西电通院郑贱平

第三章：统计信号估计

3.1 问题描述

3.2 随机参量的**Bayes**估计

3.3 ML估计

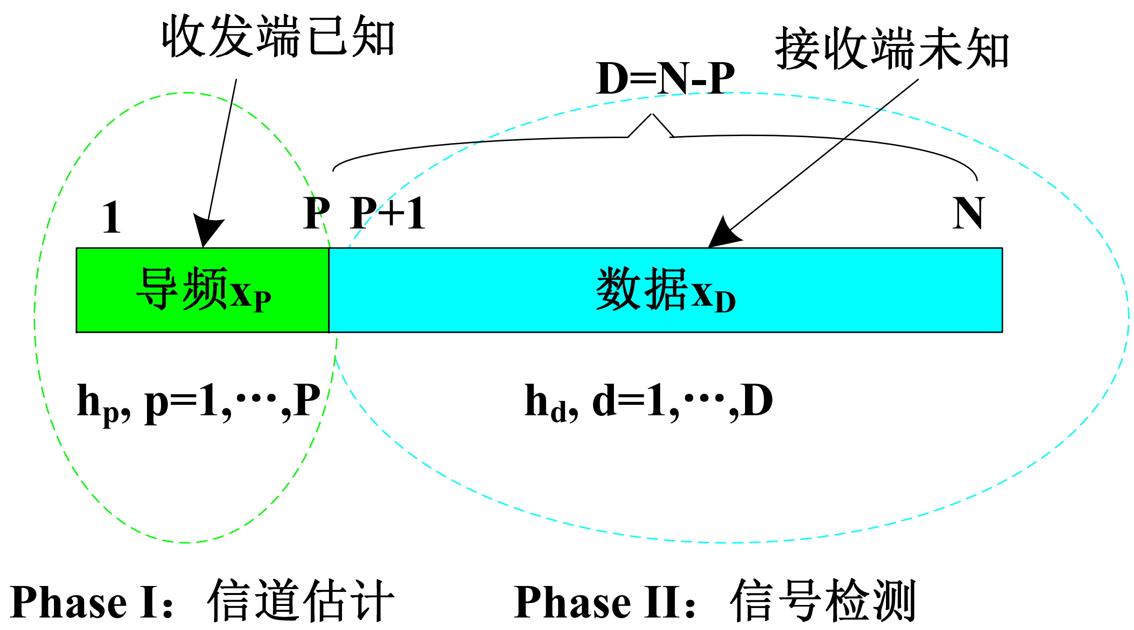
3.4 估计量的性质

3.5 线性最小均方误差估计

3.6 最小二乘估计

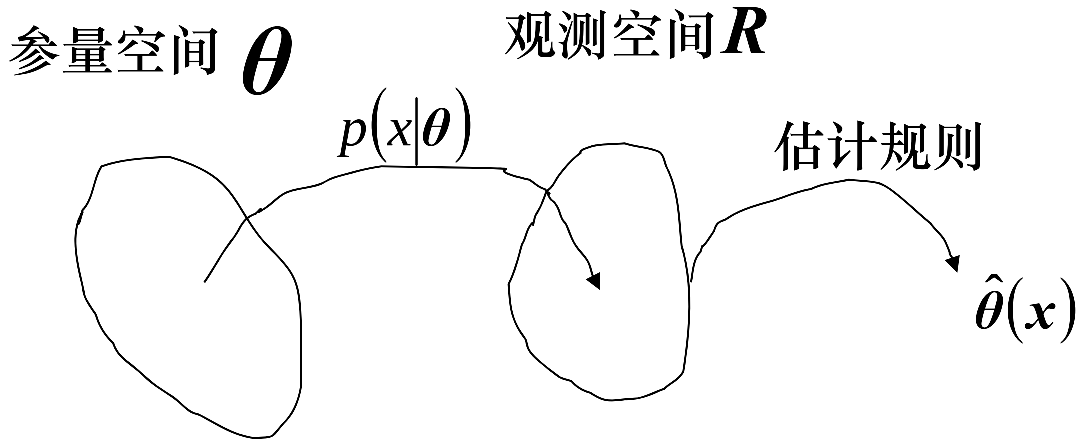
3.1 问题描述（信道估计为例）

- 数字通信数据帧结构



- 信道估计：根据y_P、x_P以及h_P的统计 信息，估计h_P，即：
  $(y_P, x_P, stat\_info(h_P)) \rightarrow h_P$ （如$y_P=h_Px_P+w$）
- 可行性：一般信道都是slowly time varying的（相干时间>>时延要求），因此$h_d \approx h_p$
- 其他估计问题：载波频率、相位、时延等

建模



➢**参量空间：** 需要接收端作出估计的参量集合

➢**观测空间：** 接收端收到的观测信号的集合

➢**概率映射：** 信源发送信号到接收端过程中，会有噪声的影响，观测信号中包含被估计矢量的信息，所以观测信号是以被估计矢量为参数的随机矢量，用 $p(x|\theta)$ 来描述。

建模

- **估计规则：** 利用被估计矢量的先验知识和观测信号的统计特性，根据指标要求，构造观测矢量的函数来定义估计量。

$$\hat{\theta}(\boldsymbol{x}) = g(\boldsymbol{x}) = g(x_1, x_2, \cdots, x_N)$$

估计量性能的评估

估计量的均值 $E\left[\hat{\boldsymbol{\theta}}(\mathbf{x})\right]$

估计量的均方误差 $E\left[\tilde{\theta}^2(\boldsymbol{x})\right] = E\left[\left(\theta - \hat{\theta}(\boldsymbol{x})\right)^2\right]$ $\qquad \tilde{\theta}(\boldsymbol{x}) = \theta - \hat{\theta}(\boldsymbol{x})$

本章的核心问题之一就是研究上述函数的构造方法，评估所构造估计量的优劣。

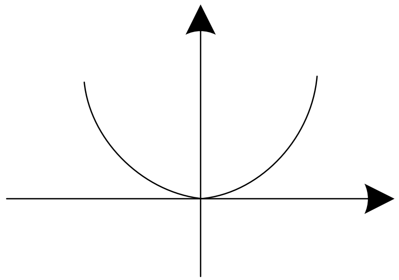
3.2 随机参量的贝叶斯估计

- 常用代价函数
- 贝叶斯估计的概念
- 最小均方误差估计
- 最大后验概率估计
- 条件中值估计
- 最佳估计的不变性

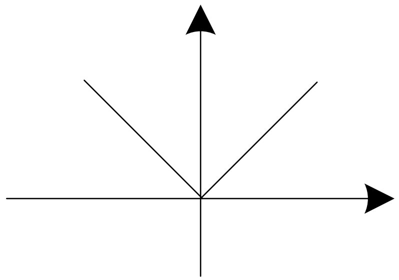
代价函数和贝叶斯估计

贝叶斯估计：使平均代价最小的一种估计准则。

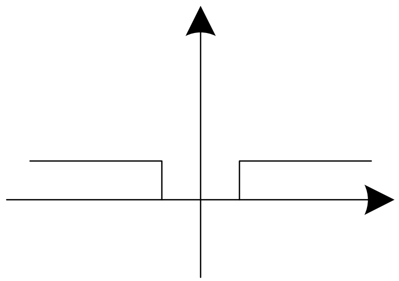
误差平方代价函数 $c(\tilde{\theta}) = (\theta - \hat{\theta})^2$



误差绝对值代价函数 $c(\tilde{\theta}) = |\theta - \hat{\theta}|$



均匀代价函数 $c(\tilde{\theta}) = c(\theta - \hat{\theta}) = \begin{cases} 1, & |\tilde{\theta}| \geq \Delta/2 \\ 0, & |\tilde{\theta}| < \Delta/2 \end{cases}$



代价函数的基本特性：非负性和 $\tilde{\theta} = 0$ 时的最小性。

平均代价

设被估计的单随机变量的先验概率密度函数为 $p(\theta)$

易知代价函数 $c(\tilde{\theta})$是随机参量θ和观测矢量$\boldsymbol{x}$的函数

平均代价C为

$$C=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c(\tilde{\theta})p(\boldsymbol{x},\theta)d\boldsymbol{x}d\theta$$

在 $p(\theta)$ 给定，选定代价函数的条件下，使平均代价最小的估计称为贝叶斯估计。

平均代价

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

$$C=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c(\tilde{\theta})p(\boldsymbol{x},\theta)d\boldsymbol{x}d\theta$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x}d\theta$$

$$=\int_{-\infty}^{\infty}p(\boldsymbol{x})\left[\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})d\theta\right]d\boldsymbol{x}$$

$\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})d\theta$ 是非负值，

因此使平均代价最小，就等价于使

$$C(\hat{\theta}|\boldsymbol{x})=\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})d\theta$$ 条件平均代价

最小。

Relation with cost in M-ary Detection

$$C=\lim_{M\to\infty}\sum_{j=0}^{M-1}\sum_{i=0}^{M-1}c_{ij}P\left(H_i\mid H_j\right)P\left(H_j\right)$$

$$=\lim_{M\to\infty}\sum_{j}\lim_{M\to\infty}\sum_{i}c_{ij}P\left(H_i\mid H_j\right)P\left(H_j\right)$$

$$=\lim_{M\to\infty}\sum_{j}\lim_{M\to\infty}\sum_{i}c_{ij}\int_{R_i}P\left(x\mid H_j\right)dxP\left(H_j\right)$$

$$=\lim_{M\to\infty}\sum_{j}\int_{R}c\left(j,x\right)P\left(x\mid H_j\right)dxP\left(H_j\right)$$

$$=\int_{\{\theta\}}\int_{R}c\left(\theta,x\right)P\left(x\mid\theta\right)dxp\left(\theta\right)d\theta$$

$$=\int_{\{\theta\}}\int_{R}c\left(\theta,x\right)P\left(x,\theta\right)dxd\theta$$

估计：参数连续取值；检测：参数取自有限个离散点集合。

检测与估计的联系

- 检测：参量的状态是有限的（M-ary检测）
- 估计：参量的状态是连续的（比如实数域，复数域）
- 当M→∞时，检测就变成了估计
- 用检测做估计：复杂度太高，不合适
- 用估计做检测：可以，实际上经常这样用
  - 比如，在衰落信道y=hx+w的信号检测中，经常对信号先进行估计得到x的估计值x1（复数域上的任意值），然后将其量化到信号星座上的某个点，即检测值x2。
- 无线通信中，有时候并不严格区分检测与估计

最小均方误差估计

选定的代价函数为

$$c(\tilde{\theta})=(\theta-\hat{\theta})^2$$

$$C(\hat{\theta}|\boldsymbol{x})=\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})d\theta=\int_{-\infty}^{\infty}(\theta-\hat{\theta})^2\,p(\theta|\boldsymbol{x})d\theta$$

