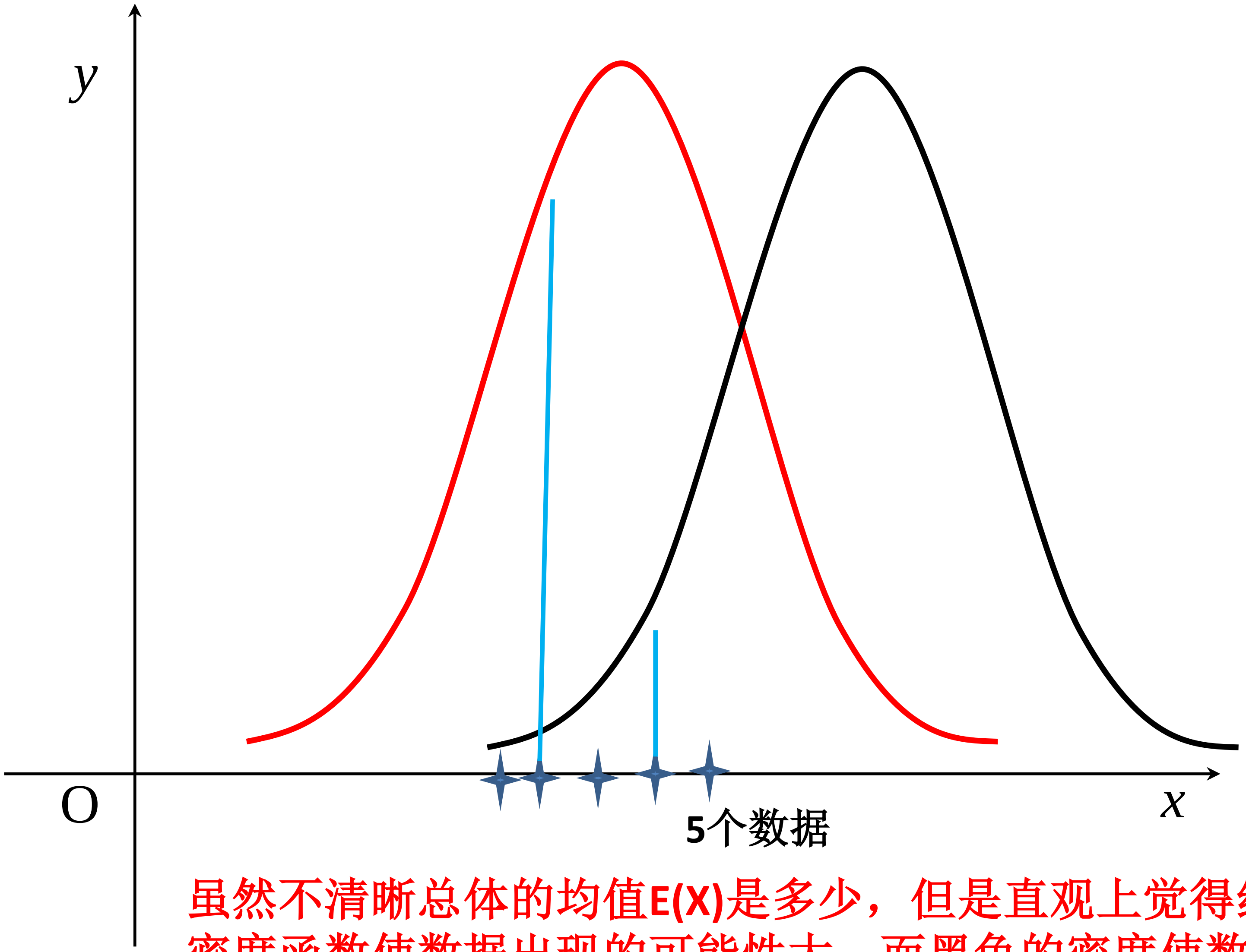
参数估计-极大似然估计

注意：极大似然估计的想法

找一个参数的值，试样本出现的可能性达到最大

因为样本已经出现（无论是具体的数值，还是具有表达式的小写x_i），那么我们倾向于认为样本是以“较大概率”地方式出现的（换句话说，小概率事件很难在一次观测中出现）

既然样本已经出现，那就是比较大的概率出现的



虽然不清晰总体的均值E(X)是多少，但是直观上觉得红色的密度函数使数据出现的可能性大，而黑色的密度使数据出现的可能性小，因此更倾向于数据来自于红色密度

极大似然估计：数据出现的可能性达到最大

（1）离散型样本$X_1,\dots,X_n$ 观测值为$x_1,\dots,x_n$ 找一个λ

使数据出现的可能性达到最大，即关于$L(\lambda)$ 找最大值（或极大值）：

$$L(\lambda)=Likelihood(\lambda)=P(X_1=x_1,\dots,X_n=x_n)=\prod_{k=1}^{n}P(X_k=x_k,\ \lambda)$$

当$\lambda=\hat{\lambda}$ 时，$L(\lambda)$达到最大（极大），$\hat{\lambda}$ 就是参数λ 的极大似然估计

（2）连续型样本$X_1,\dots,X_n$，观测值为$x_1,\dots,x_n$ 数据出现的

可能性 $L(\theta)=Likelihood(\theta)=\prod_{k=1}^{n}f(x_k,\theta)$，关于$L(\theta)$找最大值

（或极大值）使数据出现的可能性达到最大。

当$\theta=\hat{\theta}$ 时，$L(\theta)$达到最大（极大），$\hat{\theta}$ 就是参数θ 的极大似然估计

此时求解出来的往往是样本的函数，也就是统计量

什么时候求导，什么时候不能求导

（1）密度函数（连续）或概率（离散）大于0的x区域，与未知参数无关，此时就可以求导

（2）密度函数（连续）或概率（离散）大于0的x区域，与未知参数有关，此时往往不可以求导，需要对该参数考虑单调的方法找最大值。

例1：　可以求导

设 X 服从参数为 $\lambda\ (\lambda > 0)$ 的泊松分布，$X_1, X_2, \cdots, X_n$ 是来自 X 的一个样本，求 λ 的最大似然估计量.

解　因为 X 的分布律为

$$P\{X = x\} = \frac{\lambda^x}{x!}e^{-\lambda}, \quad (x = 0, 1, 2, \cdots, n)$$

所以 λ 的似然函数为

$$L(\lambda) = \prod_{i=1}^{n}\left(\frac{\lambda^{x_i}}{x_i!}e^{-\lambda}\right) = e^{-n\lambda}\frac{\lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n}(x_i!)},$$

