{\mu}^{+\infty}\frac{x^2}{\theta}e^{-\frac{x-\mu}{\theta}}dx \quad 令 y=x-\mu$$

$$=\int_0^{+\infty}(y+\mu)^2\frac{1}{\theta}e^{-\frac{y}{\theta}}dy=E[(Y+\mu)^2]$$

$$=D(Y+\mu)+[E(Y+\mu)]^2$$

$$=\theta^2+(\theta+\mu)^2$$

$$D(Y)=D(Y+\mu)$$

$$D(X)=E(X^2)-[E(X)]^2=\theta^2$$

令

或替换原则

$$\begin{cases}\theta+\mu=\bar{X}, \\ \theta^2=M_2\end{cases} \Rightarrow \begin{aligned}&\hat{\theta}=\sqrt{M_2}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2} \\ &\hat{\mu}=\bar{X}-\sqrt{M_2}\end{aligned}$$

二、极大似然估计法

(Maximum likelihood method(*M.L.E.*), 英国的Fisher)

由小概率事件原理:

概率很小的事件, 在一次试验中几乎是不可能发生的, 从而在实际中可看成不可能事件.

若随机事件有若干个可能结果: $A_1, A_2, \ldots, A_m$,

在一次试验中事件A_1出现了, 应认为A_1发生的可能性大.

试验的条件即参数θ的取值应该使A_1出现的概率最大。

不合格品率的*M.L.E.*估计

极大似然估计法基本思想:

按照最大可能性准则进行推断.

例7.3：不合格品率p 的极大似然(*M.L.E.*)估计

设总体X是抽一件产品的不合格品数，记

$p= P\{X=1\}=P\{$产品不合格$\}$

则 X的分布律可表示为:

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

解: 现抽取X的一组样本$X_1, X_2, \dots, X_n$

若干个结果$\{x_1, x_2, \dots, x_n\}$, $\{y_1, y_2, \dots, y_n\}$,…, $\{z_1, z_2, \dots, z_n\}$

做一次试验，得到一组实际观察值$x_1, x_2, \dots, x_n$,

则事件: $\{X_1=x_1, X_2=x_2, \dots, X_n=x_n\}$

出现的可能性最大, 其概率为:

例7.3.：不合格品率p 的极大似然(*M.L.E.*)估计

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

事件$\{X_1=x_1,X_2=x_2,\ldots,X_n=x_n\}$ 出现的可能性最大,其概率为:

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

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

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

比如$n=100$,抽到10个不合格,

即$p^{10}(1-p)^{90}$.

$$=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i},\ (x_i=0,1;0<p<1)$$

应选取使$L(p)$ 达到最大的值作为参数 p 的估计.

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

应选取使$L(p)$达到最大的值作为参数 p 的估计.

注意到： $L(\hat{p})=\max_{0<p<1}L(p) \Leftrightarrow \ln L(\hat{p})=\max_{0<p<1}\ln L(p)$

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

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

解得：

$$\hat{p}=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$$ (估计值)

例7. 1矩估计量：

$$\frac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$$

- 极大似然估计法基本思想：按照最大可能性准则进行推断.

对于离散型总体X：

本质：极大值点的问题

做一次试验，得样本$X_1, X_2, \ldots, X_n$的一组实际观测值 $x_1, x_2, \ldots, x_n$， 则事件: $\{X_1=x_1, X_2=x_2, \ldots, X_n=x_n\}$出现的可能性最大, 概率为

$$P\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\} = \prod_{i=1}^{n} P\{X_i = x_i\}$$

对于连续型总体X：

样本$(X_1, X_2, \ldots, X_n)$落在点 $(x_1, x_2, \ldots, x_n)$的领域(边长分别为$dx_1, dx_2, \ldots, dx_n$的n维长方体) 内的概率近似为

$$\prod_{i=1}^{n} f(x_i;\theta_1,\cdots\theta_m)dx_i = \prod_{i=1}^{n} f(x_i;\theta_1,\cdots\theta_m)\prod_{i=1}^{n} dx_i$$

$\prod_{i=1}^{n} dx_i$与参数$\theta_1,\cdots\theta_m$无关，故只考虑$\prod_{i=1}^{n} f(x_i;\theta_1,\cdots\theta_m)$

定义: 若总体 X 的概率密度函数为 $f(x,\theta)$ (θ可能为向量),

$X_1, X_2, \cdots, X_n$为来自X的一个样本, n 维随机变量

$(X_1, X_2, \cdots, X_n)$ 的联合概率密度函数记为:

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

称之为参数θ的**似然函数**.