求解方法

使条件平均代价最小的一个必要条件是对上式中 $\hat{\theta}$ 求偏导

令偏导为零来求得最佳的估计量 $\hat{\theta}$

最小均方误差估计

$$\frac{\partial C(\hat{\theta}|\boldsymbol{x})}{\partial \hat{\theta}} = \frac{\partial}{\partial \hat{\theta}} \int_{-\infty}^{\infty} (\theta - \hat{\theta})^2 p(\theta|\boldsymbol{x}) d\theta$$

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

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

$$\int_{-\infty}^{\infty} p(\theta|\boldsymbol{x}) d\theta = 1$$

$$\hat{\theta}_{mse} = \int_{-\infty}^{\infty} \theta p(\theta|\boldsymbol{x}) d\theta$$

最小均方误差估计

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

注： 1.最小均方误差估计的估计量实际是条件均值

$$\hat{\theta}_{mse} = \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{x}) d\theta = E[\theta|\mathbf{x}]$$

2.最小均方误差估计的条件平均代价实际是条件方差

$$C_{mse}(\hat{\theta}|\mathbf{x}) = \int_{-\infty}^{\infty} (\theta - \hat{\theta}_{mse})^2 p(\theta|\mathbf{x}) d\theta = \int_{-\infty}^{\infty} (\theta - E[\theta|\mathbf{x}])^2 p(\theta|\mathbf{x}) d\theta$$

3.最小均方误差估计量的另一种形式

$$\hat{\theta}_{mse} = \int_{-\infty}^{\infty} \theta p(\theta|\mathbf{x}) d\theta = \int_{-\infty}^{\infty} \theta \frac{p(\theta, \mathbf{x})}{p(\mathbf{x})} d\theta$$

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

最大后验估计

选定的代价函数为 $c(\tilde{\theta}) = c(\theta - \hat{\theta}) = \begin{cases} 1, & |\tilde{\theta}| \geq \Delta/2 \\ 0, & |\tilde{\theta}| < \Delta/2 \end{cases}$

$$C(\hat{\theta}|\boldsymbol{x}) = \int_{-\infty}^{\infty} c(\tilde{\theta}) p(\theta|\boldsymbol{x}) d\theta$$

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

$$= 1 - \int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} p(\theta|\mathbf{x}) d\theta$$

$$\int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} p(\theta|\boldsymbol{x}) d\theta \approx \Delta p(\hat{\theta}|\boldsymbol{x})$$

使条件平均代价最小，应该使 $\int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} p(\theta|\boldsymbol{x}) d\theta$ 取到最大值

当 Δ 很小时，为保证上式最大，应当选择估计量 $\hat{\theta}$ ，

使它处于后验概率密度函数 $p(\theta|\boldsymbol{x})$ 最大值的位置。

最大后验估计

根据上述分析，得到最大后验概率估计量为

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

两种等价形式

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

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

条件中值估计

选定的代价函数为 $c(\tilde{\theta})=|\theta-\hat{\theta}|$

$$C(\hat{\theta}|\boldsymbol{x})=\int_{-\infty}^{\infty}c(\tilde{\theta})p(\theta|\boldsymbol{x})d\theta=\int_{-\infty}^{\infty}|\theta-\hat{\theta}|p(\theta|\boldsymbol{x})d\theta$$

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

求解方法

使条件平均代价最小的一个必要条件是对上式中 $\hat{\theta}$ 求偏导

令偏导为零来求得最佳的估计量 $\hat{\theta}$

条件中值估计

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

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

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

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

$$\Downarrow$$

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

例1

研究在加性噪声中单随机参量θ 的估计问题。

观测方程为 $x_k = \theta + n_k, \quad k = 1, 2, \cdots, N$

其中n_k是均值为零，方差为 σ_n^2 的独立同分布高斯随机噪声

被估计量θ 是均值为零，方差为σ_θ^2 高斯随机变量

求 θ 的贝叶斯估计量(最小均方误差、最大后验和条件中值)

解： 最大后验估计

根据最大后验估计准则，估计量为满足以下方程的解，即

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

由题设，可知，给定 θ 条件下，观测信号x_k是均值为θ，方差为σ_n^2的高斯随机变量

$$p(\theta)=\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)$$

$$p(x_k|\theta)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N} p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$p(\theta)=\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)\quad p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$\frac{\partial\ln p(\boldsymbol{x}|\theta)}{\partial\theta}+\frac{\partial\ln p(\theta)}{\partial\theta}=\frac{\partial\left(-\sum_{k=1}^{N}\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)}{\partial\theta}+\frac{\partial\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)}{\partial\theta}$$

$$=\sum_{k=1}^{N}\frac{2(x_k-\theta)}{2\sigma_n^2}-\frac{2\theta}{2\sigma_\theta^2}$$

所以最大后验估计量为满足以下方程的解

$$\left.\sum_{k=1}^{N}\frac{2(x_k-\theta)}{2\sigma_n^2}-\frac{2\theta}{2\sigma_\theta^2}\right|_{\theta=\hat{\theta}_{map}}=0\qquad\left.\sum_{k=1}^{N}\frac{x_k}{\sigma_n^2}-\left(\frac{N}{\sigma_n^2}+\frac{1}{\sigma_\theta^2}\right)\theta\right|_{\theta=\hat{\theta}_{map}}=0$$

$$\hat{\theta}_{map}=\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)$$

估计量的均方误差为

$$E\left[\left(\theta-\hat{\theta}_{map}\right)^2\right]=E\left[\left(\theta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)\right)^2\right]$$

$$=E\left[\left(\theta-\frac{\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}\sum_{k=1}^{N}\left(\theta+n_k\right)\right)^2\right]$$

$$=E\left[\left(\frac{\left(N\sigma_\theta^2+\sigma_n^2\right)\theta-N\sigma_\theta^2\theta}{N\sigma_\theta^2+\sigma_n^2}+\frac{\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}\sum_{k=1}^{N}n_k\right)^2\right]$$

$$=\frac{\sigma_n^4\sigma_\theta^2}{\left(N\sigma_\theta^2+\sigma_n^2\right)^2}+\frac{N\sigma_n^2\sigma_\theta^4}{\left(N\sigma_\theta^2+\sigma_n^2\right)^2}=\frac{\sigma_n^2\sigma_\theta^2\left(N\sigma_\theta^2+\sigma_n^2\right)}{\left(N\sigma_\theta^2+\sigma_n^2\right)^2}=\frac{\sigma_n^2\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}$$

最小均方误差估计

根据最小均方误差估计准则，估计量为

$$\theta_{mse}=\int_{-\infty}^{\infty}\theta p(\theta|\boldsymbol{x})d\theta$$

由题设，可知，给定 θ 条件下，观测信号x_k是均值为θ，方差为σ_n^2 的高斯随机变量

$$p(\theta)=\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)$$

$$p(x_k|\theta)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

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

$K_1(\boldsymbol{x})$

$$= \frac{1}{p(\boldsymbol{x})}\left(\frac{1}{2\pi\sigma_n^2}\right)^{N/2}\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left(-\frac{\theta^2}{2\sigma_\theta^2}-\sum_{k=1}^{N}\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$= K_1(\boldsymbol{x})\exp\left(-\frac{1}{2}\left(\frac{\theta^2}{\sigma_\theta^2}+\sum_{k=1}^{N}\frac{x_k^2-2x_k\theta+\theta^2}{\sigma_n^2}\right)\right)$$

$$= K_1(\boldsymbol{x})\exp\left(-\frac{\sum_{k=1}^{N}x_k^2}{2\sigma_n^2}\right)\exp\left(-\frac{1}{2}\left(\frac{\sigma_n^2+N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\theta^2-2\theta\sum_{k=1}^{N}\frac{x_k}{\sigma_n^2}\right)\right)$$

$K_2(\boldsymbol{x})$

$$=K_2(\boldsymbol{x})\exp\left(-\frac{1}{2}\left(\frac{\sigma_n^2+N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\theta^2-2\theta\sum_{k=1}^{N}\frac{x_k}{\sigma_n^2}\right)\right)$$

$$=K_2(\boldsymbol{x})\exp\left(-\frac{1}{2}\frac{\sigma_n^2+N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\left(\theta^2-2\frac{\sigma_n^2\sigma_\theta^2}{\sigma_n^2+N\sigma_\theta^2}\theta\sum_{k=1}^{N}\frac{x_k}{\sigma_n^2}\right)\right)$$

$$=K_3(\boldsymbol{x})\exp\left(-\frac{1}{2}\frac{\sigma_n^2+N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\left(\theta-\frac{\sigma_\theta^2}{\sigma_n^2+N\sigma_\theta^2}\sum_{k=1}^{N}x_k\right)^2\right)$$

$$K_3(\boldsymbol{x})=K_2(\boldsymbol{x})\exp\left(\frac{N\sigma_\theta^2+\sigma_n^2}{2\sigma_n^2\sigma_\theta^2}\left[\frac{\sigma_n^2\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)\right]^2\right)$$

$$p(\theta|\boldsymbol{x}) = K_3(\boldsymbol{x})\exp\left(-\frac{1}{2}\frac{\sigma_n^2 + N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\left(\theta - \frac{\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}\sum_{k=1}^{N} x_k\right)^2\right)$$

上述分布是高斯型的，其均值为 $\frac{\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}\sum_{k=1}^{N} x_k$ **方差为** $\frac{\sigma_n^2\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}$

所以最小均方误差估计量为

$$\theta_{mse} = \frac{\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}\sum_{k=1}^{N} x_k$$

估计量的均方误差为

$$E\left[\left(\theta - \hat{\theta}_{mse}\right)^2\right] = \frac{\sigma_n^2\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}$$

条件中值估计

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

由于

$$p(\theta|\boldsymbol{x}) = K_3(\boldsymbol{x})\exp\left(-\frac{1}{2}\frac{\sigma_n^2 + N\sigma_\theta^2}{\sigma_n^2\sigma_\theta^2}\left(\theta - \frac{\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}\sum_{k=1}^{N} x_k\right)^2\right)$$

所以条件中值估计量为

$$\theta_{med} = \frac{\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}\sum_{k=1}^{N} x_k$$

估计量的均方误差为

$$E\left[\left(\theta - \hat{\theta}_{med}\right)^2\right] = \frac{\sigma_n^2\sigma_\theta^2}{\sigma_n^2 + N\sigma_\theta^2}$$

条件中值估计 $$\theta_{med}=\frac{\sigma_\theta^2}{\sigma_n^2+N\sigma_\theta^2}\sum_{k=1}^{N}x_k$$

