复习

7.2 贝叶斯估计

后验概率

$$\int_{-\infty}^{\infty} C(\theta,\hat{\theta}(z))f(\theta\,|\,z)d\theta \ = \ \text{最小}$$

代价函数

平方代价： $C(\theta,\hat{\theta})=(\theta-\hat{\theta})^2$ ⟶ $\hat{\theta}=E[\theta\,|\,z]$

绝对值代价： $C(\theta,\hat{\theta})=|\,\theta-\hat{\theta}\,|$ ⟶

$$\int_{-\infty}^{\hat{\theta}_{abs}} f(\theta\,|\,z)d\theta = \int_{\hat{\theta}_{abs}}^{\infty} f(\theta\,|\,z)d\theta$$

均匀代价： $C(\theta,\ \hat{\theta})=\begin{cases}1, & |\,\theta-\hat{\theta}\,|\geq\dfrac{\Delta}{2}\\[2mm] 0, & |\,\theta-\hat{\theta}\,|<\dfrac{\Delta}{2}\end{cases}$ ⟶

$$\left.\frac{\partial f(\theta\,|\,z)}{\partial\theta}\right|_{\theta=\hat{\theta}_{map}}=0$$

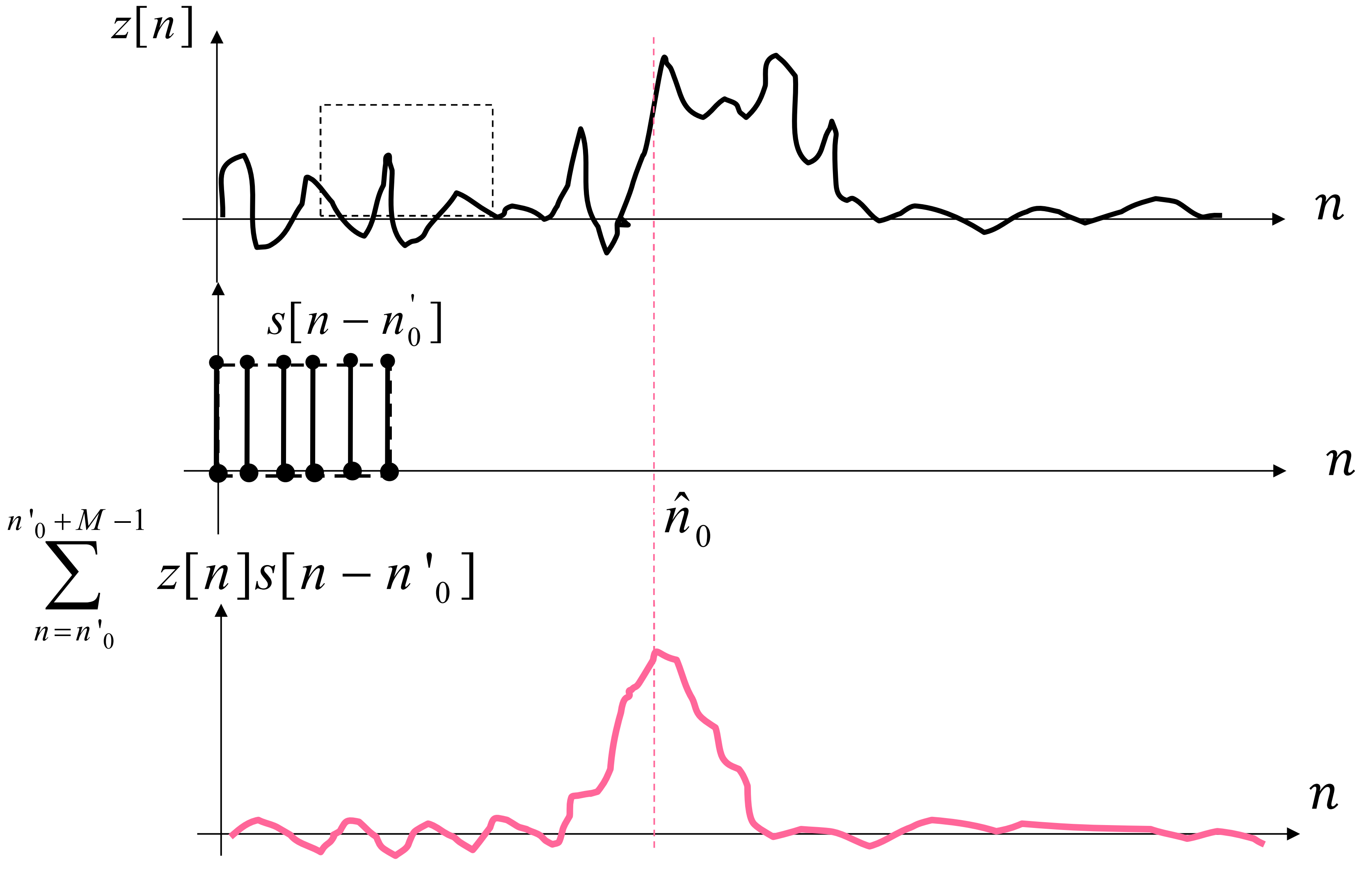
$$\left.\frac{\partial \ln f(\theta\,|\,z)}{\partial\theta}\right|_{\theta=\hat{\theta}_{map}}=0$$

