

第四章 参数估计理论

刘志文

信息与电子学院

<http://isip.bit.edu.cn>

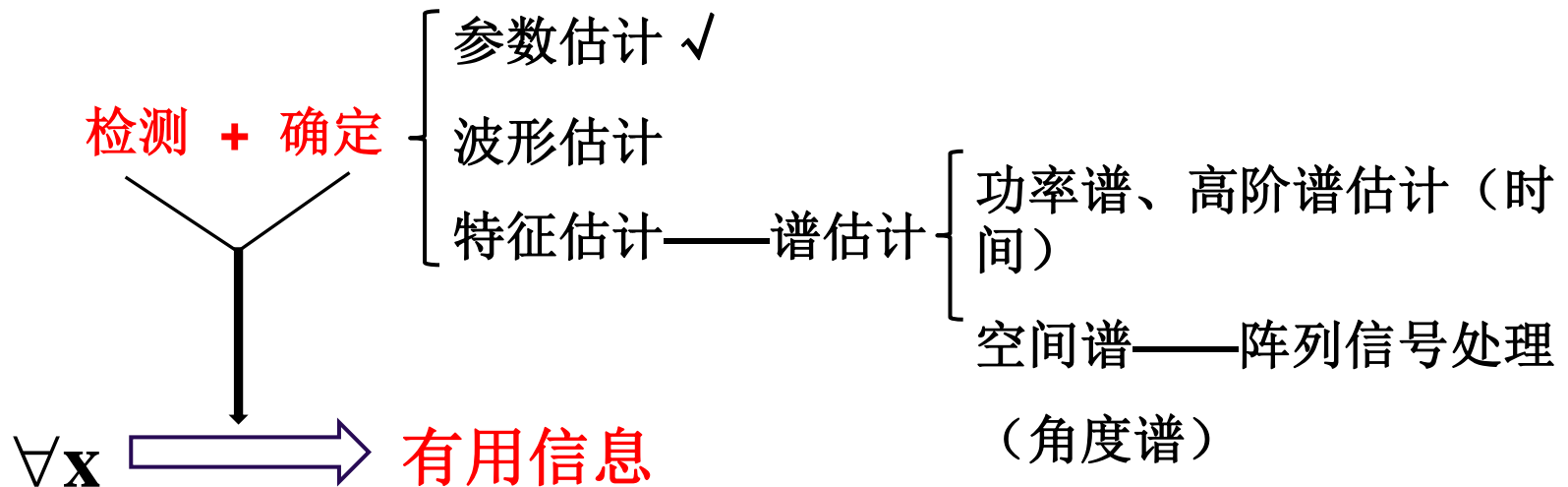
Email: zwliu@bit.edu.cn(提交作业)

ssp2014sss@163.com(下载作业题)

第一节 引言

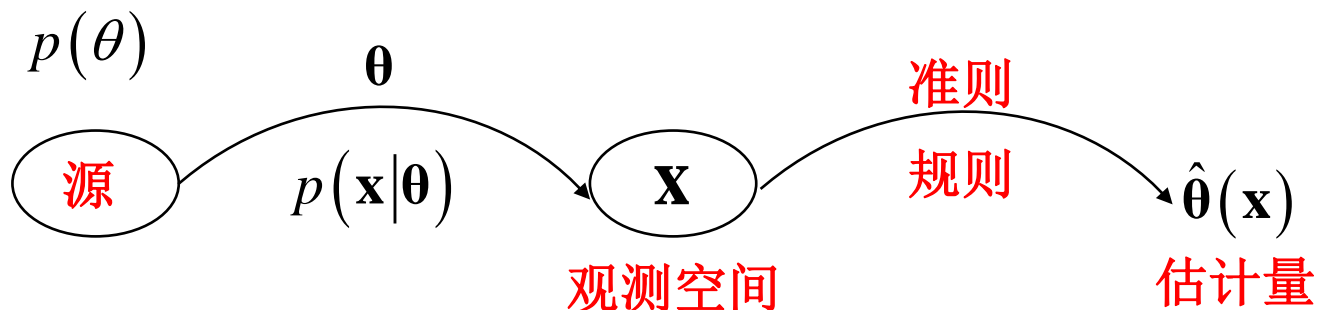
问题提出:

