数理统计
第三章
点估计

2026 年 4 月 15 日

目录

- 最大似然法是在参数分布族情形下常用的参数估计方法。

- 最大似然法是在参数分布族情形下常用的参数估计方法。
- 设有一参数分布族$\mathcal{F} = \{f(x,\theta), \theta \in \Theta\}$，其中$\Theta$为参数空间。

3.3.1 引言及定义

- 最大似然法是在参数分布族情形下常用的参数估计方法。
- 设有一参数分布族$\mathcal{F}=\{f(x,\theta),\theta\in\Theta\}$，其中$\Theta$为参数空间。
- 令$\boldsymbol{X}=(X_1,\ldots,X_n)$为从$\mathcal{F}$中抽取的简单样本，

  $f(\boldsymbol{x},\theta)=f(x_1,\ldots,x_n,\theta)$为样本$\boldsymbol{X}$的概率函数。

- 最大似然法是在参数分布族情形下常用的参数估计方法。
- 设有一参数分布族$\mathcal{F}=\{f(x,\theta),\theta\in\Theta\}$，其中$\Theta$为参数空间。
- 令$\boldsymbol{X}=(X_1,\ldots,X_n)$为从$\mathcal{F}$中抽取的简单样本，
  $f(\boldsymbol{x},\theta)=f(x_1,\ldots,x_n,\theta)$为样本$\boldsymbol{X}$的概率函数。
- 当总体分布为连续型时，$f(\boldsymbol{x},\theta)$表示样本$\boldsymbol{X}$的密度函数；

3.3.1 引言及定义

- 最大似然法是在参数分布族情形下常用的参数估计方法。
- 设有一参数分布族$\mathcal{F}=\{f(x,\theta),\theta\in\Theta\}$，其中$\Theta$为参数空间。
- 令$\boldsymbol{X}=(X_1,\ldots,X_n)$为从$\mathcal{F}$中抽取的简单样本，

  $f(\boldsymbol{x},\theta)=f(x_1,\ldots,x_n,\theta)$为样本$\boldsymbol{X}$的概率函数。
- 当总体分布为连续型时，$f(\boldsymbol{x},\theta)$表示样本$\boldsymbol{X}$的密度函数；
- 当总体分布为离散型时，$f(\boldsymbol{x},\theta)$为样本$\boldsymbol{X}$的概率分布，

  即$f(\boldsymbol{x},\theta)=P_\theta(X_1=x_1,\ldots,X_n=x_n)$。

3.3.1 引言及定义

定义 (3.3.1（似然函数）)

设$f(\boldsymbol{x},\theta)=f(x_1,\ldots,x_n,\theta)$为样本$\boldsymbol{X}=(X_1,\ldots,X_n)$的概率函数。

当$\boldsymbol{x}$固定时，把$f(\boldsymbol{x},\theta)$看成θ的函数，称为似然函数，记为

$$L(\theta,\boldsymbol{x})=f(\boldsymbol{x},\theta),\ \theta\in\Theta,\ \boldsymbol{x}\in\mathcal{X}, \tag{1}$$

其中Θ为参数空间，$\mathcal{X}$为样本空间。

称$\ell(\theta,\boldsymbol{x})=\ln L(\theta,\boldsymbol{x})$为对数似然函数。

注 3.3.1 似然函数和概率函数是同一表达（见式（1）），但表示两种不同含义。

注 3.3.1 似然函数和概率函数是同一表达（见式（1）），但表示两种不同含义。

当把θ固定，将其看成定义在样本空间$\mathcal{X}$上的函数时，称为**概率函数**；

注 3.3.1 似然函数和概率函数是同一表达（见式（1）），但表示两种不同含义。

当把θ固定，将其看成定义在样本空间$\mathcal{X}$上的函数时，称为**概率函数**；

当把x固定，将其看成定义在参数空间Θ上的函数时，称为**似然函数**。

例 (3.3.1)

设罐子里有许多黑球和红球。假定已知它们的比例是1：3，但不知道是黑球多还是红球多。也就是说抽出一个黑球的概率或者是1/4或者是3/4。如果有放回地从罐子中抽n个球，要根据抽样数据，说明抽到黑球的概率是1/4还是3/4。

例 (3.3.1)

设罐子里有许多黑球和红球。假定已知它们的比例是1：3，但不知道是黑球多还是红球多。也就是说抽出一个黑球的概率或者是1/4或者是3/4。如果有放回地从罐子中抽n个球，要根据抽样数据，说明抽到黑球的概率是1/4还是3/4。

解 将此问题用统计模型来表述。令X_i表示第i次抽球的结果，即

例 (3.3.1)

设罐子里有许多黑球和红球。假定已知它们的比例是1：3，但不知道是黑球多还是红球多。也就是说抽出一个黑球的概率或者是1/4或者是3/4。如果有放回地从罐子中抽n个球，要根据抽样数据，说明抽到黑球的概率是1/4还是3/4。

解 将此问题用统计模型来表述。令X_i表示第i次抽球的结果，即

$$X_i = \begin{cases} 1, & \text{第i次抽出为黑球}, \\ 0, & \text{否则}. \end{cases}$$

记每次抽样中抽到黑球的概率为θ，此处θ只取可能的两个值

$$\theta_1 = 1/4, \text{ 和}\theta_2 = 3/4$$

之一，参数空间$\Theta = \{\theta:\ \theta_1 = 1/4,\ \theta_2 = 3/4\}$。

记每次抽样中抽到黑球的概率为θ，此处θ只取可能的两个值

$$\theta_1 = 1/4,\ \text{和}\theta_2 = 3/4$$

之一，参数空间$\Theta = \{\theta:\ \theta_1 = 1/4,\ \theta_2 = 3/4\}$。

记$X = \sum_{i=1}^{n} X_i$，则$X \sim B(n, \theta)$。

样本分布族

$$\mathcal{F} = \{f(x,\theta):\ \theta \in \Theta\} = \{f(x,\theta_1), f(x,\theta_2)\},$$

样本分布族

$$\mathcal{F} = \{f(x,\theta):\ \theta \in \Theta\} = \{f(x,\theta_1), f(x,\theta_2)\},$$

其中$f(x,\theta_1)$为$B(n,\theta_1)$，$f(x,\theta_2)$为$B(n,\theta_2)$。

样本分布族

$$\mathcal{F} = \{f(x,\theta):\ \theta \in \Theta\} = \{f(x,\theta_1), f(x,\theta_2)\},$$

其中$f(x,\theta_1)$为$B(n,\theta_1)$，$f(x,\theta_2)$为$B(n,\theta_2)$。

要根据抽样结果对θ作出估计，即θ取值为1/4还是3/4?

样本分布族

$$\mathcal{F} = \{f(x,\theta):\ \theta \in \Theta\} = \{f(x,\theta_1), f(x,\theta_2)\},$$

其中$f(x,\theta_1)$为$B(n,\theta_1)$，$f(x,\theta_2)$为$B(n,\theta_2)$。

要根据抽样结果对θ作出估计，即θ取值为1/4还是3/4?

或者说样本来自总体$f(x,\theta_1)$还是$f(x,\theta_2)$?

显然，当观测到样本$X=x$时，似然函数

3.3.1 引言及定义

显然，当观测到样本$X = x$时，似然函数

$$L(\theta, x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \ x = 0, 1, 2, \ldots, n$$

表示了在θ对应的分布下观测到$X = x$的概率。

显然，当观测到样本$X=x$时，似然函数

$$L(\theta,x)=\binom{n}{x}\theta^x(1-\theta)^{n-x},\ x=0,1,2,\ldots,n$$

表示了在θ对应的分布下观测到$X=x$的概率。

为简单计，取$n=3$。

显然，当观测到样本$X = x$时，似然函数

$$L(\theta, x) = \binom{n}{x} \theta^x (1-\theta)^{n-x}, \ x = 0, 1, 2, \ldots, n$$

表示了在θ对应的分布下观测到$X = x$的概率。

为简单计，取$n = 3$。当$x = 0, 1, 2, 3$时似然函数取值如下表

表: 3.3.1

x	0	1	2	3
$L(\theta_1, x)$	27/64	27/64	9/64	1/64
$L(\theta_2, x)$	1/64	9/64	27/64	27/64

由表1可见:

当$x=0,1$时，$L(\theta_1,x)>L(\theta_2,x)$;

由表1可见：

当$x = 0, 1$时，$L(\theta_1, x) > L(\theta_2, x)$;

当$x = 2, 3$时，$L(\theta_2, x) > L(\theta_1, x)$。

由表1可见：

当$x=0,1$时，$L(\theta_1,x)>L(\theta_2,x)$;

当$x=2,3$时，$L(\theta_2,x)>L(\theta_1,x)$。

因此得出结论：当样本观察值$x=\sum_{i=1}^{3}x_i$取值为0，1时，认为样本来自总体$f(x,\theta_1)$，即取参数θ的估计值为$\theta_1=1/4$;

由表1可见：

当$x = 0, 1$时，$L(\theta_1, x) > L(\theta_2, x)$;

当$x = 2, 3$时，$L(\theta_2, x) > L(\theta_1, x)$。

因此得出结论：当样本观察值$x = \sum_{i=1}^{3} x_i$取值为0, 1时，认为样本来自总体$f(x, \theta_1)$，即取参数θ的估计值为$\theta_1 = 1/4$;

当$x = 2$, 3时，认为样本来自总体$f(x, \theta_2)$，即取θ的估计值为$\theta_2 = 3/4$。

由表1可见：

当$x = 0, 1$时，$L(\theta_1, x) > L(\theta_2, x)$;

当$x = 2, 3$时，$L(\theta_2, x) > L(\theta_1, x)$。

因此得出结论：当样本观察值$x = \sum_{i=1}^{3} x_i$取值为0，1时，认为样本来自总体$f(x, \theta_1)$，即取参数θ的估计值为$\theta_1 = 1/4$;

当$x = 2$，3时，认为样本来自总体$f(x, \theta_2)$，即取θ的估计值为$\theta_2 = 3/4$。

取使得事件$\{X = x\}$发生的概率最大的那个θ值，看起来“最像”是合理的。

将此例模型化如下：

将此例模型化如下：

若样本$\boldsymbol{X}=(X_1,\ldots,X_n)$为从总体$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$中抽取的简单样本，其中$\Theta=\{\theta_1,\theta_2\}$。

将此例模型化如下：

若样本$\boldsymbol{X} = (X_1, \ldots, X_n)$为从总体$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$中抽取的简单样本，其中$\Theta = \{\theta_1, \theta_2\}$。

等价地，分布族$\mathcal{F}$中只有两个总体$f(x, \theta_1)$和$f(x, \theta_2)$。

将此例模型化如下：

若样本$\boldsymbol{X} = (X_1, \ldots, X_n)$为从总体$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$中抽取的简单样本，其中$\Theta = \{\theta_1, \theta_2\}$。

等价地，分布族$\mathcal{F}$中只有两个总体$f(x, \theta_1)$和$f(x, \theta_2)$。

一旦获得了样本$\boldsymbol{x}$，如何用最大似然方法求出真参数θ的估计值呢？

上例表明，若

上例表明，若

$$L(\theta_1, \boldsymbol{x}) > L(\theta_2, \boldsymbol{x}),$$

上例表明，若

$$L(\theta_1, \boldsymbol{x}) > L(\theta_2, \boldsymbol{x}),$$

则倾向于认为样本$\boldsymbol{X}$来自总体$f(x, \theta_1)$（即真参数θ为θ_1）的理由比认为样本$\boldsymbol{X}$来自总体$f(x, \theta_2)$（即真参数θ为θ_2）的理由更充分些。

上例表明，若

$$L(\theta_1, \boldsymbol{x}) > L(\theta_2, \boldsymbol{x}),$$

则倾向于认为样本$\boldsymbol{X}$来自总体$f(x,\theta_1)$（即**真参数**θ**为**θ_1）的理由比认为样本$\boldsymbol{X}$来自总体$f(x,\theta_2)$（即真参数θ为θ_2）的理由更充分些。

或者说，真参数θ为θ_1的“似然性”更大些。

上例表明，若

$$L(\theta_1, \boldsymbol{x}) > L(\theta_2, \boldsymbol{x}),$$

则倾向于认为样本$\boldsymbol{X}$来自总体$f(x, \theta_1)$（即真参数θ为θ_1）的理由比认为样本$\boldsymbol{X}$来自总体$f(x, \theta_2)$（即真参数θ为θ_2）的理由更充分些。