7.3 最大似然估计

$$\left.\frac{\partial f(z\,|\,\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{mL}} = 0$$

$$\left.\frac{\partial \ln f(z\,|\,\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{mL}} = 0$$

复习



习题： 7.7 7.15 7.19

各种方法的适应范围

A为未知常参量：
- 最大似然估计
- 最小二乘估计

A为随机变量：
- 最小均方估计
- 最大后验估计
- 条件中位数估计

（以上三种：$f(A)$已知）

- 线性最小均方估计 —— $f(A)$未知，一、二阶特性已知

中山大學 SUN YAT-SEN UNIVERSITY 7.4 估计量的性能

复习相关概念：

1. 方差：刻画离散程度

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

总体方差：
$$s^2=\frac{1}{n}\sum_{i=1}^{n}(\theta_i-\bar{\theta})^2$$

样本方差：
$$s^2=\frac{1}{n\text{-}1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

7.4 估计量的性能

2. 均方误差（MSE）与均方根误差（RMSE）：刻画估计量与被估计量（即真值）之间的偏差

$$Mse(\hat{\theta}) = E\left[\left(\hat{\theta}-\theta\right)^2\right]$$

掌握！

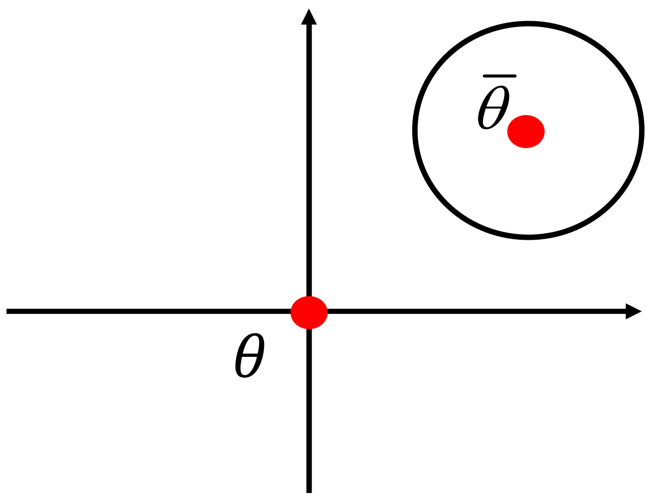
Classical: $$mse(\hat{\theta}) = \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f(\mathbf{z} \mid \theta) d\mathbf{z}$$

Bayesian: $$Bmse(\hat{\theta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 f(\mathbf{z}, \theta) d\mathbf{z} d\theta$$

【交互】：首先为什么是 $\mathbf{z},\theta$的二维积分，为什么又能化为一维？

7.4 估计量的性能

如果用作图的方式来表现方差与均方误差之间的区别：



对于无偏估计，

方差=均方差

3. 均方值(MES)与均方根值（RMES）

$$E(X^2)$$

$$\frac{1}{n}\sum_{i=1}^{n} x_i^{2}$$

7.4 估计量的性能

1 性能指标

无偏性、有效性、一致性

（1）无偏性

若估计量的均值等于被估计量的均值，则称此估计是无偏的，即 $E[\hat{\theta}(z)] = E[\theta]$

$$E[\hat{\theta}(z)] = \begin{cases} \theta & \theta\text{为确定性参量（非随机参量）} \\ E[\theta] & \theta\text{为随机参量} \end{cases}$$

