统计信号处理

第二章

最小方差无偏估计 I
——概念及其性能界

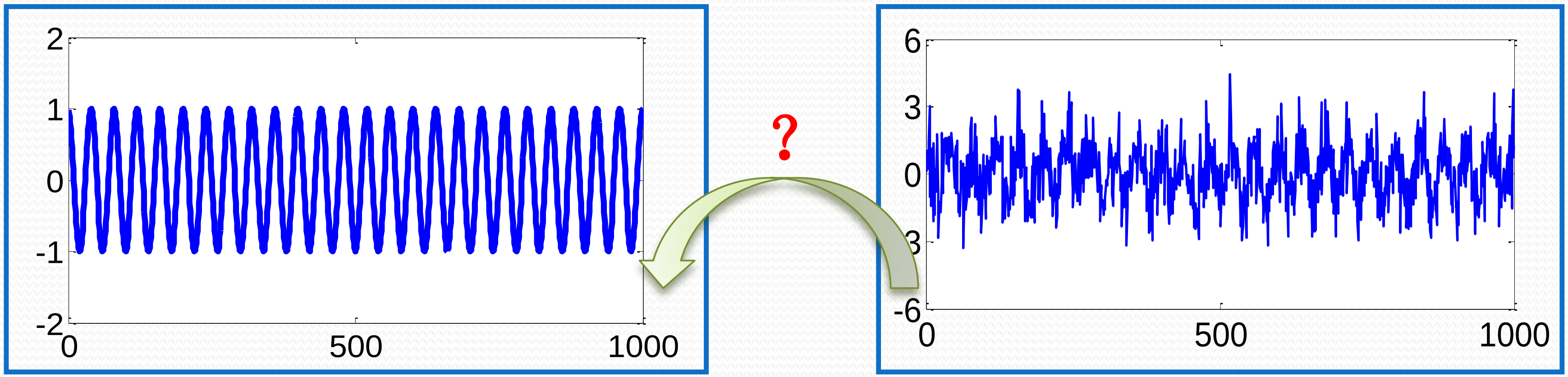
清华大学电子工程系
李洪 副教授

2023. 2

内容概要

- 一、估计的概念
- 二、最小方差无偏估计
- 三、标量参数时克拉美罗界
- 四、矢量参数时克拉美罗界
- 五、应用案例
- 六、小结

一、估计的概念



$$A\cos(2\pi f t+\phi)$$

$$A\cos(2\pi f t+\phi)+w(t)$$

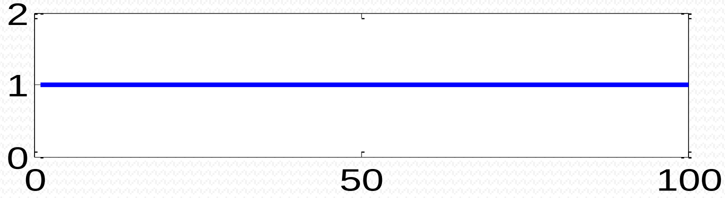
数学模型：

$$A?\quad f?\quad \phi? \xleftarrow{\text{函数 } g} \{x[0],x[1],x[2],\ldots,x[N-1]\}$$