对于离散型样本, 其似然函数为其联合分布律:

$$L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta) \quad P\{X = x_i\} = p(x_i; \theta) \quad i = 1,2,..$$

极大似然估计法

求参数θ的估计值，使似然函数达到极大值。

定义：若

$$L(x_1, x_2, \ldots, x_n; \hat{\theta}) = \max_{\theta} L(x_1, x_2, \ldots, x_n; \theta)$$

则称 $\hat{\theta}$ 为θ的极大似然估计值。

相应的估计量 $\hat{\theta} = \hat{\theta}(X_1, X_2, \ldots, X_n)$

称为参数θ 的极大似然估计量.

注：$\ln L$ 是 L 的严格单增函数，$\ln L$ 与L有相同的极大值点，一般只需求 $\ln L$ 的极大值点.

求极大似然估计的一般步骤:

求极大似然估计的一般步骤:

1. 写出似然函数: $L(x_1, x_2, \ldots, x_n; \theta) = \prod_{i=1}^{n} p(x_i; \theta_1, \theta_2, \ldots, \theta_m)$

连续型为联合概率密度,离散型为联合分布律

2. 对似然函数取对数:

$$\ln L = \sum_{i=1}^{n} \ln p(x_i; \theta_1, \theta_2, \ldots, \theta_m)$$

3. 对θ_j (j =1,…, m)分别求偏导, 建立似然方程(组):

$$\frac{\partial \ln L}{\partial \theta_j} = 0 \qquad (j = 1,2,\ldots,m)$$

不是估计量

解得 $\hat{\theta}_1, \ldots, \hat{\theta}_m$ 分别为 $\theta_1, \ldots, \theta_m$ 的极大估计值

例7.4：指数分布的点估计 指数分布的点估计

某电子管的使用寿命 X (单位：小时) 服从指数分布，

$$f_X(x;\theta)=\begin{cases}\dfrac{1}{\theta}e^{-\frac{x}{\theta}}, & x>0\\ 0, & else\end{cases} \quad (\theta>0)$$

今取一组样本，样本数据如下，问如何估计θ?

16	29	50	68	100	130	140	270	280
340	410	450	520	620	190	210	800	1100

分析 可用两种方法估计: 矩法估计和极大似然估计

（一）矩法估计

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

令 $\overline{X}=\theta$ 则可得的矩估计量

$$\hat{\theta}=\overline{X}$$

代入具体数值可得θ的估计值为:

$$\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{18}\cdot 5723\approx 318(\text{小时})$$

（二）极大似然估计

1. 构造似然函数:

设$x_1, x_2, \ldots, x_n$是$X_1, \ldots, X_n$的样本值,

$$L(x_1,\ldots,x_n;\theta)=\begin{cases}\prod_{i=1}^{n}\frac{1}{\theta}e^{-\frac{x_i}{\theta}}=\theta^{-n}\cdot e^{-\frac{1}{\theta}\sum_{i=1}^{n}x_i} & x_i>0, i=1,\ldots,n\\ 0, & else.\end{cases}$$

2. 取对数: 当 $0<x_i(i=1,2,\ldots,n)$ 时,

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

3. 建立似然方程

$$\frac{d\ln L}{d\theta}=-\frac{n}{\theta}+\frac{1}{\theta^2}\sum_{i=1}^{n}x_i=0$$

3. 建立似然方程 $\dfrac{d\ln L}{d\theta}=-\dfrac{n}{\theta}+\dfrac{1}{\theta^2}\sum_{i=1}^{n}x_i=0$

4. 求解得*M.L.E*值:

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

得*M.L.E*量: $\hat{\theta}=\dfrac{1}{n}\sum_{i=1}^{n}X_i=\bar{X}$

代入具体数值可得θ的估计值为:

$$\frac{1}{n}\sum_{i=1}^{n}x_i=\frac{1}{18}\cdot 5723\approx 318(\text{小时})$$

例 7.1.12 设总体 X 的概率分布列于表 7.1.1 中，其中 $\theta\left(0<\theta<\frac{1}{2}\right)$ 是未知参数，利用总体的如下样本值：

$$3,\ 1,\ 3,\ 0,\ 3,\ 1,\ 2,\ 3,$$

求 θ 的矩估计值和最大似然估计值.