或者说，真参数θ为θ_1的“似然性”更大些。

这样，自然把“似然性”最大（即看起来最像）的那个值作为真参数θ的估计值。这正是“最大似然”一词的由来。

更一般地，若样本$\boldsymbol{X}$的分布族$\mathcal{F} = \{f(\boldsymbol{x}, \theta) : \theta \in \Theta\}$，参数空间$\Theta$为$R$的有限子集或无限子集。

更一般地，若样本$\boldsymbol{X}$的分布族$\mathcal{F}=\{f(\boldsymbol{x},\theta):\theta\in\Theta\}$，参数空间$\Theta$为$R$的有限子集或无限子集。

当样本$\boldsymbol{x}$给定时，若$\hat{\theta}^*$使似然函数$L(\hat{\theta}^*,\boldsymbol{x})$为集合

$$\{L(\theta,\boldsymbol{x}):\text{一切}\theta\in\Theta\}$$

中最大者，

更一般地，若样本$\boldsymbol{X}$的分布族$\mathcal{F}=\{f(\boldsymbol{x},\theta):\theta\in\Theta\}$，参数空间$\Theta$为$R$的有限子集或无限子集。

当样本$\boldsymbol{x}$给定时，若$\hat{\theta}^*$使似然函数$L(\hat{\theta}^*,\boldsymbol{x})$为集合

$$\{L(\theta,\boldsymbol{x}):\text{一切}\theta\in\Theta\}$$

中最大者，即参数θ的真值为$\hat{\theta}^*$的“似然性”比参数空间Θ中任何其他参数值的“似然性”都大，

3.3.1 引言及定义

更一般地，若样本$\boldsymbol{X}$的分布族$\mathcal{F}=\{f(\boldsymbol{x},\theta):\theta\in\Theta\}$，参数空间$\Theta$为$R$的有限子集或无限子集。

当样本$\boldsymbol{x}$给定时，若$\hat{\theta}^*$使似然函数$L(\hat{\theta}^*,\boldsymbol{x})$为集合

$$\{L(\theta,\boldsymbol{x}):\text{一切}\theta\in\Theta\}$$

中最大者，即参数θ的真值为$\hat{\theta}^*$的“似然性”比参数空间Θ中任何其他参数值的“似然性”都大，则取“似然性”最大的$\hat{\theta}^*$作为θ的估计值，由这一方法得到的参数θ的估计，称为“最大似然估计”。

定义 (3.3.2（最大似然估计）)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从参数分布族$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$中抽取的简单样本，$L(\theta, \boldsymbol{x})$是似然函数，$\ell(\theta, \boldsymbol{x}) = \ln L(\theta, \boldsymbol{x})$为对数似然函数，若存在统计量$\hat{\theta}^* = \hat{\theta}^*(\boldsymbol{X})$，满足条件

定义 (3.3.2（最大似然估计）)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从参数分布族$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$中抽取的简单样本，$L(\theta, \boldsymbol{x})$是似然函数，$\ell(\theta, \boldsymbol{x}) = \ln L(\theta, \boldsymbol{x})$为对数似然函数，若存在统计量$\hat{\theta}^* = \hat{\theta}^*(\boldsymbol{X})$，满足条件

$$L(\hat{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) = \max_{\theta \in \Theta} L(\theta, \boldsymbol{x}), \ \forall \boldsymbol{x} \in \mathcal{X},$$

定义 (3.3.2（最大似然估计）（续）)

或等价地使得

定义 (3.3.2（最大似然估计）（续）)

或等价地使得

$$\ell(\hat{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) = \max_{\theta \in \Theta} \ell(\theta, \boldsymbol{x}), \ \forall \boldsymbol{x} \in \mathcal{X},$$

定义 (3.3.2（最大似然估计）（续）)

或等价地使得

$$\ell(\hat{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) = \max_{\theta\in\Theta} \ell(\theta, \boldsymbol{x}),\ \forall \boldsymbol{x} \in \mathcal{X},$$

则称$\hat{\theta}^*(\boldsymbol{X})$为$\theta$的最大似然估计（maximum likelihood estimation, MLE）。

定义 (3.3.2（最大似然估计）（续）)

或等价地使得

$$\ell(\hat{\theta}^*(\boldsymbol{x}), \boldsymbol{x}) = \max_{\theta \in \Theta} \ell(\theta, \boldsymbol{x}), \ \forall \boldsymbol{x} \in \mathcal{X},$$

则称$\hat{\theta}^*(\boldsymbol{X})$为$\theta$的最大似然估计（maximum likelihood estimation, MLE）。

若待估函数是$g(\theta)$，则定义$g(\hat{\theta}^*)$为$g(\theta)$的MLE。

注 3.3.2 最大似然估计具有不变性：

注 3.3.2 最大似然估计具有不变性：

设样本分布族如定义 3.3.2 所述，若参数θ的最大似然估计为θ^*，则θ的任一（可测）函数$g(\theta)$的最大似然估计为$g(\hat{\theta}^*)$。

注 3.3.2 最大似然估计具有不变性：

设样本分布族如定义 3.3.2 所述，若参数θ的最大似然估计为θ^*，则θ的任一（可测）函数$g(\theta)$的最大似然估计为$g(\hat{\theta}^*)$。

正因为MLE具有这一性质，因此我们将θ的函数$g(\theta)$的MLE放到定义3.3.2中。

3.3.2 最大似然估计的求法及例

1. 用微积分中求极值的方法

3.3.2 最大似然估计的求法及例

1. 用微积分中求极值的方法

设$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$为参数向量（当$k = 1$时，$\boldsymbol{\theta}$为参数$\theta$）。

1. 用微积分中求极值的方法

设$\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$为参数向量（当$k = 1$时，$\boldsymbol{\theta}$为参数$\theta$）。

若似然函数$L(\boldsymbol{\theta}, \boldsymbol{x})$是$\boldsymbol{\theta}$的连续可微函数，则可用微积分中求极值点的方法去求$\boldsymbol{\theta}$的MLE，即找使$L(\boldsymbol{\theta}, \boldsymbol{x})$达到最大时$\boldsymbol{\theta}$的值。

由于$L(\boldsymbol{\theta},\boldsymbol{x})$和对数似然函数$\ln L(\boldsymbol{\theta},\boldsymbol{x})=\ell(\boldsymbol{\theta},\boldsymbol{x})$具有相同的极值点，可用$\ell(\boldsymbol{\theta},\boldsymbol{x})$来代替$L(\boldsymbol{\theta},\boldsymbol{x})$。

由于$L(\boldsymbol{\theta}, \boldsymbol{x})$和对数似然函数$\ln L(\boldsymbol{\theta}, \boldsymbol{x}) = \ell(\boldsymbol{\theta}, \boldsymbol{x})$具有相同的极值点，可用$\ell(\boldsymbol{\theta}, \boldsymbol{x})$来代替$L(\boldsymbol{\theta}, \boldsymbol{x})$。

并且，由于$L(\boldsymbol{\theta}, \boldsymbol{x})$一般是若干个函数的乘积，而$\ell(\boldsymbol{\theta}, \boldsymbol{x})$为若干个函数之和，较易于处理。

3.3.2 最大似然估计的求法及例

若$\ell(\boldsymbol{\theta},\boldsymbol{x})$的极大值在参数空间$\Theta$的内点处（而非边界点）达到，则此点必为似然方程组

若$\ell(\boldsymbol{\theta}, \boldsymbol{x})$的极大值在参数空间$\Theta$的内点处（而非边界点）达到，则此点必为似然方程组

$$S_i(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{\partial \ell(\boldsymbol{\theta}, \boldsymbol{x})}{\partial \theta_i} = 0,\ i = 1, 2, \ldots, k$$

的解。

若$\ell(\boldsymbol{\theta},\boldsymbol{x})$的极大值在参数空间$\Theta$的内点处（而非边界点）达到，则此点必为似然方程组

$$S_i(\boldsymbol{\theta},\boldsymbol{x}) = \frac{\partial \ell(\boldsymbol{\theta},\boldsymbol{x})}{\partial \theta_i} = 0,\ i = 1, 2, \ldots, k$$

的解。

由于$S(\boldsymbol{\theta},\boldsymbol{x}) = \partial \ell(\boldsymbol{\theta},\boldsymbol{x})/\partial \boldsymbol{\theta}$的值表示了对数似然函数对参数$\boldsymbol{\theta}$的无穷小量变化的敏感程度，其在光滑对数似然函数的局部最大值处值为0，因此称为得分函数。

因此求最大似然估计首先求似然方程组的解$\hat{\theta}^*$。但此解是否一定是θ的MLE呢？

因此求最大似然估计首先求似然方程组的解$\hat{\theta}^*$。但此解是否一定是θ的MLE呢？

$\hat{\theta}^*$满足似然方程，只是MLE的必要条件，而非充分条件。

一般只有满足下列条件：

一般只有满足下列条件：

（1）似然函数的极大值在参数空间Θ内部达到；

（2）似然方程只有唯一解；

一般只有满足下列条件：

（1）似然函数的极大值在参数空间Θ内部达到；

（2）似然方程只有唯一解；

则似然方程之解$\hat{\theta}^*$必为θ的MLE。

3.3.2 最大似然估计的求法及例

一般只有满足下列条件：

（1）似然函数的极大值在参数空间Θ内部达到；

（2）似然方程只有唯一解；

则似然方程之解$\hat{\theta}^*$必为θ的MLE。

因此求出似然方程（或方程组）的解后，要验证它为θ的MLE，有时并非易事。

但对样本分布族是指数族的场合，有非常满意的结果，叙述如下。

但对样本分布族是指数族的场合，有非常满意的结果，叙述如下。

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从某总体中抽取的简单样本，$X_1$的分布为指数族。将样本分布族表示为指数族的自然形式有

但对样本分布族是指数族的场合，有非常满意的结果，叙述如下。

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从某总体中抽取的简单样本，$X_1$的分布为指数族。将样本分布族表示为指数族的自然形式有

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} \theta_i T_i(\boldsymbol{x})\right\} h(\boldsymbol{x}), \ \boldsymbol{\theta} \in \Theta^*,$$

但对样本分布族是指数族的场合，有非常满意的结果，叙述如下。

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从某总体中抽取的简单样本，$X_1$的分布为指数族。将样本分布族表示为指数族的自然形式有

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = C(\boldsymbol{\theta}) \exp\left\{\sum_{i=1}^{k} \theta_i T_i(\boldsymbol{x})\right\} h(\boldsymbol{x}),\ \boldsymbol{\theta} \in \Theta^*,$$

其中Θ^*为自然参数空间，Θ_0为Θ^*之内点集。

这时$\boldsymbol{X}$的联合密度为

这时$\boldsymbol{X}$的联合密度为

$$L(\boldsymbol{\theta},\boldsymbol{x})=\prod_{j=1}^{n}L(\boldsymbol{\theta},x_j)=C^n(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)\right\}h(\boldsymbol{x}),$$

这时$\boldsymbol{X}$的联合密度为

$$L(\boldsymbol{\theta},\boldsymbol{x})=\prod_{j=1}^{n}L(\boldsymbol{\theta},x_j)=C^n(\boldsymbol{\theta})\exp\left\{\sum_{i=1}^{k}\theta_i\sum_{j=1}^{n}T_i(x_j)\right\}h(\boldsymbol{x}),$$

此处$h(\boldsymbol{x})=\prod_{j=1}^{n}h(x_j)$。对上式取对数得对数似然函数

这时$\boldsymbol{X}$的联合密度为

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \prod_{j=1}^{n} L(\boldsymbol{\theta}, x_j) = C^n(\boldsymbol{\theta}) \exp\left\{ \sum_{i=1}^{k} \theta_i \sum_{j=1}^{n} T_i(x_j) \right\} h(\boldsymbol{x}),$$

此处$h(\boldsymbol{x}) = \prod_{j=1}^{n} h(x_j)$。对上式取对数得对数似然函数

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \ln L(\boldsymbol{\theta}, \boldsymbol{x}) = n \ln C(\boldsymbol{\theta}) + \sum_{i=1}^{k} \theta_i \sum_{j=1}^{n} T_i(x_j) + \ln h(\boldsymbol{x}).$$

定理 (3.3.1)

若对任何样本$\boldsymbol{X}=(X_1,\ldots,X_n)$，方程组

$$\frac{n}{C(\boldsymbol{\theta})}\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_i}=-\sum_{j=1}^{n}T_i(X_j),\ i=1,2,\ldots,k$$

在Θ_0内有解，则解必唯一且为$\boldsymbol{\theta}$的*MLE*。

定理 (3.3.1)

若对任何样本$\boldsymbol{X}=(X_1,\ldots,X_n)$，方程组

$$\frac{n}{C(\boldsymbol{\theta})}\frac{\partial C(\boldsymbol{\theta})}{\partial \theta_i}=-\sum_{j=1}^{n}T_i(X_j),\ i=1,2,\ldots,k$$