信号处理: 从观测中提取有用信息



第一节 引言

参数估计的图形化表示:



准则:

$N - P$	X		
MAP	✓	$p(\theta \mathbf{x})$	
$\min PE$	X		
$\min \bar{C}$	✓	$C(\hat{\theta}, \theta)$	$\min \bar{C} \rightarrow \min R \sim$ 风险
$\min\max \bar{C}$	✓		

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:

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

风险:

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

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

$p(\theta | \mathbf{x})$ 为后验概率密度函数

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

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

条件风险:

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

\Updownarrow

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

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

第二节 贝叶斯估计

三种典型代价函数:

$C(\hat{\theta}, \theta)$ 满足:

① $C(\hat{\theta}, \theta) \geq 0$

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

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

(1) 平均误差代价函数:

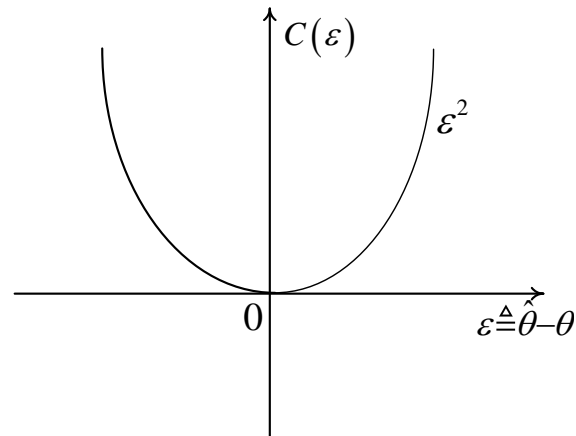
标量:

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

矢量:

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

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



第二节 贝叶斯估计

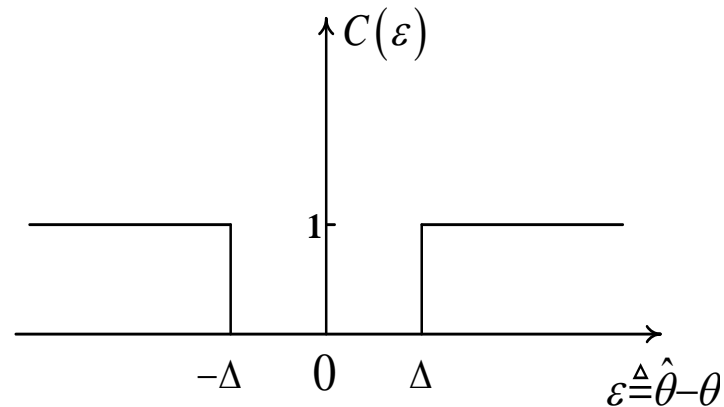
(2) 绝对误差代价函数:

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

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

(3) 均匀误差代价函数:

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



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: $\hat{\theta}_{ms}$

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

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

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

⇓

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

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

$$\therefore \hat{\theta}_B = \int_{-\infty}^{+\infty} \theta p(\theta|\mathbf{x}) d\theta = E(\theta|\mathbf{x}) \quad \text{后验均值}$$

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

第二节 贝叶斯估计

可见，在平方误差代价函数时，

$$\hat{\theta}_{ms} = E(\theta|\mathbf{x}) \quad \text{因此也叫后验均值估计}$$

矢量：

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

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

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

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

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

第二节 贝叶斯估计

2.2: 条件中值估计: $\hat{\theta}_{med}$

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

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

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

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

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

$$\therefore \int_{-\infty}^{\hat{\theta}} p(\theta|\mathbf{x}) d\theta = \int_{\hat{\theta}}^{+\infty} p(\theta|\mathbf{x}) d\theta \quad \text{条件中值}$$