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

例 7.1.12 设总体 X 的概率分布列于表 7.1.1 中，其中 $\theta\left(0<\theta<\frac{1}{2}\right)$ 是未知参数，利用总体的如下样本值：

3，1，3，0，3，1，2，3，

求 θ 的矩估计值和最大似然估计值.

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

例 7.5 均匀分布的极大似然估计

设样本$X_1, X_2, \ldots, X_n$是来自在区间$(0, \theta)$上均匀分布的总体, θ未知, 求: (1)θ的矩估计量; (2)极大似然估计量.

解：(1) $\because E(X) = \dfrac{\theta}{2}$　令 $\dfrac{\theta}{2} = \bar{X}$

θ的矩法估计量为$2\bar{X}$

(2) 设$x_1, x_2, \ldots, x_n$是$X_1, X_2, \ldots, X_n$的样本值,

总体的概率密度函数为:

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

然而可得似然函数

$$L(x_1,\dots,x_n;\theta)=\begin{cases}\dfrac{1}{\theta^n}, & 0<x_i<\theta, \quad i=1,2,\cdots n\\ 0, & \text{else}\end{cases}$$

注　　意：

该似然函数不能通过求导构造似然方程。

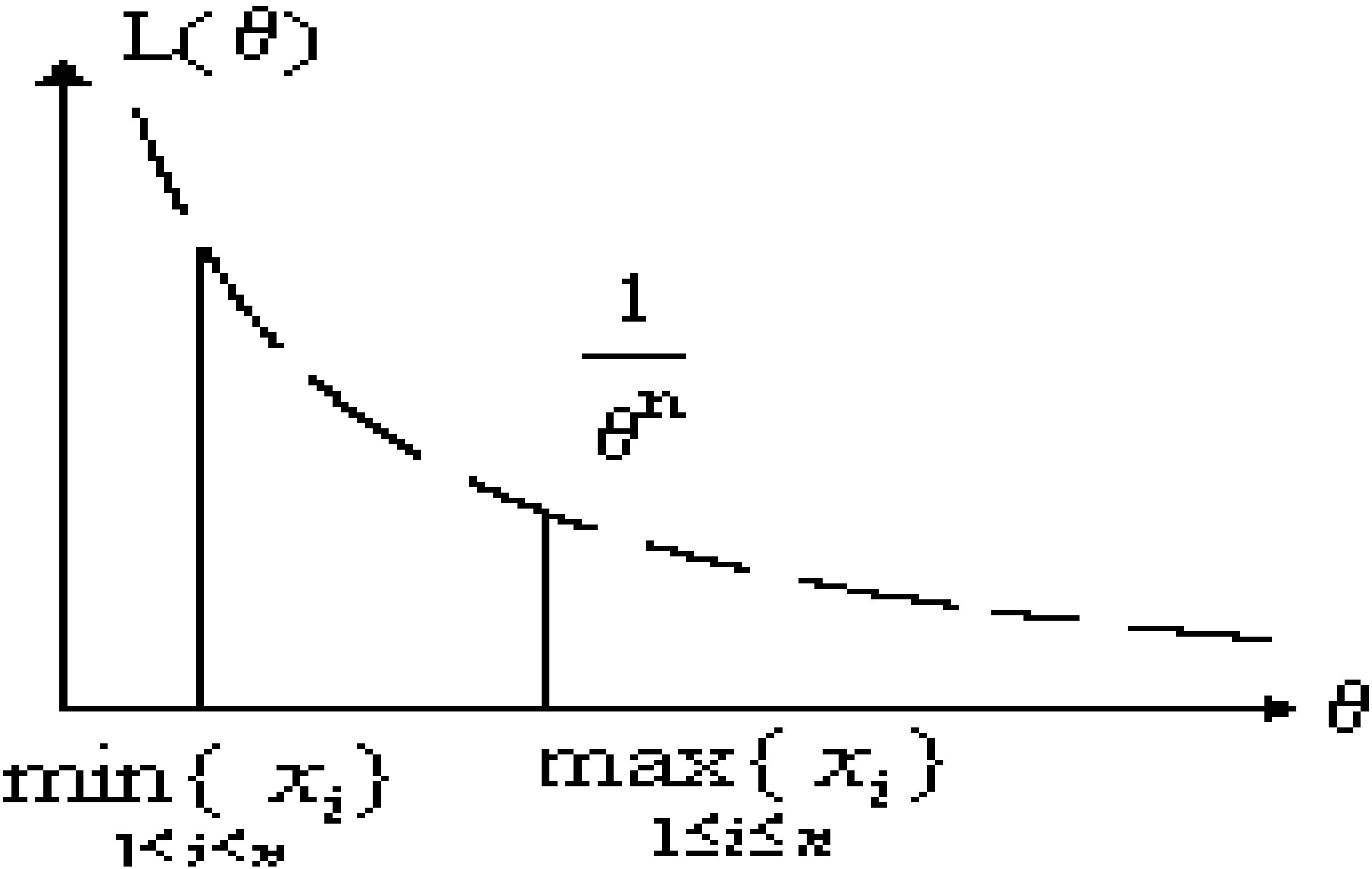
尝试用其他方法求解！

分析　θ的估计应满足：

1. θ的值尽可能小；

2. θ的值不能小于任何一个x_i。

似然函数趋势图



$$\because L=\begin{cases}\dfrac{1}{\theta^n}, & 0<\min\limits_{1\le i\le n}\{x_i\},\max\limits_{1\le i\le n}\{x_i\}<\theta\\ 0, & \text{else}\end{cases}$$