在Θ_0内有解，则解必唯一且为$\boldsymbol{\theta}$的*MLE*。

证明略。定理中的公式由$\frac{\partial \ell(\boldsymbol{\theta},\boldsymbol{x})}{\partial \theta_i}=0$可得。

因此，若样本分布为指数族，只要似然方程或似然方程组的解属于自然参数空间的内点集，则其解必为θ的MLE。

因此，若样本分布为指数族，只要似然方程或似然方程组的解属于自然参数空间的内点集，则其解必为θ的MLE。

常见的分布族，如二项分布族、Poisson分布族、正态分布族、Gamma分布族等都是指数族，定理3.3.1的条件皆成立。因此似然方程（组）的解就是有关参数的MLE。

2. 从定义出发

2. 从定义出发

当似然函数$L(\theta, x)$对参数θ不可微，甚至不连续时，似然方程一般没有意义，必须直接从定义3.3.2出发去求参数θ的最大似然估计。

例 (3.3.2)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$是从两点分布族$\{B(1,p):0<p<1\}$中抽取的简单样本，求$p$和$g(p)=p(1-p)$的MLE。

例 (3.3.2)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从两点分布族$\{B(1, p) : 0 < p < 1\}$中抽取的简单样本，求$p$和$g(p) = p(1 - p)$的MLE。

解 1（教材） 样本的似然函数为

例 (3.3.2)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从两点分布族$\{B(1,p) : 0 < p < 1\}$中抽取的简单样本，求$p$和$g(p) = p(1-p)$的MLE。

解 1（教材） 样本的似然函数为

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

取对数得对数似然函数为

取对数得对数似然函数为

$$\ell(p, \boldsymbol{x}) = \ln L(p, \boldsymbol{x}) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p).$$

对参数p求导并令该式为0可得得对数似然方程为

取对数得对数似然函数为

$$\ell(p, \boldsymbol{x}) = \ln L(p, \boldsymbol{x}) = \left(\sum_{i=1}^{n} x_i\right) \ln p + \left(n - \sum_{i=1}^{n} x_i\right) \ln(1-p).$$

对参数p求导并令该式为0可得得对数似然方程为

$$\frac{\partial \ell(p, \boldsymbol{x})}{\partial p} = \frac{1}{p} \sum_{i=1}^{n} x_i - \frac{1}{1-p} \left(n - \sum_{i=1}^{n} x_i\right) = 0.$$

求解上式并将x换成X，得到参数p的估计量为

3.3.2 最大似然估计的求法及例

求解上式并将x换成X，得到参数p的估计量为

$$\hat{p}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

求解上式并将x换成X，得到参数p的估计量为

$$\hat{p}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}.$$

由于两点分布族为指数族，当$\bar{X} \neq 0, 1$时，$\hat{p}^* = \bar{X}$属于自然参数空间$\hat{\Theta}^* = (0, 1)$的内点集，故$\hat{p}^*$为p的MLE。

它与例3.2.2中得到的p的矩估计量相同。

它与例3.2.2中得到的p的矩估计量相同。

按定义可知$g(p) = p(1-p)$的MLE为

它与例3.2.2中得到的p的矩估计量相同。

按定义可知$g(p) = p(1-p)$的MLE为

$$\hat{g}^*(\boldsymbol{X}) = \hat{p}^*(1-\hat{p}^*) = \bar{X}(1-\bar{X}).$$

注 3.3.3 当$0 < \bar{X} < 1$时，易知$\hat{p}^* = \bar{X}$为p的唯一的MLE。

注 3.3.3 当$0 < \bar{X} < 1$时，易知$\hat{p}^* = \bar{X}$为p的唯一的MLE。

而当$\sum_{i=1}^{n} X_i = 0$或n时，$\bar{X} = 0$或1，不在$\Theta = (0, 1)$内。严格意义上讲，此时p的MLE不存在。

3.3.2 最大似然估计的求法及例

注 3.3.3 当$0 < \bar{X} < 1$时，易知$\hat{p}^* = \bar{X}$为p的唯一的MLE。

而当$\sum_{i=1}^{n} X_i = 0$或n时，$\bar{X} = 0$或1，不在$\Theta = (0, 1)$内。严格意义上讲，此时p的MLE不存在。

为了克服这一缺陷，对MLE的定义加以补充：当$\hat{p}^* \notin \Theta$时，若存在一列$\{\hat{p}_n^*\}$使得$\lim\limits_{n\to\infty} \hat{p}_n^* = \hat{p}^*$，且$\hat{p}_n^* \in \Theta$，$n = 1, 2, \ldots$，则$\hat{p}^*$也称为$p$的MLE。

注 3.3.3 当$0 < \bar{X} < 1$时，易知$\hat{p}^* = \bar{X}$为p的唯一的MLE。

而当$\sum_{i=1}^{n} X_i = 0$或n时，$\bar{X} = 0$或1，不在$\Theta = (0, 1)$内。严格意义上讲，此时p的MLE不存在。

为了克服这一缺陷，对MLE的定义加以补充：当$\hat{p}^* \notin \Theta$时，若存在一列$\{\hat{p}_n^*\}$使得$\lim_{n\to\infty} \hat{p}_n^* = \hat{p}^*$，且$\hat{p}_n^* \in \Theta$，$n = 1, 2, \ldots$，则$\hat{p}^*$也称为$p$的MLE。

在给出上述补充定义后，本题中当$\bar{X} = 0, 1$时，可认为$\hat{p}^* = \bar{X}$是p的MLE。 □

解 2 已知两点分布族为指数族，且$X_i \sim B(1,p)$, $0 < p < 1$，其概率分布为$P(x_i,p) = p^{x_i}(1-p)^{1-x_i}$，$x_i = 0,1$。则参数$p$的似然函数为

解 2 已知两点分布族为指数族，且$X_i \sim B(1,p)$, $0<p<1$，其概率分布为$P(x_i,p)=p^{x_i}(1-p)^{1-x_i}$，$x_i=0,1$。则参数$p$的似然函数为

$$
\begin{aligned}
L(p,\boldsymbol{x})=\prod_{i=1}^{n}p^{x_i}(1-p)^{1-x_i}&=p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}\\
&=(1-p)^n(\frac{p}{1-p})^{\sum_{i=1}^{n}x_i}\\
&=(1-p)^n\exp\left\{\sum_{i=1}^{n}x_i\ln\frac{p}{1-p}\right\}.
\end{aligned}
$$

记$\theta = \ln \frac{p}{1-p}$，则$e^{\theta} = \frac{p}{1-p}$，$p = \frac{1}{1+e^{-\theta}}$，$1 - p = \frac{1}{1+e^{\theta}}$，可得到似然函数的自然形式为

记$\theta = \ln \frac{p}{1-p}$，则$e^{\theta} = \frac{p}{1-p}$，$p = \frac{1}{1+e^{-\theta}}$，$1 - p = \frac{1}{1+e^{\theta}}$，可得到似然函数的自然形式为

$$L(\theta, \boldsymbol{x}) = (1 + e^{\theta})^{-n} \exp\left\{\theta \sum_{i=1}^{n} x_i\right\},$$

它的自然参数空间为$\Theta^* = \{\theta : -\infty < \theta < \infty\}$。

记$\theta = \ln \frac{p}{1-p}$，则$e^{\theta} = \frac{p}{1-p}$，$p = \frac{1}{1+e^{-\theta}}$，$1 - p = \frac{1}{1+e^{\theta}}$，可得到似然函数的自然形式为

$$L(\theta, \boldsymbol{x}) = (1 + e^{\theta})^{-n} \exp\left\{\theta \sum_{i=1}^{n} x_i\right\},$$

它的自然参数空间为$\Theta^* = \{\theta : -\infty < \theta < \infty\}$。

对上式取对数可得可对数似然函数为

记$\theta=\ln\frac{p}{1-p}$，则$e^{\theta}=\frac{p}{1-p}$，$p=\frac{1}{1+e^{-\theta}}$，$1-p=\frac{1}{1+e^{\theta}}$，可得到似然函数的自然形式为

$$L(\theta,\boldsymbol{x})=(1+e^{\theta})^{-n}\exp\left\{\theta\sum_{i=1}^{n}x_i\right\},$$

它的自然参数空间为$\Theta^*=\{\theta:-\infty<\theta<\infty\}$。

对上式取对数可得可对数似然函数为

$$\ell(\theta,\boldsymbol{x})=\ln L(\theta,\boldsymbol{x})=-n\ln(1+e^{\theta})+\theta\sum_{i=1}^{n}x_i,$$

对θ求导得到对数似然方程并求解可得

对θ求导得到对数似然方程并求解可得

$$\frac{\partial \ell(\theta, \boldsymbol{x})}{\partial \theta} = -n\frac{e^\theta}{1+e^\theta} + \sum_{i=1}^n x_i = 0$$
$$\Rightarrow \frac{1}{1+e^{-\theta}} = \bar{x} \Rightarrow \hat{\theta}^* = \ln\frac{\bar{X}}{1-\bar{X}}.$$

对θ求导得到对数似然方程并求解可得

$$\frac{\partial \ell(\theta, \boldsymbol{x})}{\partial \theta} = -n\frac{e^\theta}{1+e^\theta} + \sum_{i=1}^n x_i = 0$$
$$\Rightarrow \frac{1}{1+e^{-\theta}} = \bar{x} \Rightarrow \hat{\theta}^* = \ln \frac{\bar{X}}{1-\bar{X}}.$$

当$\bar{X} \neq 0, 1$时（$\bar{X} \in (0,1)$），$\hat{\theta}^* = \ln \frac{\bar{X}}{1-\bar{X}}$存在且属于自然参数空间$\Theta^* = \{\theta : -\infty < \theta < \infty\}$的内点集，所以$\hat{\theta}^* = \ln \frac{\bar{X}}{1-\bar{X}}$为$\theta = \ln \frac{p}{1-p}$的MLE。

又因为$p=\frac{1}{1+e^{-\theta}}$是$\theta$的函数，所以$p$的MLE为

$$\hat{p}^* = \bar{X}.$$

又因为$p = \frac{1}{1+e^{-\theta}}$是$\theta$的函数，所以$p$的MLE为

$$\hat{p}^* = \bar{X}.$$

同样的，因为$g(p) = (1-p)p$是p的函数，所以$g(p)$的MLE为

$$\hat{g}^*(X) = \bar{X}(1-\bar{X}).$$

$\square$

例 (3.3.3)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从正态分布族
$\{N(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$中抽取的简单样本。
记$\boldsymbol{\theta} = (\mu, \sigma^2)$，求$\mu, \sigma^2$和$g(\boldsymbol{\theta}) = \mu/\sigma^2$的MLE。

例 (3.3.3)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从正态分布族

$\{N(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$中抽取的简单样本。

记$\boldsymbol{\theta} = (\mu, \sigma^2)$，求$\mu, \sigma^2$和$g(\boldsymbol{\theta}) = \mu/\sigma^2$的MLE。

解 1（教材）

样本$\boldsymbol{X} = (X_1, \ldots, X_n)$的分布为

例 (3.3.3)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从正态分布族 $\{N(\mu, \sigma^2) : -\infty < \mu < \infty,\ \sigma^2 > 0\}$中抽取的简单样本。记$\boldsymbol{\theta} = (\mu, \sigma^2)$，求$\mu, \sigma^2$和$g(\boldsymbol{\theta}) = \mu/\sigma^2$的MLE。

解 1（教材）

样本$\boldsymbol{X} = (X_1, \ldots, X_n)$的分布为

$$f(\boldsymbol{x}, \boldsymbol{\theta}) = (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right\}.$$

对数似然函数为

对数似然函数为

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \ln f(\boldsymbol{x}, \boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

对数似然函数为

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \ln f(\boldsymbol{x}, \boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

由对数似然方程组

对数似然函数为

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \ln f(\boldsymbol{x}, \boldsymbol{\theta}) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

由对数似然方程组

$$\begin{aligned} \frac{\partial \ell(\boldsymbol{\theta}, \boldsymbol{x})}{\partial \mu} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0, \\ \frac{\partial \ell(\boldsymbol{\theta}, \boldsymbol{x})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0, \end{aligned}$$

解得

$$\hat{\mu}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

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

解得

$$\hat{\mu}^* = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

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

由于正态分布族为指数族，可知：若令$\theta_1 = \frac{\mu}{\sigma^2}$，$\theta_2 = -\frac{1}{2\sigma^2}$，则定理3.3.1中对数似然方程组的解$\hat{\theta}_1^* = \frac{\hat{\mu}^*}{\hat{\sigma}_*^2}$，$\hat{\theta}_2^* = -\frac{1}{2\hat{\sigma}_*^2}$属于自然参数空间$\Theta^* = \{(\theta_1, \theta_2):\ \theta_1 \in \boldsymbol{R},\ \theta_2 < 0\}$的内点集，故它们分别是$\theta_1$和$\theta_2$的MLE。

