第四章 参数估计理论

刘志文

http://isip.bit.edu.cn
Email: zwliu@bit.edu.cn(提交作业)

第一节 引言

问题提出：

信号处理：　从观测中提取有用信息

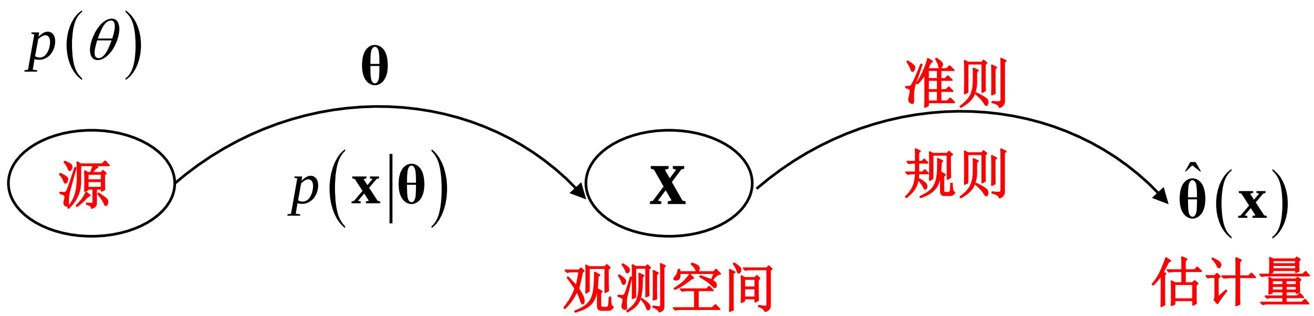
检测 + 确定

- 参数估计 √
- 波形估计
- 特征估计——谱估计
  - 功率谱、高阶谱估计（时间）
  - 空间谱——阵列信号处理（角度谱）

$\forall \mathbf{x} \Rightarrow$ 有用信息

第一节 引言

参数估计的图形化表示：



准则：

$N-P$	✗		
MAP	✓	$p(\boldsymbol{\theta}\|\mathbf{x})$	
$\min P_E$	✗		
$\min \bar{C}$	✓	$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$	$\min \bar{C} \rightarrow \min R \sim$ 风险
$\operatorname{minmax} \bar{C}$	✓		

既然其它两种准则都是贝叶斯准则的特例，下面先来讨论贝叶斯估计。

第二节 贝叶斯估计

贝叶斯估计的定义：

$$\min R = E[C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})] \longrightarrow \hat{\boldsymbol{\theta}}_B(\mathbf{x}) = ?$$

风险：

$$R = \int_{-\infty}^{+\infty}\int_{\Theta} C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) p(\mathbf{x},\boldsymbol{\theta})\, d\mathbf{x} d\boldsymbol{\theta}$$

$$= \int_{-\infty}^{+\infty}\int_{\Theta} C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x} d\boldsymbol{\theta}$$

$$= \int_{-\infty}^{+\infty} [\int_{\Theta} C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta}] \cdot p(\mathbf{x})\, d\mathbf{x}$$

$p(\boldsymbol{\theta}|\mathbf{x})$为后验概率密度函数

$$p(\boldsymbol{\theta}|\mathbf{x}) = \frac{p(\mathbf{x}|\boldsymbol{\theta})\, p(\boldsymbol{\theta})}{p(\mathbf{x})}$$

条件风险：

$$\forall \mathbf{x}, \min R(\hat{\boldsymbol{\theta}}|\mathbf{x}) = \int_{\Theta} C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x})\, d\boldsymbol{\theta} > 0$$

$$\Updownarrow$$

$$\min R(\hat{\boldsymbol{\theta}}) = \int_{-\infty}^{+\infty} R(\hat{\boldsymbol{\theta}}|\mathbf{x})\, p(\mathbf{x})\, d\mathbf{x}$$

$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}) = ?$

第二节 贝叶斯估计

三种典型代价函数：

$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})$ 满足：

① $C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})\ge 0$

② $C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=C(\boldsymbol{\theta},\hat{\boldsymbol{\theta}})$， $C(\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|)=C(-\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|)$

③ $C(\|\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\|)\ge C(\|\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\|)$，若 $\|\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\|\ge\|\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\|$

（1）平均误差代价函数：

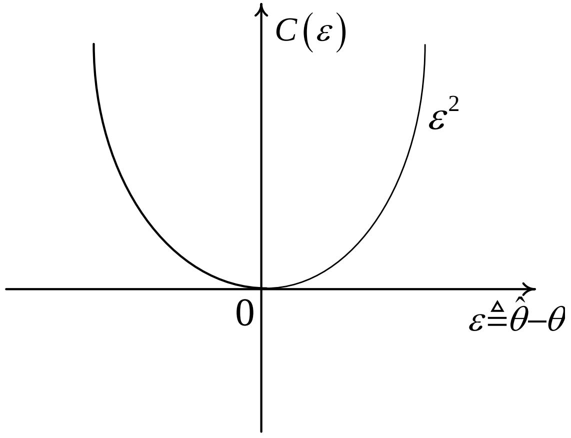
标量：

$$C(\hat{\theta},\theta)=(\hat{\theta}-\theta)^2=\varepsilon^2$$

矢量：

$$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^{\mathrm{T}}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$$

