数理统计
第三章
点估计

2026 年 4 月 19 日

目录

设有一参数分布族$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$，其中$\Theta$为参数空间。

设有一参数分布族$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$，其中$\Theta$为参数空间。

设$g(\theta)$是定义在Θ上的实函数，$\boldsymbol{X}=(X_1,\ldots,X_n)$为自$\mathcal{F}$中抽取的简单样本，$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\ldots,X_n)$为$g(\theta)$的一个估计量。

设有一参数分布族$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$，其中$\Theta$为参数空间。

设$g(\theta)$是定义在Θ上的实函数，$\boldsymbol{X}=(X_1,\ldots,X_n)$为自$\mathcal{F}$中抽取的简单样本，$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\ldots,X_n)$为$g(\theta)$的一个估计量。

如何评价$\hat{g}(\boldsymbol{X})$的优劣？

一般用$\hat{g}(\boldsymbol{X}) - g(\theta)$作为其偏差，为消除$\hat{g}(\boldsymbol{X}) - g(\theta)$取值出现“正负”可能抵消的影响，一般用$[\hat{g}(\boldsymbol{X}) - g(\theta)]^2$来代替。

一般用$\hat{g}(\boldsymbol{X})-g(\theta)$作为其偏差，为消除$\hat{g}(\boldsymbol{X})-g(\theta)$取值出现“正负”可能抵消的影响，一般用$[\hat{g}(\boldsymbol{X})-g(\theta)]^2$来代替。

由于这个量是随机的，将其求平均，即计算其均值，以得到一个整体性的指标

$$E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2,$$

这就是估计量$\hat{g}(\boldsymbol{X})$的均方误差。

定义 (3.4.1)

设$\hat{g}(\boldsymbol{X})$为$g(\theta)$的估计量，则称

$$E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2$$

为$\hat{g}(\boldsymbol{X})$的**均方误差（mean squared error，MSE）**。

定义 (3.4.1)

设$\hat{g}(\boldsymbol{X})$为$g(\theta)$的估计量，则称

$$E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2$$

为$\hat{g}(\boldsymbol{X})$的**均方误差（mean squared error，MSE）**。

设$\hat{g}_1(\boldsymbol{X})$和$\hat{g}_2(\boldsymbol{X})$为$g(\theta)$的两个不同的估计量，若

$$E_\theta[\hat{g}_1(\boldsymbol{X})-g(\theta)]^2 \geq E[\hat{g}_2(\boldsymbol{X})-g(\theta)]^2,\text{ 一切}\theta \in \Theta,$$

且不等号至少对某个$\theta \in \Theta$成立，则称在MSE准则下$\hat{g}_2(\boldsymbol{X})$优于$\hat{g}_1(\boldsymbol{X})$。

定义 (3.4.1（续）)

若存在$\hat{g}^*(\boldsymbol{X})$，使得对$g(\theta)$的任一估计量$\hat{g}(\boldsymbol{X})$，都有

$$E_\theta[\hat{g}^*(\boldsymbol{X})-g(\theta)]^2 \leq E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2,\ \text{一切}\theta \in \Theta,$$

则称$\hat{g}^*(\boldsymbol{X})$为$g(\theta)$的**一致最小均方误差估计**。

可惜的是，一致最小均方误差估计常不存在。

可惜的是，一致最小均方误差估计常不存在。

解决这个问题的办法之一是把最优性准则放宽一些，使适合这种最优性准则的估计一般能存在。

可惜的是，一致最小均方误差估计常不存在。

解决这个问题的办法之一是把最优性准则放宽一些，使适合这种最优性准则的估计一般能存在。

从直观上想，在一个大的估计类中，一致最优估计量不存在，把估计类缩小，就有可能存在一致最优的估计量。因此把估计类缩小为无偏估计类来考虑。

3.4.1 引言及定义

在无偏估计类中，估计量的均方误差就变为其方差。

在无偏估计类中，估计量的均方误差就变为其方差。

即当$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计时，$E(\hat{g}(\boldsymbol{X}))=g(\theta)$，其均方误差为

在无偏估计类中，估计量的均方误差就变为其方差。

即当$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计时，$E(\hat{g}(\boldsymbol{X}))=g(\theta)$，其均方误差为

$$
\begin{aligned}
MSE[\hat{g}(\boldsymbol{X})] =& E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2 \\
=&
\end{aligned}
$$

在无偏估计类中，估计量的均方误差就变为其方差。

即当$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计时，$E(\hat{g}(\boldsymbol{X}))=g(\theta)$，其均方误差为

$$\begin{aligned}MSE[\hat{g}(\boldsymbol{X})] =& E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2\\ =& E_\theta[\hat{g}(\boldsymbol{X})-E(\hat{g}(\boldsymbol{X}))]^2\\ =& \end{aligned}$$

在无偏估计类中，估计量的均方误差就变为其方差。

即当$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计时，$E(\hat{g}(\boldsymbol{X}))=g(\theta)$，其均方误差为

$$\begin{aligned} MSE[\hat{g}(\boldsymbol{X})] =& E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2 \\ =& E_\theta[\hat{g}(\boldsymbol{X})-E(\hat{g}(\boldsymbol{X}))]^2 \\ =& Var_\theta[\hat{g}(\boldsymbol{X})], \end{aligned}$$

在无偏估计类中，估计量的均方误差就变为其方差。

即当$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计时，$E(\hat{g}(\boldsymbol{X}))=g(\theta)$，其均方误差为

$$\begin{aligned} MSE[\hat{g}(\boldsymbol{X})] =& E_\theta[\hat{g}(\boldsymbol{X})-g(\theta)]^2 \\ =& E_\theta[\hat{g}(\boldsymbol{X})-E(\hat{g}(\boldsymbol{X}))]^2 \\ =& Var_\theta[\hat{g}(\boldsymbol{X})], \end{aligned}$$

此处$Var_\theta[\hat{g}(\boldsymbol{X})]$表示$\hat{g}(\boldsymbol{X})$的方差。

3.4.1 引言及定义

存在这样的情形，参数$g(\theta)$的无偏估计不存在。请看下例。

存在这样的情形，参数$g(\theta)$的无偏估计不存在。请看下例。

例 (3.4.1)

设样本$X\sim$二项分布$B(n,p)$，n已知而p未知。令$g(p)=1/p$，则参数$g(p)$的无偏估计不存在。

存在这样的情形，参数$g(\theta)$的无偏估计不存在。请看下例。

例 (3.4.1)

设样本$X\sim$二项分布$B(n,p)$，n已知而p未知。令$g(p)=1/p$，则参数$g(p)$的无偏估计不存在。

证： 采用反证法。

存在这样的情形，参数$g(\theta)$的无偏估计不存在。请看下例。

例 (3.4.1)

设样本$X\sim$二项分布$B(n,p)$，n已知而p未知。令$g(p)=1/p$，则参数$g(p)$的无偏估计不存在。

证： 采用反证法。

若不然，存在$g(p)$有无偏估计$\hat{g}(X)$。由于X只取$0,1,\ldots,n$这些值，令$\hat{g}(X)$的取值用$\hat{g}(i)=a_i$表示，$i=0,1,\ldots,n$。