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

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

第二节 贝叶斯估计

2.3: 最大后验概率估计: $\hat{\theta}_{map}$

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

$$R(\hat{\theta} | \mathbf{x}) = \int_{\|\hat{\theta} - \theta\| > \Delta} p(\theta | \mathbf{x}) d\theta = 1 - \int_{\|\hat{\theta} - \theta\| \leq \Delta} p(\theta | \mathbf{x}) d\theta$$

$$\min R(\hat{\theta} | \mathbf{x}) \Leftrightarrow \max \int_{\|\hat{\theta} - \theta\| \leq \Delta} p(\theta | \mathbf{x}) d\theta \sim \Delta$$

$$\Delta \rightarrow 0, \quad \theta \rightarrow \hat{\theta} \Rightarrow \int_{\|\hat{\theta} - \theta\| \leq \Delta} p(\theta | \mathbf{x}) d\theta = p(\theta | \mathbf{x}) d\theta$$

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

\Updownarrow

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

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

后验方程

第二节 贝叶斯估计

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

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

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

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

第三节 极小极大估计

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

第四节 最大似然估计

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

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

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

$\hat{\boldsymbol{\theta}}_{ml}$ 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:

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

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

后验方程

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

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

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

似然方程

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

第五节 估计量的评价/性质

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

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

\Downarrow

$$E[(\hat{\theta} - \theta)] \quad \text{估计误差均值}$$

$$E[(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T] \quad \text{估计的均方误差}$$

\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 $E(\hat{\theta} - \theta) = \mathbf{0}$, 即 $E(\hat{\theta}) = E(\theta)$,
则称 $\hat{\theta}$ 为 θ 的**无偏估计**

如果 $\lim_{N \rightarrow \infty} E(\hat{\theta}_N - \theta) = 0$ 则称**渐近无偏估计**

5.2: 有效性

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

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

则称 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 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

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

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

⇓

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

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

则称 $\hat{\theta}_E$ 为有效估计。

有效估计 $\hat{\theta}_E$: 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 $\hat{\theta}$ 已经充分利用了 \mathbf{x} 中有关 θ 的信息。

数学上如何表示呢？

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

5.4: 一致性

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

则称 $\hat{\theta}_N$ 为 θ 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

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

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

C-R不等式

C-R不等式取等号的充要条件:

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

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

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

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

第五节 估计量的评价/性质

② θ 为随机量

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

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

上述不等式取等号的充要条件:

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

两边积分

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

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

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:

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

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

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

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

上式取等号的充要条件:

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

两边积分

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

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

第六节 线性最小均方误差估计

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

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

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

\Downarrow

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

$$J(\mathbf{B}, \mathbf{a}) = E \left[(\mathbf{B}\mathbf{x} + \mathbf{a} - \boldsymbol{\theta})(\mathbf{B}\mathbf{x} + \mathbf{a} - \boldsymbol{\theta})^T \right]$$

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

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

$$= E \left[(\mathbf{B}(\mathbf{x} - E(\mathbf{x}))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{B}^T - \mathbf{B}(\mathbf{x} - E(\mathbf{x}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T + \mathbf{B}(\mathbf{x} - E(\mathbf{x})))\mathbf{b} \right.$$

$$\left. - (\boldsymbol{\theta} - E(\boldsymbol{\theta}))(\mathbf{x} - E(\mathbf{x}))^T \mathbf{B}^T + (\boldsymbol{\theta} - E(\boldsymbol{\theta}))(\boldsymbol{\theta} - E(\boldsymbol{\theta}))^T - (\boldsymbol{\theta} - E(\boldsymbol{\theta}))\mathbf{b}^T \right.$$

$$\left. + \mathbf{b}(\mathbf{x} - E(\mathbf{x}))^T \mathbf{B}^T - \mathbf{b}(\boldsymbol{\theta} - E(\boldsymbol{\theta})) + \mathbf{b}\mathbf{b}^T \right]$$

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

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

第六节 线性最小均方误差估计

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

$$= (\mathbf{B} - \text{cov}(\boldsymbol{\theta}, \mathbf{x}) \text{Var}^{-1}(\mathbf{x})) \text{Var}(\mathbf{x}) (\mathbf{B} - \text{cov}(\boldsymbol{\theta}, \mathbf{x}) \text{Var}^{-1}(\mathbf{x}))^T \\ + \text{Var}(\boldsymbol{\theta}) - \text{cov}(\boldsymbol{\theta}, \mathbf{x}) \text{Var}^{-1}(\mathbf{x}) \text{cov}(\mathbf{x}, \boldsymbol{\theta}) + \mathbf{b} \mathbf{b}^T$$