$$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})^{\mathrm{T}}$$



第二节　贝叶斯估计

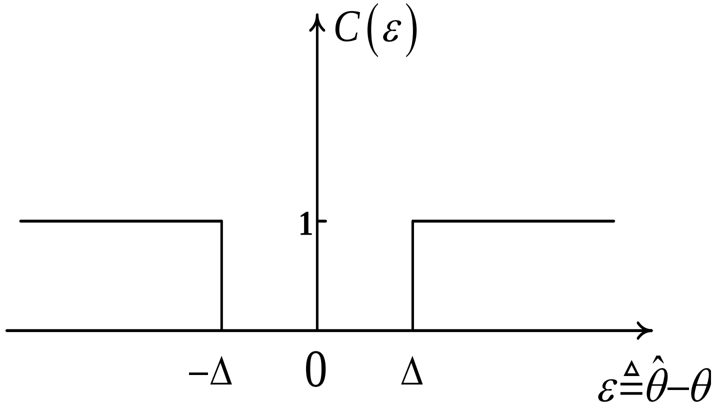
（2）绝对误差代价函数：

标量：$$C(\hat{\theta},\theta)=\left|\hat{\theta}-\theta\right|=|\varepsilon|$$

矢量：$$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=\left|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right|=\sum_{k=1}^{N}\left|\hat{\boldsymbol{\theta}}_k-\boldsymbol{\theta}_k\right|$$

（3）均匀误差代价函数：

$$C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})=\begin{cases}0, & \left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|\le\Delta\\ 1, & \left\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right\|>\Delta\end{cases}$$



三种代价函数对应的贝叶斯估计：

第二节　贝叶斯估计

2.1：最小均方误差估计：$\underline{\hat{\theta}_{ms}}$

$$\hat{\boldsymbol{\theta}}_{ms}=\arg\{\min_{\hat{\boldsymbol{\theta}}} E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]\}$$

标量：

$$\min R\left(\hat{\theta}|\mathbf{x}\right)=\int_{-\infty}^{+\infty}\left(\hat{\theta}-\theta\right)^{2}p\left(\theta|\mathbf{x}\right)d\theta$$

$$\frac{\partial R\left(\hat{\theta}|\mathbf{x}\right)}{\partial\hat{\theta}}=\int_{-\infty}^{+\infty}2\left(\hat{\theta}-\theta\right)p\left(\theta|\mathbf{x}\right)d\theta=0$$

$$\Downarrow$$

$$\hat{\theta}\int_{-\infty}^{+\infty}p\left(\theta|\mathbf{x}\right)d\theta-\int_{-\infty}^{+\infty}\theta p\left(\theta|\mathbf{x}\right)d\theta=0$$

$$\because\int_{-\infty}^{+\infty}p\left(\theta|\mathbf{x}\right)d\theta=1,$$

$$\therefore\hat{\theta}_{B}=\int_{-\infty}^{+\infty}\theta p\left(\theta|\mathbf{x}\right)d\theta=E\left(\theta|\mathbf{x}\right)$$ 后验均值

$$\therefore\min R(\theta|\mathbf{x})\Leftrightarrow\min R=E\left[C\left(\hat{\theta},\theta\right)\right]=E\left[\left(\hat{\theta}-\theta\right)^{2}\right]$$

第二节　贝叶斯估计

可见，在平方误差代价函数时，

$$\hat{\theta}_{ms} = E(\theta|\mathbf{x})$$

因此也叫后验均值估计

矢量：

$$R(\hat{\boldsymbol{\theta}}|\mathbf{x}) = \int_{-\infty}^{+\infty} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$$

$$\frac{\partial R(\hat{\boldsymbol{\theta}}|\mathbf{x})}{\partial \hat{\boldsymbol{\theta}}} = \int_{-\infty}^{+\infty} \frac{\partial (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^{\mathrm{T}} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})}{\partial \hat{\boldsymbol{\theta}}} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$$

$$= \int_{-\infty}^{+\infty} 2(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta} = \mathbf{0}$$

$$\Rightarrow \int_{-\infty}^{+\infty} 2(\hat{\theta}_i - \theta_i) p(\theta_i|\mathbf{x}) d\theta_i = 0, i = 1, 2, \ldots, M$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_{ms} = E(\boldsymbol{\theta}|\mathbf{x}) = \int_{-\infty}^{+\infty} \boldsymbol{\theta} p(\boldsymbol{\theta}|\mathbf{x}) d\boldsymbol{\theta}$$

后验均值

2.2：条件中值估计：$\hat{\boldsymbol{\theta}}_{med}$

$$\hat{\boldsymbol{\theta}}_{med} = \arg\{\min_{\hat{\boldsymbol{\theta}}} E\left(\left|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right|\right)\}$$

标量：

$$\min_{\theta} R\left(\hat{\theta}|\mathbf{x}\right) = \int_{-\infty}^{+\infty} \left|\hat{\theta} - \theta\right| p\left(\theta|\mathbf{x}\right) d\theta$$

$$\frac{\partial R\left(\hat{\theta}|\mathbf{x}\right)}{\partial\hat{\theta}} = \int_{-\infty}^{+\infty} \frac{\partial\left|\hat{\theta} - \theta\right|}{\partial\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta = 0$$

$$\Rightarrow \int_{-\infty}^{\hat{\theta}} \frac{\partial\left(\hat{\theta} - \theta\right)}{\partial\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta + \int_{\hat{\theta}}^{+\infty} \frac{\partial\left(\theta - \hat{\theta}\right)}{\partial\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta$$

$$= \int_{-\infty}^{\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta - \int_{\hat{\theta}}^{+\infty} p\left(\theta|\mathbf{x}\right) d\theta = 0$$

$$\therefore \int_{-\infty}^{\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta = \int_{\hat{\theta}}^{+\infty} p\left(\theta|\mathbf{x}\right) d\theta$$