最小均方误差估计 $$\theta_{mse}=\frac{\sigma_\theta^2}{\sigma_n^2+N\sigma_\theta^2}\sum_{k=1}^{N}x_k$$

最大后验估计 $$\hat{\theta}_{map}=\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)$$

结论：如果被估计量的后验概率密度函数是高斯型的，在三种典型代价函数下，使平均代价最小的估计量相同，都等于最小均方误差估计量，估计量的均方误差都是最小的 ——最佳估计的不变性。

例2

研究在加性噪声中单随机参量 s 的估计问题。

观测方程为 $x = s + n,$

其中***n***是均值为零，方差为 σ_n^2 的独立同分布高斯随机噪声

被估计量 s 在$(-S_M, S_M)$之间均匀分布的随机变量

求 s的贝叶斯估计量(最小均方误差和最大后验)

解：　最大后验估计

根据最大后验估计准则，估计量为满足以下方程的解，即

$$\left[\frac{\partial \ln p(x|s)}{\partial s}+\frac{\partial \ln p(s)}{\partial s}\right]\Bigg|_{s=\hat{s}_{map}}=0$$

由题设，可知，给定 s 条件下，观测信号x_k是均值为 s ，方差为σ_n^2的高斯随机变量

$$p(s)=\begin{cases}\dfrac{1}{2S_M}, & -S_M\le s\le S_M\\ 0, & \text{其他}\end{cases}$$

$$p(x|s)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right)$$

$$p(s)=\begin{cases}\frac{1}{2S_M}, & -S_M \le s \le S_M \\ 0, & \text{其他}\end{cases} \qquad p(x|s)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right)$$

$$\frac{\partial \ln p(x|s)}{\partial s}+\frac{\partial \ln p(s)}{\partial s}=\frac{\partial\left(-\frac{(x-s)^2}{2\sigma_n^2}\right)}{\partial s}+\frac{\partial\left(\frac{1}{2S_M}\right)}{\partial s}$$

$$=\frac{2(x-s)}{2\sigma_n^2}=\frac{x-s}{\sigma_n^2}$$

所以最大后验估计量为满足以下方程的解

$$\left.\frac{x-s}{\sigma_n^2}\right|_{s=\hat{s}_{map}}=0 \qquad \hat{s}_{map}=x$$

由于s在($-S_M, S_M$)之间取值，所以

$$\hat{s}_{map} = \begin{cases} -S_M, & x < -S_M \\ x, & -S_M \le x \le S_M \\ S_M, & x > S_M \end{cases}$$

最小均方误差估计

根据最小均方误差估计准则，估计量为

$$\hat{s}_{mse} = \int_{-\infty}^{\infty} s p(s|x) ds$$

$$= \frac{\int_{-\infty}^{\infty} s p(x|s) p(s) ds}{\int_{-\infty}^{\infty} p(x|s) p(s) ds}$$

$$= \frac{\int_{-s_M}^{s_M} s \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right) \frac{1}{2s_M} ds}{\int_{-s_M}^{s_M} \frac{1}{\sqrt{2\pi\sigma_n^2}} \exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right) \frac{1}{2s_M} ds}$$

$$\frac{\int_{-s_M}^{s_M} s\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right)\frac{1}{2s_M}ds}{\int_{-s_M}^{s_M} \frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x-s)^2}{2\sigma_n^2}\right)\frac{1}{2s_M}ds}$$

$$=\frac{\int_{s_M+x}^{s_M-x}(x-u)\exp\left(-\frac{u^2}{2\sigma_n^2}\right)du}{\int_{s_M+x}^{s_M-x}\exp\left(-\frac{u^2}{2\sigma_n^2}\right)du}, u=x-s$$

$$=x-\frac{\sigma_n^2\int_{d+v}^{d-v}t\exp\left(-\frac{t^2}{2}\right)dt}{\sigma_n\int_{d+v}^{d-v}\exp\left(-\frac{t^2}{2}\right)dt},t=u/\sigma_n,v=x/\sigma_n,d=s_M/\sigma_n$$

$$=x-\frac{\sigma_n\left(\exp\left(-(d-v)^2/2\right)-\exp\left(-(d+v)^2/2\right)\right)}{\int_{d+v}^{d-v}\exp\left(-\frac{t^2}{2}\right)dt}$$

$$=x-\frac{\sigma_n\left(\exp\left(-(d-v)^2/2\right)-\exp\left(-(d+v)^2/2\right)\right)}{\sqrt{2\pi}\left(Q(d+v)-Q(d-v)\right)}$$

3.3 最大似然估计

- **ML估计：先验等概下的MAP估计**
- 出发点：若先验概率 $p(\theta)$ 未知，或者**θ**为非随机的未知量，此时**MAP**不适用。
- 构造：

$$\hat{\theta}_{ML} = \arg\max_{\theta \in \mathcal{A}} p\left(\boldsymbol{x} \middle| \theta\right)$$

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

例1

- 如果参量θ的观测方程为

$$x_k = \theta + n_k, \quad k = 1,2,\cdots,N$$

其中n_k是均值为零，方差为 σ_n^2 的独立同分布高斯随机噪声；θ是均值为零，方差为 σ_θ^2 的高斯变量。求 $\hat{\theta}_{ML}$并与比较$\hat{\theta}_b$

$$p\left(x_k\middle|\theta\right)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{\left(x_k-\theta\right)^2}{2\sigma_n^2}\right)$$

$$p\left(\boldsymbol{x}\middle|\theta\right)=\prod_{k=1}^{N}p\left(x_k\middle|\theta\right)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{\left(x_k-\theta\right)^2}{2\sigma_n^2}\right)$$

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

$$\frac{\partial\ln p\left(\boldsymbol{x}\middle|\theta\right)}{\partial\theta}=\frac{\partial\left(-\sum_{k=1}^{N}\frac{\left(x_k-\theta\right)^2}{2\sigma_n^2}\right)}{\partial\theta}=\frac{2\sum_{k=1}^{N}\left(x_k-\theta\right)}{2\sigma_n^2}$$

$$\hat{\theta}_{ml}=\frac{1}{N}\sum_{k=1}^{N}x_k$$

均方误差

$$E\left[\left(\theta-\hat{\theta}_{ml}\right)^2\right]=E\left[\left(\theta-\frac{1}{N}\sum_{k=1}^{N}x_k\right)^2\right]$$

$$=E\left[\left(\theta-\frac{1}{N}\sum_{k=1}^{N}(\theta+n_k)\right)^2\right]=E\left[\left(\frac{1}{N}\sum_{k=1}^{N}n_k\right)^2\right]$$

$$=\frac{1}{N^2}\sum_{k=1}^{N}E\left[n_k^2\right]=\frac{\sigma_n^2}{N}$$

$$\hat{\theta}_b = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2 / N}\left(\frac{1}{N}\sum_{k=1}^{N} x_k\right)$$

$$E\left[\left(\theta - \hat{\theta}_b\right)^2\right] = \frac{1}{\sigma_n^2 / \sigma_\theta^2 + N}\sigma_n^2$$

$$\le \frac{1}{N}\sigma_n^2 = E\left[\left(\theta - \hat{\theta}_{ML}\right)^2\right]$$

ML估计的不变性

$$\alpha = g(\theta)$$

$$\hat{\alpha}_{ML} \xleftarrow{g(\cdot)} \hat{\theta}_{ML} \longleftarrow x(\theta)$$

- 若 $\alpha = g(\theta)$ 是一对一变换，有 $\hat{\alpha}_{ML} = g\left(\hat{\theta}_{ML}\right)$
- ……………是一对**J(J>1)**变换，

$$p(x|\theta) \Rightarrow p_j(x|\alpha); j = 1, \ldots, J$$

$$p(x|\alpha) = \max\left\{p_j(x|\alpha), j = 1, \ldots, J\right\}$$

$$\hat{\alpha}_{ML} = \arg\max_{\alpha} p(x|\alpha)$$

例2

- 同例1，求 $\alpha = \exp(\theta)$ 的ML估计

解：

由题设，可知，给定 θ 条件下，观测信号x_k是均值为θ，方差为σ_n^2的高斯随机变量

$$p(x_k|\theta)=\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

由于$\alpha=\exp(\theta)$ 是 θ 的一对一变换，即是单调函数，因此可得

$$p(\boldsymbol{x}|\alpha)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\ln\alpha)^2}{2\sigma_n^2}\right)$$

由最大似然估计原理，得最大似然估计量为满足以下方程的解。

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

$$\frac{\partial \ln p(\boldsymbol{x}|\alpha)}{\partial \alpha} = \frac{\partial\left(-\sum_{k=1}^{N}\frac{(x_k-\ln\alpha)^2}{2\sigma_n^2}\right)}{\partial \alpha} = \frac{2\sum_{k=1}^{N}(x_k-\ln\alpha)}{2\sigma_n^2}\cdot\frac{1}{\alpha}$$

所以最大似然估计量为

$$\hat{\alpha}_{ml} = \exp\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right) = \exp\left(\hat{\theta}_{ml}\right)$$

3.4 估计量的性质：无偏性

- 非随机变量

$$E\left[\hat{\theta}\right]=\int_{-\infty}^{\infty}\hat{\theta}p\left(\boldsymbol{x}\middle|\theta\right)d\boldsymbol{x}=\theta+b\left(\theta\right)$$

➢ 无偏估计

$$\text{if } b\left(\theta\right)=0, i.e., E\left[\hat{\theta}\right]=\theta$$

➢ 有偏估计

$$\text{if } b\left(\theta\right)\neq 0$$

➢ 已知偏差的有偏估计

$$\text{if } b\left(\theta\right)=b\neq 0$$

$\hat{\theta}-b$ 为无偏估计

估计量的性质：无偏性

- 随机变量

$$E\left[\hat{\theta}\right]=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\hat{\theta}p\left(\boldsymbol{x},\theta\right)d\boldsymbol{x}d\theta$$

  - 无偏估计

$$\text{if } E\left[\hat{\theta}\right]=E\left[\theta\right]$$

  - 有偏估计

$$\text{if } E\left[\hat{\theta}\right]\neq E\left[\theta\right]$$

- 渐近无偏估计

$$\lim_{N\to\infty}E\left[\hat{\theta}\left(x_1^N\right)\right]=\begin{cases}\theta, & \text{非RV}\\ E\left(\theta\right), & \text{RV}\end{cases}$$

有效性

对于被估计量 θ 的任意无偏估计 $\hat{\theta}_1$ 和 $\hat{\theta}_2$，若估计的均方误差

$$E\left[\left(\theta-\theta_1\right)^2\right] < E\left[\left(\theta-\theta_2\right)^2\right]$$

则称估计量 $\hat{\theta}_1$ 比 $\hat{\theta}_2$ 更有效。

如果 θ 的无偏估计量 $\hat{\theta}$ 小于其他任意无偏估计量的均方误差，则称该估计量为最小均方误差估计量。

问题：能否确定一个均方误差的下界?

