统计信号处理

第八章

贝叶斯原理与一般贝叶斯估计

清华大学电子工程系
李洪 副教授

2023.4

内容概要

- 一、贝叶斯原理
- 二、一般贝叶斯估计(MMSE)
- 三、贝叶斯一般线性模型
- 四、MMSE的性质
- 五、MMSE性能评估
- 六、贝叶斯风险
- 七、小结

一、贝叶斯原理

例：白噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为信号幅度A，$w[n]$为高斯白噪声，且其方差为σ^2，即$w[n]\sim N(0,\sigma^2)$。

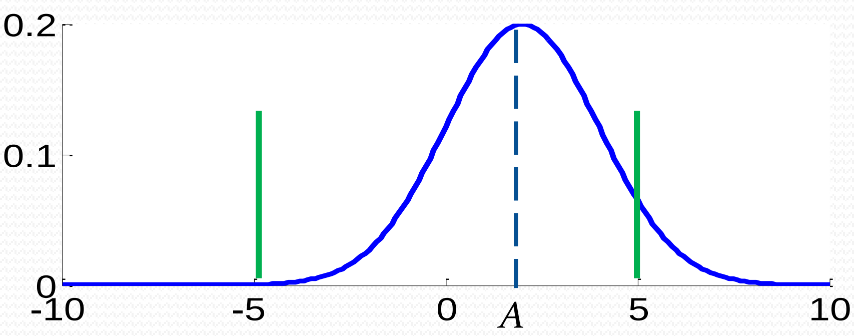
其MVU估计量为：

$$\hat{A}=\bar{x}=\frac{1}{N}\sum_{n=0}^{N-1}x[n]$$

估计量服从

$$\hat{A}\sim N\left(A,\frac{\sigma^2}{N}\right)$$

是否合理?



物理条件制约，如已知：$-A_0\le A\le A_0$

- **第一个问题：如何将一些合理的“约束条件”考虑进去?**

例：白噪声中电平估计问题：

$$x[n] = A + w[n],\ n = 0,1,\ldots,N-1$$

待估计参数为信号幅度A，$w[n]$为高斯白噪声，且其方差为σ^2，即$w[n] \sim N(0,\sigma^2)$。

合理的“约束条件”会带来什么变化?

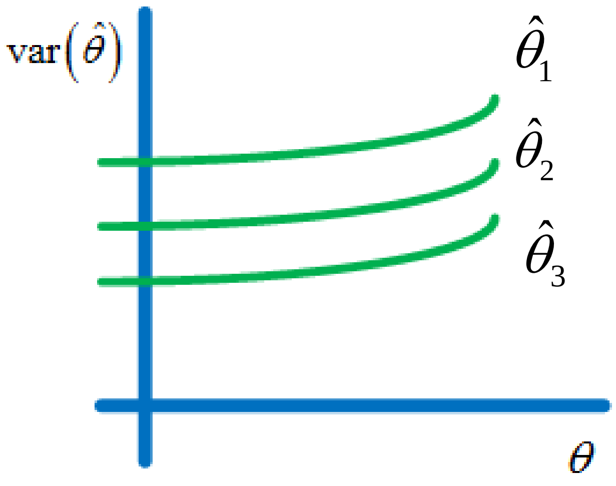
$$\hat{A} = \bar{x} = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$$

物理条件约束：$-A_0 \le A \le A_0$

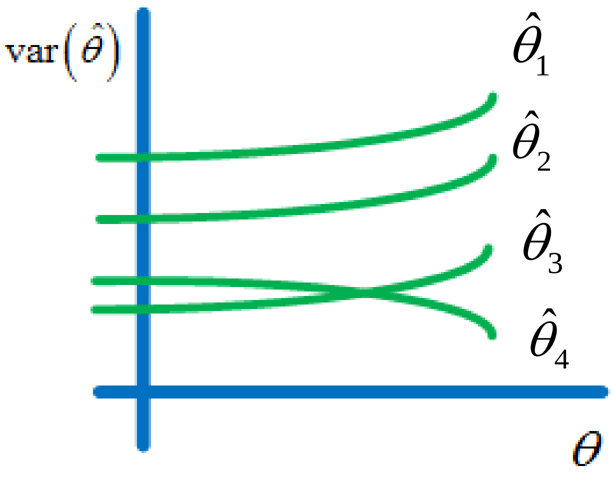
新估计量：$\breve{A} = \begin{cases} -A_0, & \bar{x} < -A_0 \\ \bar{x}, & -A_0 \le \bar{x} \le A_0 \\ A_0, & \bar{x} > A_0 \end{cases}$