条件中值

$$\therefore \hat{\theta}_{med} : \int_{-\infty}^{\hat{\theta}} p\left(\theta|\mathbf{x}\right) d\theta = \int_{\hat{\theta}}^{+\infty} p\left(\theta|\mathbf{x}\right) d\theta$$

矢量： $\hat{\boldsymbol{\theta}}_{med} = (\hat{\theta}_{med}^{(1)}, \hat{\theta}_{med}^{(2)}, \ldots, \hat{\theta}_{med}^{(M)})^T, \hat{\theta}_{med}^{(k)} : \int_{-\infty}^{\hat{\theta}_{med}^{(k)}} p\left(\theta|\mathbf{x}\right) d\theta = \int_{\hat{\theta}_{med}^{(k)}}^{+\infty} p\left(\theta|\mathbf{x}\right) d\theta$

第二节 贝叶斯估计

2.3：最大后验概率估计：$\hat{\boldsymbol{\theta}}_{map}$

$$\hat{\boldsymbol{\theta}}_{map} = \arg\left\{\min_{\hat{\boldsymbol{\theta}}} R\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{x}\right)\right\}$$

$$R\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{x}\right) = \int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|>\Delta} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)d\boldsymbol{\theta} = 1 - \int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\le\Delta} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)d\boldsymbol{\theta}$$

$$\min R\left(\hat{\boldsymbol{\theta}}\middle|\mathbf{x}\right) \Leftrightarrow \max \int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\le\Delta} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)d\boldsymbol{\theta} \quad \sim \quad \Delta$$

$$\Delta \to 0,\quad \boldsymbol{\theta} \to \hat{\boldsymbol{\theta}} \quad \Rightarrow \int_{\|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\|\le\Delta} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)d\boldsymbol{\theta} = p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)d\boldsymbol{\theta}$$

$$\therefore \quad \min_{\hat{\boldsymbol{\theta}}} R(\hat{\boldsymbol{\theta}}\,|\,\mathbf{x}) \Leftrightarrow \max_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right) \quad \Rightarrow \quad \hat{\boldsymbol{\theta}}_{map} = \arg\left\{\max_{\boldsymbol{\theta}} p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)\right\}$$

$$\Updownarrow$$

$$\left.\frac{\partial p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0} \qquad \left.\frac{\partial \ln p\left(\boldsymbol{\theta}\middle|\mathbf{x}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0}$$

$$\left.\frac{\partial p\left(\boldsymbol{\theta},\mathbf{x}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0} \qquad \left.\frac{\partial \ln p\left(\boldsymbol{\theta},\mathbf{x}\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0}$$

后验方程

第二节 贝叶斯估计

小结： $\hat{\theta}_B$: $\min R(\hat{\boldsymbol{\theta}})=\mathrm{E}[C(\hat{\boldsymbol{\theta}},\boldsymbol{\theta})]$ ~ $p(\boldsymbol{\theta}), p(\mathbf{x}|\boldsymbol{\theta})$，$C$（$\hat{\boldsymbol{\theta}}$，$\boldsymbol{\theta}$）

$$\left.\begin{array}{l} \hat{\theta}_{ms}: \min E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right] \\ \hat{\theta}_{med}: \min E\left(\left|\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right|\right) \\ \hat{\theta}_{map}: \max p\left(\boldsymbol{\theta}|\mathbf{x}\right) \end{array}\right\} \sim p\left(\boldsymbol{\theta}|\mathbf{x}\right) \Leftrightarrow p\left(\boldsymbol{\theta}\right), p\left(\mathbf{x}|\boldsymbol{\theta}\right)$$

注意：$p\left(\mathbf{x}\right)$对于构造估计量$\hat{\boldsymbol{\theta}}$无任何信息！！！

问题：若$p\left(\mathbf{x}|\boldsymbol{\theta}\right)$，$C$（$\hat{\boldsymbol{\theta}}$，$\boldsymbol{\theta}$）已知，但是$p\left(\boldsymbol{\theta}\right)$未知时，$\hat{\boldsymbol{\theta}}$=？

第三节 极小极大估计

$$\hat{\boldsymbol{\theta}}_{mm}=\arg\{\min_{\hat{\boldsymbol{\theta}}} \max_{p(\boldsymbol{\theta})} R\left(\hat{\boldsymbol{\theta}}, p\right)=E\left[C\left(\hat{\boldsymbol{\theta}},\boldsymbol{\theta}\right)\right]\}$$

第四节 最大似然估计

仅$p(\mathbf{x}|\boldsymbol{\theta})$ 已知时

$$\forall x_1 \Rightarrow \left.\begin{matrix} p(x_1|\theta_1) \\ p(x_1|\theta_2) \\ \vdots \\ p(x_1|\theta_M) \end{matrix}\right\} \max_{\boldsymbol{\theta}} p(x_1|\boldsymbol{\theta}) \Rightarrow \hat{\boldsymbol{\theta}}_{ml}$$

$$\hat{\boldsymbol{\theta}}_{ml} = \arg\left\{\max_{\boldsymbol{\theta}} p(\mathbf{x}|\boldsymbol{\theta})\right\}$$