一致性

假设根据N次观测量构造的估计量为 $\hat{\theta}(\boldsymbol{x}_N)$

若

$$\lim_{N\to\infty} P\left[\left|\theta-\hat{\theta}(\boldsymbol{x}_N)\right|>\varepsilon\right]=0$$

则称估计量$\hat{\theta}(\boldsymbol{x}_N)$是一致收敛的估计量。

若

$$\lim_{N\to\infty} E\left[\left(\theta-\hat{\theta}(\boldsymbol{x}_N)\right)^2\right]=0$$

则称估计量 $\hat{\theta}(\boldsymbol{x}_N)$是均方一致收敛的估计量。

充分性

- 若被估计量 θ 的估计量为$\hat{\theta}(x)$，**x**是观测量。如果以 θ 为参量的似然函数$p(x|\theta)$ 能够表示为：

$$p(x|\theta)=g(\hat{\theta}(x)|\theta)h(x), \quad h(x)\geq 0$$

  则称 $\hat{\theta}(x)$ 为充分估计量。

  其中，$g(\hat{\theta}(x)|\theta)$是通过 $\hat{\theta}(x)$ 才与**x**有关的函数，并且以 θ为参量。

- 有效估计量必然是充分估计量

Cramer-Rao界：非RV

- 非RV情况：设 $\hat{\theta}$ 是非随机参量 θ 的无偏估计，则有

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

当且仅当对任意的 θ 和 $\boldsymbol{x}$，均满足

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

时，不等式取等号。

证明

设 $\hat{\theta}$ 是非随机参量 θ 的无偏估计，则有

$$E\left[\hat{\theta}\right]=\theta$$

$$E\left[\left(\theta-\hat{\theta}\right)\right]=\int_{-\infty}^{\infty}\left(\theta-\hat{\theta}\right)p\left(\boldsymbol{x}\middle|\theta\right)d\boldsymbol{x}=0$$

对上式求偏导， 得

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

$$=0$$

证明

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

$$\int_{-\infty}^{\infty}p\left(\boldsymbol{x}\middle|\theta\right)d\boldsymbol{x}=1 \qquad \frac{\partial p\left(\boldsymbol{x}\middle|\theta\right)}{\partial\theta}=\frac{\partial\ln p\left(\boldsymbol{x}\middle|\theta\right)}{\partial\theta}p\left(\boldsymbol{x}\middle|\theta\right)$$

上式改写为

$$\int_{-\infty}^{\infty}\left(\theta-\hat{\theta}\right)\frac{\partial\ln p\left(\boldsymbol{x}\middle|\theta\right)}{\partial\theta}p\left(\boldsymbol{x}\middle|\theta\right)d\boldsymbol{x}=-1$$

$$\left[\int_{-\infty}^{\infty}\left(\theta-\hat{\theta}\right)\frac{\partial\ln p\left(\boldsymbol{x}\middle|\theta\right)}{\partial\theta}p\left(\boldsymbol{x}\middle|\theta\right)d\boldsymbol{x}\right]^2=1$$

证明

根据柯西-施瓦滋不等式

$$\left[\int_{-\infty}^{\infty} w(x)g(x)h(x)d\boldsymbol{x}\right]^2 \leq \int_{-\infty}^{\infty} w(x)g^2(x)d\boldsymbol{x} + \int_{-\infty}^{\infty} w(x)h^2(x)d\boldsymbol{x}$$

当且仅当 $g(x)=kh(x)$ 时，上式等号成立。

$$w(\boldsymbol{x})=p(\boldsymbol{x}|\theta) \qquad g(\boldsymbol{x})=\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta} \qquad h(\boldsymbol{x})=\theta-\hat{\theta}$$

证明

$$1=\left[\int_{-\infty}^{\infty}\left(\theta-\hat{\theta}\right)\frac{\partial \ln p\left(\boldsymbol{x}|\theta\right)}{\partial\theta}p\left(\boldsymbol{x}|\theta\right)d\boldsymbol{x}\right]^2$$

$$\le\int_{-\infty}^{\infty}\left(\theta-\hat{\theta}\right)^2 p\left(\boldsymbol{x}|\theta\right)d\boldsymbol{x}\int_{-\infty}^{\infty}\left[\frac{\partial \ln p\left(\boldsymbol{x}|\theta\right)}{\partial\theta}\right]^2 p\left(\boldsymbol{x}|\theta\right)d\boldsymbol{x}$$

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

等号成立条件

$$g\left(\boldsymbol{x}\right)=k\left(\theta\right)h\left(\boldsymbol{x}\right)$$

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

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

证明

克拉美-罗不等式的另一种形式 $\int_{-\infty}^{\infty} p(\boldsymbol{x}|\theta)dx = 1$

求偏导 $\int_{-\infty}^{\infty} \frac{\partial p(\boldsymbol{x}|\theta)}{\partial\theta}d\boldsymbol{x} = 0 = \int_{-\infty}^{\infty} \frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}p(\boldsymbol{x}|\theta)d\boldsymbol{x}$

再求一次偏导

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

$$= \int_{-\infty}^{\infty} \frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial\theta^2}p(\boldsymbol{x}|\theta)d\boldsymbol{x} + \int_{-\infty}^{\infty} \left(\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}\right)^2 p(\boldsymbol{x}|\theta)d\boldsymbol{x}$$

$$= E\left[\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial\theta^2}\right] + E\left[\left(\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}\right)^2\right]$$

证明

克拉美-罗不等式的另一种形式

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

所以

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

Remarks

非随机参量情况下的克拉美-罗不等式的含义和用途

(1) 非随机参量 θ 的任意无偏估计量 $\hat{\theta}$ 的方差$Var\left[\hat{\theta}\right]$，即均方误差恒不小于

$$-1\Big/E\left[\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial\theta^2}\right]=1\Big/E\left[\left(\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}\right)^2\right]$$

(2) 若非随机参量 θ 的无偏估计量 $\hat{\theta}$ 满足 $\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}=(\theta-\hat{\theta})k(\theta)$

则估计量的方差 $Var[\hat{\theta}]=E\left[(\theta-\hat{\theta})^2\right]$取到最小值，即取到克拉美-罗界。

Remarks

(3) 若非随机参量 θ 的无偏估计量 $\hat{\theta}$ 满足 $\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta} = (\theta - \hat{\theta})k(\theta)$

则无偏估计量 $\hat{\theta}$ 是有效的，否则是无效的。

(4) 若非随机参量 θ 的无偏估计量 $\hat{\theta}$ 是有效的，则估计量的方差，即均方误差可由克拉美-罗界取得。

$$Var[\hat{\theta}] = E\left[(\theta - \hat{\theta})^2\right] = -1 \Big/ E\left[\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial \theta^2}\right] = 1 \Big/ E\left[\left(\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta}\right)^2\right]$$

Remarks

(5) 若非随机参量 θ 的无偏有效估计量 $\hat{\theta}$ 存在，它必定是 θ 的最大似然估计量 $\hat{\theta}_{ml}$，且可由最大似然方程解得。

最大似然估计量为 $$\left.\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ml}} = 0$$

由 $$\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta} = (\theta - \hat{\theta})k(\theta) \quad \Rightarrow \quad \left.(\theta - \hat{\theta})k(\theta)\right|_{\theta=\hat{\theta}_{ml}} = 0 \quad \Rightarrow \quad \hat{\theta} = \hat{\theta}_{ml}$$

(6) 非随机参量 θ 的最大似然估计量 $\hat{\theta}_{ml}$ 不一定是无偏有效的。

均方误差

若非随机参量 θ 的无偏估计量 $\hat{\theta}$ 也是有效的，则其均方误差为

$$Var(\hat{\theta}) = E\left[(\theta-\hat{\theta})^2\right] = \frac{1}{-k(\theta)}$$

由 $\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta} = (\theta-\hat{\theta})k(\theta)$ ➡ $\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial \theta^2} = k(\theta) + (\theta-\hat{\theta})\frac{\partial k(\theta)}{\partial \theta}$

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

➡ $$Var(\hat{\theta}) = E\left[(\theta-\hat{\theta})^2\right] = -1\Big/ E\left[\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial \theta^2}\right] = \frac{1}{-k(\theta)}$$

例1

如果参量 θ 的观测方程为

$$x_k = \theta + n_k, \quad k = 1,2,\cdots,N$$

其中$\boldsymbol{n_k}$是均值为零，方差为 σ_n^2 的独立同分布高斯随机噪声，

试讨论估计量 θ 的最大似然估计量 $\hat{\theta}_{ml} = \frac{1}{N}\sum_{k=1}^{N} x_k$ 的无偏性、有效性和一致性。

由题设，

$$E\left[\hat{\theta}_{ml}\right]=E\left[\frac{1}{N}\sum_{k=1}^{N}x_k\right]=E\left[\frac{1}{N}\sum_{k=1}^{N}(\theta+n_k)\right]=\theta$$

$\hat{\theta}_{ml}$是θ的无偏估计量。

由于

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}=\frac{\partial\left(-\sum_{k=1}^{N}\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)}{\partial\theta}=\frac{2\sum_{k=1}^{N}(x_k-\theta)}{2\sigma_n^2}=\frac{-N\theta+\sum_{k=1}^{N}x_k}{\sigma_n^2}$$

$$=\frac{-N\left(\theta-\frac{1}{N}\sum_{k=1}^{N}x_k\right)}{\sigma_n^2}=\frac{-N\left(\theta-\hat{\theta}_{ml}\right)}{\sigma_n^2}$$

最大似然估计量 $\hat{\theta}_{ml}$ 是 θ 的有效估计量，且估计量的均方误差为

$$E\left[\left(\theta-\hat{\theta}_{ml}\right)^2\right]=\frac{1}{-k(\theta)}=\frac{\sigma_n^2}{N}$$

$$\lim_{N\to\infty}P\left(\left|\theta-\hat{\theta}_{ml}(\boldsymbol{x}_N)\right|>\varepsilon\right)=\lim_{N\to\infty}P\left(\left|\theta-\frac{1}{N}\sum_{k=1}^{N}(\theta+n_k)\right|>\varepsilon\right)$$
$$=\lim_{N\to\infty}P\left(\left|\frac{1}{N}\sum_{k=1}^{N}n_k\right|>\varepsilon\right)=0$$