由$\hat{g}(X)$的无偏性，应有

由$\hat{g}(X)$的无偏性，应有

$$\begin{aligned} E_p(\hat{g}(X)) &= \sum_{i=0}^{n} a_i P(\hat{g}(X) = a_i) = \sum_{i=0}^{n} a_i P(X = i) \\ &= \sum_{i=0}^{n} a_i \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{p},\ 0 < p < 1. \end{aligned}$$

由$\hat{g}(X)$的无偏性，应有

$$\begin{aligned} E_p(\hat{g}(X)) &= \sum_{i=0}^{n} a_i P(\hat{g}(X) = a_i) = \sum_{i=0}^{n} a_i P(X = i) \\ &= \sum_{i=0}^{n} a_i \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{p},\ 0 < p < 1. \end{aligned}$$

于是有

由$\hat{g}(X)$的无偏性，应有

$$
\begin{aligned}
E_p(\hat{g}(X)) &= \sum_{i=0}^{n} a_i P(\hat{g}(X) = a_i) = \sum_{i=0}^{n} a_i P(X = i) \\
&= \sum_{i=0}^{n} a_i \binom{n}{i} p^i (1-p)^{n-i} = \frac{1}{p},\ 0 < p < 1.
\end{aligned}
$$

于是有

$$
\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1} (1-p)^{n-i} - 1 = 0,\ 0 < p < 1.
$$

但上式左端是p的$n+1$次多项式，即

但上式左端是p的$n+1$次多项式，即

$$\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i} - 1$$

$$=$$

但上式左端是p的$n+1$次多项式，即

$$\begin{aligned}&\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i}-1\\ =\ & a_0\binom{n}{0}p(1-p)^n + a_1\binom{n}{1}p^2(1-p)^{n-1}\\ &+\cdots+a_n\binom{n}{n}p^{n+1}(1-p)^0-1,\end{aligned}$$

但上式左端是p的$n+1$次多项式，即

$$\begin{aligned}
&\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i} - 1 \\
= a_0 &\binom{n}{0} p(1-p)^n + a_1 \binom{n}{1} p^2(1-p)^{n-1} \\
&+ \cdots + a_n \binom{n}{n} p^{n+1}(1-p)^0 - 1,
\end{aligned}$$

它最多在$(0,1)$区间有$n+1$个实根。

但上式左端是p的$n+1$次多项式，即

$$
\begin{aligned}
&\sum_{i=0}^{n} a_i \binom{n}{i} p^{i+1}(1-p)^{n-i} - 1 \\
= a_0 &\binom{n}{0} p(1-p)^n + a_1 \binom{n}{1} p^2(1-p)^{n-1} \\
&+ \cdots + a_n \binom{n}{n} p^{n+1}(1-p)^0 - 1,
\end{aligned}
$$

它最多在$(0,1)$区间有$n+1$个实根。可无偏性要求对$(0,1)$中的任一实数p上式都成立，这个矛盾说明$g(p)=1/p$的无偏估计不存在。 □

今后把不存在无偏估计的参数除外。

今后把不存在无偏估计的参数除外。

参数的无偏估计若存在，则称此参数为**可估参数**；

今后把不存在无偏估计的参数除外。

参数的无偏估计若存在，则称此参数为**可估参数**；

若参数函数的无偏估计存在，则称此函数为**可估函数**（estimable function）。因此可估函数的无偏估计类是非空的。

今后把不存在无偏估计的参数除外。

参数的无偏估计若存在，则称此参数为**可估参数**；

若参数函数的无偏估计存在，则称此函数为**可估函数**（estimable function）。因此可估函数的无偏估计类是非空的。

假如可估函数的无偏估计类中的无偏估计不止一个，怎样比较它们的优劣？

定义 (3.4.2（一致最小方差无偏估计）)

设$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$是一个参数分布族，其中$\Theta$为参数空间，$g(\theta)$为定义在$\Theta$上的可估函数。

定义 (3.4.2（一致最小方差无偏估计）)

设$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$是一个参数分布族，其中$\Theta$为参数空间，$g(\theta)$为定义在$\Theta$上的可估函数。

设$\hat{g}^*(\boldsymbol{X}) = \hat{g}^*(X_1, \ldots, X_n)$为$g(\theta)$的一个无偏估计，若对$g(\theta)$的任一无偏估计$\hat{g}(\boldsymbol{X}) = \hat{g}(X_1, \ldots, X_n)$，都有

定义 (3.4.2（一致最小方差无偏估计）)

设$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$是一个参数分布族，其中$\Theta$为参数空间，$g(\theta)$为定义在$\Theta$上的可估函数。

设$\hat{g}^*(\boldsymbol{X})=\hat{g}^*(X_1,\ldots,X_n)$为$g(\theta)$的一个无偏估计，若对$g(\theta)$的任一无偏估计$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\ldots,X_n)$，都有

$$Var_\theta[\hat{g}^*(\boldsymbol{X})]\le Var_\theta[\hat{g}(\boldsymbol{X})],\ \text{一切}\theta\in\Theta,$$

定义 (3.4.2（一致最小方差无偏估计）)

设$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$是一个参数分布族，其中$\Theta$为参数空间，$g(\theta)$为定义在$\Theta$上的可估函数。

设$\hat{g}^*(\boldsymbol{X})=\hat{g}^*(X_1,\ldots,X_n)$为$g(\theta)$的一个无偏估计，若对$g(\theta)$的任一无偏估计$\hat{g}(\boldsymbol{X})=\hat{g}(X_1,\ldots,X_n)$，都有

$$Var_\theta[\hat{g}^*(\boldsymbol{X})]\leq Var_\theta[\hat{g}(\boldsymbol{X})],\ \text{一切}\theta\in\Theta,$$

则称$\hat{g}^*(\boldsymbol{X})$是$g(\theta)$的**一致最小方差无偏估计**（uniformly minimum variance unbiased estimation, **UMVUE**）。

对给定参数分布族，如何寻找可估函数的UMVUE呢？

对给定参数分布族，如何寻找可估函数的UMVUE呢？

有三种方法：

3.4.1 引言及定义

对给定参数分布族，如何寻找可估函数的UMVUE呢？

有三种方法：

1. 零无偏估计法（不做要求）；

对给定参数分布族，如何寻找可估函数的UMVUE呢？

有三种方法：

1. 零无偏估计法（不做要求）；
2. 充分完全统计量法（不做要求）；

对给定参数分布族，如何寻找可估函数的UMVUE呢？

有三种方法：

1. 零无偏估计法（不做要求）；
2. 充分完全统计量法（不做要求）；
3. Cramer-Rao不等式法（3.5节）.