$\hat{\boldsymbol{\theta}}_{ml}$的构造：

$$(1)\quad \left.\frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}$$

$$(2)\quad \left.\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}$$

似然方程

第四节 最大似然估计

与最大后验概率估计的比较

$$\hat{\boldsymbol{\theta}}_{ml} = \arg\left\{\max_{\boldsymbol{\theta}} p(\mathbf{x}|\boldsymbol{\theta})\right\} \sim p(\mathbf{x}|\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}}_{map} = \arg\left\{\max_{\boldsymbol{\theta}} p(\boldsymbol{\theta}|\mathbf{x})\right\} \sim p(\boldsymbol{\theta}|\mathbf{x}) \Leftrightarrow p(\mathbf{x}|\boldsymbol{\theta}), p(\boldsymbol{\theta})$$

$\hat{\boldsymbol{\theta}}_{ml}$的构造：

(1) $$\left.\frac{\partial p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}$$

似然方程

(2) $$\left.\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}$$

$\hat{\boldsymbol{\theta}}_{map}$的构造：

$$\left.\frac{\partial p(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0} \qquad \left.\frac{\partial \ln p(\boldsymbol{\theta}|\mathbf{x})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0}$$

后验方程

$$\left.\frac{\partial p(\boldsymbol{\theta}, \mathbf{x})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0} \qquad \left.\frac{\partial \ln p(\boldsymbol{\theta}, \mathbf{x})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}} = \mathbf{0}$$

第四节 最大似然估计

与最大后验概率估计的比较(续)

$$p(\boldsymbol{\theta},\mathbf{x})=p(\mathbf{x}|\boldsymbol{\theta})\,p(\boldsymbol{\theta})$$

$$\ln p(\boldsymbol{\theta},\mathbf{x})=\ln p(\mathbf{x}|\boldsymbol{\theta})+\ln p(\boldsymbol{\theta})$$

后验方程

$$\left.\frac{\partial \ln p(\boldsymbol{\theta},\mathbf{x})}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}}=\left[\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}+\frac{\partial \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}}=\mathbf{0}$$

$$\Downarrow\leftarrow\quad \frac{\partial \ln p(\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}}=\mathbf{0}$$

? ? ?

似然方程

$$\frac{\partial \ln p(\mathbf{x}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}}=\mathbf{0}\Rightarrow\hat{\boldsymbol{\theta}}_{map}=\hat{\boldsymbol{\theta}}_{ml}$$

X ! ! !

第五节 估计量的评价/性质

估计量 $\hat{\boldsymbol{\theta}}$：$\forall \mathbf{x} \to \hat{\boldsymbol{\theta}}(\mathbf{x}) \sim \mathbf{x}$ 注意：不是$\boldsymbol{\theta}$的函数！！！

估计误差：$\boldsymbol{\varepsilon}=\hat{\boldsymbol{\theta}}-\boldsymbol{\theta} \sim \mathbf{x},\boldsymbol{\theta}$ $\to$随机矢量$\to$如何描述？

$$\Downarrow$$

$E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right]$ 估计误差均值

$E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$ 估计的均方误差

$$\Downarrow$$

估计量具备什么样的特性才是我们需要的？

$$\Downarrow$$

显然，估计误差越小越好

估计量要尽可能多地利用观测中的有用信息

估计量的特性要随着可利用观测的增加其性能不断提高

第五节　估计量的评价/性质

5.1：无偏性

如果 $E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=\mathbf{0}$，即 $E\left(\hat{\boldsymbol{\theta}}\right)=E\left(\boldsymbol{\theta}\right)$，

则称 $\hat{\boldsymbol{\theta}}$ 为 $\boldsymbol{\theta}$ 的无偏估计

如果 $\lim\limits_{N\to\infty} E\left(\hat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right)=0$ 则称渐近无偏估计

5.2：有效性

$\forall\hat{\boldsymbol{\theta}}_1,\hat{\boldsymbol{\theta}}_2 : E\left(\hat{\boldsymbol{\theta}}_1\right)=E\left(\hat{\boldsymbol{\theta}}_2\right)=E\left(\boldsymbol{\theta}\right)$，即均为无偏估计

如果$E\left[\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_1-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]<E\left[\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_2-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$，

则称$\hat{\boldsymbol{\theta}}_1$比$\hat{\boldsymbol{\theta}}_2$更有效。

那么，是否存在一个最有效的估计呢？

第五节 估计量的评价/性质

$$\min_{\hat{\boldsymbol{\theta}}} E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right] \quad \rightarrow \quad \hat{\boldsymbol{\theta}}_E \quad ?$$

$$s.t. \quad E\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=\mathbf{0}$$

$$\Downarrow$$

$$\forall \hat{\boldsymbol{\theta}}_E, \hat{\boldsymbol{\theta}}, \quad E\left[\hat{\boldsymbol{\theta}}_E\right]=E\left[\hat{\boldsymbol{\theta}}\right]=E\left[\boldsymbol{\theta}\right]$$

如果 $E\left[\left(\hat{\boldsymbol{\theta}}_E-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_E-\boldsymbol{\theta}\right)^{\mathrm{T}}\right] \leq E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$,