$$\ln L(\lambda) = -n\lambda + \left(\sum_{i=1}^{n} x_i\right)\ln\lambda - \sum_{i=1}^{n}(x_i!),$$

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

解得λ的最大似然估计值 $$\lambda = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x},$$

λ的最大似然估计量为 $$\hat{\lambda} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

这一估计量与矩估计量是相同的.

连续分布（正态分布）：求导方法

例2 对正态总体$N(\mu,\sigma^2)$，$\vartheta=(\mu,\sigma^2)$是二维参数，设有样本 $x_1, x_2, \ldots, x_n$，则似然函数及其对数分别为

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

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

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

将 ln$L(\mu,\sigma^2)$ 分别关于两个分量求偏导并令其为0, 即得到似然方程组

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

(6.1.9)

$$\Rightarrow \sum_{i=1}^{n}(x_i-\mu)=0 \Rightarrow \sum_{i=1}^{n}x_i=n\mu, \Rightarrow \mu=\overline{x}$$

$$\frac{\partial \ln L(\mu,\sigma^2)}{\partial \sigma^2}=\frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2-\frac{n}{2\sigma^2}=0$$

(6.1.10)

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

解此方程组，由(6.1.9)可得μ的极大似然估计为

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

将之代入(6.1.10)，得出σ^2的极大似然估计

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

利用二阶导函数矩阵（海塞矩阵）的非正定性可以说明上述估计使得似然函数取极大值(本教材无需计算)。

本例中极大似然估计与某些矩估计也是相同的

例3： 单调方法

设总体 X在$[a,b]$上服从均匀分布,其中a, b未知, $x_1,x_2,\cdots,x_n$是来自总体 X的一个样本值, 求a,b的最大似然估计量.