目录

3.5.1 引言

Cramer-Rao 不等式（简称 C-R 不等式）是判别一个无偏估计量是否为 UMVUE 的主要方法之一。

这一方法的思想如下：

Cramer-Rao 不等式（简称 C-R 不等式）是判别一个无偏估计量是否为 UMVUE 的主要方法之一。

这一方法的思想如下：

设$\mathcal{U}_g$是$g(\theta)$的一切无偏估计构成的类。$\mathcal{U}_g$中估计量的方差有一个下界，如果$g(\theta)$的一个无偏估计量$\hat{g}$的方差达到这个下界，则$\hat{g}$就是$g(\theta)$的一个UMVUE，当然样本分布族和$\hat{g}$要满足一定的正则条件。

这个不等式是由C. R. Rao 和 H. Cramer 在 1945 年和 1946 年分别证明的，故称此不等式为 Cramer-Rao 不等式（简称 C-R 不等式）。

这个不等式是由C. R. Rao 和 H. Cramer 在 1945 年和 1946 年分别证明的，故称此不等式为 Cramer-Rao 不等式（简称 C-R 不等式）。

以后一些统计学者将此不等式的条件作了一些改进和精确化，但结果的基本形式并无重大变化。

这一方法的缺陷是:

这一方法的缺陷是：

由于 C-R 不等式确定的下界（称为 C-R 下界）常比真下界小。在一些场合，虽然$g(\theta)$的 UMVUE $\hat{g}$存在，但其方差大于 C-R 下界。

这一方法的缺陷是：

由于 C-R 不等式确定的下界（称为 C-R 下界）常比真下界小。在一些场合，虽然$g(\theta)$的 UMVUE $\hat{g}$存在，但其方差大于 C-R 下界。

在这一情况下，用 C-R 不等式就无法判定 $g(\theta)$的 UMVUE 存在。因此这一方法的适用范围不广。

C-R 不等式除用于判别$g(\theta)$的UMVUE之外，它在数理统计理论上还要其他的用处，如估计的效率和有效估计的概念以及 Fisher 信息量都与之有关。

C-R 不等式除用于判别$g(\theta)$的UMVUE之外，它在数理统计理论上还要其他的用处，如估计的效率和有效估计的概念以及 Fisher 信息量都与之有关。

C-R 不等式成立需要样本分布族满足一些正则条件，适合这些条件的分布族称为 C-R 正则分布族，下面给出其定义。

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）)

若单参数概率函数族$\mathcal{F} = \{f(x, \theta) : \theta \in \Theta\}$满足下列条件：

（1）参数空间Θ是直线上的某个开区间；

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）)

若单参数概率函数族$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$满足下列条件：

（1）参数空间Θ是直线上的某个开区间；

（2）对任何$x\in\mathcal{X}$及$\theta\in\Theta$，$f(x,\theta)>0$，即分布族具有共同支撑；

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）)

若单参数概率函数族$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$满足下列条件：

（1）参数空间Θ是直线上的某个开区间；

（2）对任何$x\in\mathcal{X}$及$\theta\in\Theta$，$f(x,\theta)>0$，即分布族具有共同支撑；

（3）对任何$x\in\mathcal{X}$及$\theta\in\Theta$，$\partial f(x,\theta)/\partial\theta$存在；

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）)

若单参数概率函数族$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$满足下列条件：

（1）参数空间Θ是直线上的某个开区间；

（2）对任何$x \in \mathcal{X}$及$\theta \in \Theta$，$f(x,\theta) > 0$，即分布族具有共同支撑；

（3）对任何$x \in \mathcal{X}$及$\theta \in \Theta$，$\partial f(x,\theta)/\partial\theta$存在；

（4）概率函数$f(x,\theta)$的积分与微分运算可交换，即

$$\frac{\partial}{\partial\theta}\int f(x,\theta)dx = \int \frac{\partial}{\partial\theta} f(x,\theta)dx,$$

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）)

若单参数概率函数族$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$满足下列条件：

（1）参数空间Θ是直线上的某个开区间；

（2）对任何$x \in \mathcal{X}$及$\theta \in \Theta$，$f(x,\theta) > 0$，即分布族具有共同支撑；

（3）对任何$x \in \mathcal{X}$及$\theta \in \Theta$，$\partial f(x,\theta)/\partial\theta$存在；

（4）概率函数$f(x,\theta)$的积分与微分运算可交换，即

$$\frac{\partial}{\partial\theta}\int f(x,\theta)dx = \int \frac{\partial}{\partial\theta} f(x,\theta)dx,$$

若$f(x,\theta)$为离散随机变量的概率分布，上述条件改为无穷级数和微分运算可交换；

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）续)

（5）下列数学期望存在，且

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）续)

（5）下列数学期望存在，且

$$0 < I(\theta) = E_\theta\left[\frac{\partial \ln f(X,\theta)}{\partial \theta}\right]^2 < \infty, \tag{1}$$

定义 (3.5.1（C-R 正则分布族，C-R 正则条件）续)

（5）下列数学期望存在，且

$$0 < I(\theta) = E_\theta \left[\frac{\partial \ln f(X,\theta)}{\partial \theta} \right]^2 < \infty, \tag{1}$$

则称该分布族为C-R 正则分布族，其中（1）-（5）称为C-R 正则条件。

$I(\theta)$称为该分布的Fisher 信息量（或称为**Fisher 信息函数**）。

3.5.2 单参数 C-R 不等式

1. C-R 不等式及例

1. C-R 不等式及例

定理 (3.5.1)

设$\mathcal{F}=\{f(x,\theta):\theta\in\Theta\}$是 *C-R* 正则分布族，$g(\theta)$是定义于参数空间$\Theta$上的可微函数。设$\boldsymbol{X}=(X_1,\ldots,X_n)$是由总体$f(x,\theta)\in\mathcal{F}$中抽取的简单样本，$\hat{g}(\boldsymbol{X})$是$g(\theta)$的任一无偏估计，且满足下列条件：

1. C-R 不等式及例

定理 (3.5.1)

设$\mathcal{F} = \{f(x,\theta) : \theta \in \Theta\}$是 *C-R* 正则分布族，$g(\theta)$是定义于参数空间$\Theta$上的可微函数。设$\boldsymbol{X} = (X_1, \ldots, X_n)$是由总体$f(x,\theta) \in \mathcal{F}$中抽取的简单样本，$\hat{g}(\boldsymbol{X})$是$g(\theta)$的任一无偏估计，且满足下列条件：

（*6*）积分

$$\int \cdots \int \hat{g}(\boldsymbol{x}) f(\boldsymbol{x}, \theta) d\boldsymbol{x}$$

可在积分号下对θ求导数（此处$d\boldsymbol{x} = dx_1 \cdots dx_n$），

定理 (3.5.1（续）)

则有

$$Var_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}, \ \forall\theta \in \Theta. \tag{2}$$

此不等式称为 **Cramer-Rao 不等式**，简称 **C-R 不等式**。

注 3.5.1 在 C-R 不等式中出现的$I(\theta)$ 是由式（1）

$$I(\theta) = E_\theta\left[\frac{\partial \ln f(X,\theta)}{\partial \theta}\right]^2$$