7.4 估计量的性能

当观测是多次测量时，这时估计量可表示为$\hat{\theta}=\hat{\theta}(\boldsymbol{z}_N)$，其中观测矢量为$\boldsymbol{z}_N=[z_1,z_2,...,z_N]^T$，一般说来，观测数据越多，估计的性能越好，对于有偏估计，如果

$$\lim_{N\to\infty} E[\hat{\theta}(\boldsymbol{z}_N)]=\begin{cases}\theta & \theta\text{为未知常量}\\ E(\theta) & \theta\text{为随机变量}\end{cases}$$

则称$\hat{\theta}(\boldsymbol{z}_N)$为渐近无偏估计。

7.4 估计量的性能

（2）有效性

对于无偏估计量，不同的估计一般具有不同的均方误差，均方误差越小越好。

若估计量的均方误差能达到最小方差，则此估计称为有效估计（ee）。

对于非随机量，有： $Var(\hat{\theta}) = E\{[\hat{\theta} - \theta]^2\}$

对于随机量，有： $Var(\hat{\theta}) = E\{[\hat{\theta} - E(\hat{\theta})]^2\}$

（3）一致性

即对于任意小数ε，若有：

$$\lim_{N\to\infty} P\left(\left|\hat{\theta}(z_1, z_2, \cdots, z_N) - \theta\right| < \varepsilon\right) = 1$$

则估计量 $\hat{\theta}$ 为一致估计量。

若满足 $\lim_{N\to\infty} E\left\{\left[\theta - \hat{\theta}(\mathbf{z}_N)\right]^2\right\} = 0$

则称 $\hat{\theta}$ 为均方一致估计量。

7.4 估计量的性能

例：高斯白噪声中未知电平的估计

$$z_i = A + v_i \qquad i = 1, 2, \ldots, N \qquad A \sim U(-A_0, A_0)$$

$\hat{A}_{map} = \begin{cases} -A_0 & \bar{z} < -A_0 \\ \bar{z} & A_0 \le \bar{z} \le A_0 \\ A_0 & \bar{z} > A_0 \end{cases}$	$\hat{A}_{ml} = \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$
$E(\hat{A}_{map}) \ne A$	$E(\hat{A}_{ml}) = E\left(\frac{1}{N}\sum_{i=1}^{N} z_i\right) = A$
$Mse(\hat{A}_{ml}) > Mse(\hat{A}_{map})$	

先验信息的利用，将有利于提高估计的性能

7.4 估计量的性能

2、无偏估计量的性能边界

（1）非随机参量

假定满足正则条件

$$E\left\{\frac{\partial \ln f(\mathbf{z};\theta)}{\partial \theta}\right\}=0$$

$$Var(\hat{\theta})=E\{[\theta-\hat{\theta}]^2\}\geq I(\theta)^{-1}$$

克拉美-罗限

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

等号成立的条件：

$$\frac{\partial \ln f(\mathbf{z};\theta)}{\partial \theta}=k(\theta)(\hat{\theta}-\theta)$$

7.4 估计量的性能

【证明】设$\hat{\theta}$是无偏估计，则有$E\left(\hat{\theta}-\theta\right)=0$，按照定义展开，即

$$E\left(\hat{\theta}-\theta\right)=\int_{-\infty}^{\infty}\left(\hat{\theta}-\theta\right)f\left(z|\theta\right)dz=0$$

对θ求导，得到

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}\left(\hat{\theta}-\theta\right)f\left(z|\theta\right)dz=\int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\left[\left(\hat{\theta}-\theta\right)f(z|\theta)\right]dz=0,$$

进一步有 $$-\int_{-\infty}^{\infty}f(z|\theta)\,dz+\int_{-\infty}^{\infty}\frac{\partial f(z|\theta)}{\partial\theta}\left(\hat{\theta}-\theta\right)\mathrm{d}z=0$$