最大似然估计量 $\hat{\theta}_{ml}$ 是一致收敛估计量。

$$\lim_{N\to\infty}E\left[\left(\theta-\hat{\theta}_{ml}\right)^2\right]=\lim_{N\to\infty}\frac{\sigma_n^2}{N}=0$$

最大似然估计量 $\hat{\theta}_{ml}$ 是均方一致收敛估计量

Cramer-Rao界：RV

设 $\hat{\theta}$ 是随机参量 θ 的无偏估计，则有

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

或

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

克拉美-罗不等式

当且仅当 $\frac{\partial\ln p(\boldsymbol{x},\theta)}{\partial\theta}=\left(\theta-\hat{\theta}\right)k$ 时，上述两式取等号。

克拉美-罗不等式取等号的条件

Remarks

(1) 由于

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

所以

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

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

Remarks

(2) 随机参量 θ 的任意无偏估计量 $\hat{\theta}$ 的方差 $Var\left[\hat{\theta}\right]$，即均方误差恒不小于

$$-1\Big/E\left[\frac{\partial^2 \ln p(\boldsymbol{x},\theta)}{\partial\theta^2}\right]=1\Big/E\left[\left(\frac{\partial \ln p(\boldsymbol{x},\theta)}{\partial\theta}\right)^2\right]$$

(3) 若随机参量 θ 的无偏估计量 $\hat{\theta}$ 满足

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

则估计量的方差 $Var\left[\hat{\theta}\right]=E\left[\left(\theta-\hat{\theta}\right)^2\right]$ 取到最小值，即取到克拉美-罗界。

Remarks

(4) 若随机参量 θ 的无偏估计量 $\hat{\theta}$ 满足

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

则无偏估计量 $\hat{\theta}$ 是有效的，否则是无效的。

(5) 若随机参量 θ 的无偏估计量 $\hat{\theta}$ 是有效的，则估计量的方差，即均方误差可由克拉美-罗界取得。

$$Var[\hat{\theta}] = E\left[(\theta - \hat{\theta})^2\right] = -1 \Big/ E\left[\frac{\partial^2 \ln p(\boldsymbol{x},\theta)}{\partial \theta^2}\right] = 1 \Big/ E\left[\left(\frac{\partial \ln p(\boldsymbol{x},\theta)}{\partial \theta}\right)^2\right]$$

Remarks

(6) 若随机参量 θ 的无偏有效估计量 $\hat{\theta}$ 存在，它必定是 θ 的最大后验估计量 $\hat{\theta}_{map}$。

最大后验估计量为

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

$$\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta}+\frac{\partial \ln p(\theta)}{\partial \theta}=(\theta-\hat{\theta})k \quad \Rightarrow \quad (\theta-\hat{\theta})k\Big|_{\theta=\hat{\theta}_{map}}=0$$

$$\Rightarrow \quad \hat{\theta}=\hat{\theta}_{map}$$

均方误差

若随机参量 θ 的无偏估计量 $\hat{\theta}$ 也是有效的，则其均方误差为

$$Var(\hat{\theta}) = E\left[(\theta - \hat{\theta})^2\right] = \frac{1}{-k}$$

由
$$\frac{\partial \ln p(\boldsymbol{x}, \theta)}{\partial \theta} = (\theta - \hat{\theta})k \Rightarrow \frac{\partial^2 \ln p(\boldsymbol{x}, \theta)}{\partial \theta^2} = k$$

$$\Rightarrow E\left[\frac{\partial^2 \ln p(\boldsymbol{x}, \theta)}{\partial \theta^2}\right] = k$$

$$\Rightarrow Var(\hat{\theta}) = E\left[(\theta - \hat{\theta})^2\right] = -1 \Big/ E\left[\frac{\partial^2 \ln p(\boldsymbol{x}, \theta)}{\partial \theta^2}\right] = \frac{1}{-k}$$

例2

同例1。试讨论估计量θ 的贝叶斯估计量

$$\hat{\theta}_b = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_n^2 / N}\left(\frac{1}{N}\sum_{k=1}^{N} x_k\right)$$

的无偏性、有效性和一致性。

由题设，

$$E\left[\hat{\theta}_b\right]=E\left[\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)\right]=E\left[\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}(\theta+n_k)\right)\right]$$

$$=0=E[\theta]$$

$\hat{\theta}_b$是θ的无偏估计量。

由于

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$\frac{\partial\ln p(\boldsymbol{x}|\theta)}{\partial\theta}=\frac{\partial\left(-\sum_{k=1}^{N}\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)}{\partial\theta}=\frac{2\sum_{k=1}^{N}(x_k-\theta)}{2\sigma_n^2}=\frac{-N\theta+\sum_{k=1}^{N}x_k}{\sigma_n^2}$$

由于 $$p(\theta)=\frac{1}{\sqrt{2\pi\sigma_\theta^2}}\exp\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)$$

$$\frac{\partial \ln p(\theta)}{\partial\theta}=\frac{\partial\left(-\frac{\theta^2}{2\sigma_\theta^2}\right)}{\partial\theta}=-\frac{\theta}{\sigma_\theta^2}$$

$$\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}+\frac{\partial \ln p(\theta)}{\partial\theta}=\frac{-N\theta+\sum_{k=1}^{N}x_k}{\sigma_n^2}-\frac{\theta}{\sigma_\theta^2}=-\left(\frac{N\sigma_\theta^2+\sigma_n^2}{\sigma_n^2\sigma_\theta^2}\right)\theta+\frac{1}{\sigma_n^2}\sum_{k=1}^{N}x_k$$

$$=-\left(\frac{N\sigma_\theta^2+\sigma_n^2}{\sigma_n^2\sigma_\theta^2}\right)\left(\theta-\frac{\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}\sum_{k=1}^{N}x_k\right)=-\left(\frac{N\sigma_\theta^2+\sigma_n^2}{\sigma_n^2\sigma_\theta^2}\right)\left(\theta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)\right)$$

$$=-\left(\frac{N\sigma_\theta^2+\sigma_n^2}{\sigma_n^2\sigma_\theta^2}\right)\left(\theta-\hat{\theta}_b\right)$$

贝叶斯估计量 $\hat{\theta}_b$ 是 θ 的有效估计量，且估计量的均方误差为

$$E\left[\left(\theta-\hat{\theta}_b\right)^2\right]=\frac{1}{-k}=\frac{\sigma_n^2\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}$$

$$\lim_{N\to\infty}P\left(\left|\theta-\hat{\theta}_b\left(\boldsymbol{x}_N\right)\right|>\varepsilon\right)=\lim_{N\to\infty}P\left(\left|\theta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}x_k\right)\right|>\varepsilon\right)$$

$$=\lim_{N\to\infty}P\left(\left|\frac{\sigma_n^2/N}{\sigma_\theta^2+\sigma_n^2/N}\theta-\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_n^2/N}\left(\frac{1}{N}\sum_{k=1}^{N}n_k\right)\right|>\varepsilon\right)=0$$

贝叶斯估计量 $\hat{\theta}_b$ 是一致收敛估计量。

$$\lim_{N\to\infty}E\left[\left(\theta-\hat{\theta}_b\right)^2\right]=\lim_{N\to\infty}\frac{\sigma_n^2\sigma_\theta^2}{N\sigma_\theta^2+\sigma_n^2}=0$$

贝叶斯估计量 $\hat{\theta}_b$ 是均方一致收敛估计量

非随机参量函数的CRLB

设 非随机参量 θ 的函数 $\alpha = g(\theta)$，其估计量 $\hat{\alpha}$ 是 α 的任意无偏估计，则有

$$Var[\hat{\alpha}] = E\left[(\alpha - \hat{\alpha})^2\right] \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{E\left[\left(\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial \theta}\right)^2\right]}$$

或

$$Var[\hat{\alpha}] = E\left[(\alpha - \hat{\alpha})^2\right] \geq \frac{\left(\frac{\partial g(\theta)}{\partial \theta}\right)^2}{-E\left[\frac{\partial^2 \ln p(\boldsymbol{x}|\theta)}{\partial \theta^2}\right]}$$

克拉美-罗不等式

当且仅当

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

时，上述两式取等号。

克拉美-罗不等式取等号的条件

例3

同例1。 $\alpha = b\theta$ 求 $\hat{\alpha}_{ml}$ 的无偏性和有效性，并求估计的均方误差。

解

由于

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

易知

$$\hat{\theta}_{ml}=\frac{1}{N}\sum_{k=1}^{N}x_k$$

根据最大似然估计的不变性，得到

$$\hat{\alpha}_{ml}=b\hat{\theta}_{ml}=\frac{b}{N}\sum_{k=1}^{N}x_k$$

$$E[\hat{\alpha}_{ml}]=E\left[\frac{b}{N}\sum_{k=1}^{N}x_k\right]=b\theta=\alpha$$

$\hat{\alpha}_{ml}$是α的无偏估计量

$$p(\boldsymbol{x}|\theta)=\prod_{k=1}^{N}p(x_k|\theta)=\prod_{k=1}^{N}\frac{1}{\sqrt{2\pi\sigma_n^2}}\exp\left(-\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)$$

$$\frac{\partial \ln p(\boldsymbol{x}|\theta)}{\partial\theta}=\frac{\partial\left(-\sum_{k=1}^{N}\frac{(x_k-\theta)^2}{2\sigma_n^2}\right)}{\partial\theta}=\frac{1}{\sigma_n^2}\sum_{k=1}^{N}(x_k-\theta)$$

$$=\left(b\theta-\frac{b}{N}\sum_{k=1}^{N}x_k\right)\left(-\frac{N}{b\sigma_n^2}\right)=(\alpha-\hat{\alpha}_{ml})k$$

$\hat{\alpha}_{ml}$是α的有效估计量

$$Var(\hat{\alpha}_{ml})=\frac{\left(\frac{\partial(b\theta)}{\partial\theta}\right)^2}{-E\left[\frac{\partial^2\ln p(\boldsymbol{x}|\theta)}{\partial\theta^2}\right]}=\frac{b^2}{\frac{N}{\sigma_n^2}}=\frac{b^2\sigma_n^2}{N}$$

3.5 LMMSE估计

- **Model**

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

- **MMSE、MAP估计：需要后验概率信息** $p(\theta|\boldsymbol{x})$
- **ML估计：需要先验概率信息** $p(\boldsymbol{x}|\theta)$
- 若仅已知前二阶距信息：观测信号和被估计随机矢量的均值矢量、协方差矩阵和互协方差矩阵。

  ---采用LMMSE估计

LMMSE估计准则

- 线性最小均方误差估计准则

  首先，构造的估计矢量 $\hat{\boldsymbol{\theta}}$ 是观测矢量x的线性函数，即：

$$\hat{\boldsymbol{\theta}} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x}$$