定义的 Fisher 信息量。

注 3.5.1 在 C-R 不等式中出现的$I(\theta)$ 是由式（1）

$$I(\theta)=E_\theta\left[\frac{\partial\ln f(X,\theta)}{\partial\theta}\right]^2$$

定义的 Fisher 信息量。

- $I(\theta)$可以解释成单个样品（即样本容量为1的样本）提供的信息量，它反映了总体分布的性质。

注 3.5.1 在 C-R 不等式中出现的$I(\theta)$ 是由式（1）

$$I(\theta)=E_\theta\left[\frac{\partial\ln f(X,\theta)}{\partial\theta}\right]^2$$

定义的 Fisher 信息量。

- $I(\theta)$可以解释成单个样品（即样本容量为1的样本）提供的信息量，它反映了总体分布的性质。
- 由于$X_1,\ldots,X_n$是$i.i.d.$的，它们的地位是平等的，故每个样品X_i提供同样多的信息$I(\theta)$，故整个样本$(X_1,\ldots,X_n)$所含的信息量为$nI(\theta)$，它与样本容量n和总体分布的信息量$I(\theta)$成正比。

- 因此，式（2）

$$Var_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

右端的 C-R 下界表明，当样本的容量n越大或总体分布的信息量$I(\theta)$越大，C-R 下界就越小，UMVUE的精度就越高。这与人们的直观认知相符。

证 由于$X_1,\ldots,X_n$为$i.i.d.$样本，故有$f(\boldsymbol{x},\theta)=\prod_{i=1}^{n}f(x_i,\theta)$。记

证 由于$X_1,\dots,X_n$为$i.i.d.$样本，故有$f(\boldsymbol{x},\theta)=\prod_{i=1}^n f(x_i,\theta)$。记

$$S(\boldsymbol{x},\theta)=\frac{\partial \ln f(\boldsymbol{x},\theta)}{\partial\theta}=\sum_{i=1}^n \frac{\partial \ln f(x_i,\theta)}{\partial\theta},$$

因此由正则条件（3）和（4）可知

证 由于$X_1,\ldots,X_n$为$i.i.d.$样本，故有$f(\boldsymbol{x},\theta)=\prod_{i=1}^n f(x_i,\theta)$。记

$$S(\boldsymbol{x},\theta)=\frac{\partial \ln f(\boldsymbol{x},\theta)}{\partial\theta}=\sum_{i=1}^n\frac{\partial \ln f(x_i,\theta)}{\partial\theta},$$

因此由正则条件（3）和（4）可知

$$\begin{aligned}E_\theta\{S(\boldsymbol{X},\theta)\}&=\sum_{i=1}^n E_\theta\left\{\frac{\partial \ln f(X_i,\theta)}{\partial\theta}\right\}\\&=\sum_{i=1}^n\int\frac{1}{f(x_i,\theta)}\frac{\partial f(x_i,\theta)}{\partial\theta}\cdot f(x_i,\theta)dx_i\\&=\sum_{i=1}^n\int\frac{\partial f(x_i,\theta)}{\partial\theta}dx_i=\sum_{i=1}^n\frac{\partial}{\partial\theta}\int f(x_i,\theta)dx=0.\end{aligned}$$

由$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计和正则条件（5）和（6）可知

由$\hat{g}(\boldsymbol{X})$为$g(\theta)$的无偏估计和正则条件（5）和（6）可知

$$\begin{aligned}
Cov_\theta[\hat{g}(\boldsymbol{X}), S(\boldsymbol{X},\theta)] =& E_\theta\{\hat{g}(\boldsymbol{X})\cdot S(\boldsymbol{X},\theta)\} \\
=& \int\cdots\int \hat{g}(\boldsymbol{x})\left[\frac{1}{f(\boldsymbol{x},\theta)}\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}\right] f(\boldsymbol{x},\theta)d\boldsymbol{x} \\
=& \int\cdots\int \hat{g}(\boldsymbol{x})\frac{\partial f(\boldsymbol{x},\theta)}{\partial\theta}d\boldsymbol{x} \\
=& \frac{\partial}{\partial\theta}\int\cdots\int \hat{g}(\boldsymbol{x})f(\boldsymbol{x},\theta)d\boldsymbol{x} = \frac{\partial g(\theta)}{\partial\theta} \triangleq g'(\theta),
\end{aligned}$$

$$\begin{aligned} Var_\theta(S(\boldsymbol{X},\theta)) &= \sum_{i=1}^{n} Var_\theta\left\{\frac{\partial \ln f(X_i,\theta)}{\partial\theta}\right\} \\ &= \sum_{i=1}^{n} E_\theta\left\{\frac{\partial \ln f(X_i,\theta)}{\partial\theta}\right\}^2 = nI(\theta). \end{aligned} \tag{3}$$

由 Cauchy-Schwarz 不等式，得

$$
\begin{aligned}
Var_\theta(S(\boldsymbol{X},\theta)) &= \sum_{i=1}^{n} Var_\theta\left\{\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right\} \\
&= \sum_{i=1}^{n} E_\theta\left\{\frac{\partial \ln f(X_i,\theta)}{\partial \theta}\right\}^2 = nI(\theta).
\end{aligned} \tag{3}
$$

由 Cauchy-Schwarz 不等式，得

$$
Var_\theta\{\hat{g}(\boldsymbol{X})\} \cdot Var_\theta\{S(\boldsymbol{X},\theta)\} \geq [Cov_\theta(\hat{g}(\boldsymbol{X}), S(\boldsymbol{X},\theta))]^2 = [g'(\theta)]^2,
$$

将式（3）代入上式可得

将式（3）代入上式可得

$$Var_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}, \ \forall\theta \in \Theta.$$

$\square$

由式

$$Var_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}, \forall\theta \in \Theta$$

给出的 C-R 不等式可视为验证某一无偏估计是否为 UMVUE 的方法。

由式

$$Var_\theta[\hat{g}(\boldsymbol{X})] \geq \frac{[g'(\theta)]^2}{nI(\theta)}, \ \forall\theta \in \Theta$$

给出的 C-R 不等式可视为验证某一无偏估计是否为 UMVUE 的方法。

用 C-R 不等式寻找 $g(\theta)$ 的 UMVUE 时，首先要验证样本分布族是否满足 C-R 正则条件（1）-（5）和（6），然后再计算 Fisher 信息量$I(\theta)$和无偏估计$\hat{g}(\boldsymbol{X})$的方差$Var_\theta[\hat{g}(\boldsymbol{X})]$，看其是否达到 C-R 下界。

验证正则条件（1）-（5）和条件（6）十分麻烦。

验证正则条件（1）-（5）和条件（6）十分麻烦。

但幸运的是对指数族，上述正则条件（1）-（5）皆成立，而条件（6）正好是由定理1.5.2给出的指数族的一条重要性质。因此对于指数族，定理3.5.1的条件皆成立。

但要注意一点的是:

但要注意一点的是:

若$Var_\theta[\hat{g}(\boldsymbol{X})]$达不到 C-R 下界，并不能得出结论说 $g(\theta)$的 UMVUE 不存在，而只能说此法判别失效。