由微分法则$\left(lng(x)\right)'=\frac{g'(x)}{g(x)}$，以及$\int_{-\infty}^{\infty}f(z|\theta)\,dz=1$，上式可以写为

$$\int_{-\infty}^{\infty}\frac{\partial lnf(z|\theta)}{\partial\theta}f(z|\theta)\left(\hat{\theta}-\theta\right)\mathrm{d}z=1$$

7.4 估计量的性能

也可表示成

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(z|\theta)}{\partial \theta} \sqrt{f(z|\theta)}\sqrt{f(z|\theta)}\left(\hat{\theta}-\theta\right) \mathrm{d}z=1$$

利用柯西–许瓦兹不等式

$$\left[\int_{-\infty}^{\infty} f(x)g(x)dx\right]^2 \le \int_{-\infty}^{\infty} f^2(x)dx \int_{-\infty}^{\infty} g^2(x)dx$$

可以有不等式成立：

$$1 \le \underbrace{\int_{-\infty}^{\infty} \left[\frac{\partial \ln f(z|\theta)}{\partial \theta}\right]^2 f(z|\theta)}_{E\left\{\left[\frac{\partial \ln f(z|\theta)}{\partial \theta}\right]^2\right\} = I(\theta)} \times \underbrace{\int_{-\infty}^{\infty} \left(\hat{\theta}-\theta\right)^2 f(z|\theta)\, \mathrm{d}z}_{\mathrm{Var}(\hat{\theta})}$$

7.4 估计量的性能

由柯西-许瓦兹不等式“=”成立的条件，即

$$f(x) = kg(x)$$

可知，CRLB达到的条件，即等号成立的条件是

$$\frac{\partial lnf(z|\theta)}{\partial\theta} = k(\theta)\left(\hat{\theta} - \theta\right)$$

【证毕】

此外，有

$$E\left\{\left[\frac{\partial lnf(z|\theta)}{\partial\theta}\right]^2\right\}=-E\left\{\frac{\partial^2 lnf(z|\theta)}{\partial\theta^2}\right\}$$

【证明】

$$\int_{\{Z\}} f(\mathbf{z}\,|\,\theta)d\mathbf{z} = 1 \xRightarrow{\text{第一次求导}} \int_{\{Z\}} \frac{\partial f(\mathbf{z}\,|\,\theta)}{\partial\theta} d\mathbf{z} = \int_{\{Z\}} \frac{\partial \ln f(\mathbf{z}\,|\,\theta)}{\partial\theta} f(\mathbf{z}\,|\,\theta)d\mathbf{z} = 0$$

$$\xRightarrow{\text{第二次求导}} \int_{\{Z\}} \frac{\partial^2 \ln f(\mathbf{z}\,|\,\theta)}{\partial\theta^2} f(\mathbf{z}\,|\,\theta)d\mathbf{z} + \int_{\{Z\}} \left[\frac{\partial \ln f(\mathbf{z}\,|\,\theta)}{\partial\theta}\right]^2 f(\mathbf{z}\,|\,\theta)d\mathbf{z} = 0$$

$$\Rightarrow E\left\{\left[\frac{\partial \ln f(\mathbf{z}\,|\,\theta)}{\partial\theta}\right]^2 \middle|\, \theta\right\} = -E\left[\frac{\partial^2 \ln f(\mathbf{z}\,|\,\theta)}{\partial\theta^2} \middle|\, \theta\right]$$

【证毕】

7.4 估计量的性能

几点说明：

（1） 定理成立的条件

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}(\hat{\theta}-\theta)f(z|\theta)dz \stackrel{?}{=} \int_{-\infty}^{\infty}\frac{\partial}{\partial\theta}\left[(\hat{\theta}-\theta)f(z|\theta)\right]dz=0$$

- 定理成立的条件是求导和积分可交换
- 如果积分限与被估计量θ有关，则求导和积分就不能交换。
- CRLB定理不成立的情况：概率密度非零的区间与被估计量有关，如：概率密度为(0,θ)上均匀分布，而θ是被估计量。