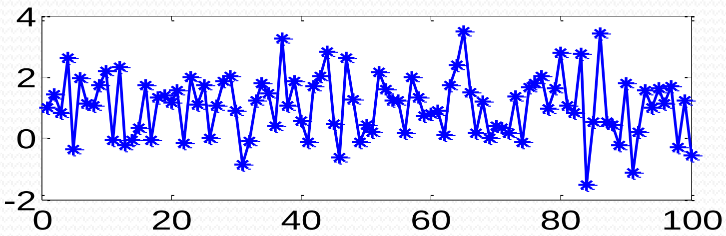
估计：根据观测数据来确定未知参数（待估计参数）的规则

估计理论分类

1. 第一类



$$s[n]=A$$

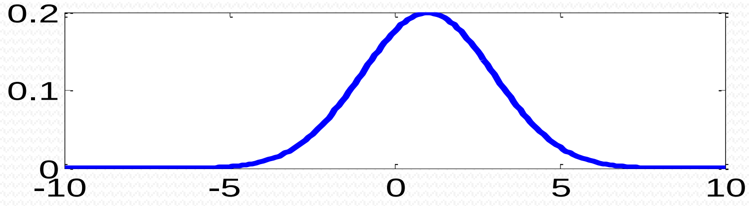


$$x[n]=A+w[n]$$

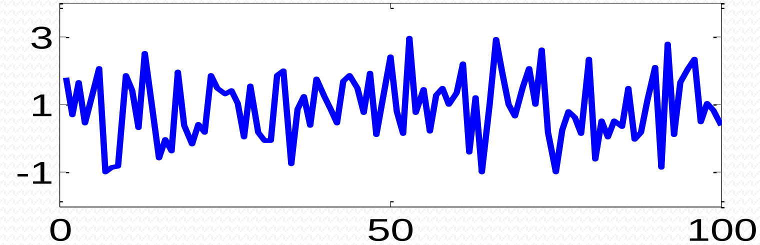
$$p(\mathbf{x};A)$$

经典估计

2. 第二类



$$p(A)$$

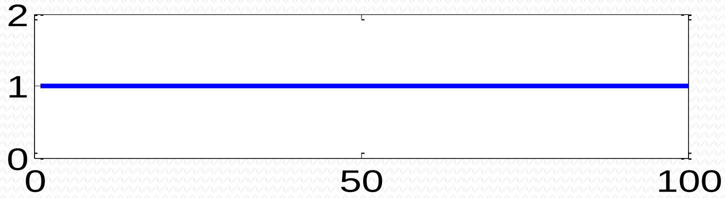


$$x[n]=A+w[n]$$

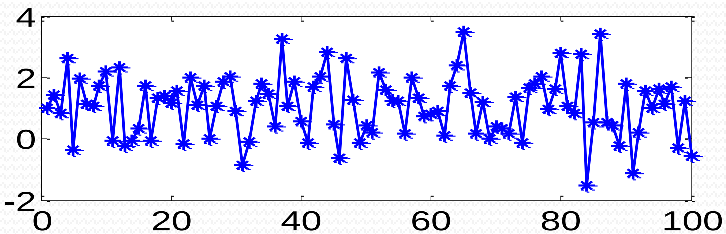
$$p(\mathbf{x},A)$$

贝叶斯估计

● 估计量特性及选择



$$s[n] = A$$



$$x[n] = A + w[n]$$

$$\hat{A}_1 = \frac{1}{N}\sum_{n=0}^{N-1} x[n] = \frac{1}{N}\sum_{n=0}^{N-1}\left(A + w[n]\right)$$

估计量是**随机变量**

$$\hat{A}_2 = x[0]$$

$$\hat{A}_3 = \frac{1}{2}\left(x[0] + x[1]\right)$$

$$\hat{A}_4 = \frac{2}{3N}\sum_{n=0}^{N/2-1} x[n] + \frac{4}{3N}\sum_{n=N/2}^{N-1} x[n]$$

孰优孰劣?

如何评判?

均方误差(mean square error, MSE)准则

$$\mathrm{mse}\left(\hat{\theta}\right)=E\left\{\left(\hat{\theta}-\theta\right)^2\right\}$$

偏差：$b\left(\theta\right)=E\left(\hat{\theta}\right)-\theta$

$$=E\left\{\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)+\left(E\left(\hat{\theta}\right)-\theta\right)\right)^2\right\}$$

$$=E\left\{\underline{\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2}+\underline{\left(E\left(\hat{\theta}\right)-\theta\right)^2}+\underline{2\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)\left(E\left(\hat{\theta}\right)-\theta\right)}\right\}$$

$$=\mathrm{var}\left(\hat{\theta}\right)+b^2\left(\theta\right)$$

$\min\left\{\mathrm{mse}\left(\hat{\theta}\right)\right\}$ ：与真值有关，一般不可实现

——尽管如此，却是后续大量估计理论方法的出发点

二、最小方差无偏估计(MVU)

- **退而求其次 ——最小方差无偏估计 (MVU)**

MVU: Minimum variance unbiased

1. 第一层含义——无偏

无偏估计

$$E\left(\hat{\theta}\right)=\theta,\ \ a<\theta<b$$

(1) 要求对 θ 取值范围内的**所有值**均成立

(2) 是否一定存在无偏估计？

例: 现有一观测数据 $x[0]$，其服从均匀分布 $U\left[0,\frac{1}{\theta}\right]$。对参数$\theta$是否存在无偏估计?