因此$\hat{\mu}^*$和$\hat{\sigma}_*^2$分别是μ和σ^2的MLE，它们也分别是μ和σ^2的矩估计量。前者是μ的无偏估计，后者不是σ^2的无偏估计。

因此$\hat{\mu}^*$和$\hat{\sigma}_*^2$分别是μ和σ^2的MLE，它们也分别是μ和σ^2的矩估计量。前者是μ的无偏估计，后者不是σ^2的无偏估计。

可见最大似然估计不一定具有无偏性。

因此$\hat{\mu}^*$和$\hat{\sigma}_*^2$分别是μ和σ^2的MLE，它们也分别是μ和σ^2的矩估计量。前者是μ的无偏估计，后者不是σ^2的无偏估计。

可见最大似然估计不一定具有无偏性。

又由定义可知$g(\boldsymbol{\theta}) = \frac{\mu}{\sigma^2}$的MLE为

$$\hat{g}^*(\boldsymbol{X}) = \frac{\hat{\mu}^*}{\hat{\sigma}_*^2} = \frac{\bar{X}}{S_n^2}.$$

$\square$

解 2

已知正态分布为指数族，且$X_i \sim N(\mu, \sigma^2)$，$-\infty < \mu < \infty$，$\sigma^2 > 0$，它的概率密度函数为

解 2

已知正态分布为指数族，且$X_i \sim N(\mu, \sigma^2)$, $-\infty < \mu < \infty$, $\sigma^2 > 0$，它的概率密度函数为

$$f(x_i, \boldsymbol{\theta}) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{(x_i-\mu)^2}{2\sigma^2}\right\}, \quad -\infty < x < \infty.$$

参数θ的似然函数为

3.3.2 最大似然估计的求法及例

参数$\boldsymbol{\theta}$的似然函数为

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta})$$

$$=$$

参数$\boldsymbol{\theta}$的似然函数为

$$
\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{x}) &= \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{ \frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2} \right\} \\
&=
\end{aligned}
$$

参数$\boldsymbol{\theta}$的似然函数为

$$
\begin{aligned}
L(\boldsymbol{\theta}, \boldsymbol{x}) &= \prod_{i=1}^{n} f(x_i, \boldsymbol{\theta}) \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right\} \\
&= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n} x_i^2 + \frac{\mu}{\sigma^2}\sum_{i=1}^{n} x_i\right\}.
\end{aligned}
$$

令$\theta_1 = -\frac{1}{2\sigma^2}$, $\theta_2 = \frac{\mu}{\sigma^2}$，则$\mu = -\frac{\theta_2}{2\theta_1}$，$\sigma^2 = -\frac{1}{2\theta_1}$，可得到似然函数的指数族的自然形式为

令$\theta_1=-\frac{1}{2\sigma^2}$, $\theta_2=\frac{\mu}{\sigma^2}$，则$\mu=-\frac{\theta_2}{2\theta_1}$，$\sigma^2=-\frac{1}{2\theta_1}$，可得到似然函数的指数族的自然形式为

$$L(\theta_1,\theta_2,\boldsymbol{x})=\left(-\frac{\pi}{\theta_1}\right)^{-\frac{n}{2}}\exp\left\{\frac{n\theta_2^2}{4\theta_1}\right\}\exp\left\{\theta_1\sum_{i=1}^n x_i^2+\theta_2\sum_{i=1}^n x_i\right\},$$

令$\theta_1=-\frac{1}{2\sigma^2}$, $\theta_2=\frac{\mu}{\sigma^2}$，则$\mu=-\frac{\theta_2}{2\theta_1}$，$\sigma^2=-\frac{1}{2\theta_1}$，可得到似然函数的指数族的自然形式为

$$L(\theta_1,\theta_2,\boldsymbol{x})=\left(-\frac{\pi}{\theta_1}\right)^{-\frac{n}{2}}\exp\left\{\frac{n\theta_2^2}{4\theta_1}\right\}\exp\left\{\theta_1\sum_{i=1}^n x_i^2+\theta_2\sum_{i=1}^n x_i\right\},$$

它的自然参数空间为$\Theta^*=\{(\theta_1,\theta_2):-\infty<\theta_1<0,\ -\infty<\theta_2<\infty\}$。

对似然函数$L(\theta_1, \theta_2, \boldsymbol{x})$取对数可得到对数似然函数

对似然函数$L(\theta_1, \theta_2, \boldsymbol{x})$取对数可得到对数似然函数

$$\ell(\theta_1, \theta_2, \boldsymbol{x}) = -\frac{n}{2} \ln\left(-\frac{\pi}{\theta_1}\right) + \frac{n\theta_2^2}{4\theta_1} + \theta_1 \sum_{i=1}^{n} x_i^2 + \theta_2 \sum_{i=1}^{n} x_i.$$

3.3.2 最大似然估计的求法及例

对似然函数$L(\theta_1,\theta_2,\boldsymbol{x})$取对数可得到对数似然函数

$$\ell(\theta_1,\theta_2,\boldsymbol{x}) = -\frac{n}{2}\ln\left(-\frac{\pi}{\theta_1}\right) + \frac{n\theta_2^2}{4\theta_1} + \theta_1\sum_{i=1}^{n} x_i^2 + \theta_2\sum_{i=1}^{n} x_i.$$

上式对θ_1和θ_2分别求导可得到对数似然方程组

对似然函数$L(\theta_1, \theta_2, \boldsymbol{x})$取对数可得到对数似然函数

$$\ell(\theta_1, \theta_2, \boldsymbol{x}) = -\frac{n}{2}\ln\left(-\frac{\pi}{\theta_1}\right) + \frac{n\theta_2^2}{4\theta_1} + \theta_1\sum_{i=1}^{n} x_i^2 + \theta_2\sum_{i=1}^{n} x_i.$$

上式对θ_1和θ_2分别求导可得到对数似然方程组

$$\begin{cases} \frac{\partial \ell(\theta_1,\theta_2,\boldsymbol{x})}{\partial \theta_1} = -\frac{n}{2}\frac{1}{-\frac{\pi}{\theta_1}}\frac{\pi}{\theta_1^2} + (-\frac{n\theta_2^2}{4\theta_1^2}) + \sum_{i=1}^{n} x_i^2 = 0, \\ \frac{\partial \ell(\theta_1,\theta_2,\boldsymbol{x})}{\partial \theta_2} = \frac{n\theta_2}{2\theta_1} + \sum_{i=1}^{n} x_i = 0. \end{cases}$$

令$S_n^2=\sum_{i=1}^n(X_i-\bar{X})^2/n$，并用$X_i$代替$x_i$，求解上述对数似然方程组，可得到$\theta_1$和$\theta_2$的估计量分别为

令$S_n^2=\sum_{i=1}^n(X_i-\bar{X})^2/n$，并用$X_i$代替$x_i$，求解上述对数似然方程组，可得到$\theta_1$和$\theta_2$的估计量分别为

$$\begin{aligned}\hat{\theta}_2^* &= \frac{n\bar{X}}{\sum_{i=1}^n(X_i-\bar{X})^2} = \frac{\bar{X}}{S_n^2},\\ \hat{\theta}_1^* &= -\frac{\hat{\theta}_2^{\,*}}{2\bar{X}} = -\frac{1}{2S_n^2}.\end{aligned}$$

由于$\hat{\theta}_1^*$和$\hat{\theta}_2^*$属于自然参数空间Θ^*的内点集，所以它们分别为θ_1和θ_2的MLE。

由于$\hat{\theta}_1^*$和$\hat{\theta}_2^*$属于自然参数空间Θ^*的内点集，所以它们分别为θ_1和θ_2的MLE。

因为$\mu = -\frac{\theta_2}{2\theta_1}$和$\sigma^2 = -\frac{1}{2\theta_1}$为$\theta_1$和$\theta_2$的函数，所以$\hat{\mu} = \bar{X}$和$\hat{\sigma}_2^* = S_n^2$分别为$\mu$和$\sigma^2$的MLE。

由于$\hat{\theta}_1^*$和$\hat{\theta}_2^*$属于自然参数空间Θ^*的内点集，所以它们分别为θ_1和θ_2的MLE。

因为$\mu = -\frac{\theta_2}{2\theta_1}$和$\sigma^2 = -\frac{1}{2\theta_1}$为$\theta_1$和$\theta_2$的函数，所以$\hat{\mu} = \bar{X}$和$\hat{\sigma}_2^* = S_n^2$分别为$\mu$和$\sigma^2$的MLE。

又因为$g(\theta) = \frac{\mu}{\sigma^2} = \theta_2$，所以$\hat{g}^*(X) = \hat{\theta}_2^* = \frac{\bar{X}}{S_n^2}$为$g(\theta)$的MLE。 □

例 (3.3.4)

设元件寿命X服从下列指数分布$Exp(\lambda)$，其密度函数为

$$f(x,\lambda)=\begin{cases}\lambda e^{-\lambda x}, & x>0,\\ 0, & x\le 0.\end{cases}$$

例 (3.3.4（续）)

设$X_1, \ldots, X_n$分别表示接受试验的n个元件寿命。

例 (3.3.4（续）)

设$X_1, \ldots, X_n$分别表示接受试验的n个元件寿命。

由于受时间的限制，试验实际上只进行到有r（$r \le n$）个元件失效时就停止了。

例 (3.3.4（续）)

设$X_1, \ldots, X_n$分别表示接受试验的n个元件寿命。

由于受时间的限制，试验实际上只进行到有r（$r \leq n$）个元件失效时就停止了。

以$X_{(1)} \leq X_{(2)} \leq \cdots \leq X_{(r)}$记这$r$个元件的寿命，即只观察到了样本$X_1, \ldots, X_n$的前$r$个次序统计量$X_{(1)}, \ldots, X_{(r)}$。

例 (3.3.4（续）)

设$X_1,\ldots,X_n$分别表示接受试验的n个元件寿命。

由于受时间的限制，试验实际上只进行到有r（$r\le n$）个元件失效时就停止了。

以$X_{(1)}\le X_{(2)}\le\cdots\le X_{(r)}$记这$r$个元件的寿命，即只观察到了样本$X_1,\ldots,X_n$的前$r$个次序统计量$X_{(1)},\ldots,X_{(r)}$。

基于这前 r 个次序统计量，求 λ 和 $g(\lambda)=1/\lambda$ 的MLE。

引理 (2.4.1（回顾）)

设总体X的分布函数为$F(x)$，密度函数为$f(x)$，$X_1,\ldots,X_n$为从总体X中抽取的简单样本，则有

$$\underset{a<x_1<\cdots<x_n<b}{\int\cdots\int} f(x_1)\cdots f(x_n)dx_1\cdots dx_n = \frac{1}{n!}[F(b)-F(a)]^n,$$

此处a可取有限数或$-\infty$，b可取有限数或∞。

定理 (2.4.1（回顾）)

设总体X有密度函数$f(x)$，$-\infty < x < \infty$，$X_1, X_2, \ldots, X_n$为从总体X中抽取的简单样本，如前所述$(X_{(1)}, X_{(2)}, \ldots, X_{(n)})$为样本$(X_1, \ldots, X_n)$的次序统计量。令$Y_i = X_{(i)}$，$i = 1, \ldots, n$，则次序统计量$(Y_1, \ldots, Y_n)$的联合密度函数为

$$g(y_1, \ldots, y_n) = \begin{cases} n! f(y_1) f(y_2) \cdots f(y_n), & y_1 < y_2 < \cdots < y_n, \\ 0, & \text{其他}. \end{cases}$$

解 1（教材）

记$t_i = x_{(i)}$，$i = 1, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。

解 1（教材）

记$t_i = x_{(i)}$，$i = 1, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。

下面利用引理2.4.1和2.4.1小节所介绍的方法求$(X_{(1)}, X_{(2)}, \cdots, X_{(r)})$的联合密度如下：

解 1（教材）

记$t_i = x_{(i)}$，$i = 1, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。

下面利用引理2.4.1和2.4.1小节所介绍的方法求$(X_{(1)}, X_{(2)}, \cdots, X_{(r)})$的联合密度如下：

$$
\begin{aligned}
p(t_1, \ldots, t_r; \lambda) &= \int_{t_r<t_{r+1}<\cdots<t_n<\infty} \cdots \int n! f(t_1, \lambda) \cdots f(t_n, \lambda) dt_{r+1} \cdots dt_n \\
&= \frac{n!}{(n-r)!} f(t_1, \lambda) \cdots f(t_r, \lambda) [1 - F(t_r)]^{n-r} \\
&= \frac{n!}{(n-r)!} \lambda^r \exp\left\{ -\lambda \left(\sum_{i=1}^{r} t_i + (n-r) t_r \right) \right\}.
\end{aligned}
$$