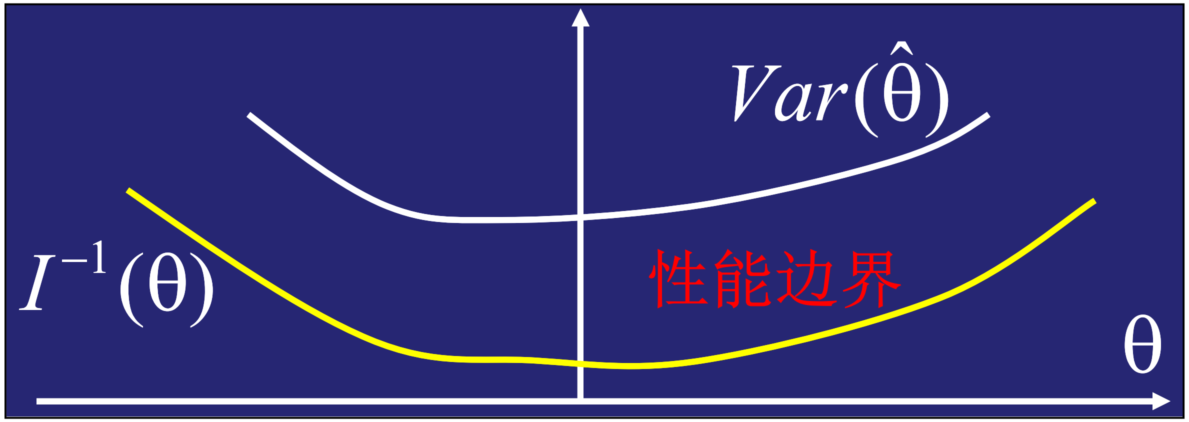
7.4 估计量的性能

正则条件保证了上述要求，即

$$0=E\left\{\frac{\partial \ln f(\mathbf{z}|\theta)}{\partial\theta}\right\}=\int_{-\infty}^{\infty}\frac{\partial \ln f(\mathbf{z}|\theta)}{\partial\theta}f(\mathbf{z}|\theta)d\mathbf{z}=\int_{-\infty}^{\infty}\frac{\partial f(\mathbf{z}|\theta)}{\partial\theta}d\mathbf{z}$$

且 $\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(\mathbf{z}|\theta)d\mathbf{z}=0$ 所以 $\int_{-\infty}^{\infty}\frac{\partial f(\mathbf{z}/\theta)}{\partial\theta}d\mathbf{z}=\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty}f(\mathbf{z}/\theta)d\mathbf{z}$

（2） I(θ)称为Fisher信息，I越大，越有可能得到好的估计。



7.4 估计量的性能

（3）如果有效估计量存在，则该有效估计量一定是最大似然估计。

因为如果有效估计量存在，则表明满足

$$\frac{\partial \ln f(z|\theta)}{\partial \theta} = k(\theta)\left(\hat{\theta} - \theta\right)$$

而最大似然估计满足 $\left.\frac{\partial \ln f(z|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ml}} = 0$

所以有 $k(\hat{\theta}_{ml})\left(\hat{\theta} - \hat{\theta}_{ml}\right)=0$ 即$\hat{\theta} = \hat{\theta}_{ml}$

（4）如果有效估计量不存在，则最大似然估计的方差不一定是最小的。

7.4 估计量的性能

例**7.8**　高斯白噪声中的DC电平。例子7.4中的ML估计是否达到了CRLB,估计方差是多少?

$$z_i = A + v_i \qquad i = 1, 2, \ldots, N \qquad v_i \sim N(0, \sigma^2)$$

$$\hat{A}_{ml} = \bar{z} = \frac{1}{N}\sum_{i=1}^{N} z_i$$

$$E[\hat{A}_{ml}] = E\left[\frac{1}{N}\sum_{i=1}^{N} z_i\right] = \frac{1}{N}\sum_{i=1}^{N} E[z_i] = A$$　无偏估计

$$\frac{\partial \ln f(\mathbf{z}; A)}{\partial A} = \frac{N}{\sigma^2}\left(\frac{1}{N}\sum_{i=1}^{N} z_i - A\right) = \frac{N}{\sigma^2}(\hat{A}_{ml} - A)$$　达到了CRLB

$$Var(\hat{A}_{ml}) = -\frac{1}{I} = \frac{\sigma^2}{N}$$

7.4 估计量的性能

2、无偏估计量的性能边界

（2）随机参量

任何无偏估计量的均方误差满足

克拉美-罗限

$$Mse(\hat{\theta}) = E\{[\theta-\hat{\theta}]^2\} \geq J^{-1}$$

$$J = E\left\{\left[\frac{\partial \ln f(z,\theta)}{\partial \theta}\right]^2\right\} = -E\left\{\frac{\partial^2 \ln f(z,\theta)}{\partial \theta^2}\right\}$$