很明显，L的最大值出现在$L\neq 0$的区域，

此时，$\theta \geq \max\limits_{1\le i\le n}\{x_i\}$

如图所示：似然函数L是关于θ的减函数.

所以，当$\theta=\max\limits_{1\le i\le n}\{x_i\}$时，$L$达到极大值.

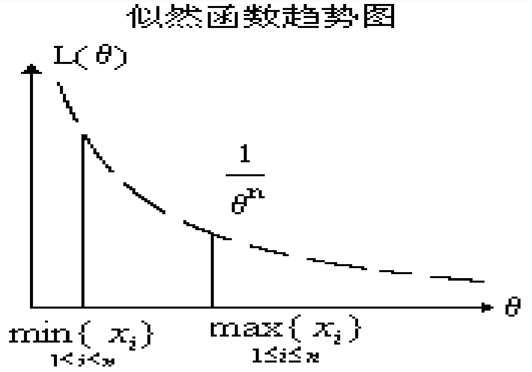
故θ的极大似然估计量为：$\hat{\theta}=\max\limits_{1\le i\le n}\{X_i\}$

续例(教材例7.1.8) 均匀分布的极大似然估计

设样本X_1，X_2，…，X_n是来自在区间(a, b)上均匀分布的总体, a, b未知, 求a,b的极大似然估计量.

可得似然函数

$$L(x_1,\dots,x_n;a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a<\min\limits_{1\le i\le n}\{x_i\},\max\limits_{1\le i\le n}\{x_i\}<b,\quad i=1,2,\cdots n\\ 0, & \text{else}\end{cases}$$



$$\text{显然}:\frac{1}{(b-a)^n}\le\frac{1}{(\max\limits_{1\le i\le n}\{x_i\}-\min\limits_{1\le i\le n}\{x_i\})^n}$$

a,b的极大似然估计值: $\hat{a}=x_{(1)}=\min\limits_{1\le i\le n}\{x_i\}$ $\hat{b}=x_{(n)}=\max\limits_{1\le i\le n}\{x_i\}$

a,b的极大似然估计量: $\hat{a}=X_{(1)}=\min\limits_{1\le i\le n}\{X_i\}$ $\hat{b}=X_{(n)}=\max\limits_{1\le i\le n}\{X_i\}$

小 结

1. 矩法估计量与极大似然估计量不一定相同；

2. 用矩法估计参数比较简单，但有信息量损失；

3. 极大似然估计法精度较高，但运算较复杂；

4. 不是所有*M.L.E* 都需要建立似然方程求解.