$$\begin{aligned} \mathrm{mse}\left(\breve{A}\right) &= E\left(\left(\breve{A} - A\right)^2\right) \\ &= \int_{-\infty}^{-A_0} \left(-A_0 - A\right)^2 p_{\hat{A}}\left(u;A\right)du + \int_{-A_0}^{+A_0} \left(u - A\right)^2 p_{\hat{A}}\left(u;A\right)du \\ &\quad + \int_{A_0}^{+\infty} \left(A_0 - A\right)^2 p_{\hat{A}}\left(u;A\right)du \\ &< \mathrm{mse}\left(\hat{A}\right) = \int_{-\infty}^{+\infty} \left(u - A\right)^2 p_{\hat{A}}\left(u;A\right)du \end{aligned}$$

- 合理的“约束条件”使MSE更小
- 类似“约束条件”被称为先验信息

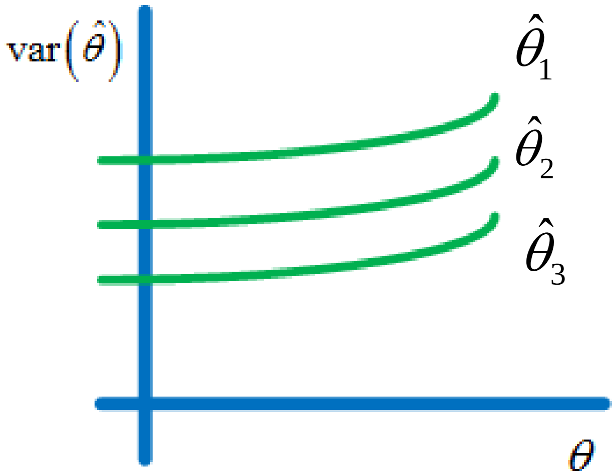


$\hat{\theta}_3$ 最优

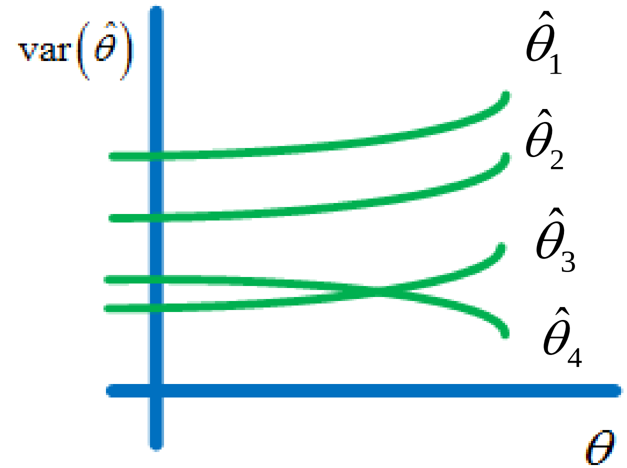


谁最优？

- MVU最优的含义：一致最小方差无偏估计量
- 第二个问题：无“一致”最小时该如何是好?



$\hat{\theta}_3$ 最优



谁最优？

$$\int E\left(\left(\hat{\theta}-\theta\right)^2\right)p(\theta)d\theta$$

经典估计中： $E\left(\left(\hat{\theta}-\theta\right)^2\right)=\int\left(\hat{\theta}-\theta\right)^2 p(\mathbf{x};\theta)d\mathbf{x}$

$$\iint\left(\hat{\theta}-\theta\right)^2 p(\mathbf{x};\theta)d\mathbf{x}\,p(\theta)d\theta \Rightarrow \iint\left(\hat{\theta}-\theta\right)^2 p(\mathbf{x};\theta)p(\theta)d\mathbf{x}d\theta \Rightarrow$$

$$\iint\left(\hat{\theta}-\theta\right)^2 p(\mathbf{x}\mid\theta)p(\theta)d\mathbf{x}d\theta=\underline{\iint\left(\theta-\hat{\theta}\right)^2 p(\mathbf{x},\theta)d\mathbf{x}d\theta}=\mathrm{Bmse}\left(\hat{\theta}\right)$$

常记为: $E\left(\left(\theta-\hat{\theta}\right)^2\right)$ ——贝叶斯均方误差

➢ 贝叶斯MSE Vs 经典理论MSE

贝叶斯MSE:

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

$$\text{Bmse}\left(\hat{\theta}\right)=\iint\left(\theta-\hat{\theta}\right)^2 p\left(\mathbf{x},\theta\right)d\mathbf{x}d\theta$$

Vs

经典估计MSE:

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

$$\text{mse}\left(\hat{\theta}\right)=\int\left(\hat{\theta}-\theta\right)^2 p\left(\mathbf{x};\theta\right)d\mathbf{x}$$

- 核心是两种估计思想/理念的不同!