解 记 $x_{(1)}=\min(x_1,x_2,\cdots,x_n)$,

$x_{(n)}=\max(x_1,x_2,\cdots,x_n)$,

X的概率密度为

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

密度大于0的区域与位置的a,b有关

因为$a \le x_1, x_2, \cdots, x_n \le b$等价于$a \le x_{(1)}, x_{(n)} \le b$,

作为a, b的函数的似然函数为

$$L(a,b)=\begin{cases}\dfrac{1}{(b-a)^n}, & a \le x_{(1)},\ b \ge x_{(n)} \\ 0, & \text{其它}\end{cases}$$

于是对于满足条件 $a \le x_{(1)},\ b \ge x_{(n)}$的任意$a,b$有

$$L(a,b)=\frac{1}{(b-a)^n} \le \frac{1}{(x_{(n)}-x_{(1)})^n},$$

L(a,b)关于a,单调递增，所以当a取到最大值的时候，L(a,b)关于a达到最大。

L(a,b)关于b,单调递减少，所以当b取到最小值的时候，L(a,b)关于b达到最大。

即似然函数 $L(a, b)$ 在 $a = x_{(1)}$, $b = x_{(n)}$ 时

取到最大值 $(x_{(n)} - x_{(1)})^{-n}$,

a, b 的最大似然估计值

$$\hat{a} = x_{(1)} = \min_{1 \le i \le n} x_i, \qquad \hat{b} = x_{(n)} = \max_{1 \le i \le n} x_i,$$

a, b 的最大似然估计量

$$\hat{a} = \min_{1 \le i \le n} X_i, \qquad \hat{b} = \max_{1 \le i \le n} X_i .$$

连续分布（均匀分布）：单调方法

虽然求导函数是求极大似然估计最常用的方法，但并不是在所有场合求导都是有效的。

例4 设 $x_1, x_2, \ldots, x_n$ 是来自均匀总体 $U[0, \theta]$(闭区间)的样本，试求 θ 的极大似然估计。

$$L(\theta)=\prod_{k=1}^{n} f(x_k,\theta)=\prod_{k=1}^{n}\begin{cases}\dfrac{1}{\theta} & 0\le x_k\le\theta \\ 0 & \text{其他}\end{cases}=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_1,\ldots x_n\le\theta \\ 0 & \text{其他}\end{cases}$$

$$=\begin{cases}\dfrac{1}{\theta^n} & 0\le\min\{x_1,\ldots x_n\}\le\max\{x_1,\ldots x_n\}\le\theta \\ 0 & \text{其他}\end{cases}=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_{(1)}\le x_{(n)}\le\theta \\ 0 & \text{其他}\end{cases}$$