  同时要求估计矢量的均方误差最小，即为

$$\varepsilon_{\hat{\theta}}^2 = E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})] = Tr\{E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T]\}$$

  最小，式中 $Tr(\cdot)$表示矩阵的迹。

- 所以，线性最小均方误差估计的估计规则，就是把估计量构造成观测量的线性函数，同时要求估计量的均方误差最小。

LMMSE估计构造

$$\varepsilon_{\hat{\theta}}^2 = E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})] = Tr\{E[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^T]\}$$

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

$$\frac{\partial E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)^T\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)\right]}{\partial \boldsymbol{a}} = E\left\{\frac{\partial\left[\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)^T\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)\right]}{\partial \boldsymbol{a}}\right\}$$

$$= -2E\left[\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)\right] = 2\left(\boldsymbol{a}+\boldsymbol{B}\mu_x-\boldsymbol{\mu}_\theta\right)$$

令 $\boldsymbol{a}+\boldsymbol{B}\mu_x-\boldsymbol{\mu}_\theta = 0 \Rightarrow \boldsymbol{a} = \boldsymbol{\mu}_\theta - \boldsymbol{B}\mu_x$

LMMSE估计构造

$$\frac{\partial E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)^T\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}\right)\right]}{\partial \boldsymbol{B}}=\frac{\partial E\left[\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)^T\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)\right]}{\partial \boldsymbol{B}}$$

$$=E\left\{\frac{\partial Tr\left[\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)\left(\boldsymbol{\theta}-\boldsymbol{a}-\boldsymbol{B}\boldsymbol{x}\right)^T\right]}{\partial \boldsymbol{B}}\right\}$$

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

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

$$=2\boldsymbol{a}E\left[\boldsymbol{x}^T\right]+2\boldsymbol{B}E\left[\boldsymbol{x}\boldsymbol{x}^T\right]-2E\left[\boldsymbol{\theta}\boldsymbol{x}^T\right]$$

LMMSE估计构造

- **Lemma**

$$\frac{d\left(\mathrm{tr}\left(\boldsymbol{AX}\right)\right)}{d\boldsymbol{X}}=\frac{d\left(\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{A}^T\right)\right)}{d\boldsymbol{X}}=\boldsymbol{A}^T$$

$$\frac{d\left(\mathrm{tr}\left(\boldsymbol{X}^T\boldsymbol{AX}\right)\right)}{d\boldsymbol{X}}=\left(\boldsymbol{A}+\boldsymbol{A}^T\right)\boldsymbol{X}$$

$$\frac{d\left(\mathrm{tr}\left(\boldsymbol{XAX}^T\right)\right)}{d\boldsymbol{X}}=\boldsymbol{X}\left(\boldsymbol{A}+\boldsymbol{A}^T\right)$$

LMMSE估计构造

注意到

$$\boldsymbol{a} = \boldsymbol{\mu}_\theta - \boldsymbol{B}\mu_x$$

$$2\left(\boldsymbol{\mu}_\theta - \boldsymbol{B}\mu_x\right)E\left[\boldsymbol{x}^T\right] + 2\boldsymbol{B}E\left[\boldsymbol{x}\boldsymbol{x}^T\right] - 2E\left[\boldsymbol{\theta}\boldsymbol{x}^T\right]$$

$$= 2\boldsymbol{B}\left(E\left[\boldsymbol{x}\boldsymbol{x}^T\right] - \mu_x E\left[\boldsymbol{x}^T\right]\right) - 2\left(E\left[\boldsymbol{\theta}\boldsymbol{x}^T\right] - \boldsymbol{\mu}_\theta E\left[\boldsymbol{x}^T\right]\right)$$

$$= 0 \Rightarrow$$

$$BC_x - C_{\theta x} = 0 \Rightarrow$$

$$B = C_{\theta x} C_x^{-1}$$

LMMSE估计构造

解得

$$\boldsymbol{B} = \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}$$

$$\boldsymbol{a} = \boldsymbol{\mu}_\theta - \boldsymbol{B}\mu_x = \boldsymbol{\mu}_\theta - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\mu_x$$

所以

$$\hat{\theta}_{lmse} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{x} = \boldsymbol{\mu}_\theta - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\mu_x + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\boldsymbol{x}$$

$$= \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)$$

LMMSE估计的物理解释

$$\hat{\theta}_{lmse} = \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)$$

$$\Delta\theta = \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right) = \boldsymbol{C}_\theta^{1/2}\Delta\bar{\boldsymbol{\theta}}$$

均值（先验）　　新息（观测提供的信息）

观测量的信息

$$\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{u}_x = \boldsymbol{C}_x^{1/2}\Delta\bar{\boldsymbol{x}}$$

参量的信息

$$\Delta\bar{\boldsymbol{\theta}} = \frac{\mathrm{Cov}\left(\Delta\bar{\boldsymbol{\theta}}, \Delta\bar{\boldsymbol{x}}\right)}{\sqrt{\mathrm{Var}\left(\Delta\bar{\boldsymbol{\theta}}\right)}\sqrt{\mathrm{Var}\left(\Delta\bar{\boldsymbol{x}}\right)}}\Delta\bar{\boldsymbol{x}} = \mathrm{Cov}\left(\Delta\bar{\boldsymbol{\theta}}, \Delta\bar{\boldsymbol{x}}\right)\Delta\bar{\boldsymbol{x}}$$

$$= \mathrm{Cov}\left(\boldsymbol{C}_\theta^{-1/2}\Delta\boldsymbol{\theta}, \boldsymbol{C}_x^{-1/2}\Delta\boldsymbol{x}\right)\Delta\bar{\boldsymbol{x}} = \boldsymbol{C}_\theta^{-1/2}\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1/2}\Delta\bar{\boldsymbol{x}}$$

$$= \boldsymbol{C}_\theta^{-1/2}\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1/2}\boldsymbol{C}_x^{-1/2}\Delta\boldsymbol{x} = \boldsymbol{C}_\theta^{-1/2}\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\Delta\boldsymbol{x}$$

$$\Delta\boldsymbol{\theta} = \boldsymbol{C}_\theta^{1/2}\Delta\bar{\boldsymbol{\theta}} = \boldsymbol{C}_\theta^{1/2}\boldsymbol{C}_\theta^{-1/2}\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\Delta\boldsymbol{x} = \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\Delta\boldsymbol{x}$$

$$\hat{\theta}_{lmse} = \boldsymbol{\mu}_\theta + \Delta\boldsymbol{\theta} = \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\Delta\boldsymbol{x} = \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)$$

LMMSE估计的性质

(1) 估计矢量是观测矢量的线性函数

(2) 线性最小均方误差估计矢量是无偏估计

$$\hat{\boldsymbol{\theta}}_{lmse} = \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)$$

$$E\left[\hat{\boldsymbol{\theta}}_{lmse}\right] = E\left[\boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)\right]$$

$$= \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}E\left[\left(\boldsymbol{x} - \mu_x\right)\right] = \boldsymbol{\mu}_\theta$$

所以 $\hat{\boldsymbol{\theta}}_{lmse}$ 是 $\boldsymbol{\theta}$ 无偏估计

$$M_{\hat{\theta}_{lmse}} = E\left((\boldsymbol{\theta} - \hat{\theta}_{lmse})(\boldsymbol{\theta} - \hat{\theta}_{lmse})^T\right)$$

$$= E\left\{\left(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)\right)\left(\boldsymbol{\theta} - \boldsymbol{\mu}_\theta - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x} - \mu_x\right)\right)^T\right\}$$

$$= C_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\boldsymbol{C}_x\boldsymbol{C}_x^{-1}C_{\theta x}^T - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}C_{\theta x}^T - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}C_{\theta x}^T$$

$$= \boldsymbol{C}_\theta - \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}C_{x\theta} = \boldsymbol{C}_\theta - \boldsymbol{C}_{\theta|x}$$

Recall

$$\boldsymbol{H}\left(\tilde{\theta}\right) = \boldsymbol{H}\left(\theta \mid \hat{\theta}\right) = \boldsymbol{H}\left(\theta \mid x\right) = H\left(\theta\right) - I\left(\theta, x\right)$$

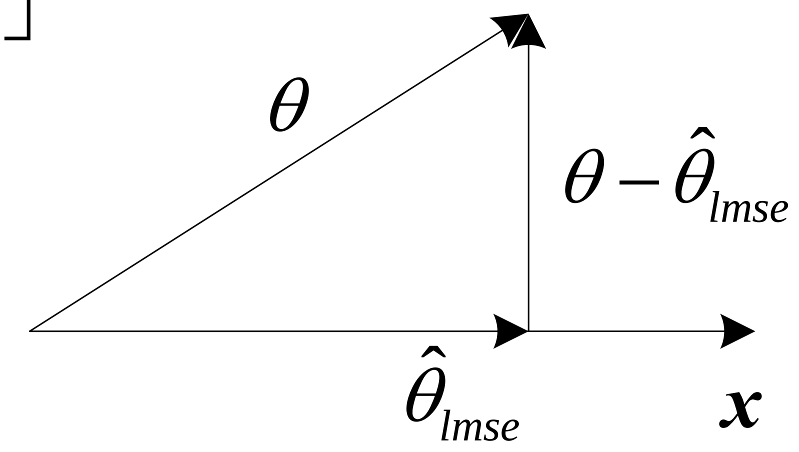
LMMSE估计的性质

(3)估计的误差矢量与观测矢量的正交性

被估计矢量 $\boldsymbol{\theta}$ 与线性最小均方误差估计矢量 $\hat{\boldsymbol{\theta}}_{lmse}$ 之间的误差矢量

$\tilde{\boldsymbol{\theta}} = \boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{lmse}$ **与观测矢量x是正交的，即**

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



LMMSE估计的性质

由于 $\hat{\boldsymbol{\theta}}_{lmse}$ 是$\boldsymbol{\theta}$ 无偏估计

$$E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{lmse}\right)\boldsymbol{\mu}_x^T\right]=0$$

$$\begin{aligned}
E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{lmse}\right)\boldsymbol{x}^T\right] &= E\left[\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{lmse}\right)\left(\boldsymbol{x}^T-\boldsymbol{\mu}_x^T\right)\right] \\
&= E\left[\left(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta-\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)\right)\left(\boldsymbol{x}^T-\boldsymbol{\mu}_x^T\right)\right] \\
&= E\left[\left(\boldsymbol{\theta}-\boldsymbol{\mu}_\theta\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)^T\right]-\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}E\left[\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_x\right)^T\right] \\
&= \boldsymbol{C}_{\theta x}-\boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}\boldsymbol{C}_x \\
&= 0
\end{aligned}$$