假定存在无偏估计:

$$\hat{\theta}=g\left(x[0]\right)$$

则要求: $E\left(\hat{\theta}\right)=\theta$

按期望的数学定义有:

$$E\left(\hat{\theta}\right)=\int\hat{\theta}p\left(x;\theta\right)dx$$

$$=\int_0^{1/\theta}g\left(x\right)\theta dx$$

$$\Rightarrow\int_0^{1/\theta}g\left(x\right)dx=1$$

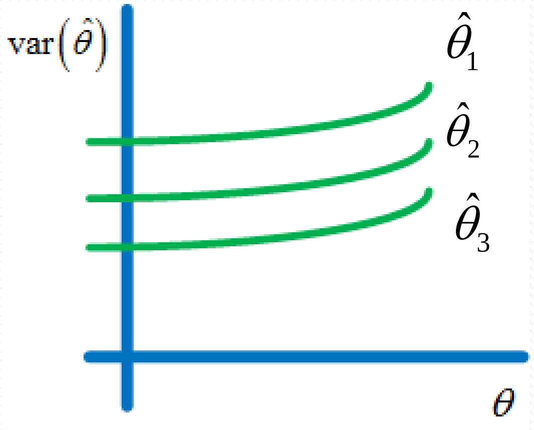
$$\Rightarrow\begin{cases}\forall\theta_1,\theta_2\\ \int_0^{1/\theta_1}g\left(x\right)dx=1\\ \int_0^{1/\theta_2}g\left(x\right)dx=1\end{cases}\Rightarrow$$

$$\int_{1/\theta_1}^{1/\theta_2}g\left(x\right)dx=0\Rightarrow g\left(x\right)=0$$

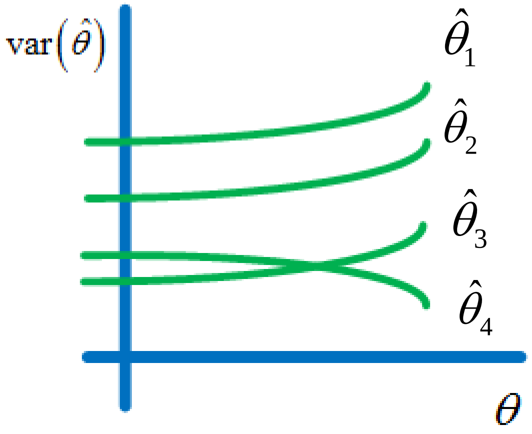
即，不存在无偏估计量!

若分布函数为$U\left[0,\theta\right]$？

2. 第二层含义——方差最小



一致最小方差无偏估计



局部最小方差无偏估计

- MVU要求无论待估计参数是多少，其方差都是最小的，因此要求**一致最小方差无偏估计**
- 是否一定存在一致最小方差无偏估计？

例： 假定两观测量 $x[0]$和 $x[1]$ 相互独立，且

$$x[0] \sim N(\theta,1)$$

$$x[1] \sim \begin{cases} N(\theta,1), & \theta \ge 0 \\ N(\theta,2), & \theta < 0 \end{cases}$$

待估计参数为 θ。

$$\begin{cases} \text{估计量1:} \ \hat{\theta}_1 = \frac{1}{2}\left(x[0]+x[1]\right) \\ \text{估计量2:} \ \hat{\theta}_2 = \frac{2}{3}x[0]+\frac{1}{3}x[1] \end{cases}$$