则称 $\hat{\boldsymbol{\theta}}_E$ 为有效估计。

有效估计 $\hat{\boldsymbol{\theta}}_E$： 具有最小均方误差的无偏估计！

第五节 估计量的评价/性质

5.3：充分性

是指 $\hat{\boldsymbol{\theta}}$ 已经充分利用了 $\mathbf{x}$ 中有关 $\boldsymbol{\theta}$ 的信息。

数学上如何表示呢？

如果 $p\left(\mathbf{x}\middle|\boldsymbol{\theta}\right)=g\left(\hat{\boldsymbol{\theta}}\middle|\boldsymbol{\theta}\right)h\left(\mathbf{x}\right)$，则称估计量$\hat{\boldsymbol{\theta}}$为充分估计量。

5.4：一致性

$$\forall\varepsilon>0,\lim_{N\to\infty}P\left\{\left\|\hat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right\|<\varepsilon\right\}=1 \text{ 或 } \lim_{N\to\infty}E\left[\left\|\hat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}\right\|^2\right]=0$$

则称 $\hat{\boldsymbol{\theta}}_N$ 为 $\boldsymbol{\theta}$ 的一致估计量。

第五节 估计量的评价/性质

5.5：估计误差下限--Cramer-Rao界（CRB）

① 待估计量为非随机矢量

$$\forall\hat{\boldsymbol{\theta}},\ E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right]=0,\text{有}\ E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]\geq\mathbf{F}^{-1}$$

$$\mathbf{F}\triangleq E\left[\frac{\partial\ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}\frac{\partial\ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\right]=-E\left[\frac{\partial^2\ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^{\mathrm{T}}}\right]$$

C-R不等式

C-R不等式取等号的充要条件：

$$\frac{\partial\ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial\boldsymbol{\theta}}=\mathbf{K}(\boldsymbol{\theta})\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\ \rightarrow\ \hat{\boldsymbol{\theta}}_E\text{: 有效估计}$$

可见，如果 $\hat{\boldsymbol{\theta}}_E$存在，则 $\hat{\boldsymbol{\theta}}_E=\hat{\boldsymbol{\theta}}_{ml}$ 最大似然估计！

两边积分

$$\ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)=\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{K}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\Rightarrow p\left(\mathbf{x}|\boldsymbol{\theta}\right)=c\left(\mathbf{x}\right)e^{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{K}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)}\ \rightarrow\text{充分估计}$$

标量情况：$E\left[\left(\theta-\theta\right)^2\right]\geq\left\{E\left[\left(\frac{\partial\ln p\left(\mathbf{x}|\theta\right)}{\partial\theta}\right)^2\right]\right\}^{-1}=\left\{-E\left[\frac{\partial^2\ln p\left(\mathbf{x}|\theta\right)}{\partial\theta^2}\right]\right\}^{-1}$

② $\boldsymbol{\theta}$ 为随机量

标量： $\forall\hat{\theta}, E\left(\hat{\theta}\right)=E\left(\theta\right), E\left[\left(\hat{\theta}-\theta\right)\right]=0$

$$E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^2\right]\geq E\left\{\left[\left(\frac{\partial\ln p\left(\mathbf{x}|\theta\right)}{\partial\theta}\right)^2\right]+E\left[\left(\frac{\partial\ln p\left(\theta\right)}{\partial\theta}\right)^2\right]\right\}^{-1}$$

上述不等式取等号的充要条件：

$$\frac{\partial\ln p\left(\theta|\mathbf{x}\right)}{\partial\theta}=k\left(\hat{\theta}-\theta\right)$$

两边积分

$$\ln p\left(\theta|\mathbf{x}\right)=\int k\left(\hat{\theta}-\theta\right)d\theta+C$$

$$p\left(\theta|\mathbf{x}\right)=C'e^{k\left(\hat{\theta}-\theta\right)^2}$$ ——高斯分布

可见，如果 $\hat{\theta}_E$ 存在，则 $\hat{\theta}_E=\hat{\theta}_{map}=\hat{\theta}_{ms}=\hat{\theta}_{med}$

第五节 估计量的评价/性质

矢量：

$$\forall \hat{\boldsymbol{\theta}}, E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right]=0$$

$$E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]\geq \mathbf{F}_t^{-1}$$

$$\mathbf{F}_d \triangleq E\left[\frac{\partial \ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\frac{\partial \ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]\triangleq -E\left[\frac{\partial^2 \ln p\left(\mathbf{x}|\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]$$

$$\mathbf{F}_P \triangleq E\left[\frac{\partial \ln p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}}\frac{\partial \ln p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]= -E\left[\frac{\partial^2 \ln p\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]$$

上式取等号的充要条件：

$$\frac{\partial \ln p\left(\boldsymbol{\theta}|\mathbf{x}\right)}{\partial \boldsymbol{\theta}}=\mathbf{K}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$$

两边积分

$$\ln p\left(\boldsymbol{\theta}\,|\,\mathbf{x}\right)=\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{K}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right) \Rightarrow p\left(\boldsymbol{\theta}\,|\,\mathbf{x}\right)=c\left(\mathbf{x}\right)e^{\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\mathbf{K}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)} \rightarrow \text{高斯分布}$$

可见，如果 $\hat{\boldsymbol{\theta}}_E$ 存在，则 $\hat{\boldsymbol{\theta}}_E=\hat{\boldsymbol{\theta}}_{map}=\hat{\boldsymbol{\theta}}_{ms}=\hat{\boldsymbol{\theta}}_{med}$