但要注意一点的是：

若$Var_\theta[\hat{g}(\boldsymbol{X})]$达不到 C-R 下界，并不能得出结论说 $g(\theta)$的 UMVUE 不存在，而只能说此法判别失效。

存在这样的例子，$\hat{g}(\boldsymbol{X})$是$g(\theta)$的UMVUE，但其方差大于 C-R下界。

例 (3.5.1)

设$\boldsymbol{X} = (X_1, \dots, X_n)$为从两点分布$B(1, p)$中抽取的简单样本，利用$C-R$不等式证明样本均值$\bar{X} = \sum_{i=1}^{n} X_i/n$为$p$的UMVUE。

证:

设随机变量$X \sim B(1, p)$，则其概率分布为

$$f(x, p) = p^x (1-p)^{1-x},\ x = 0, 1,\ 0 < p < 1.$$

证:

设随机变量$X \sim B(1,p)$，则其概率分布为

$$f(x,p) = p^x(1-p)^{1-x},\ x = 0,1,\ 0 < p < 1.$$

由于两点分布族是指数族，C-R正则条件成立。

Fisher信息函数为

Fisher信息函数为

$$
\begin{aligned}
I(p) =& E_p\left[\frac{\partial \ln f(X,p)}{\partial p}\right]^2 = E_p\left[\frac{X-p}{p(1-p)}\right]^2 \\
=& \frac{Var_p(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)},
\end{aligned}
$$

Fisher信息函数为

$$
\begin{aligned}
I(p) =& E_p\left[\frac{\partial \ln f(X,p)}{\partial p}\right]^2 = E_p\left[\frac{X-p}{p(1-p)}\right]^2 \\
=& \frac{Var_p(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)},
\end{aligned}
$$

因此C-R下界为$1/[nI(p)] = p(1-p)/n$。

Fisher信息函数为

$$
\begin{aligned}
I(p) =& E_p\left[\frac{\partial \ln f(X,p)}{\partial p}\right]^2 = E_p\left[\frac{X-p}{p(1-p)}\right]^2 \\
=& \frac{Var_p(X)}{p^2(1-p)^2} = \frac{1}{p(1-p)},
\end{aligned}
$$

因此C-R下界为$1/[nI(p)] = p(1-p)/n$。

而$\bar{X}$为p的无偏估计，其方差$Var_p(\bar{X}) = p(1-p)/n$达到C-R下界。

故$\bar{X}$为p的UMVUE。 □

例 (3.5.2)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从Poisson分布$P(\lambda)$中抽取的简单样本，用C-R不等式验证样本均值$\bar{X}=\sum_{i=1}^{n}X_i/n$为$\lambda$的UMVUE。

证：

设随机变量$X \sim P(\lambda)$，其概率分布为

证:

设随机变量$X \sim P(\lambda)$，其概率分布为

$$f(x, \lambda) = e^{-\lambda}\lambda^x / x!,\ x = 0, 1, \ldots;\ \lambda > 0.$$

证：

设随机变量$X \sim P(\lambda)$，其概率分布为

$$f(x, \lambda) = e^{-\lambda}\lambda^x/x!,\ x = 0, 1, \ldots;\ \lambda > 0.$$

由于Poisson分布族为指数族，故C-R正则条件成立。

Fisher信息函数为

Fisher信息函数为

$$
\begin{aligned}
I(\lambda) =& E_\lambda \left[\frac{\partial \ln f(X,\lambda)}{\partial \lambda}\right]^2 = E_\lambda \left(\frac{X-\lambda}{\lambda}\right)^2 \\
=& \frac{1}{\lambda^2} Var_\lambda(X) = \frac{1}{\lambda},
\end{aligned}
$$

Fisher信息函数为

$$\begin{aligned} I(\lambda) =& E_\lambda\left[\frac{\partial \ln f(X,\lambda)}{\partial \lambda}\right]^2 = E_\lambda\left(\frac{X-\lambda}{\lambda}\right)^2 \\ =& \frac{1}{\lambda^2} Var_\lambda(X) = \frac{1}{\lambda}, \end{aligned}$$

C-R下界为$1/[nI(\lambda)] = \lambda/n$。

Fisher信息函数为

$$\begin{aligned} I(\lambda) =& E_\lambda \left[\frac{\partial \ln f(X,\lambda)}{\partial \lambda}\right]^2 = E_\lambda \left(\frac{X-\lambda}{\lambda}\right)^2 \\ =& \frac{1}{\lambda^2} Var_\lambda(X) = \frac{1}{\lambda}, \end{aligned}$$

C-R下界为$1/[nI(\lambda)] = \lambda/n$。

$\hat{g}(\boldsymbol{X}) = \bar{X}$为$g(\lambda) = \lambda$的无偏估计，且方差$Var_\lambda(\bar{X}) = \lambda/n$达到C-R下界。故$\bar{X}$为$\lambda$的UMVUE。 □

例 (3.5.3)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从指数分布$Exp(\lambda)$中抽取的简单样本，用C-R不等式验证$\hat{g}(\boldsymbol{X})=\bar{X}$为$g(\lambda)=1/\lambda$的UMVUE。

证:

指数分布$Exp(\lambda)$的密度函数为

证：

指数分布$Exp(\lambda)$的密度函数为

$$f(x,\lambda)=\lambda e^{-\lambda x}I_{(0,\infty)}(x),\ \lambda>0.$$

证:

指数分布$Exp(\lambda)$的密度函数为

$$f(x,\lambda) = \lambda e^{-\lambda x} I_{(0,\infty)}(x),\ \lambda > 0.$$

指数分布族是指数族，故C-R正则条件成立。

3.5.2 单参数 C-R 不等式

Fisher信息函数为

Fisher信息函数为

$$
\begin{aligned}
I(\lambda) =& E_\lambda \left[\frac{\partial \ln f(X,\lambda)}{\partial \lambda}\right]^2 = E_\lambda \left(\frac{1}{\lambda} - X\right)^2 \\
=& Var_\lambda(X) = \frac{1}{\lambda^2},
\end{aligned}
$$

Fisher信息函数为

$$
\begin{aligned}
I(\lambda) =& E_\lambda \left[\frac{\partial \ln f(X,\lambda)}{\partial \lambda}\right]^2 = E_\lambda \left(\frac{1}{\lambda} - X\right)^2 \\
=& Var_\lambda(X) = \frac{1}{\lambda^2},
\end{aligned}
$$

故C-R下界为

Fisher信息函数为

$$
\begin{aligned}
I(\lambda) =& E_\lambda \left[\frac{\partial \ln f(X, \lambda)}{\partial \lambda} \right]^2 = E_\lambda \left(\frac{1}{\lambda} - X \right)^2 \\
=& Var_\lambda(X) = \frac{1}{\lambda^2},
\end{aligned}
$$