记$T=\sum_{i=1}^{r} t_i+(n-r)t_r$，故对数似然函数为

记$T=\sum\limits_{i=1}^{r} t_i+(n-r)t_r$，故对数似然函数为

$$\ell(\lambda, t_1, \ldots, t_r)=\ln c+r\ln\lambda-\lambda T,$$

其中$c=n!/(n-c)!$。

记$T=\sum_{i=1}^{r} t_i+(n-r)t_r$，故对数似然函数为

$$\ell(\lambda, t_1, \ldots, t_r) = \ln c + r\ln\lambda - \lambda T,$$

其中$c=n!/(n-c)!$。对λ求导，得似然方程为

$$\frac{\partial \ell}{\partial \lambda} = \frac{r}{\lambda} - T = 0,$$

解得

$$\hat{\lambda}^* = \frac{r}{T} = r\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right]^{-1}.$$

解得

$$\hat{\lambda}^* = \frac{r}{T} = r\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right]^{-1}.$$

似然函数$L(\lambda, x_{(1)}, \ldots, x_{(r)})$是指数族的形式，且$\hat{\lambda}^*$属于自然参数空间$\Theta^* = \{\lambda : \lambda > 0\} = (0, \infty)$的内点集，故$\hat{\lambda}^*$为$\lambda$的MLE。

解得

$$\hat{\lambda}^* = \frac{r}{T} = r\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right]^{-1}.$$

似然函数$L(\lambda, x_{(1)}, \dots, x_{(r)})$是指数族的形式，且$\hat{\lambda}^*$属于自然参数空间$\Theta^* = \{\lambda : \lambda > 0\} = (0, \infty)$的内点集，故$\hat{\lambda}^*$为$\lambda$的MLE。

由定义可知$g(\lambda) = 1/\lambda$的MLE为

解得

$$\hat{\lambda}^* = \frac{r}{T} = r\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right]^{-1}.$$

似然函数$L(\lambda, x_{(1)}, \ldots, x_{(r)})$是指数族的形式，且$\hat{\lambda}^*$属于自然参数空间$\Theta^* = \{\lambda : \lambda > 0\} = (0, \infty)$的内点集，故$\hat{\lambda}^*$为$\lambda$的MLE。

由定义可知$g(\lambda) = 1/\lambda$的MLE为

$$\hat{g}^*(X_{(1)}, \ldots, X_{(r)}) = \frac{T}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right].$$

□

解 2

记$t_i = x_{(i)}$，$i = 1, 2, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。下面利用引理2.4.1的结果和2.4.1小节所介绍的方法求$(X_{(1)}, \ldots, X_{(r)})$的联合密度函数如下：

解 2

记$t_i = x_{(i)}$，$i = 1, 2, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。下面利用引理2.4.1的结果和2.4.1小节所介绍的方法求$(X_{(1)}, \ldots, X_{(r)})$的联合密度函数如下：

$$
\begin{aligned}
p(t_1, \ldots, t_r; \lambda) &= \int \cdots \int_{t_r < t_{r+1} < \ldots, t_n < \infty} n! f(t_1, \lambda) \cdots f(t_n, \lambda) dt_{r+1} \cdots dt_n \\
&=
\end{aligned}
$$

解 2

记$t_i = x_{(i)}$，$i = 1, 2, \ldots, n$，则有$t_1 \le t_2 \le \cdots \le t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。下面利用引理2.4.1的结果和2.4.1小节所介绍的方法求$(X_{(1)}, \ldots, X_{(r)})$的联合密度函数如下：

$$
\begin{aligned}
p(t_1, \ldots, t_r; \lambda) &= \int \cdots \int_{t_r < t_{r+1} < \ldots, t_n < \infty} n! f(t_1, \lambda) \cdots f(t_n, \lambda) dt_{r+1} \cdots dt_n \\
&= \frac{n!}{(n-r)!} f(t_1, \lambda) \cdots f(t_r, \lambda) [1 - F(t_r)]^{n-r} \\
&=
\end{aligned}
$$

解 2

记$t_i = x_{(i)}$，$i = 1, 2, \ldots, n$，则有$t_1 \leq t_2 \leq \cdots \leq t_n$。设$F(x)$为$X$的分布函数，易知$F(x) = 1 - e^{-\lambda x}$。下面利用引理2.4.1的结果和2.4.1小节所介绍的方法求$(X_{(1)}, \ldots, X_{(r)})$的联合密度函数如下：

$$
\begin{aligned}
p(t_1, \ldots, t_r; \lambda) &= \underset{t_r < t_{r+1} < \ldots, t_n < \infty}{\int \cdots \int} n! f(t_1, \lambda) \cdots f(t_n, \lambda) dt_{r+1} \cdots dt_n \\
&= \frac{n!}{(n-r)!} f(t_1, \lambda) \cdots f(t_r, \lambda) [1 - F(t_r)]^{n-r} \\
&= \frac{n!}{(n-r)!} \lambda^r \exp\left\{ -\lambda (\sum_{i=1}^{r} t_i + (n-r) t_r) \right\}.
\end{aligned}
$$

记$T=\sum_{i=1}^{r}t_i+(n-r)t_r$，则$2T\lambda\sim\chi^2_{2r}$，$T\sim\Gamma(r,\lambda)$属于指数族。似然函数为

3.3.2 最大似然估计的求法及例

记$T=\sum_{i=1}^{r} t_i+(n-r)t_r$，则$2T\lambda\sim\chi^2_{2r}$，$T\sim\Gamma(r,\lambda)$属于指数族。似然函数为

$$L(\lambda,t)=\frac{\lambda^r}{\Gamma(r)}t^{r-1}\exp\{-\lambda t\},$$

记$T=\sum_{i=1}^{r}t_i+(n-r)t_r$，则$2T\lambda\sim\chi^2_{2r}$，$T\sim\Gamma(r,\lambda)$属于指数族。似然函数为

$$L(\lambda,t)=\frac{\lambda^r}{\Gamma(r)}t^{r-1}\exp\{-\lambda t\},$$

令$\theta=-\lambda$，则其指数族的自然形式为

记$T=\sum_{i=1}^{r}t_i+(n-r)t_r$，则$2T\lambda\sim\chi^2_{2r}$，$T\sim\Gamma(r,\lambda)$属于指数族。似然函数为

$$L(\lambda,t)=\frac{\lambda^r}{\Gamma(r)}t^{r-1}\exp\{-\lambda t\},$$

令$\theta=-\lambda$，则其指数族的自然形式为

$$L(\theta,t)=\frac{t^{r-1}}{\Gamma(r)}(-\theta)^r\exp\{\theta t\},$$

自然参数空间为$\Theta^*=\{\theta:-\infty<\theta<0\}.$

故对数似然函数为

故对数似然函数为

$$\ell(\theta, t) = \ln L(\theta, t) = \ln \frac{t^{r-1}}{\Gamma(r)} + r \ln(-\theta) + \theta t.$$

对数似然方程为

故对数似然函数为

$$\ell(\theta, t) = \ln L(\theta, t) = \ln \frac{t^{r-1}}{\Gamma(r)} + r \ln(-\theta) + \theta t.$$

对数似然方程为

$$\frac{\partial \ell(\theta, t)}{\partial \theta} = r \frac{-1}{-\theta} + t = 0.$$

将t换为T，再将$t_i = x_{(i)}$换为$X_{(i)}$，$i = 1, \ldots, r$，求解上式，可得到θ的估计量为

将t换为T，再将$t_i = x_{(i)}$换为$X_{(i)}$，$i = 1, \ldots, r$，求解上式，可得到θ的估计量为

$$
\begin{aligned}
\hat{\theta}^*(X_{(1)}, \ldots, X_{(n)}) &= \frac{-r}{T} = \frac{-r}{\sum_{i=1}^{r} t_i + (n-r)t_r} \\
&= \frac{-r}{\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}}.
\end{aligned}
$$

因为$\hat{\theta}^*$属于自然参数空间Θ^*的内点集，所以$\hat{\theta}^*$为θ的MLE。

又因为$\lambda = -\theta$为θ的函数，所以

又因为$\lambda = -\theta$为θ的函数，所以

$$\begin{aligned}\hat{\lambda}^*(X_{(1)}, \ldots, X_{(r)}) =& -\hat{\theta}^*(X_{(1)}, \ldots, X_{(r)}) \\ =& \frac{r}{T} = \frac{r}{\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}},\end{aligned}$$

为λ的MLE。

由题知$g(\lambda)=1/\lambda$为λ的函数，所以$g(\lambda)$的MLE为

由题知$g(\lambda) = 1/\lambda$为λ的函数，所以$g(\lambda)$的MLE为

$$\hat{g}^*(X_{(1)}, \ldots, X_{(r)}) = \frac{T}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right].$$

□

由题知$g(\lambda) = 1/\lambda$为λ的函数，所以$g(\lambda)$的MLE为

$$\hat{g}^*(X_{(1)}, \ldots, X_{(r)}) = \frac{T}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right].$$

□

本例中所述元件寿命试验进行到第r个元件失效时就终止，这种试验称为**定数截尾试验**。

由题知$g(\lambda) = 1/\lambda$为λ的函数，所以$g(\lambda)$的MLE为

$$\hat{g}^*(X_{(1)}, \ldots, X_{(r)}) = \frac{T}{r} = \frac{1}{r}\left[\sum_{i=1}^{r} X_{(i)} + (n-r)X_{(r)}\right].$$

□

本例中所述元件寿命试验进行到第r个元件失效时就终止，这种试验称为定数截尾试验。

另一种方式是先定下一个时间$T > 0$，当试验进行T时试验就终止，这种试验称为定时截尾试验。

例 (3.3.5)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(0, \theta] : \theta > 0\}$中抽取的简单样本。

（1）求θ的MLE $\hat{\theta}^*$。

例 (3.3.5)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(0,\theta] : \theta > 0\}$中抽取的简单样本。

（1）求θ的MLE $\hat{\theta}^*$。

（2）说明$\hat{\theta}^*$是否为θ的无偏估计。若不然，作适当修正获得θ的无偏估计$\hat{\theta}_1^*$。

例 (3.3.5)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(0,\theta] : \theta > 0\}$中抽取的简单样本。

（1）求θ的MLE $\hat{\theta}^*$。

（2）说明$\hat{\theta}^*$是否为θ的无偏估计。若不然，作适当修正获得θ的无偏估计$\hat{\theta}_1^*$。

（3）试将$\hat{\theta}_1^*$与θ的矩估计$\hat{\theta}_1$进行比较，看哪一个有效？

例 (3.3.5)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(0, \theta] : \theta > 0\}$中抽取的简单样本。

（1）求θ的MLE $\hat{\theta}^*$。

（2）说明$\hat{\theta}^*$是否为θ的无偏估计。若不然，作适当修正获得θ的无偏估计$\hat{\theta}_1^*$。

（3）试将$\hat{\theta}_1^*$与θ的矩估计$\hat{\theta}_1$进行比较，看哪一个有效？

（4）证明θ的最大似然估计$\hat{\theta}^*$是θ的弱相合估计。

解 （1）样本$\boldsymbol{X} = (X_1, \ldots, X_n)$的联合密度为

解 （1）样本$\boldsymbol{X}=(X_1,\ldots,X_n)$的联合密度为

$$f(\boldsymbol{x},\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_{(1)}\leq x_{(n)}\leq\theta,\\ 0, & \text{其他}.\end{cases} \tag{2}$$

解 （1）样本$\boldsymbol{X}=(X_1,\ldots,X_n)$的联合密度为

$$f(\boldsymbol{x},\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_{(1)}\le x_{(n)}\le\theta,\\ 0, & \text{其他}.\end{cases} \tag{2}$$

因为均匀分布$U(0,\theta]$的支撑集$\{x:f(x,\theta)>0\}=\{x:0<x\le\theta\}$依赖于$\theta$，它不属于指数族。

解 （1）样本$\boldsymbol{X}=(X_1,\ldots,X_n)$的联合密度为

$$f(\boldsymbol{x},\theta)=\begin{cases}\frac{1}{\theta^n}, & 0<x_{(1)}\le x_{(n)}\le\theta,\\ 0, & \text{其他}.\end{cases} \tag{2}$$