第六节 线性最小均方误差估计

$\hat{\boldsymbol{\theta}}_{LMS}$ 的定义：

$$\min_{\hat{\boldsymbol{\theta}}} E\left[\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$$

$$s.t.\ \hat{\boldsymbol{\theta}}=\mathbf{Bx}+\mathbf{a}$$

$$\Updownarrow$$

$$\min_{\mathbf{B},\mathbf{a}} E\left[\left(\mathbf{Bx}+\mathbf{a}-\boldsymbol{\theta}\right)\left(\mathbf{Bx}+\mathbf{a}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$$

$$\begin{aligned}
J\left(\mathbf{B},\mathbf{a}\right)=&E\left[\left(\mathbf{Bx}+\mathbf{a}-\boldsymbol{\theta}\right)\left(\mathbf{Bx}+\mathbf{a}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]\\
=&E\left[\left(\mathbf{Bx}-\mathbf{B}E\left(\mathbf{x}\right)+\mathbf{B}E\left(\mathbf{x}\right)+\mathbf{a}-\boldsymbol{\theta}+E\left(\boldsymbol{\theta}\right)-E\left(\boldsymbol{\theta}\right)\right)\left(\ldots\right)^{\mathrm{T}}\right]\\
=&E\left[\left(\mathbf{B}\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)-\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)+\mathbf{B}E\left(\mathbf{x}\right)-E\left(\boldsymbol{\theta}\right)+\mathbf{a}\right)\left(\ldots\right)^{\mathrm{T}}\right]\\
=&E[\Big(\mathbf{B}\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}-\mathbf{B}\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)^{\mathrm{T}}+\mathbf{B}\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)\Big)\mathbf{b}\\
&-\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}+\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)^{\mathrm{T}}-\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)\mathbf{b}^{\mathrm{T}}\\
&+\mathbf{b}\left(\mathbf{x}-E\left(\mathbf{x}\right)\right)^{\mathrm{T}}\mathbf{B}^{\mathrm{T}}-\mathbf{b}\left(\boldsymbol{\theta}-E\left(\boldsymbol{\theta}\right)\right)+\mathbf{bb}^{\mathrm{T}}\\
=&\mathbf{B}Var\left(\mathbf{x}\right)\mathbf{B}^{\mathrm{T}}-\mathbf{B}\operatorname{cov}\left(\mathbf{x},\boldsymbol{\theta}\right)-\operatorname{cov}\left(\boldsymbol{\theta},\mathbf{x}\right)\mathbf{B}^{\mathrm{T}}+Var\left(\boldsymbol{\theta}\right)+\mathbf{bb}^{\mathrm{T}}
\end{aligned}$$

其中，$\mathbf{b}=\mathbf{B}E\left(\mathbf{x}\right)+\mathbf{a}-E\left(\boldsymbol{\theta}\right)$

第六节 线性最小均方误差估计

$$\min_{\mathbf{B},\mathbf{a}} J(\mathbf{B},\mathbf{a}) = \mathbf{B}Var(\mathbf{x})\mathbf{B}^{\mathrm{T}} - \mathbf{B}\operatorname{cov}(\mathbf{x},\boldsymbol{\theta}) - \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})\mathbf{B}^{\mathrm{T}} + Var(\boldsymbol{\theta}) + \mathbf{b}\mathbf{b}^{\mathrm{T}}$$

$$= \left(\mathbf{B} - \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\right)Var(\mathbf{x})\left(\mathbf{B} - \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\right)^{\mathrm{T}}$$

$$+ Var(\boldsymbol{\theta}) - \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\operatorname{cov}(\mathbf{x},\boldsymbol{\theta}) + \mathbf{b}\mathbf{b}^{\mathrm{T}}$$

$$\because \quad \mathbf{b}\mathbf{b}^{\mathrm{T}} \geq 0 \quad \Leftrightarrow \mathbf{v}^{\mathrm{T}}\mathbf{b}\mathbf{b}^{\mathrm{T}}\mathbf{v} = \left(\mathbf{b}^{\mathrm{T}}\mathbf{v}\right)^{\mathrm{T}}\left(\mathbf{b}^{\mathrm{T}}\mathbf{v}\right) \geq 0$$

$$\therefore \quad \mathbf{b} = \mathbf{0} \rightarrow \mathbf{B}^{*}E(\mathbf{x}) + \mathbf{a}^{*} - E(\boldsymbol{\theta}) = \mathbf{0}$$

$$\because \mathbf{D} > 0,\quad \mathbf{C}\mathbf{D}^{\mathrm{T}}\mathbf{C} \geq \mathbf{0} \quad \therefore \quad \rightarrow \mathbf{C} = \mathbf{0} \rightarrow \begin{cases} \mathbf{B}^{*} = \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x}) \\ \mathbf{a}^{*} = E(\boldsymbol{\theta}) - \mathbf{B}^{*}E(\mathbf{x}) \end{cases}$$

$$\Updownarrow$$

$$\hat{\boldsymbol{\theta}}_{LMS} = E(\boldsymbol{\theta}) + \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\left(\mathbf{x} - E(\mathbf{x})\right)$$

线性最小均方误差估计

第六节 线性最小均方误差估计

$\hat{\boldsymbol{\theta}}_{LMS}$的特性：

$$\hat{\boldsymbol{\theta}}_{LMS}=E(\boldsymbol{\theta})+\mathrm{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})(\mathbf{x}-E(\mathbf{x}))$$