故C-R下界为

$$
\frac{g'(\lambda))^2}{nI(\lambda)} = \frac{\left(\frac{\partial \frac{1}{\lambda}}{\lambda} \right)^2}{nI(\lambda)} = \frac{\frac{1}{\lambda^4}}{\frac{n}{\lambda^2}} = \frac{1}{n\lambda^2}.
$$

而$\hat{g}(\bar{X}) = \bar{X}$为$g(\lambda) = 1/\lambda$的无偏估计，其方差$Var_\lambda(\bar{X}) = \frac{1}{n\lambda^2}$达到C-R下界，故$\bar{X}$为$g(\lambda) = 1/\lambda$的UMVUE。 □

例 (3.5.4)

设$\boldsymbol{X} = (X_1, \ldots, X_n)$为从$N(a, \sigma^2)$中抽取的简单样本，其中$\sigma^2$已知，用C-R不等式验证$\bar{X}$为$a$的UMVUE。

证:

由于正态分布族是指数族，C-R正则条件皆成立。正态分布$N(a,\sigma^2)$的密度函数为

证：

由于正态分布族是指数族，C-R正则条件皆成立。正态分布$N(a,\sigma^2)$的密度函数为

$$f(x,a)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}.$$

证:

由于正态分布族是指数族，C-R正则条件皆成立。正态分布$N(a,\sigma^2)$的密度函数为

$$f(x,a)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}.$$

Fisher信息函数为

证：

由于正态分布族是指数族，C-R正则条件皆成立。正态分布$N(a,\sigma^2)$的密度函数为

$$f(x,a)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{1}{2\sigma^2}(x-a)^2\right\}.$$

Fisher信息函数为

$$\begin{aligned}I(a)=&E_a\left[\frac{\partial\ln f(X,a)}{\partial a}\right]^2=E_a\left[\frac{(X-a)^2}{\sigma^4}\right]\\=&\frac{1}{\sigma^4}Var_\alpha(X)=\frac{1}{\sigma^2}.\end{aligned}$$

故C-R下界为$1/[nI(a)] = \sigma^2/n$。而$Var_a(\bar{X}) = \sigma^2/n$达到C-R下界，故$\bar{X}$为$a$的UMVUE。 $\square$

故C-R下界为$1/[nI(a)] = \sigma^2/n$。而$Var_a(\bar{X}) = \sigma^2/n$达到C-R下界，故$\bar{X}$为$a$的UMVUE。 □

同样，若本题中改a为已知，但σ^2未知，容易验证

$$S_a^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - a)^2$$

为σ^2的UMVUE（习题）。

以上讨论的都是参数θ为一维的情形，对$\boldsymbol{\theta}$为高维的情形也可以建立类似的结果。为此先引进一些记号。

以上讨论的都是参数θ为一维的情形，对$\boldsymbol{\theta}$为高维的情形也可以建立类似的结果。为此先引进一些记号。

设$\boldsymbol{A} = (a_{ij})$和$\boldsymbol{B} = (b_{ij})$是同阶的非负定方阵，若$\boldsymbol{A} - \boldsymbol{B}$是非负定的，则记为$\boldsymbol{A} \geq \boldsymbol{B}$。这时必有$a_{ii} \geq b_{ii}$，对一切$i$。

以上讨论的都是参数θ为一维的情形，对$\boldsymbol{\theta}$为高维的情形也可以建立类似的结果。为此先引进一些记号。

设$\boldsymbol{A}=(a_{ij})$和$\boldsymbol{B}=(b_{ij})$是同阶的非负定方阵，若$\boldsymbol{A}-\boldsymbol{B}$是非负定的，则记为$\boldsymbol{A}\geq\boldsymbol{B}$。这时必有$a_{ii}\geq b_{ii}$，对一切$i$。

现设$\boldsymbol{\theta}=(\theta_1,\ldots,\theta_k)$，总体概率函数记为$f(x,\boldsymbol{\theta})$，$\boldsymbol{X}=(X_1,\ldots,X_n)$为从总体中抽取的简单样本。

3.5.3 多参数 C-R 不等式简介

设$\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{X}) = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$是$\boldsymbol{\theta}$的一个无偏估计。以$Cov_{\boldsymbol{\theta}}(\hat{\theta})$记$\hat{\boldsymbol{\theta}}$的协方差阵，它是一个$k$阶非负定方阵，其$(i, j)$元为$E_\theta[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$，则在类似于$\boldsymbol{\theta}$为一维的正则条件下，可以证明

设$\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{X}) = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$是$\boldsymbol{\theta}$的一个无偏估计。以$Cov_{\boldsymbol{\theta}}(\hat{\theta})$记$\hat{\boldsymbol{\theta}}$的协方差阵，它是一个$k$阶非负定方阵，其$(i,j)$元为$E_\theta[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$，则在类似于$\boldsymbol{\theta}$为一维的正则条件下，可以证明

$$Cov_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geq (n\boldsymbol{I}(\boldsymbol{\theta}))^{-1}, \tag{4}$$

3.5.3 多参数 C-R 不等式简介

设$\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{X}) = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$是$\boldsymbol{\theta}$的一个无偏估计。以$Cov_{\boldsymbol{\theta}}(\hat{\theta})$记$\hat{\boldsymbol{\theta}}$的协方差阵，它是一个$k$阶非负定方阵，其$(i, j)$元为$E_\theta[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$，则在类似于$\boldsymbol{\theta}$为一维的正则条件下，可以证明

$$Cov_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geq (n\boldsymbol{I}(\boldsymbol{\theta}))^{-1}, \tag{4}$$

其中$\boldsymbol{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$是总体的Fisher信息矩阵（样本量是1时的信息阵），其中

设$\hat{\boldsymbol{\theta}} = \hat{\theta}(\boldsymbol{X}) = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$是$\boldsymbol{\theta}$的一个无偏估计。以$Cov_{\boldsymbol{\theta}}(\hat{\theta})$记$\hat{\boldsymbol{\theta}}$的协方差阵，它是一个$k$阶非负定方阵，其$(i,j)$元为$E_\theta[(\hat{\theta}_i - \theta_i)(\hat{\theta}_j - \theta_j)]$，则在类似于$\theta$为一维的正则条件下，可以证明

$$Cov_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) \geq (n\boldsymbol{I}(\boldsymbol{\theta}))^{-1}, \tag{4}$$

其中$\boldsymbol{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$是总体的Fisher信息矩阵（样本量是1时的信息阵），其中

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[\left(\frac{\partial \ln f(X,\boldsymbol{\theta})}{\partial \theta_i}\right)\left(\frac{\partial \ln f(X,\boldsymbol{\theta})}{\partial \theta_j}\right)\right],\ i,j = 1,2,\ldots,k, \tag{5}$$

则式（4）就是多维的C-R不等式。

则式（4）就是多维的C-R不等式。

若记$[\boldsymbol{I}(\boldsymbol{\theta})]^{-1}=\boldsymbol{I}^*(\boldsymbol{\theta})=(I_{ij}^*(\boldsymbol{\theta}))$，故由式（4）得