因为均匀分布$U(0,\theta]$的支撑集$\{x:f(x,\theta)>0\}=\{x:0<x\le\theta\}$依赖于$\theta$，它不属于指数族。

且它的似然函数$L(\theta,\boldsymbol{x})=f(\boldsymbol{x},\theta)$在$\theta$处不可微，所以不能用对数似然函数求微商的方法去求$\theta$的MLE。只能从MLE的定义出发来讨论。

为使$L(\theta,\boldsymbol{x})$达到极大，由式（2）可见，应使分母上的θ尽可能地小，但θ又不能太小以致L为0。这个界限就在

为使$L(\theta, \boldsymbol{x})$达到极大，由式（2）可见，应使分母上的$\theta$尽可能地小，但$\theta$又不能太小以致$L$为0。这个界限就在

$$\hat{\theta}^* = \max\{X_1, \ldots, X_n\} = X_{(n)}.$$

为使$L(\theta,\boldsymbol{x})$达到极大，由式（2）可见，应使分母上的θ尽可能地小，但θ又不能太小以致L为0。这个界限就在

$$\hat{\theta}^* = \max\{X_1,\ldots,X_n\} = X_{(n)}.$$

故当$\theta > \hat{\theta}^*$时，$L = \theta^{-n} > 0$，但$L < L(\hat{\theta}^*,\boldsymbol{x}) = 1/(x_{(n)})^n$；当$\theta < \hat{\theta}^*$时，$L = 0$。

3.3.2 最大似然估计的求法及例

为使$L(\theta,\boldsymbol{x})$达到极大，由式（2）可见，应使分母上的θ尽可能地小，但θ又不能太小以致L为0。这个界限就在

$$\hat{\theta}^* = \max\{X_1,\ldots,X_n\} = X_{(n)}.$$

故当$\theta > \hat{\theta}^*$时，$L = \theta^{-n} > 0$，但$L < L(\hat{\theta}^*,\boldsymbol{x}) = 1/(x_{(n)})^n$；当$\theta < \hat{\theta}^*$时，$L = 0$。

因此$\hat{\theta}^* = X_{(n)}$是唯一使L达到最大的θ值，即θ的MLE，此时$L = L(\hat{\theta}^*,\boldsymbol{x}) = 1/(x_{(n)})^n$。

（2）为求$E[X_{(n)}]$，就要算出$T = X_{(n)}$的密度函数，易求T的密度函数

（2）为求$E[X_{(n)}]$，就要算出$T = X_{(n)}$的密度函数，易求T的密度函数

$$g(t,\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0 \le t \le \theta, \\ 0, & \text{其他}, \end{cases} \tag{3}$$

故有

（2）为求$E[X_{(n)}]$，就要算出$T = X_{(n)}$的密度函数，易求T的密度函数

$$g(t,\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0 \le t \le \theta, \\ 0, & \text{其他}, \end{cases} \tag{3}$$

故有

$$E(\hat{\theta}^*) = E(T) = \int_0^\theta \frac{nt^n}{\theta^n} dt = \frac{n}{n+1}\theta,$$

3.3.2 最大似然估计的求法及例

（2）为求$E[X_{(n)}]$，就要算出$T = X_{(n)}$的密度函数，易求T的密度函数

$$g(t,\theta) = \begin{cases} \frac{nt^{n-1}}{\theta^n}, & 0 \le t \le \theta, \\ 0, & \text{其他}, \end{cases} \tag{3}$$

故有

$$E(\hat{\theta}^*) = E(T) = \int_0^\theta \frac{nt^n}{\theta^n} dt = \frac{n}{n+1}\theta,$$

所以$\hat{\theta}^* = X_{(n)}$不是θ的无偏估计。

显见

显见

$$\hat{\theta}_1^* = \frac{n+1}{n} X_{(n)}$$

为θ的无偏估计。

（3）θ的矩估计$\hat{\theta}_1 = 2\bar{X}$是$\theta$的无偏估计。由于

（3）θ的矩估计$\hat{\theta}_1=2\bar{X}$是θ的无偏估计。由于

$$Var(\hat{\theta}_1^*)=\frac{\theta^2}{n(n+2)},\ \ Var(\hat{\theta}_1)=\frac{\theta^2}{3n},$$

（3）θ的矩估计$\hat{\theta}_1=2\bar{X}$是θ的无偏估计。由于

$$Var(\hat{\theta}_1^*)=\frac{\theta^2}{n(n+2)},\ \ Var(\hat{\theta}_1)=\frac{\theta^2}{3n},$$

所以在$n\geq 2$时，$\hat{\theta}_1^*$比$\hat{\theta}_1$有效。

（3）θ的矩估计$\hat{\theta}_1 = 2\bar{X}$是$\theta$的无偏估计。由于

$$Var(\hat{\theta}_1^*) = \frac{\theta^2}{n(n+2)}, \quad Var(\hat{\theta}_1) = \frac{\theta^2}{3n},$$

所以在$n \geq 2$时，$\hat{\theta}_1^*$比$\hat{\theta}_1$有效。

在$n = 1$时，有$Var(\hat{\theta}_1^*) = Var(\hat{\theta}_1)$，即这两个估计的效果是相同的。

（4）已知T的密度函数由式（3）给出，故对任给的$\epsilon > 0$，有

（4）已知T的密度函数由式（3）给出，故对任给的$\epsilon > 0$，有

$$P(|\hat{\theta}^{*} - \theta| \geq \epsilon)$$

$$=$$

（4）已知T的密度函数由式（3）给出，故对任给的$\epsilon > 0$，有

$$
\begin{aligned}
& P(|\hat{\theta}^* - \theta| \geq \epsilon) \\
= & 1 - P(|\hat{\theta}^* - \theta| < \epsilon) = 1 - P(\theta - \epsilon < T < \theta + \epsilon) \\
= &
\end{aligned}
$$

（4）已知T的密度函数由式（3）给出，故对任给的$\epsilon > 0$，有

$$
\begin{aligned}
& P(|\hat{\theta}^* - \theta| \geq \epsilon) \\
= & 1 - P(|\hat{\theta}^* - \theta| < \epsilon) = 1 - P(\theta - \epsilon < T < \theta + \epsilon) \\
= & 1 - \int_{\theta-\epsilon}^{\theta} \frac{nt^{n-1}}{\theta^n} dt = 1 - \frac{1}{\theta^n}[\theta^n - (\theta - \epsilon)^n] = \left(1 - \frac{\epsilon}{\theta}\right)^n .
\end{aligned}
$$

因此有

（4）已知T的密度函数由式（3）给出，故对任给的$\epsilon > 0$，有

$$
\begin{aligned}
& P(|\hat{\theta}^* - \theta| \geq \epsilon) \\
= & 1 - P(|\hat{\theta}^* - \theta| < \epsilon) = 1 - P(\theta - \epsilon < T < \theta + \epsilon) \\
= & 1 - \int_{\theta-\epsilon}^{\theta} \frac{nt^{n-1}}{\theta^n} dt = 1 - \frac{1}{\theta^n}[\theta^n - (\theta - \epsilon)^n] = \left(1 - \frac{\epsilon}{\theta}\right)^n.
\end{aligned}
$$

因此有

$$
\lim_{n\to\infty} P(|\hat{\theta}^* - \theta| \geq \epsilon) = \lim_{n\to\infty} \left(1 - \frac{\epsilon}{\theta}\right)^n = 0,
$$

故知$\hat{\theta}^* = X_{(n)}$为θ的弱相合估计。 □

注 3.3.4 在例 3.3.5 中将均匀分布族$\{U(0,\theta] : \theta > 0\}$改为均匀分布族$\{U(0,\theta) : \theta > 0\}$，则结论如何？

注 3.3.4 在例 3.3.5 中将均匀分布族$\{U(0,\theta]:\theta>0\}$改为均匀分布族$\{U(0,\theta):\theta>0\}$，则结论如何？

容易发现此时参数θ的MLE不存在（θ取不到$X_{(n)}$）。

注 3.3.4 在例 3.3.5 中将均匀分布族$\{U(0,\theta]:\theta>0\}$改为均匀分布族$\{U(0,\theta):\theta>0\}$，则结论如何？

容易发现此时参数θ的MLE不存在（θ取不到$X_{(n)}$）。

但是只要做如下的简单修改，问题就可以解决：

注 3.3.4 在例 3.3.5 中将均匀分布族$\{U(0,\theta]:\theta>0\}$改为均匀分布族$\{U(0,\theta):\theta>0\}$，则结论如何？

容易发现此时参数θ的MLE不存在（θ取不到$X_{(n)}$）。

但是只要做如下的简单修改，问题就可以解决：

由于连续型随机变量取一点的概率为0，因此可以修改均匀分布的支撑集为区间$(0,\theta]$，如例 3.3.5（1）所述θ的MLE就很容易求得了。

例 (3.3.6)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(\theta, \theta+1) : -\infty < \theta < \infty\}$中抽取的简单样本，求$\theta$的MLE。

例 (3.3.6)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$是从均匀分布族$\{U(\theta,\theta+1):-\infty<\theta<\infty\}$中抽取的简单样本，求$\theta$的MLE。

解

给定样本x时，θ的似然函数为

例 (3.3.6)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$是从均匀分布族$\{U(\theta, \theta+1) : -\infty < \theta < \infty\}$中抽取的简单样本，求$\theta$的MLE。

解

给定样本$\boldsymbol{x}$时，θ的似然函数为

$$L(\theta, \boldsymbol{x}) = \begin{cases} 1, & \theta < x_{(1)} \le x_{(n)} < \theta + 1, \\ 0, & \text{其他}. \end{cases} = \begin{cases} 1, & x_{(n)} - 1 < \theta < x_{(1)}, \\ 0, & \text{其他}. \end{cases}$$

这时，似然函数只取1和0两个值，只要$x_{(n)} - 1 < \theta < x_{(1)}$都可使$L$达到极大，故$\theta$的MLE不止一个。若取

这时，似然函数只取1和0两个值，只要$x_{(n)}-1<\theta<x_{(1)}$都可使L达到极大，故θ的MLE不止一个。若取

$$\hat{\theta}_1^*(\boldsymbol{X})=X_{(1)},\ \hat{\theta}_2^*(\boldsymbol{X})=X_{(n)}-1,$$

这时，似然函数只取1和0两个值，只要$x_{(n)}-1<\theta<x_{(1)}$都可使L达到极大，故θ的MLE不止一个。若取

$$\hat{\theta}_1^*(\boldsymbol{X})=X_{(1)},\ \hat{\theta}_2^*(\boldsymbol{X})=X_{(n)}-1,$$

则对任给的$0\leq\lambda\leq1$，

这时，似然函数只取1和0两个值，只要$x_{(n)} - 1 < \theta < x_{(1)}$都可使$L$达到极大，故$\theta$的MLE不止一个。若取

$$\hat{\theta}_1^*(\boldsymbol{X}) = X_{(1)},\ \hat{\theta}_2^*(\boldsymbol{X}) = X_{(n)} - 1,$$

则对任给的$0 \leq \lambda \leq 1$，

$$\begin{aligned}\hat{\theta}^*(\boldsymbol{X}) =& \lambda\hat{\theta}_1^*(\boldsymbol{X}) + (1-\lambda)\hat{\theta}_2^*(\boldsymbol{X}) \\ =& \lambda X_{(1)} + (1-\lambda)(X_{(n)} - 1)\end{aligned}$$

都是θ的MLE，故知θ的MLE有无穷多个。 $\square$

3.3.2 最大似然估计的求法及例

例 (3.3.7)

设k个事件$A_1, \cdots, A_k$构成完备事件群，事件A_i发生的概率为$0 < p_i < 1,\ i = 1, \ldots, k$ 且 $\sum_{i=1}^{k} p_i = 1$。

例 (3.3.7)

设k个事件$A_1, \cdots, A_k$构成完备事件群，事件A_i发生的概率为$0 < p_i < 1$, $i = 1, \ldots, k$ 且 $\sum_{i=1}^k p_i = 1$。

将试验独立重复n次，以X_i记A_i发生的次数，$i = 1, 2, \ldots, k$，$\sum_{i=1}^k X_i = n$，则$\boldsymbol{X} = (X_1, \ldots, X_k)$服从多项分布$M(n, p_1, \ldots, p_k)$，$\boldsymbol{X} = (X_1, \ldots, X_k)$的密度函数为

例 (3.3.7)

设k个事件$A_1, \cdots, A_k$构成完备事件群，事件A_i发生的概率为$0 < p_i < 1$, $i = 1, \ldots, k$ 且 $\sum_{i=1}^k p_i = 1$。