（1）$E\left(\hat{\boldsymbol{\theta}}_{LMS}\right)=E(\boldsymbol{\theta})$ 无偏估计

（2）$E\left[\left(\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]=Var(\boldsymbol{\theta})-cov(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})cov(\mathbf{x},\boldsymbol{\theta})$

$$\Downarrow\leftarrow\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}=\underbrace{\mathrm{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})}_{\mathbf{B}^*}(\mathbf{x}-E(\mathbf{x}))-(\boldsymbol{\theta}-E(\boldsymbol{\theta}))$$

$$=E\left\{\left[\mathbf{B}^*(\mathbf{x}-E(\mathbf{x}))-(\boldsymbol{\theta}-E(\boldsymbol{\theta}))\right]\left[\mathbf{B}^*(\mathbf{x}-E(\mathbf{x}))-(\boldsymbol{\theta}-E(\boldsymbol{\theta}))\right]^{\mathrm{T}}\right\}$$

$$=\mathrm{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\mathrm{cov}(\mathbf{x},\boldsymbol{\theta})-\mathrm{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\mathrm{cov}(\mathbf{x},\boldsymbol{\theta})$$

$$-\mathrm{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\mathrm{cov}(\mathbf{x},\boldsymbol{\theta})+Var(\boldsymbol{\theta})$$

(3) $\underline{\hat{\theta}}_{LMS}$满足正交原理

$$E\left[\left(\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}\right)\mathbf{x}^{\mathrm{T}}\right]=\mathbf{0}$$

第六节 线性最小均方误差估计

$\hat{\boldsymbol{\theta}}_{LMS}$ 特性: (续)

(4) 线性观测方程/模型时的 $\hat{\boldsymbol{\theta}}_{LMS}$

$$\mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n}$$

$$\mathbf{n} \sim N(\mathbf{0}, \mathbf{R}), E(\mathbf{n}^{\mathrm{T}}) = \mathbf{0}, Var(\mathbf{n}) = \mathbf{R},$$

$$E(\boldsymbol{\theta}) = \boldsymbol{\mu}, Var(\boldsymbol{\theta}) = \mathbf{Q}, E(\boldsymbol{\theta}\mathbf{n}^{\mathrm{T}}) = \mathbf{0}$$

$$\hat{\boldsymbol{\theta}}_{LMS} = ?$$

$$E[\mathbf{x}] = E[\mathbf{H}\boldsymbol{\theta} + \mathbf{n}] = \mathbf{H}E[\boldsymbol{\theta}] = \mathbf{H}\boldsymbol{\mu}$$

$$Var(\mathbf{x}) = E\left[\left(\mathbf{x} - E(\mathbf{x})\right)\left(\mathbf{x} - E(\mathbf{x})\right)^{\mathrm{T}}\right]$$

$$= E\left[\left(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}) + \mathbf{n}\right)\left(\mathbf{H}(\boldsymbol{\theta} - \boldsymbol{\mu}) + \mathbf{n}\right)^{\mathrm{T}}\right]$$

$$= \mathbf{H}Var(\boldsymbol{\theta})\mathbf{H}^{\mathrm{T}} + \mathbf{R}$$

$$= \mathbf{HQH}^{\mathrm{T}} + \mathbf{R}$$

第六节 线性最小均方误差估计

$$\operatorname{cov}(\boldsymbol{\theta},\mathbf{x}) = E\left[\left(\boldsymbol{\theta}-E(\boldsymbol{\theta})\right)\left(\mathbf{x}-E(\mathbf{x})\right)^{\mathrm{T}}\right]$$

$$= E\left[\left(\boldsymbol{\theta}-E(\boldsymbol{\theta})\right)\left(\mathbf{H}\left(\boldsymbol{\theta}-E(\boldsymbol{\theta})\right)+\mathbf{n}\right)^{\mathrm{T}}\right]$$

$$= \mathbf{Q}\mathbf{H}^{\mathrm{T}}$$

$$\hat{\boldsymbol{\theta}}_{LMS} = E(\boldsymbol{\theta}) + \operatorname{cov}(\boldsymbol{\theta},\mathbf{x})Var^{-1}(\mathbf{x})\left(\mathbf{x}-E(\mathbf{x})\right)$$

$$= \boldsymbol{\mu} + \mathbf{Q}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1}\left(\mathbf{x}-\mathbf{H}^{\mathrm{T}}\boldsymbol{\mu}\right)$$

均方误差：

$$E\left[\left(\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{LMS}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$$

$$= \mathbf{Q} - \mathbf{Q}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}}+\mathbf{R}\right)^{-1}\mathbf{H}\mathbf{Q} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}+\mathbf{Q}^{-1}\right)^{-1}$$

$$< (\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}$$

第七节 最小二乘估计

7.1 线性最小二乘估计（$\hat{\underline{\theta}}_{LS}$）

$$\forall \mathbf{x} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \Rightarrow \hat{\boldsymbol{\theta}}_{Ls} = ? \qquad \underline{\mathbf{H}}\text{已知}$$

$$\min J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^{\mathrm{T}} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

$$\hat{\boldsymbol{\theta}}_{Ls} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{x}$$