谁是MVU？

均无偏

$$\mathrm{var}\left(\hat{\theta}_1\right) = \frac{1}{4}\mathrm{var}\left(x[0]\right)+\frac{1}{4}\mathrm{var}\left(x[1]\right)$$

$$= \begin{cases} \frac{18}{36}, \theta \ge 0 \\ \frac{27}{36}, \theta < 0 \end{cases}$$

$$\mathrm{var}\left(\hat{\theta}_2\right) = \frac{4}{9}\mathrm{var}\left(x[0]\right)+\frac{1}{9}\mathrm{var}\left(x[1]\right)$$

$$= \begin{cases} \frac{20}{36}, \theta \ge 0 \\ \frac{24}{36}, \theta < 0 \end{cases}$$

二者中不存在MVU

● MVU的内涵

MVU估计量 $\hat{\theta}$：

$$s.t.\ E\left(\hat{\theta}\right)=\theta$$

$$\min\left\{\operatorname{var}\left(\hat{\theta}\right)\right\} \qquad \operatorname{var}\left(\hat{\theta}\right)=E\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right)$$

——使各估计值发散程度尽量小

$$\operatorname{var}\left(\hat{\theta}\right)=E\left(\left(\hat{\theta}-\theta\right)^2\right)=\operatorname{mse}\left(\hat{\theta}\right)$$

——体现与最小MSE相同思想

$$s.t.\ E\left(\hat{\theta}\right)=\theta$$

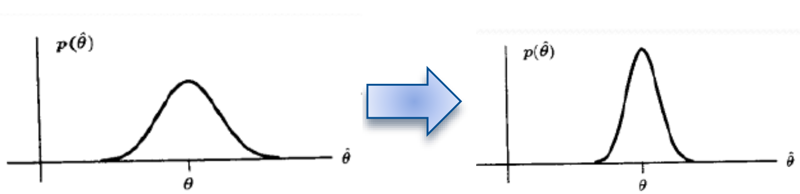
——使估计值平均意义上靠近真值

使估计值与真值尽量靠近

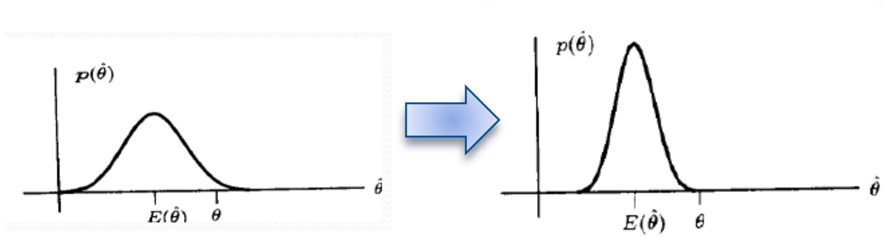
某种程度上，MVU是对无法实现的最小MSE准则“迂回”实现：

先无偏，再方差最小

无偏时减小方差



有偏时减小方差



多选题 1分

设置

此题未设置答案，请点击右侧设置按钮

下面有关最小方差无偏估计的含义，说法正确的是？

A 是所有估计量中方差最小的

B 该估计量一定是无偏的

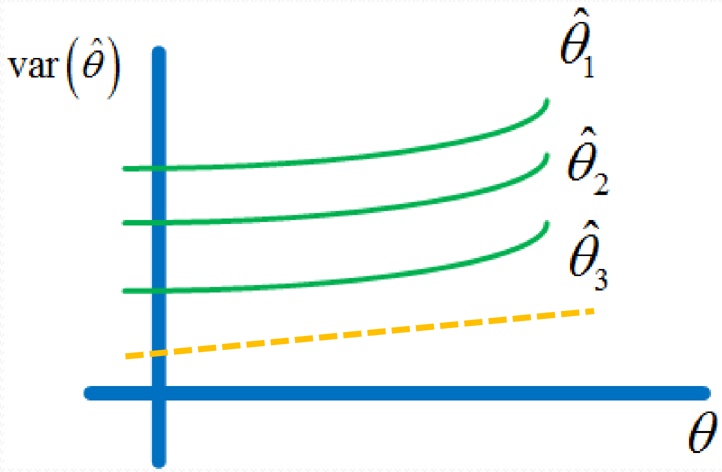
C 该估计量可能不存在

D 是无偏估计量中方差最小的

E 是无偏估计量中均方误差最小的

提交

MVU如何找？



- $\hat{\theta}_3$ 是这三个中最好的
- 是否还有更好的？
- **方差是否有界？**
- **若有，是什么样的界？**

✓ **克拉美罗界**

- Cramer-Rao Lower Bound (CRLB), 也称克拉美罗下界
- 到目前，最容易确定的性能界
- 若存在可达到该性能界的MVU，可求之。否则，即使不存在，也为评价估计量性能提供了“参考”