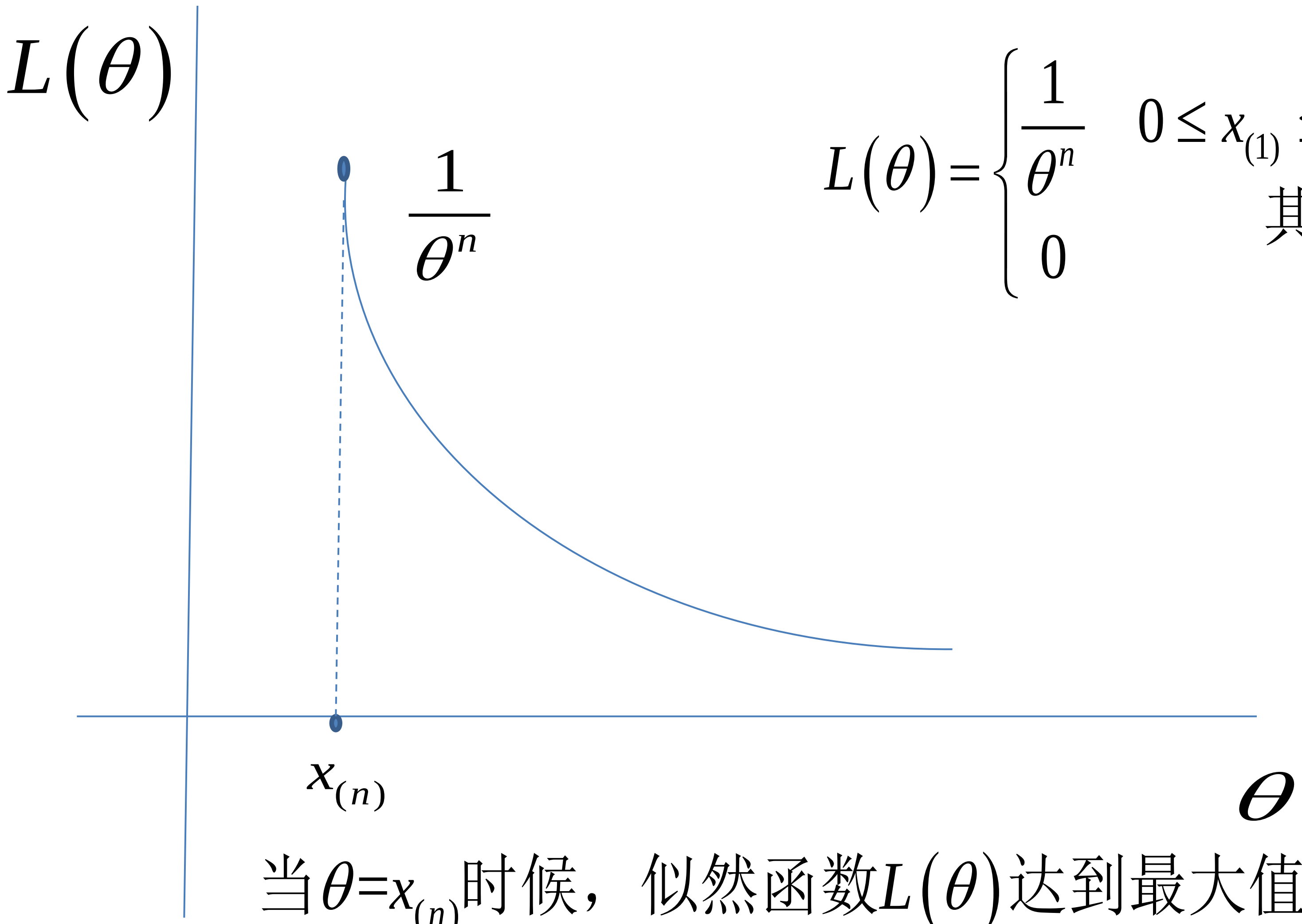
似然函数

$$L(\theta)=\begin{cases}\dfrac{1}{\theta^n} & 0\le x_{(1)}\le x_{(n)}\le\theta \\ 0 & \text{其他}\end{cases}$$

要使$L(\theta)$达到最大，首先一点是示性函数取值应该为1，其次是$1/\theta^n$尽可能大。由于$1/\theta^n$是θ的单调减函数，所以θ的取值应尽可能小，但示性函数为1决定了θ不能小于$x_{(n)}$，由此给出θ的极大似然估计：$\hat{\theta}=x_{(n)}$。

可看下图

$$\hat{\theta}=x_{(n)}$$



$$L(\theta)=\begin{cases}\frac{1}{\theta^n} & 0\le x_{(1)}\le x_{(n)}\le\theta \\ 0 & \text{其他}\end{cases}$$

当θ=$x_{(n)}$时候，似然函数$L(\theta)$达到最大值，因此$x_{(n)}$是θ的最大似然估计（此时能取到最大值）

注意到，本例中，也不是不可以求导。找θ使得$L(\theta)$达到最大，很明显$\frac{1}{\theta^n}>0$. 因此最大值不可能在$L(\theta)$=0处出现，

只需讨论$x_{(n)}\le\theta$范围内的$L(\theta)=\frac{1}{\theta^n}$，取对数$\ln L(\theta)=-n\ln\theta$,

$\frac{dL(\theta)}{d\theta}=-\frac{n}{\theta}<0$(不可能等于0，严格小于0)

由于导数 $\frac{dL(\theta)}{d\theta}<0,\ln L(\theta)$严格单调递减，

在范围内$x_{(n)}\le\theta$，当θ=$x_{(n)}$时达到最大值。

实际上，很容易看出$\frac{1}{\theta^n}$单调减，求导有点多此一举了。