$\hat{\boldsymbol{\theta}}_{Ls}$ 的特性：

① $E\left[\hat{\boldsymbol{\theta}}_{Ls}\right] = E[\boldsymbol{\theta}] = m$ 无偏性

$$\begin{aligned}
E\left[\hat{\boldsymbol{\theta}}_{Ls}\right] &= E\left[\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}}\underline{\mathbf{x}}\right] \\
&= E\left[\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}}(\mathbf{H}\boldsymbol{\theta} + \mathbf{n})\right] \\
&= E\left[\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{\theta} + \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{n}\right] \\
&= E[\boldsymbol{\theta}] + \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1} \mathbf{H}^{\mathrm{T}} E[\mathbf{n}] \\
&= E[\boldsymbol{\theta}]
\end{aligned}$$

$\hat{\boldsymbol{\theta}}_{LS}$特性：

② $E\left[\left(\hat{\boldsymbol{\theta}}_{Ls}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{Ls}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$

$$\hat{\boldsymbol{\theta}}_{Ls}-\boldsymbol{\theta}=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{x}-\boldsymbol{\theta}$$

$$=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{x}-\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{H}\boldsymbol{\theta}$$

$$=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\left(\mathbf{x}-\mathbf{H}\boldsymbol{\theta}\right)$$

$$=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{n}$$

$$\therefore E\left[\left(\hat{\boldsymbol{\theta}}_{Ls}-\boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{Ls}-\boldsymbol{\theta}\right)^{\mathrm{T}}\right]$$

$$=E\left[\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{n}\mathbf{n}^{\mathrm{T}}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\right]$$

$$=\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{R}\mathbf{H}\left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}$$

第七节 最小二乘估计

7.1 加权最小二乘估计（$\hat{\boldsymbol{\theta}}_{WLS}$）

$$\min J(\boldsymbol{\theta}) = (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^{\mathrm{T}} \mathbf{w} (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial (\mathbf{x} - \mathbf{H}\boldsymbol{\theta})^{\mathrm{T}}}{\partial \boldsymbol{\theta}} (\mathbf{w} + \mathbf{w}^{\mathrm{T}})(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})$$

$$= -\mathbf{H}^{\mathrm{T}} (2\mathbf{w})(\mathbf{x} - \mathbf{H}\boldsymbol{\theta})\Big|_{\hat{\boldsymbol{\theta}}_{WLS}} = \mathbf{0}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_{WLS} = (\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{H})^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{x}$$

$\hat{\boldsymbol{\theta}}_{WLS}$ 特性：

① $\hat{\boldsymbol{\theta}}_{WLS}$为 $\boldsymbol{\theta}$ 的无偏估计

$$E\left[\hat{\boldsymbol{\theta}}_{WLS}\right] = E\left[(\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{H})^{-1} \mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{x}\right] = E[\boldsymbol{\theta}]$$

$$\Rightarrow E\left[\hat{\boldsymbol{\theta}}_{WLS} - \boldsymbol{\theta}\right] = 0$$

$\hat{\boldsymbol{\theta}}_{WLS}$ 的特性：

$$\min J(\boldsymbol{\theta}) = (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^{\mathrm{T}} \mathbf{w} (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

$$\frac{\partial J(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{\partial (\mathbf{x}-\mathbf{H}\boldsymbol{\theta})^{\mathrm{T}}}{\partial \boldsymbol{\theta}} (\mathbf{w}+\mathbf{w}^{\mathrm{T}})(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})$$

$$= -\mathbf{H}^{\mathrm{T}}(2\mathbf{w})(\mathbf{x}-\mathbf{H}\boldsymbol{\theta})\Big|_{\hat{\boldsymbol{\theta}}_{WLS}} = \mathbf{0}$$

$$\Rightarrow \hat{\boldsymbol{\theta}}_{WLS} = (\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{x}$$

$$E\left[(\hat{\boldsymbol{\theta}}_{WLS}-\boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{WLS}-\boldsymbol{\theta})^{\mathrm{T}}\right]$$

$$= (\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{H})^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{R}\mathbf{w}\mathbf{H}(\mathbf{H}^{\mathrm{T}}\mathbf{w}\mathbf{H})$$

$$\geq (\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H})^{-1}$$

上式中，$\mathbf{w} = \mathbf{R}^{-1}$ 时取等号

$\hat{\boldsymbol{\theta}}_{WLS}$ 与 $\hat{\boldsymbol{\theta}}_{LMS}$ 的比较：

① 均为无偏估计；

② $$\hat{\boldsymbol{\theta}}_{LMS} = E[\theta] + \operatorname{cov}(\boldsymbol{\theta}, \mathbf{x}) Var^{-1}(\mathbf{x})[\mathbf{x} - E(\mathbf{x})]$$
$$= \mathbf{m} + \mathbf{Q}\mathbf{H}^{\mathrm{T}}\left(\mathbf{H}\mathbf{Q}\mathbf{H}^{\mathrm{T}} + \mathbf{R}\right)^{-1}(\mathbf{x} - \mathbf{H}\mathbf{m})$$

$$E\left[\left(\hat{\boldsymbol{\theta}}_{lms} - \boldsymbol{\theta}\right)\left(\hat{\boldsymbol{\theta}}_{lms} - \boldsymbol{\theta}\right)^{\mathrm{T}}\right]$$
$$= \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H} + \mathbf{Q}^{-1}\right)^{-1} \le \left(\mathbf{H}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{H}\right)^{-1}$$

$\therefore \hat{\boldsymbol{\theta}}_{LMS}$ 利用先验知识换取了较小的误差。