——为我们确定估计量性能及寻找最佳估计量找到了评估手段、方向

三、标量参数时克拉美罗界

- 1. 标量参数克拉美罗界（CRLB）定理
- 2. 参数变化下CRLB

1. 标量参数的CRLB

核心问题： $\mathrm{var}\left(\hat{\theta}\right)=E\left(\left(\hat{\theta}-E\left(\hat{\theta}\right)\right)^2\right) \Leftrightarrow \mathrm{var}\left(\hat{\theta}\right)=E\left(\left(\hat{\theta}-\theta\right)^2\right)$

假定 $\hat{\theta}$ 是 θ 的无偏估计，

$$\int \hat{\theta} p\left(\mathbf{x};\theta\right)\, d\mathbf{x}=\theta$$

$$\Rightarrow \frac{\partial \int \hat{\theta} p\left(\mathbf{x};\theta\right)\, d\mathbf{x}}{\partial\theta}=1$$

$$\Rightarrow \int \hat{\theta} \frac{\partial p\left(\mathbf{x};\theta\right)}{\partial\theta}\, d\mathbf{x}=1$$

$$\frac{\partial p\left(\mathbf{x};\theta\right)}{\partial\theta}=\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} p\left(\mathbf{x};\theta\right)$$

即 $\int \hat{\theta} \frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} p\left(\mathbf{x};\theta\right)\, d\mathbf{x}=1$ （1）

恒等式： $\int p\left(\mathbf{x};\theta\right)\, d\mathbf{x}=1$

$$\Rightarrow \frac{\partial \int p\left(\mathbf{x};\theta\right)\, d\mathbf{x}}{\partial\theta}=0$$

$$\Rightarrow \int \frac{\partial p\left(\mathbf{x};\theta\right)}{\partial\theta}\, d\mathbf{x}=0$$

即 $\int \frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} p\left(\mathbf{x};\theta\right)\, d\mathbf{x}=0$ （2）

$\int \theta \frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} p\left(\mathbf{x};\theta\right)\, d\mathbf{x}=0$ （3）

由式（1）、（3）做差可得

$$\int\left(\hat{\theta}-\theta\right)\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\,p\left(\mathbf{x};\theta\right)\,d\mathbf{x}=1 \qquad （4）$$

Cauchy-Schwarz不等式

$$\left[\int w\left(\mathbf{x}\right)g\left(\mathbf{x}\right)h\left(\mathbf{x}\right)d\mathbf{x}\right]^2\le\int w\left(\mathbf{x}\right)g^2\left(\mathbf{x}\right)d\mathbf{x}\int w\left(\mathbf{x}\right)h^2\left(\mathbf{x}\right)d\mathbf{x}$$

根据Cauchy-Schwarz不等式有

$$\overbrace{\left[\int\left(\hat{\theta}-\theta\right)\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\,p\left(\mathbf{x};\theta\right)\,d\mathbf{x}\right]^2}^{1}\le\overbrace{\int\left(\hat{\theta}-\theta\right)^2 p\left(\mathbf{x};\theta\right)\,d\mathbf{x}}^{\operatorname{var}\left(\hat{\theta}\right)}\times\overbrace{\int\left(\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\right)^2 p\left(\mathbf{x};\theta\right)\,d\mathbf{x}}^{E\left[\left(\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\right)^2\right]}$$

$$\operatorname{var}\left(\hat{\theta}\right)\ge\frac{1}{E\left[\left(\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\right)^2\right]}$$