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

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

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



$$\hat{\boldsymbol{\theta}}_{LMS} = E(\boldsymbol{\theta}) + \text{cov}(\boldsymbol{\theta}, \mathbf{x}) \text{Var}^{-1}(\mathbf{x}) (\mathbf{x} - E(\mathbf{x})) \quad \text{线性最小均方误差估计}$$

第六节 线性最小均方误差估计

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$$= \mathbf{H}\mathbf{Q}\mathbf{H}^T + \mathbf{R}$$

第六节 线性最小均方误差估计

$$\begin{aligned}\text{cov}(\boldsymbol{\theta}, \mathbf{x}) &= E\left[(\boldsymbol{\theta} - E(\boldsymbol{\theta}))(\mathbf{x} - E(\mathbf{x}))^T\right] \\ &= E\left[(\boldsymbol{\theta} - E(\boldsymbol{\theta}))(\mathbf{H}(\boldsymbol{\theta} - E(\boldsymbol{\theta})) + \mathbf{n})^T\right] \\ &= \mathbf{QH}^T\end{aligned}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{LMS} &= E(\boldsymbol{\theta}) + \text{cov}(\boldsymbol{\theta}, \mathbf{x})\text{Var}^{-1}(\mathbf{x})(\mathbf{x} - E(\mathbf{x})) \\ &= \boldsymbol{\mu} + \mathbf{QH}^T(\mathbf{HQH}^T + \mathbf{R})^{-1}(\mathbf{x} - \mathbf{H}^T\boldsymbol{\mu})\end{aligned}$$

均方误差:

$$\begin{aligned}& E\left[(\hat{\boldsymbol{\theta}}_{LMS} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{LMS} - \boldsymbol{\theta})^T\right] \\ &= \mathbf{Q} - \mathbf{QH}^T(\mathbf{HQH}^T + \mathbf{R})^{-1}\mathbf{HQ} = (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H} + \mathbf{Q}^{-1})^{-1} \\ &< (\mathbf{H}^T\mathbf{R}^{-1}\mathbf{H})^{-1}\end{aligned}$$

第七节 最小二乘估计

7.1 线性最小二乘估计 ($\hat{\theta}_{LS}$)

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

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

$$\hat{\boldsymbol{\theta}}_{LS} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$$

$\hat{\boldsymbol{\theta}}_{LS}$ 的特性:

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

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

第七节 最小二乘估计

$\hat{\boldsymbol{\theta}}_{LS}$ 特性:

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

$$\begin{aligned} \hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta} &= \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{x} - \boldsymbol{\theta} \\ &= \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{x} - \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{H} \boldsymbol{\theta} \\ &= \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \left(\mathbf{x} - \mathbf{H} \boldsymbol{\theta} \right) \\ &= \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{n} \end{aligned}$$

$$\begin{aligned} \therefore E \left[\left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta} \right) \left(\hat{\boldsymbol{\theta}}_{LS} - \boldsymbol{\theta} \right)^T \right] \\ &= E \left[\left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{n} \mathbf{n}^T \mathbf{H} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \right] \\ &= \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{R} \mathbf{H} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \end{aligned}$$

第七节 最小二乘估计

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

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

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

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

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

$\hat{\boldsymbol{\theta}}_{WLS}$ 特性:

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

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

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

第七节 最小二乘估计

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

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

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

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

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

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

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

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

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

第七节 最小二乘估计

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

① 均为无偏估计;

$$\begin{aligned}\text{② } \hat{\boldsymbol{\theta}}_{LMS} &= E[\boldsymbol{\theta}] + \text{cov}(\boldsymbol{\theta}, \mathbf{x}) \text{Var}^{-1}(\mathbf{x}) [\mathbf{x} - E(\mathbf{x})] \\ &= \mathbf{m} + \mathbf{QH}^T (\mathbf{HQH}^T + \mathbf{R})^{-1} (\mathbf{x} - \mathbf{Hm})\end{aligned}$$

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

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