LMMSE估计的性质

(4)最小均方误差估计与线性最小均方误差估计的关系

随机矢量的最小均方误差估计矢量可以是观测矢量的非线性函数，而线性最小均方误差估计的估计矢量一定是观测矢量的线性函数。

当观测矢量与被估计矢量是联合高斯分布时，最小均方误差估计与线性最小均方误差估计两者相同

例

设M维被估计随机矢量的均值矢量和协方差矩阵分别为 $\boldsymbol{\mu}_\theta$ 和 $\boldsymbol{C}_\theta$，观测方程为

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

且已知

$$E[\boldsymbol{n}] = 0 \quad E[\boldsymbol{n}\boldsymbol{n}^T] = \boldsymbol{C}_n \qquad E[\boldsymbol{\theta}\boldsymbol{n}^T] = 0$$

求 $\boldsymbol{\theta}$ 的线性最小均方误差估计矢量 $\hat{\boldsymbol{\theta}}_{lms}$ 和估计矢量的均方误差阵 $\boldsymbol{M}_{\hat{\boldsymbol{\theta}}_{lms}}$

解：

由 $$\hat{\boldsymbol{\theta}}_{lmse} = \boldsymbol{\mu}_\theta + \boldsymbol{C}_{\theta x}\boldsymbol{C}_x^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_x)$$

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

$$\boldsymbol{C}_x = E\left[(\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)^T\right]$$

$$\boldsymbol{C}_{\boldsymbol{x}} = E\left[\left(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}}\right)^{T}\right]$$

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

$$= \boldsymbol{H}\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{H}^{T}+\boldsymbol{C}_{\boldsymbol{n}}$$

$$\boldsymbol{C}_{\boldsymbol{\theta x}} = E\left[\left(\boldsymbol{\theta}-\boldsymbol{\mu}_{\boldsymbol{\theta}}\right)\left(\boldsymbol{x}-\boldsymbol{\mu}_{\boldsymbol{x}}\right)^{T}\right]$$

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

$$= \boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{H}^{T}$$

$$\hat{\boldsymbol{\theta}}_{lmse} = \boldsymbol{\mu}_{\boldsymbol{\theta}}+\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{H}^{T}\left(\boldsymbol{H}\boldsymbol{C}_{\boldsymbol{\theta}}\boldsymbol{H}^{T}+\boldsymbol{C}_{\boldsymbol{n}}\right)^{-1}\left(\boldsymbol{x}-\boldsymbol{H}\boldsymbol{\mu}_{\boldsymbol{\theta}}\right)$$

$$\hat{\boldsymbol{\theta}}_{lmse} = \boldsymbol{\mu}_{\boldsymbol{\theta}} + \boldsymbol{C}_{\boldsymbol{\theta}} \boldsymbol{H}^T \left(\boldsymbol{H} \boldsymbol{C}_{\boldsymbol{\theta}} \boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{n}} \right)^{-1} \left(\boldsymbol{x} - \boldsymbol{H} \boldsymbol{\mu}_{\boldsymbol{\theta}} \right)$$

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

$$= \boldsymbol{C}_{\boldsymbol{\theta}} - \boldsymbol{C}_{\boldsymbol{\theta}} \boldsymbol{H}^T \left(\boldsymbol{H} \boldsymbol{C}_{\boldsymbol{\theta}} \boldsymbol{H}^T + \boldsymbol{C}_{\boldsymbol{n}} \right)^{-1} \boldsymbol{H} \boldsymbol{C}_{\boldsymbol{\theta}}$$

随机矢量函数的线性最小均方误差估计

- **线性变换上的可转换性**

θ 的线性MMSE估计矢量为 $\hat{\theta}_{lmse}$

θ 的线性函数 $\alpha = A\theta + b$ 的线性MMSE估计矢量 $\hat{\alpha}_{lmse}$ 为：

$$\hat{\alpha}_{lmse} = A\hat{\theta}_{lmse} + b$$

证明：$\hat{\alpha}_{lmse} = \mu_a + C_{ax}C_x^{-1}(x - \mu_x)$

$$\mu_a = E(a) = E(A\theta + b) = A\mu_\theta + b$$

$$C_{ax} = E\left[(a - \mu_a)(x - \mu_x)^T\right]$$

$$= E\left[(A\theta + b - A\mu_\theta - b)(x - \mu_x)^T\right] = AC_{\theta x}$$

$$\hat{\alpha}_{lmse} = A\mu_\theta + b + AC_{\theta x}C_x^{-1}(x - \mu_x) = A\hat{\theta}_{lmse} + b$$

随机矢量函数的线性最小均方误差估计

线性变换上的可转换性

$\boldsymbol{\theta}$ 的线性函数 $\boldsymbol{\alpha} = \boldsymbol{A\theta} + \boldsymbol{b}$ 的线性MMSE估计矢量 $\hat{\boldsymbol{\alpha}}_{lmse}$ 为：

$$\hat{\boldsymbol{\alpha}}_{lmse} = \boldsymbol{A}\hat{\boldsymbol{\theta}}_{lmse} + \boldsymbol{b}$$

无偏性： $E(\hat{\boldsymbol{\alpha}}_{lmse}) = E(\boldsymbol{\alpha})$

均方误差阵： $\boldsymbol{M}_{\hat{\boldsymbol{a}}_{lmse}} = E[(\boldsymbol{a} - \hat{\boldsymbol{a}}_{lmse})(\boldsymbol{a} - \hat{\boldsymbol{a}}_{lmse})^T] = \boldsymbol{A}\boldsymbol{M}_{\hat{\boldsymbol{\theta}}_{lmse}}\boldsymbol{A}^T$

随机矢量函数的线性最小均方误差估计

- **线性MMSE估计的可叠加性**

若 $\hat{\theta}_{1lmse}$ 和 $\hat{\theta}_{2lmse}$ 分别是同维随机矢量 θ_1 和 θ_2 的线性MMSE估计矢量，那么 $\alpha=\theta_1+\theta_2$ 的线性MMSE估计矢量 $\hat{\alpha}_{lmse}$ 为：

$$\hat{\alpha}_{lmse}=\hat{\theta}_{1lmse}+\hat{\theta}_{2lmse}$$

无偏性： $E(\hat{\alpha}_{lmse})=E(\alpha)$

均方误差阵： $\boldsymbol{M}_{\hat{a}_{lmse}}=\boldsymbol{M}_{\hat{\theta}_{1lmse}}+\boldsymbol{M}_{\hat{\theta}_{2lmse}}+\boldsymbol{M}_{\hat{\theta}_1\hat{\theta}_2lmse}+\boldsymbol{M}_{\hat{\theta}_2\hat{\theta}_1lmse}$

$$\boldsymbol{M}_{\hat{\theta}_{1lmse}}=\boldsymbol{C}_{\theta_1}-\boldsymbol{C}_{\theta_1\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\theta_1\boldsymbol{x}}^T$$

$$\boldsymbol{M}_{\hat{\theta}_{2lmse}}=\boldsymbol{C}_{\theta_2}-\boldsymbol{C}_{\theta_2\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\theta_2\boldsymbol{x}}^T$$

$$\boldsymbol{M}_{\hat{\theta}_1\hat{\theta}_{2lmse}}=\boldsymbol{C}_{\theta_1\theta_2}-\boldsymbol{C}_{\theta_1\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\theta_2\boldsymbol{x}}^T$$

$$\boldsymbol{M}_{\hat{\theta}_2\hat{\theta}_{1lmse}}=\boldsymbol{C}_{\theta_2\theta_1}-\boldsymbol{C}_{\theta_2\boldsymbol{x}}\boldsymbol{C}_{\boldsymbol{x}}^{-1}\boldsymbol{C}_{\theta_1\boldsymbol{x}}^T$$

随机矢量函数的线性最小均方误差估计

- 线性MMSE估计的可叠加性可以推广到任意有限L个同维矢量的情况

若 $\hat{\boldsymbol{\theta}}_{jlmse}(j=1,2,\cdots,L)$ 是随机矢量 $\boldsymbol{\theta}_j$ 的线性MMSE估计矢量，则

$$\boldsymbol{\alpha}=\sum_{j=1}^{L}\boldsymbol{\theta}_j$$

线性MMSE估计矢量为

$$\hat{\boldsymbol{\alpha}}_{lmse}=\sum_{j=1}^{L}\hat{\boldsymbol{\theta}}_{jlmse}$$

3.6 最小二乘估计

- 不需要任何先验信息，只需知道关于被估计量的观测信号模型
- 系统模型

$$\boldsymbol{\theta} \to \boldsymbol{s}(\boldsymbol{\theta}) \to \boldsymbol{x}$$

$\boldsymbol{s}(\boldsymbol{\theta})$ 被估计量的信号模型

- 误差平方和最小

$$\hat{\boldsymbol{\theta}}_{ls} = \arg\min_{\hat{\theta}} J\left(\hat{\boldsymbol{\theta}}\right)$$

$$J\left(\hat{\boldsymbol{\theta}}\right) = \left(\boldsymbol{x} - \boldsymbol{s}\left(\hat{\boldsymbol{\theta}}\right)\right)^T \left(\boldsymbol{x} - \boldsymbol{s}\left(\hat{\boldsymbol{\theta}}\right)\right)$$

线性最小二乘估计

- 系统模型

$$\boldsymbol{x} = \boldsymbol{H\theta} + \boldsymbol{n} \qquad \boldsymbol{s}(\boldsymbol{\theta}) = \boldsymbol{H\theta}$$

- 最小二乘估计误差

$$\hat{\boldsymbol{\theta}}_{lls} = \arg\min_{\hat{\boldsymbol{\theta}}} J(\hat{\boldsymbol{\theta}})$$

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

估计量的构造

$$\left.\frac{\partial J\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}}\right|_{\hat{\theta}=\hat{\theta}_{lls}} = 0 \Rightarrow \frac{\partial}{\partial \hat{\boldsymbol{\theta}}}\left[\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)^T\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)\right]$$

$$= -2\boldsymbol{H}^T\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right) = 0 \Rightarrow \hat{\boldsymbol{\theta}}_{lls} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{x}$$

$$\frac{\partial^2 J\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}^2} = \frac{\partial}{\partial \hat{\boldsymbol{\theta}}}\left[-2\boldsymbol{H}^T\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)\right]$$

$$= 2\boldsymbol{H}^T\boldsymbol{H} \succ 0$$

$$\therefore \hat{\boldsymbol{\theta}}_{lls}\text{使}J\left(\hat{\boldsymbol{\theta}}\right)\text{最小}$$