则式（4）就是多维的C-R不等式。

若记$[\boldsymbol{I}(\boldsymbol{\theta})]^{-1} = \boldsymbol{I}^*(\boldsymbol{\theta}) = (I^*_{ij}(\boldsymbol{\theta}))$，故由式（4）得

$$Var_{\boldsymbol{\theta}}(\hat{\theta}_i) \geq \frac{I^*_{ii}(\boldsymbol{\theta})}{n}, \ i = 1, 2, \ldots, k.$$

则式（4）就是多维的C-R不等式。

若记$[\boldsymbol{I}(\boldsymbol{\theta})]^{-1}=\boldsymbol{I}^*(\boldsymbol{\theta})=(I^*_{ij}(\boldsymbol{\theta}))$，故由式（4）得

$$Var_{\boldsymbol{\theta}}(\hat{\theta}_i)\geq\frac{I^*_{ii}(\boldsymbol{\theta})}{n},\ i=1,2,\ldots,k.$$

这给出了$\boldsymbol{\theta}$的每个分量θ_i的无偏估计$\hat{\theta}_i$的方差的下限。

注 3.5.2

注 3.5.2

在上述多参数 C-R 不等式中，Fisher信息阵$\boldsymbol{I}(\boldsymbol{\theta})=(I_{ij}(\boldsymbol{\theta}))$，其元素$I_{ij}(\boldsymbol{\theta})$由式（5）给出，以后我们会看到$I_{ij}(\boldsymbol{\theta})$还有另一种表示方式，即

注 3.5.2

在上述多参数 C-R 不等式中，Fisher信息阵$\boldsymbol{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$，其元素$I_{ij}(\boldsymbol{\theta})$由式（5）给出，以后我们会看到$I_{ij}(\boldsymbol{\theta})$还有另一种表示方式，即

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[-\left(\frac{\partial^2 \ln f(X, \boldsymbol{\theta})}{\partial\theta_i \partial\theta_j}\right)\right], \quad i, j = 1, 2, \ldots, k. \tag{6}$$

注 3.5.2

在上述多参数 C-R 不等式中，Fisher信息阵$\boldsymbol{I}(\boldsymbol{\theta}) = (I_{ij}(\boldsymbol{\theta}))$，其元素$I_{ij}(\boldsymbol{\theta})$由式（5）给出，以后我们会看到$I_{ij}(\boldsymbol{\theta})$还有另一种表示方式，即

$$I_{ij}(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left[-\left(\frac{\partial^2 \ln f(X,\boldsymbol{\theta})}{\partial\theta_i\partial\theta_j}\right)\right], \quad i,j = 1,2,\ldots,k. \tag{6}$$

可以证明：在C-R正则条件下，由式（5）和式（6）给出的Fisher信息阵是相等的。

例 (3.5.6)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从正态总体$N(a,\sigma^2)$中抽取的简单样本，记$\boldsymbol{\theta}=(a,\sigma^2)$，其中$\theta_1=a$，$\theta_2=\sigma^2$。求$\boldsymbol{\theta}$的两个分量无偏估计方差的C-R下界，并将其与$\theta_1$和$\theta_2$的无偏估计$\bar{X}$和$S^2$的方差进行比较。

解

解

正态随机变量的密度函数为

解

正态随机变量的密度函数为

$$f(x,\boldsymbol{\theta}) = (2\pi\theta_2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\},$$

可知

解

正态随机变量的密度函数为

$$f(x,\boldsymbol{\theta})=(2\pi\theta_2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\},$$

可知

$$\frac{\partial\ln f(x,\boldsymbol{\theta})}{\partial\theta_1}=\frac{x-\theta_1}{\theta_2},$$

解

正态随机变量的密度函数为

$$f(x,\boldsymbol{\theta})=(2\pi\theta_2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2\theta_2}(x-\theta_1)^2\right\},$$

可知

$$\frac{\partial\ln f(x,\boldsymbol{\theta})}{\partial\theta_1}=\frac{x-\theta_1}{\theta_2},$$
$$\frac{\partial\ln f(x,\boldsymbol{\theta})}{\partial\theta_2}=\frac{-\theta_2+(x-\theta_1)^2}{2\theta_2^2}.$$

由此算出

由此算出

$$I_{11}(\boldsymbol{\theta}) = \frac{1}{\sigma^2},\ I_{22}(\boldsymbol{\theta}) = \frac{1}{2\sigma^4},\ I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0,$$

故有

由此算出

$$I_{11}(\boldsymbol{\theta}) = \frac{1}{\sigma^2},\ I_{22}(\boldsymbol{\theta}) = \frac{1}{2\sigma^4},\ I_{12}(\boldsymbol{\theta}) = I_{21}(\boldsymbol{\theta}) = 0,$$

故有

$$n\boldsymbol{I}(\boldsymbol{\theta}) = \begin{pmatrix} \frac{n}{\sigma^2} & 0 \\ 0 & \frac{n}{2\sigma^4} \end{pmatrix},\ (n\boldsymbol{I}(\boldsymbol{\theta}))^{-1} = \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^4}{n}, \end{pmatrix},$$

若记

若记

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \hat{\theta}_2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

若记

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \hat{\theta}_2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

则由多维C-R不等式可知

若记

$$\hat{\theta}_1 = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,\ \hat{\theta}_2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

则由多维C-R不等式可知

$$Cov_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = Cov_{\boldsymbol{\theta}}\begin{pmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \end{pmatrix} \geq \begin{pmatrix} \frac{\sigma^2}{n} & 0 \\ 0 & \frac{2\sigma^2}{n} \end{pmatrix}.$$

而$Var_\theta(\hat{\theta}_1) = \sigma^2/n$达到C-R下界，故它是$\theta_1 = a$的UMVUE。

而$Var_\theta(\hat{\theta}_1) = \sigma^2/n$达到C-R下界，故它是$\theta_1 = a$的UMVUE。

利用$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$及$Var_\theta[(n-1)S^2/\sigma^2] = 2(n-1)$可得

而$Var_\theta(\hat{\theta}_1) = \sigma^2/n$达到C-R下界，故它是$\theta_1 = a$的UMVUE。

利用$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$及$Var_\theta[(n-1)S^2/\sigma^2] = 2(n-1)$可得

$$Var_\theta(\hat{\theta}_2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}.$$