——无偏估计的性能界

常用恒等式

由式（2）有： $\int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0$

$$\Rightarrow \quad \frac{\partial}{\partial \theta} \int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0$$

$$\Rightarrow \quad \int \left(\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) + \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \frac{\partial p(\mathbf{x};\theta)}{\partial \theta} \right) d\mathbf{x} = 0$$

$$\Rightarrow \quad \int \left(\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} p(\mathbf{x};\theta) + \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta) \right) d\mathbf{x} = 0$$

$$E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} \right] = -E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \right)^2 \right]$$

$$\Rightarrow \quad \operatorname{var}(\hat{\theta}) \geq \frac{1}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} \right)^2 \right]} = \frac{1}{\boxed{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} \right]}}$$

观测数据的Fisher信息

➢ CRLB什么时候取等号？

对Cauchy-Schwarz不等式

$$\left[\int\left(\hat{\theta}-\theta\right)\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}p\left(\mathbf{x};\theta\right)\,d\mathbf{x}\right]^2 \le \int\left(\hat{\theta}-\theta\right)^2 p\left(\mathbf{x};\theta\right)\,d\mathbf{x}\int\left(\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta}\right)^2 p\left(\mathbf{x};\theta\right)\,d\mathbf{x}$$

当且仅当对某个参数 $I(\theta)$，使得

$$\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} = I(\theta)\left(\hat{\theta}-\theta\right) \text{ 时}$$

上述等号才会成立

即，当且仅当

$$\frac{\partial \ln p\left(\mathbf{x};\theta\right)}{\partial\theta} = \underline{I(\theta)}\left(\hat{\theta}-\theta\right)$$

?

此时估计量 $\hat{\theta}$ 是MVU，且是达到下限的MVU

是观察数据的函数，不妨记为g

对 $\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(\hat{\theta}-\theta)$ 两边求导，可得

$$\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2} = -I(\theta) + \frac{\partial I(\theta)}{\partial \theta}(\hat{\theta}-\theta)$$

两边求期望，可得

$$I(\theta) = -E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]$$

Fisher信息

当且仅当某两个函数 I 和 g 满足

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = I(\theta)(g(\mathbf{x})-\theta)$$

才可得到达到下限的MVU：$\hat{\theta} = g(\mathbf{x})$，此时的估计量方差为

$$\operatorname{var}(\hat{\theta}) = \frac{1}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right]} = \frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right]} = \frac{1}{I(\theta)}$$

回顾推导过程：

假定 $\hat{\theta}$ 是 θ 的无偏估计，

按无偏估计定义有：

$$\int \hat{\theta}\, p(\mathbf{x};\theta)\, d\mathbf{x} = \theta$$

求导

$$\frac{\partial \int \hat{\theta}\, p(\mathbf{x};\theta)\, d\mathbf{x}}{\partial \theta} = 1$$

交换求导与积分顺序

$$\int \hat{\theta} \underline{\frac{\partial p(\mathbf{x};\theta)}{\partial \theta}}\, d\mathbf{x} = 1$$

$$\frac{\partial p(\mathbf{x};\theta)}{\partial \theta} = \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)$$

即 $$\int \hat{\theta} \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 1 \quad （1）$$

恒等式

$$\int p(\mathbf{x};\theta)\, d\mathbf{x} = 1$$

求导

$$\partial \int p(\mathbf{x};\theta)\, d\mathbf{x}$$

正则条件：

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right] = 0$$，对所有 θ

即

$$\underline{\int \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0} \quad （2）$$

$$\int \theta \frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} p(\mathbf{x};\theta)\, d\mathbf{x} = 0 \quad （3）$$

假定概率密度函数 $p(\mathbf{x};\theta)$ 满足“正则条件”

$$E\left[\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right]=0\text{，对所有}\theta$$

1 那么，任意无偏估计 $\hat{\theta}$ 的方差必定满足

$$\operatorname{var}(\hat{\theta})\geq\frac{1}{E\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right]}=\frac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]}$$

2 当且仅当某两个函数 I 和 g 满足

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}=I(\theta)\left(g(\mathbf{x})-\theta\right)$$

标量参数的 Cramer-Rao Lower Bound (CRLB)定理

才可得到达到上述下限的MVU: $\hat{\theta}=g(\mathbf{x})$，且此时的方差为 $1/I(\theta)$ 称为**有效估计量!**