等号成立的条件：$$\frac{\partial \ln f(\mathbf{z},\theta)}{\partial \theta} = k\cdot(\hat{\theta}-\theta)$$ （式7.4.29）

注意，k不是$\mathbf{z}$或者θ的函数

7.4 估计量的性能

如果有某个无偏估计达到CRLB，那么该估计必定是**最大后验概率**估计．而最小均方估计的均方误差也是最小的，所以这时最小均方估计与最大后验概率估计等价。

例7.10 高斯白噪声中的直流电平估计-高斯先验分布。设有N次独立观测$z_i=A+v_i$，$i=1,2,\cdots.N$，其中$v\sim N(0,\ \sigma^2)$，$A\sim N(\mu_A,\sigma_A^2)$，求A的估计的CRLB。

【解】 令$\mathbf{z}=[z_1,z_2,z_3,\ldots z_N]$ 求后验概率

SUN YAT-SEN UNIVERSITY 中山大學

7.4 估计量的性能

$$f(\mathbf{z},A)=f(\mathbf{z}\,|\,A)f(A)$$

$$=\frac{1}{(2\pi\sigma^2)^{N/2}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i-A)^2\right]\frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\left[-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right]$$

$$\ln f(\mathbf{z},A)=-\frac{N}{2}\ln(2\pi\sigma^2)-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(z_i-A)^2-\frac{1}{2}\ln(2\pi\sigma_A^2)-\frac{1}{2\sigma_A^2}(A-\mu_A)^2$$

$$\frac{\partial\ln f(\mathbf{z},A)}{\partial A}=\frac{1}{\sigma^2}\sum_{i=1}^{N}(z_i-A)-\frac{1}{\sigma_A^2}(A-\mu_A)$$

$$\overset{\text{参考7.3}}{=}\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}\right)\left(\frac{\dfrac{N}{\sigma^2}\bar{z}+\dfrac{\mu_A}{\sigma_A^2}}{\dfrac{N}{\sigma^2}+\dfrac{1}{\sigma_A^2}}-A\right)=\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}\right)\left(\hat{A}_{map}-A\right)$$

(7. 4. 34)

7.4 估计量的性能

其中$\hat{A}_{map}$是例7.3中求得的A的最大后验概率估计，它可表示为

$$\hat{A}_{map}=\frac{\dfrac{N}{\sigma^2}\bar{z}+\dfrac{\mu_A}{\sigma_A^2}}{\dfrac{N}{\sigma^2}+\dfrac{1}{\sigma_A^2}}$$

它的均值为：

$$E(\hat{A}_{map})=E\left(\frac{\dfrac{N}{\sigma^2}\bar{z}+\dfrac{\mu_A}{\sigma_A^2}}{\dfrac{N}{\sigma^2}+\dfrac{1}{\sigma_A^2}}\right)=\mu_A=E(A)$$

可见$\hat{A}_{map}$是无偏估计。由(7.4.34)式可以看出，$\hat{A}_{map}$满足(7.4.29)式，因此，它的均方误差等于CRLB。又

7.4 估计量的性能

$$\frac{\partial^2 \ln f(\mathbf{z}, A)}{\partial A^2} = -\frac{N}{\sigma^2} - \frac{1}{\sigma_A^2}$$

$$Mse(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = \left(\frac{N}{\sigma^2} + \frac{1}{\sigma_A^2}\right)^{-1} = \frac{\sigma_A^2 \sigma^2}{N\sigma_A^2 + \sigma^2}$$

总结：有效估计量是建立在无偏的基础上的，因为克拉美一罗不等式取等号的条件，都是在任意无偏估计量基础上导出的。