将试验独立重复n次，以X_i记A_i发生的次数，$i = 1, 2, \ldots, k$，$\sum_{i=1}^k X_i = n$，则$\boldsymbol{X} = (X_1, \ldots, X_k)$服从多项分布$M(n, p_1, \ldots, p_k)$，$\boldsymbol{X} = (X_1, \ldots, X_k)$的密度函数为

$$f(\boldsymbol{x}; \boldsymbol{p}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, \ \boldsymbol{p} = (p_1, \ldots, p_k).$$

例 (3.3.7)

设k个事件$A_1, \cdots, A_k$构成完备事件群，事件A_i发生的概率为$0 < p_i < 1$，$i = 1, \ldots, k$ 且 $\sum_{i=1}^k p_i = 1$。

将试验独立重复n次，以X_i记A_i发生的次数，$i = 1, 2, \ldots, k$，$\sum_{i=1}^k X_i = n$，则$\boldsymbol{X} = (X_1, \ldots, X_k)$服从多项分布$M(n, p_1, \ldots, p_k)$，$\boldsymbol{X} = (X_1, \ldots, X_k)$的密度函数为

$$f(\boldsymbol{x}; \boldsymbol{p}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}, \ \boldsymbol{p} = (p_1, \ldots, p_k).$$

求$p_1, \ldots, p_k$的MLE。

解 1（教材）:

记$\boldsymbol{p} = (p_1, \ldots, p_k)$。给定样本$\boldsymbol{x}$时，对数似然函数为

解 1（教材）：

记$\boldsymbol{p} = (p_1, \ldots, p_k)$。给定样本$\boldsymbol{x}$时，对数似然函数为

$$\ell(\boldsymbol{p}, \boldsymbol{x}) = \ln n! - \sum_{i=1}^{k} \ln x_i! + \sum_{i=1}^{k-1} x_i \ln p_i + x_k \ln\left(1 - \sum_{i=1}^{k-1} p_i\right).$$

解 1（教材）：

记$\boldsymbol{p} = (p_1, \ldots, p_k)$。给定样本$\boldsymbol{x}$时，对数似然函数为

$$\ell(\boldsymbol{p}, \boldsymbol{x}) = \ln n! - \sum_{i=1}^{k} \ln x_i! + \sum_{i=1}^{k-1} x_i \ln p_i + x_k \ln \left(1 - \sum_{i=1}^{k-1} p_i\right).$$

对p_i求偏导数，得似然方程组

解 1（教材）：

记$\boldsymbol{p}=(p_1,\ldots,p_k)$。给定样本$\boldsymbol{x}$时，对数似然函数为

$$\ell(\boldsymbol{p},\boldsymbol{x})=\ln n!-\sum_{i=1}^{k}\ln x_i!+\sum_{i=1}^{k-1}x_i\ln p_i+x_k\ln\left(1-\sum_{i=1}^{k-1}p_i\right).$$

对p_i求偏导数，得似然方程组

$$\frac{\partial\ell(\boldsymbol{p},\boldsymbol{x})}{\partial p_i}=\frac{x_i}{p_i}-\frac{x_k}{p_k}=0,\ i=1,2,\ldots,k-1.$$

若令

$$\frac{x_1}{p_1}=\frac{x_2}{p_2}=\cdots=\frac{x_k}{p_k}=\lambda,$$

3.3.2 最大似然估计的求法及例

若令

$$\frac{x_1}{p_1} = \frac{x_2}{p_2} = \cdots = \frac{x_k}{p_k} = \lambda,$$

则有

$$x_i = \lambda p_i,\ i = 1, 2, \ldots, k.$$

若令

$$\frac{x_1}{p_1} = \frac{x_2}{p_2} = \cdots = \frac{x_k}{p_k} = \lambda,$$

则有

$$x_i = \lambda p_i,\ i = 1, 2, \ldots, k.$$

将这k个等式两边分别相加得

$$n = \sum_{i=1}^{k} x_i = \lambda \sum_{i=1}^{k} p_i = \lambda.$$

由于多项分布族属于指数族，因此由定理3.3.1和式子$x_i = \lambda p_i$, $i = 1, 2, \ldots, k$可知，p_i的MLE如下：

$$\hat{p}_i^* = \frac{X_i}{n},\ i = 1, 2, \ldots, k.$$

□

解 2：

记$\boldsymbol{p} = (p_1, \ldots, p_k)$，$\boldsymbol{x} = (x_1, \ldots, x_k)$的概率分布为

解 2：

记$\boldsymbol{p} = (p_1, \ldots, p_k)$，$\boldsymbol{x} = (x_1, \ldots, x_k)$的概率分布为

$$P(X = x; \boldsymbol{p}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

解 2：

记$\boldsymbol{p} = (p_1, \ldots, p_k)$，$\boldsymbol{x} = (x_1, \ldots, x_k)$的概率分布为

$$P(X = x; \boldsymbol{p}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}.$$

由题知，$\sum_{i=1}^{k} p_i = 1$，$p_k = 1 - \sum_{i=1}^{k-1} p_i$，只需讨论$k-1$个参数的极大似然估计。

给定样本x时，其似然函数为

给定样本$\boldsymbol{x}$时，其似然函数为

$$
\begin{aligned}
L(\boldsymbol{p},\boldsymbol{x}) =& \frac{n!}{x_1!\cdots x_k!} p_1^{x_1}\cdots p_{k-1}^{x_{k-1}}\left(1-\sum_{i=1}^{k-1}\right)^{x_k} \\
=& \frac{n!}{x_1!\cdots x_k!}\exp\left\{\sum_{i=1}^{k-1} x_i \ln p_i + x_k \ln\left(1-\sum_{i=1}^{k-1} p_i\right)\right\}.
\end{aligned}
$$

记$\theta_i = \ln p_i$，$\boldsymbol{\theta} = (\theta_1,\ldots,\theta_{k-1})$，则

有$p_i = e^{\theta_i}$, $i = 1,\ldots,k-1$, $0 < p_i < 1$。

似然函数的自然形式为

似然函数的自然形式为

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{n!}{x_1! \cdots x_k!} \exp\left\{ \sum_{i=1}^{k-1} x_i\theta_i + x_k \ln\left(1 - \sum_{i=1}^{k-1} e^{\theta_i}\right) \right\}.$$

似然函数的自然形式为

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{n!}{x_1! \cdots x_k!} \exp\left\{ \sum_{i=1}^{k-1} x_i \theta_i + x_k \ln\left(1 - \sum_{i=1}^{k-1} e^{\theta_i}\right)\right\}.$$

自然参数空间为$\Theta^* = \{-\infty < \theta_i < 0,\ i = 1, \ldots, k-1\}$。对数似然函数为

似然函数的自然形式为

$$L(\boldsymbol{\theta}, \boldsymbol{x}) = \frac{n!}{x_1! \cdots x_k!} \exp\left\{\sum_{i=1}^{k-1} x_i\theta_i + x_k \ln\left(1 - \sum_{i=1}^{k-1} e^{\theta_i}\right)\right\}.$$

自然参数空间为$\Theta^* = \{-\infty < \theta_i < 0,\ i = 1, \ldots, k-1\}$。对数似然函数为

$$\ell(\boldsymbol{\theta}, \boldsymbol{x}) = \ln\frac{n!}{x_1! \cdots x_k!} + \sum_{i=1}^{k-1} x_i\theta_i + x_k \ln\left(1 - \sum_{i=1}^{k-1} e^{\theta_i}\right).$$

对θ求偏导可得到似然方程组为

对θ求偏导可得到似然方程组为

$$\left\{\begin{array}{l}\dfrac{\partial\ell}{\partial\theta_1}=x_1+x_k\dfrac{-e^{\theta_1}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}}=0,\\ \dfrac{\partial\ell}{\partial\theta_2}=x_2+x_k\dfrac{-e^{\theta_2}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}}=0,\\ \vdots\\ \dfrac{\partial\ell}{\partial\theta_{k-1}}=x_{k-1}+x_k\dfrac{-e^{\theta_{k-1}}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}}=0.\end{array}\right. \Longrightarrow \left\{\begin{array}{l}x_1=x_k\dfrac{e^{\theta_1}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}},\\ x_2=x_k\dfrac{e^{\theta_2}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}},\\ \vdots\\ x_{k-1}=x_k\dfrac{e^{\theta_{k-1}}}{1-\sum\limits_{i=1}^{k-1}e^{\theta_i}}.\end{array}\right. \tag{4}$$

经计算得

$$\sum_{i=1}^{k-1} x_i = x_k \frac{\sum\limits_{i=1}^{k-1} e^{\theta_i}}{1 - \sum\limits_{i=1}^{k-1} e^{\theta_i}},$$

经计算得

$$\sum_{i=1}^{k-1} x_i = x_k \frac{\sum\limits_{i=1}^{k-1} e^{\theta_i}}{1 - \sum\limits_{i=1}^{k-1} e^{\theta_i}},$$

因此有

$$n = \sum_{i=1}^{k-1} x_i + x_k = x_k \frac{1}{1 - \sum\limits_{i=1}^{k-1} e^{\theta_i}} \Longrightarrow \frac{1}{1 - \sum\limits_{i=1}^{k-1} e^{\theta_i}} = \frac{n}{x_k}. \tag{5}$$

将式（5）代入似然方程组（4）可得到

$$\begin{cases} e^{\theta_1} = \dfrac{x_1}{n}, \\ e^{\theta_2} = \dfrac{x_2}{n}, \\ \vdots \\ e^{\theta_{k-1}} = \dfrac{x_{k-1}}{n}. \end{cases}$$

将式（5）代入似然方程组（4）可得到

$$\begin{cases} e^{\theta_1} = \frac{x_1}{n}, \\ e^{\theta_2} = \frac{x_2}{n}, \\ \vdots \\ e^{\theta_{k-1}} = \frac{x_{k-1}}{n}. \end{cases}$$

即$e^{\theta_i} = \frac{x_{i-1}}{n} \in (0,1)$，$\hat{\theta}_i^* = \ln \frac{x_i}{n} \in (-\infty, 0)$。

将式（5）代入似然方程组（4）可得到

$$\begin{cases} e^{\theta_1} = \frac{x_1}{n}, \\ e^{\theta_2} = \frac{x_2}{n}, \\ \vdots \\ e^{\theta_{k-1}} = \frac{x_{k-1}}{n}. \end{cases}$$

即$e^{\theta_i} = \frac{x_{i-1}}{n} \in (0,1)$，$\hat{\theta}_i^* = \ln \frac{x_i}{n} \in (-\infty, 0)$。

因为$\hat{\theta}_i^*$属于自然参数空间$\Theta^* = \{-\infty < \theta_i < 0,\ i = 1, \ldots, k-1\}$，所以$\hat{\theta}_i^*$为$\theta_i$的MLE。

又因为$\theta_i = \ln p_i$，p_i为θ_i的函数，所以$\hat{p}_i^* = e^{\hat{\theta}_i^*} = e^{\ln \frac{x_i}{n}} = \frac{x_i}{n}$为$p_i$的MLE，$i = 1, \ldots, k-1$。

又因为$\theta_i = \ln p_i$，p_i为θ_i的函数，所以$\hat{p}_i^* = e^{\hat{\theta}_i^*} = e^{\ln \frac{x_i}{n}} = \frac{x_i}{n}$为$p_i$的MLE，$i = 1, \ldots, k-1$。

$\hat{p}_k^* = 1 - \sum_{i=1}^{k-1} \hat{p}_i^* = 1 - \sum_{i=1}^{k-1} \frac{x_i}{n} = \frac{x_k}{n}$为$p_k$的MLE。 $\square$

例 (3.3.8)

设从总体X分布为

X_i	0	1	2	3
p_i	$\theta/2$	θ	$3\theta/2$	1-3θ

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从总体$X$中抽取的简单样本。

例 (3.3.8)

设从总体X分布为

X_i	0	1	2	3
p_i	$\theta/2$	θ	$3\theta/2$	1-3θ

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从总体$X$中抽取的简单样本。

（1）求θ的矩估计量$\hat{\theta}_M$和最大似然估计量$\hat{\theta}_{MLE}$，

在$\boldsymbol{X} = (0, 3, 1, 1, 0, 2, 2, 2, 3, 2)$时求出相应估计量的值。

例 (3.3.8)

设从总体X分布为

X_i	0	1	2	3
p_i	$\theta/2$	θ	$3\theta/2$	1-3θ

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从总体$X$中抽取的简单样本。

（1）求θ的矩估计量$\hat{\theta}_M$和最大似然估计量$\hat{\theta}_{MLE}$，

在$\boldsymbol{X}=(0,3,1,1,0,2,2,2,3,2)$时求出相应估计量的值。

（2）上述估计量是否是无偏的？若不是，请作修正。

例 (3.3.8)