线性最小二乘估计误差

$$J\left(\hat{\theta}_{lls}\right)=\left(\boldsymbol{x}-\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{x}\right)^T\left(\boldsymbol{x}-\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{x}\right)$$

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

$$=\boldsymbol{x}^T\left(\boldsymbol{I}-\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T-\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T+\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\right)\boldsymbol{x}$$

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

估计量的性质

- 估计矢量是观测矢量的线性函数
- 若噪声矢量均值为0，LLS估计是无偏估计

$$E\left(\hat{\boldsymbol{\theta}}_{lls}\right)=\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T E\left(\boldsymbol{x}\right)=\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T E\left(\boldsymbol{H\theta}+\boldsymbol{n}\right)$$

$$=\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{H}E\left(\boldsymbol{\theta}\right)+\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T E\left(\boldsymbol{n}\right)=E\left(\boldsymbol{\theta}\right)$$

均方误差矩阵

$$M_{\hat{\theta}_{lls}} = E\left\{\left(\theta-\hat{\theta}_{lls}\right)\left(\theta-\hat{\theta}_{lls}\right)^T\right\}$$

$$= E\left\{\left(\theta-\left(H^TH\right)^{-1}H^Tx\right)\left(\theta-\left(H^TH\right)^{-1}H^Tx\right)^T\right\}$$

$$= E\left\{\left(\theta-\left(H^TH\right)^{-1}H^T\left(H\theta+n\right)\right)\left(\theta-\left(H^TH\right)^{-1}H^T\left(H\theta+n\right)\right)^T\right\}$$

$$= E\left\{\begin{aligned}&\theta\theta^T-\theta\left(H\theta+n\right)^TH\left(H^TH\right)^{-1}-\left(H^TH\right)^{-1}H^T\left(H\theta+n\right)\theta^T\\&+\left(H^TH\right)^{-1}H^T\left(H\theta+n\right)\left(H\theta+n\right)^TH\left(H^TH\right)^{-1}\end{aligned}\right\}$$

$$E\left(\theta n^T\right)=0$$

$$C_n = E\left(nn^T\right)$$

$$= E\left\{\begin{aligned}&\theta\theta^T-\theta\theta^T-\theta\theta^T+\left(H^TH\right)^{-1}H^TH\theta\theta^TH^TH\left(H^TH\right)^{-1}\\&+\left(H^TH\right)^{-1}H^Tnn^TH\left(H^TH\right)^{-1}\end{aligned}\right\}$$

$$=\left(H^TH\right)^{-1}H^TC_nH\left(H^TH\right)^{-1}$$

例1

$$\boldsymbol{x}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} n_{11} \\ n_{12} \end{bmatrix}$$

$$x_2 = 4 = \begin{bmatrix} 1 & 2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + n_2$$

求$\hat{\boldsymbol{\theta}}_{lls}$

解

$$\begin{bmatrix} \boldsymbol{x}_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \boldsymbol{H}_1 \\ \boldsymbol{h}_2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{n}_1 \\ n_2 \end{bmatrix} \rightarrow \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix} + \begin{bmatrix} n_{11} \\ n_{12} \\ n_2 \end{bmatrix}$$

$$\hat{\boldsymbol{\theta}}_{lls} = \left(\boldsymbol{H}^T \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{x} = \left[\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix} \right]^{-1} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 2 \end{bmatrix}^T \begin{bmatrix} 2 \\ 1 \\ 4 \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{4}{3} \end{bmatrix}$$

加权估计

- 给观测噪声较小的观测量以较大的权值，以提高估计的精度
- 加权矩阵**W**：对称正定矩阵
- 二乘加权估计误差

$$J_W\left(\hat{\boldsymbol{\theta}}\right)=\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)^T\boldsymbol{W}\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)$$

- 最小二乘加权估计

$$\hat{\boldsymbol{\theta}}_{llsw}=\arg\min_{\hat{\theta}} J_W\left(\hat{\boldsymbol{\theta}}\right)$$

估计量的构造

$$\left.\frac{\partial J_W\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}}\right|_{\hat{\theta}=\hat{\theta}_{llsw}} = 0 \Rightarrow \frac{\partial}{\partial \hat{\boldsymbol{\theta}}}\left[\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)^T \boldsymbol{W}\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)\right]$$

$$= -2\boldsymbol{H}^T\boldsymbol{W}\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right) = 0 \Rightarrow \hat{\boldsymbol{\theta}}_{llsw} = \left(\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{x}$$

$$\frac{\partial^2 J_W\left(\hat{\boldsymbol{\theta}}\right)}{\partial \hat{\boldsymbol{\theta}}^2} = \frac{\partial}{\partial \hat{\boldsymbol{\theta}}}\left[-2\boldsymbol{H}^T\boldsymbol{W}\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}\right)\right]$$

$$= 2\boldsymbol{H}^T\boldsymbol{W}\boldsymbol{H} \succ 0$$

$\therefore \hat{\boldsymbol{\theta}}_{llsw}$是使$J_W\left(\hat{\boldsymbol{\theta}}\right)$最小的估量

$$J_W\left(\hat{\boldsymbol{\theta}}\right) = \left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}_{llsw}\right)^T \boldsymbol{W}\left(\boldsymbol{x}-\boldsymbol{H}\hat{\boldsymbol{\theta}}_{llsw}\right)$$

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

估计量的性质

- 估计矢量是观测矢量的线性函数
- 若噪声矢量均值为0，**LLS**估计是无偏估计

$$E\left(\hat{\boldsymbol{\theta}}_{llsw}\right)=\left(\boldsymbol{H}^{T}\boldsymbol{W}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{T}\boldsymbol{W}E\left(\boldsymbol{x}\right)=E\left(\boldsymbol{\theta}\right)$$

- 均方误差矩阵

$$\boldsymbol{M}_{\hat{\boldsymbol{\theta}}_{llsw}}=E\left\{\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{llsw}\right)\left(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}_{llsw}\right)^{T}\right\}$$

$$=\left(\boldsymbol{H}^{T}\boldsymbol{W}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^{T}\boldsymbol{W}\boldsymbol{C}_{n}\boldsymbol{W}\boldsymbol{H}\left(\boldsymbol{H}^{T}\boldsymbol{W}\boldsymbol{H}\right)^{-1}$$

最佳加权矩阵的设计

- **Lemma：设A和B分别是M*N和N*K的任意两个矩阵,且AA^T的逆矩阵存在，则有矩阵不等式**

$$B^T B \geq (AB)^T \left(AA^T\right)^{-1} AB$$

- 令

$$A = H^T C_n^{-1/2} \quad B = C_n^{1/2} C^T \quad C = \left(H^T W H\right)^{-1} H^T W$$

有

$$\boldsymbol{M}_{\hat{\theta}_{llsw}} = C C_n C^T \geq \left(H^T C^T\right)^T \left(H^T C_n^{-1} H\right)^{-1} \left(H^T C^T\right)$$

$$= CH\left(H^T C_n^{-1} H\right)^{-1} (CH)^T = \left(H^T C_n^{-1} H\right)^{-1}$$

因此

$$\boldsymbol{W}_{opt} = \boldsymbol{C}_n^{-1} \qquad \hat{\theta}_{llsw} = \left(\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H}\right)^{-1} \boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{x}$$

$$\boldsymbol{M}_{\hat{\theta}_{llsw}} = \left(\boldsymbol{H}^T \boldsymbol{C}_n^{-1} \boldsymbol{H}\right)^{-1}$$

例2

$$\begin{cases} 216 = \theta + n_1 \\ 220 = \theta + n_2 \end{cases} \quad E(\boldsymbol{n}) = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \quad E(\boldsymbol{n}\boldsymbol{n}^T) = \begin{pmatrix} 16 & 0 \\ 0 & 4 \end{pmatrix}$$

求 $\hat{\theta}_{lls}, \hat{\theta}_{llsw}$

解：

$$\hat{\theta}_{lls} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{x} = 218$$

$$\hat{\theta}_{llsw} = \left(\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{x} = 219.2$$

$$\boldsymbol{M}_{\hat{\theta}_{lls}} = \left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1}\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{n}}\boldsymbol{H}\left(\boldsymbol{H}^T\boldsymbol{H}\right)^{-1} = 5$$

$$\boldsymbol{M}_{\hat{\theta}_{llsw}} = \left(\boldsymbol{H}^T\boldsymbol{C}_{\boldsymbol{n}}^{-1}\boldsymbol{H}\right)^{-1} = 3.2$$

非线性最小二乘估计

- 参量变换方法

$$\alpha = g(\theta)$$

$$s(\theta(\alpha)) = s(g^{-1}(\alpha)) = H\alpha$$

$$\hat{\alpha}_{ls} = (H^T H)^{-1} H^T x$$

$$\hat{\theta}_{ls} = g^{-1}(\hat{\alpha}_{ls})$$

非线性最小二乘估计

- 参量分离方法
  - 一般模型： $s(\theta)=H(\alpha)\beta \qquad \boldsymbol{\theta}=(\boldsymbol{\alpha},\boldsymbol{\beta})^T$
  - 目标：使得下式最小

$$J\left(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}\right)=\left(\boldsymbol{x}-\boldsymbol{H}\left(\hat{\boldsymbol{\alpha}}\right)\hat{\boldsymbol{\beta}}\right)^T\left(\boldsymbol{x}-\boldsymbol{H}\left(\hat{\boldsymbol{\alpha}}\right)\hat{\boldsymbol{\beta}}\right)$$

  - 算法：对于给定的 $\hat{\boldsymbol{\alpha}}$ ，计算使上式达到最小的 $\hat{\boldsymbol{\beta}}_{ls}$
    此时的估计误差为

$$J\left(\hat{\boldsymbol{\alpha}},\hat{\boldsymbol{\beta}}_{ls}\right)=\boldsymbol{x}^T\left(\boldsymbol{I}-\boldsymbol{H}\left(\hat{\boldsymbol{\alpha}}\right)\left(\boldsymbol{H}^T\left(\hat{\boldsymbol{\alpha}}\right)\boldsymbol{H}\left(\hat{\boldsymbol{\alpha}}\right)\right)^{-1}\boldsymbol{H}^T\left(\hat{\boldsymbol{\alpha}}\right)\right)\boldsymbol{x}$$

然后选择 $\hat{\boldsymbol{\alpha}}_{ls}$ 使得上式最小

作业3

- 信道估计问题（Slide 2）
- Rayleigh, slow fading channel

  $y=hx+w$
- 1）分别采用LMMSE估计和LS估计时，给出MSE随长度P的变化曲线。(同时给出C-R界)
- 2）分别采用LMMSE估计和LS估计时，给出BER随信道估计负载比(P/N）的变化曲线。（不同负载比情况下仍要求每帧传输速率相同）