而$Var_\theta(\hat{\theta}_1) = \sigma^2/n$达到C-R下界，故它是$\theta_1 = a$的UMVUE。

利用$(n-1)S^2/\sigma^2 \sim \chi^2_{n-1}$及$Var_\theta[(n-1)S^2/\sigma^2] = 2(n-1)$可得

$$Var_\theta(\hat{\theta}_2) = \frac{2\sigma^4}{n-1} > \frac{2\sigma^4}{n}.$$

因此，$S^2 = \hat{\theta}_2$的方差大于C-R下界。

而$Var_\theta(\hat{\theta}_1)=\sigma^2/n$达到C-R下界，故它是$\theta_1=a$的UMVUE。

利用$(n-1)S^2/\sigma^2\sim\chi^2_{n-1}$及$Var_\theta[(n-1)S^2/\sigma^2]=2(n-1)$可得

$$Var_\theta(\hat{\theta}_2)=\frac{2\sigma^4}{n-1}>\frac{2\sigma^4}{n}.$$

因此，$S^2=\hat{\theta}_2$的方差大于C-R下界。

在例3.4.10中已证明了$\bar{X}$和S^2分别是a和σ^2的UMVUE。 $\square$

由本例看出，即使在正态总体方差这样简单的场合，方差σ^2的UMVUE S^2也达不到C-R下界。

由本例看出，即使在正态总体方差这样简单的场合，方差σ^2的UMVUE S^2也达不到C-R下界。

因此在一些问题中C-R下界常比真下界小。这表明以C-R不等式作为寻找UMVUE的方法是不够理想的。

3.5.4 有效估计和估计的效率

定义（（有效估计））

设$\hat{g}_n(\boldsymbol{X})$为$g(\theta)$的无偏估计，比值

$$e_{\hat{g}_n}(\theta) = \frac{[g'(\theta)]^2/(nI(\theta))}{Var_\theta[\hat{g}_n(\boldsymbol{X})]}$$

称为无偏估计$\hat{g}_n(\boldsymbol{X})$的**效率**。

定义（（有效估计））

设$\hat{g}_n(\boldsymbol{X})$为$g(\theta)$的无偏估计，比值

$$e_{\hat{g}_n}(\theta) = \frac{[g'(\theta)]^2/(nI(\theta))}{Var_\theta[\hat{g}_n(\boldsymbol{X})]}$$

称为无偏估计$\hat{g}_n(\boldsymbol{X})$的**效率**。

显然$0 < e_{\hat{g}_n}(\theta) \leq 1$，当$e_{\hat{g}_n}(\theta) = 1$时，称$\hat{g}_n(\boldsymbol{X})$是$g(\theta)$的**有效估计**。

定义（（有效估计））

设$\hat{g}_n(\boldsymbol{X})$为$g(\theta)$的无偏估计，比值

$$e_{\hat{g}_n}(\theta) = \frac{[g'(\theta)]^2/(nI(\theta))}{Var_\theta[\hat{g}_n(\boldsymbol{X})]}$$

称为无偏估计$\hat{g}_n(\boldsymbol{X})$的**效率**。

显然$0 < e_{\hat{g}_n}(\theta) \le 1$，当$e_{\hat{g}_n}(\theta) = 1$时，称$\hat{g}_n(\boldsymbol{X})$是$g(\theta)$的**有效估计**。

若$\hat{g}_n(\boldsymbol{X})$不是$g(\theta)$的有效估计，但$\lim\limits_{n\to\infty} e_{\hat{g}_n}(\theta) = 1$，则称$\hat{g}_n(\boldsymbol{X})$是$g(\theta)$的**渐近有效估计**。

3.5.4 有效估计和估计的效率

- 这一概念有其不足之处：有效估计是无偏估计类中最好的估计，当然希望使用它。可惜，有效估计是不多的，但渐近有效估计却不少。

- 这一概念有其不足之处：有效估计是无偏估计类中最好的估计，当然希望使用它。可惜，有效估计是不多的，但渐近有效估计却不少。
- 从有效估计的定义看，有效估计一定是UMVUE，但很多UMVUE不是有效估计。这是因为C-R下界偏小，在很多场合UMVUE的方差达不到C-R下界。

3.5.4 有效估计和估计的效率

- 这一概念有其不足之处：有效估计是无偏估计类中最好的估计，当然希望使用它。可惜，有效估计是不多的，但渐近有效估计却不少。
- 从有效估计的定义看，有效估计一定是UMVUE，但很多UMVUE不是有效估计。这是因为C-R下界偏小，在很多场合UMVUE的方差达不到C-R下界。
- 另外C-R不等式成立的前提是要求样本分布族满足C-R正则条件。当这些条件不成立时，C-R不等式的结果可以不对，这时依据它所提供的C-R下界去定义估计的效率或求有效估计就不合理了。

例 (3.5.8)

由例3.5.1-例3.5.4给出的有关参数的无偏估计，其方差都能达到C-R下界，因此它们都是相应参数的有效估计。

例 (3.5.9)

设$\boldsymbol{X}=(X_1,\ldots,X_n)$为从$N(a,\sigma^2)$中抽取的简单样本。

（1）当a未知时，证明样本方差S^2不是σ^2的有效估计，但是渐近有效估计。

（2）当a已知时，求σ^2的有效估计。

解:

解：

（1）设$g(\sigma^2)=\sigma^2$，由例3.5.6知，当a未知时，

$S^2=\sum\limits_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$的方差为

解：

（1）设$g(\sigma^2)=\sigma^2$，由例3.5.6知，当a未知时，

$S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$的方差为

$$Var_{\sigma}^2(S^2)=2\sigma^4/(n-1),$$

解：

（1）设$g(\sigma^2)=\sigma^2$，由例3.5.6知，当a未知时，

$S^2=\sum_{i=1}^{n}(X_i-\bar{X})^2/(n-1)$的方差为

$$Var_{\sigma}^{2}(S^2)=2\sigma^4/(n-1),$$

达不到C-R下界$2\sigma^4/n$，故它不是σ^2的有效估计。

3.5.4 有效估计和估计的效率

估计的效率为

3.5.4 有效估计和估计的效率

估计的效率为

$$e_{S^2}(\sigma^2) = \frac{\left[\frac{\partial g(\sigma^2)}{\partial \sigma^2}\right] / (nI(\sigma^2))}{Var_\sigma^2(S^2)} = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} = \frac{n-1}{n} < 1,$$

估计的效率为

$$e_{S^2}(\sigma^2) = \frac{\left[\frac{\partial g(\sigma^2)}{\partial \sigma^2}\right] / (nI(\sigma^2))}{Var_\sigma^2(S^2)} = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} = \frac{n-1}{n} < 1,$$

但是

$$\lim_{n\to\infty} e_{S^2}(\sigma^2) = \lim_{n\to\infty} \frac{n-1}{n} = 1,$$

估计的效率为

$$e_{S^2}(\sigma^2) = \frac{\left[\frac{\partial g(\sigma^2)}{\partial \sigma^2}\right] / (nI(\sigma^2))}{Var_\sigma^2(S^2)} = \frac{\frac{2\sigma^4}{n}}{\frac{2\sigma^4}{n-1}} = \frac{n-1}{n} < 1,$$

但是

$$\lim_{n\to\infty} e_{S^2}(\sigma^2) = \lim_{n\to\infty} \frac{n-1}{n} = 1,$$

因此S^2是σ^2的渐近有效估计。

（2）由于a已知，令$S_a^2 = \sum_{i=1}^{n}(X_i - a)^2/n$，由于$nS_a^2/\sigma^2 \sim \chi_n^2$，故$Var(nS_a^2/\sigma^2) = 2n$，因此有$Var(S_a^2) = 2\sigma^4/n$，它达到C-R下界，故$S_a^2$为$\sigma^2$的有效估计。 $\square$