设从总体X分布为

X_i	0	1	2	3
p_i	$\theta/2$	θ	$3\theta/2$	1-3θ

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从总体$X$中抽取的简单样本。

（1）求θ的矩估计量$\hat{\theta}_M$和最大似然估计量$\hat{\theta}_{MLE}$，

在$\boldsymbol{X}=(0,3,1,1,0,2,2,2,3,2)$时求出相应估计量的值。

（2）上述估计量是否是无偏的？若不是，请作修正。

（3）比较修正后的两个估计量，指出哪个更有效。

解

表格第一行表示r.v.X取值只有0，1，2，3这四种可能，第二行表示X取相应值的概率p_i，$i=0,1,2,3$。

解

表格第一行表示r.v.X取值只有0，1，2，3这四种可能，第二行表示X取相应值的概率p_i，$i=0,1,2,3$。

与上例相似这是一个多项分布，令样本$X_1,\ldots,X_n$中取值为i $(i=0,1,2,3)$的频数为n_i，则n_i是随机变量，满足条件$\sum\limits_{i=0}^{3} n_i = n$。

这里

解

表格第一行表示r.v.X取值只有0，1，2，3这四种可能，第二行表示X取相应值的概率p_i，$i=0,1,2,3$。

与上例相似这是一个多项分布，令样本$X_1,\ldots,X_n$中取值为i $(i=0,1,2,3)$的频数为n_i，则n_i是随机变量，满足条件$\sum\limits_{i=0}^{3} n_i = n$。这里

$$\begin{aligned}(n_0,n_1,n_2,n_3) \sim & M(n,(p_0,p_1,p_2,p_3)) \\ = & M(n,(\theta/2,\theta,3\theta/2,1-3\theta))\end{aligned}$$

（1）由题知

$$E(X) = \sum_{i=0}^{3} X_i p_i = 0 \times \frac{\theta}{2} + 1 \times \theta + 2 \times \frac{3\theta}{2} + 3 \times (1 - 3\theta) = 3 - 5\theta.$$

（1）由题知

$$E(X)=\sum_{i=0}^{3}X_ip_i=0\times\frac{\theta}{2}+1\times\theta+2\times\frac{3\theta}{2}+3\times(1-3\theta)=3-5\theta.$$

将$E(X)$用$\bar{X}$代替，可得到θ的矩估计为$\hat{\theta}_M=\frac{3-\bar{X}}{5}$。

令θ的似然函数为

令θ的似然函数为

$$
\begin{aligned}
L(\theta, \boldsymbol{x}) =& \frac{n!}{n_0! \cdots n_3!} p_0^{n_0} \cdots p_3^{n_3} \\
=& \frac{n!}{n_0! \cdots n_3!} \left(\frac{\theta}{2}\right)^{n_0} \theta^{n_1} \left(\frac{3\theta}{2}\right)^{n_2} (1-3\theta)^{n_3} \\
=& \frac{n!}{n_0! \cdots n_3!} \frac{3^{n_2}}{2^{n_0+n_2}} \theta^{n_0+n_1+n_2} (1-3\theta)^{n_3} \\
=& c \times \theta^{n_0+n_1+n_2} (1-3\theta)^{n_3},
\end{aligned}
$$

此处$c = \frac{n!}{n_0! \cdots n_3!} \frac{3^{n_2}}{2^{n_0+n_2}}$。

令

$$
\begin{aligned}
\frac{\partial \ln L(\theta, \boldsymbol{x})}{\partial \theta} &= \frac{\partial}{\partial \theta}\{\ln c + (n_0 + n_1 + n_2)\theta + n_3 \ln(1 - 3\theta)\} \\
&= \frac{n_0 + n_1 + n_2}{\theta} + \frac{n_3}{1 - 3\theta}(-3) = 0
\end{aligned}
$$

令

$$\frac{\partial \ln L(\theta, \boldsymbol{x})}{\partial \theta} = \frac{\partial}{\partial \theta}\{\ln c + (n_0 + n_1 + n_2)\theta + n_3 \ln(1 - 3\theta)\}$$
$$= \frac{n_0 + n_1 + n_2}{\theta} + \frac{n_3}{1 - 3\theta}(-3) = 0$$

解得θ的MLE为$\hat{\theta}_{MLE} = \frac{n_0+n_1+n_2}{3n} = \frac{n-n_3}{3n}$。

由样本$(X_1, \dots, X_n) = (0, 3, 1, 1, 0, 2, 2, 2, 3, 2)$可得到下表

X_i	0	1	2	3
n_i	2	2	4	2

这里$\sum_{i=0}^{3} n_i = 10$。

经计算

$$\bar{X} = \frac{\sum_{j=1}^{10} X_j}{10} = \frac{1}{10}(0+3+1+1+0+2+2+2+3+2) = 1.6,$$

经计算

$$\bar{X} = \frac{\sum_{j=1}^{10} X_j}{10} = \frac{1}{10}(0+3+1+1+0+2+2+2+3+2) = 1.6,$$

可得到θ的矩估计和最大似然估计分别为

$$\hat{\theta}_M = \frac{3-\bar{X}}{5} = \frac{3-1.6}{5} = 0.28,$$
$$\hat{\theta}_{MLE} = \frac{10-n_3}{3n} = \frac{10-2}{30} = \frac{4}{15} \approx 0.27.$$

（2）由于$n_3 \sim B(n, 1-3\theta)$，经计算

（2）由于$n_3 \sim B(n, 1-3\theta)$，经计算

$$E(\hat{\theta}_M) = E\left(\frac{3-\bar{X}}{5}\right) = \frac{3-E(\bar{X})}{5} = \frac{3-(3-5\theta)}{5} = \theta,$$

（2）由于$n_3 \sim B(n, 1-3\theta)$，经计算

$$
\begin{aligned}
E(\hat{\theta}_M) =& E\left(\frac{3-\bar{X}}{5}\right) = \frac{3-E(\bar{X})}{5} = \frac{3-(3-5\theta)}{5} = \theta, \\
E(\hat{\theta}_{MLE}) =& E\left(\frac{\sum\limits_{i=0}^{2} n_i}{3n}\right) = E\left(\frac{n-n_3}{3n}\right) \\
=& \frac{n-n(1-3\theta)}{3n} = \frac{n\times 3\theta}{3n} = \theta,
\end{aligned}
$$

（2）由于$n_3 \sim B(n, 1-3\theta)$，经计算

$$E(\hat{\theta}_M) = E\left(\frac{3-\bar{X}}{5}\right) = \frac{3-E(\bar{X})}{5} = \frac{3-(3-5\theta)}{5} = \theta,$$

$$E(\hat{\theta}_{MLE}) = E\left(\frac{\sum\limits_{i=0}^{2} n_i}{3n}\right) = E\left(\frac{n-n_3}{3n}\right)$$

$$= \frac{n-n(1-3\theta)}{3n} = \frac{n\times 3\theta}{3n} = \theta,$$

因此这两个估计均为无偏估计。

（3）经计算

$$
\begin{aligned}
E(X^2) =& \theta + 4\frac{3\theta}{2} + 9(1-3\theta) = 9 - 20\theta, \\
Var(X) =& E(X^2) - (EX)^2 = (9-20\theta) - (3-5\theta)^2 = 5\theta(2-5\theta),
\end{aligned}
$$

（3）经计算

$$
\begin{aligned}
E(X^2) =& \theta + 4\frac{3\theta}{2} + 9(1-3\theta) = 9 - 20\theta, \\
Var(X) =& E(X^2) - (EX)^2 = (9-20\theta) - (3-5\theta)^2 = 5\theta(2-5\theta), \\
Var(\hat{\theta}_M) =& Var\left(\frac{3-\bar{X}}{5}\right) = \frac{Var(\bar{X})}{25} = \frac{Var(X)}{25n} \\
=& \frac{5\theta(2-5\theta)}{25n} = \frac{\theta(2-5\theta)}{5n}, \\
Var(\hat{\theta}_{MLE}) =& Var\left(\frac{n-n_3}{3n}\right) = \frac{Var(n_3)}{9n^2} \\
=& \frac{n3\theta(1-3\theta)}{9n^2} = \frac{\theta(1-3\theta)}{3n}.
\end{aligned}
$$

并且

$$Var(\hat{\theta}_M) - Var(\hat{\theta}_{MLE}) = \frac{\theta(2-5\theta)}{5n} - \frac{\theta(1-3\theta)}{3n} = \frac{\theta}{15n} > 0.$$

并且

$$Var(\hat{\theta}_M) - Var(\hat{\theta}_{MLE}) = \frac{\theta(2-5\theta)}{5n} - \frac{\theta(1-3\theta)}{3n} = \frac{\theta}{15n} > 0.$$

所以$\hat{\theta}_{MLE}$比$\hat{\theta}_M$更有效。 $\square$

1. 最大似然估计的无偏性

1. 最大似然估计的无偏性

在前面的例 3.3.3 中已指出最大似然估计不一定是无偏的。因此最大似然估计可以是无偏的，但也可以是有偏的，要视具体情况而定。

2. 最大似然估计与充分统计量

2. 最大似然估计与充分统计量

定理 (3.3.2)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为自总体$\{f(x,\theta):\theta\in\Theta\}$中抽取的简单样本。若$T=T(X_1,\ldots,X_n)$是$\theta$的充分统计量，且$\theta$的最大似然估计$\hat{\theta}^*$唯一存在，则它必为$T$的函数。

证 由因子分解定理可知样本$\boldsymbol{X}$的概率函数，即似然函数可表示为

证 由因子分解定理可知样本$\boldsymbol{X}$的概率函数，即似然函数可表示为

$$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x}),$$

此处$h(\boldsymbol{x})$与θ无关。若θ的MLE唯一存在，记为$\hat{\theta}^*$，则

证 由因子分解定理可知样本$\boldsymbol{X}$的概率函数，即似然函数可表示为

$$L(\theta, \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i, \theta) = g(T(\boldsymbol{x}), \theta) h(\boldsymbol{x}),$$

此处$h(\boldsymbol{x})$与θ无关。若θ的MLE唯一存在，记为$\hat{\theta}^*$，则

$$L(\hat{\theta}^*, \boldsymbol{x}) = \max_{\theta \in \Theta} L(\theta, \boldsymbol{x}) \Longleftrightarrow g(T(\boldsymbol{x}), \hat{\theta}^*) = \max_{\theta \in \Theta} g(T(\boldsymbol{x}), \theta),$$

因此使$\max\limits_{\theta} L(\theta, \boldsymbol{x})$达到上确界唯一之点$\hat{\theta}^*$，即为使$\max\limits_{\theta} g(T(\boldsymbol{x}), \theta)$达到上确界唯一之点，它必为$T(\boldsymbol{x})$的函数。 $\square$

因此使$\max\limits_{\theta} L(\theta, \boldsymbol{x})$达到上确界唯一之点$\hat{\theta}^*$，即为使$\max\limits_{\theta} g(T(\boldsymbol{x}), \theta)$达到上确界唯一之点，它必为$T(\boldsymbol{x})$的函数。 □

此性质说明θ的最大似然估计$\hat{\theta}^* = \hat{\theta}^*(X_1, \ldots, X_n)$可表示为充分统计量$T(\boldsymbol{X})$的函数，即

$$\hat{\theta}^* = \hat{\theta}^*(T(X_1, \ldots, X_n)).$$

因此使$\max\limits_{\theta} L(\theta, \boldsymbol{x})$达到上确界唯一之点$\hat{\theta}^*$，即为使$\max\limits_{\theta} g(T(\boldsymbol{x}), \theta)$达到上确界唯一之点，它必为$T(\boldsymbol{x})$的函数。 $\square$

此性质说明θ的最大似然估计$\hat{\theta}^* = \hat{\theta}^*(X_1, \ldots, X_n)$可表示为充分统计量$T(\boldsymbol{X})$的函数，即

$$\hat{\theta}^* = \hat{\theta}^*(T(X_1, \ldots, X_n)).$$

如例 3.3.2 ——例 3.3.7中的最大似然估计皆为充分统计量的函数。

注 3.3.5* 最大似然估计的上述性质告诉我们：假如充分统计量存在，θ的最大似然估计量$\hat{\theta}^*$为充分统计量的函数。

注 3.3.5* 最大似然估计的上述性质告诉我们：假如充分统计量存在，θ的最大似然估计量$\hat{\theta}^*$为充分统计量的函数。

若此最大似然估计本身也是充分的，由于此充分统计量$\hat{\theta}^*$是其他充分统计量的函数，由定义可知$\hat{\theta}^*$是极小充分的。