- 将待估计参数作为随机变量
- 待估计参数不同取值处对应的MSE的“平均”
  - ✓ 引入了待估计参数分布情况
  - ✓ 可方便地考虑先验信息的影响
  - ✓ 无需“一致”最小
    ——只需“平均”最小
  - ✓ 可认为是对经典理论MSE的“平均”

Vs

- 将待估计参数作为确定性参数
  - ✓ 未对待估计参数做任何假设
  - ✓ 不考虑先验信息
  - ✓ 需“一致”最小
    ——往往导致不存在最优估计量

二、一般贝叶斯估计

一般贝叶斯估计： $\min\left\{\mathrm{Bmse}\left(\hat{\theta}\right)\right\}$

$$
\begin{aligned}
\mathrm{Bmse}\left(\hat{\theta}\right) &= E\left(\left(\theta-\hat{\theta}\right)^2\right) \\
&= \iint\left(\theta-\hat{\theta}\right)^2 p\left(\mathbf{x},\theta\right)d\mathbf{x}d\theta \\
&= \iint\left(\theta-\hat{\theta}\right)^2 p\left(\theta\,|\,\mathbf{x}\right)p\left(\mathbf{x}\right)d\mathbf{x}d\theta \\
&= \int\underline{\left\{\int\left(\theta-\hat{\theta}\right)^2 p\left(\theta\,|\,\mathbf{x}\right)d\theta\right\}}p\left(\mathbf{x}\right)d\mathbf{x}
\end{aligned}
$$

$$
\min\left\{\mathrm{Bmse}\left(\hat{\theta}\right)\right\} \Leftrightarrow \min\left\{\int\left(\theta-\hat{\theta}\right)^2 p\left(\theta\,|\,\mathbf{x}\right)d\theta\right\}
$$

$$\min\left\{\int\left(\theta-\hat{\theta}\right)^2 p(\theta|\mathbf{x})d\theta\right\}$$

$$\Updownarrow$$

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

$$\frac{\partial}{\partial\hat{\theta}}\int\left(\theta-\hat{\theta}\right)^2 p(\theta|\mathbf{x})d\theta$$

$$=\int\frac{\partial}{\partial\hat{\theta}}\left(\theta-\hat{\theta}\right)^2 p(\theta|\mathbf{x})d\theta$$

$$=\int-2\left(\theta-\hat{\theta}\right)p(\theta|\mathbf{x})d\theta$$

$$=-2\int\theta p(\theta|\mathbf{x})d\theta+2\hat{\theta}\underbrace{\int p(\theta|\mathbf{x})d\theta}_{1}$$

贝叶斯公式:

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

$$\Rightarrow\quad \hat{\theta}=\int\theta \underline{p(\theta|\mathbf{x})}d\theta$$

$$\boxed{\text{即}\quad \hat{\theta}=E(\theta|\mathbf{x})}$$

- ✓ 使贝叶斯MSE最小的估计量是**后验概率均值**
- ✓ 相应的估计量称为**最小均方误差估计量**（minimum mean square error, **MMSE**）

例：白噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\dots,N-1$$

待估计参数为信号幅度A。$w[n]$为高斯白噪声，且其方差为 σ^2，即 $w[n]\sim N\left(0,\sigma^2\right)$。其MMSE估计量是?

$$\text{MMSE:}\quad \hat{A}=\int Ap\left(A\mid\mathbf{x}\right)dA$$

$$p\left(A\mid\mathbf{x}\right)=\frac{p\left(\mathbf{x}\mid A\right)p\left(A\right)}{\int p\left(\mathbf{x}\mid A\right)p\left(A\right)dA}$$

$$p_x\left(x[n]\|A\right)=p_w\left(x[n]-A|A\right)$$

若 A 与 $w[n]$ 无关

$$p_x\left(x[n]\|A\right)=p_w\left(x[n]-A\right)$$

$$=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{1}{2\sigma^2}\left(x[n]-A\right)^2\right\}$$

$$p\left(\mathbf{x}\mid A\right)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}$$

1. 若信号幅度服从均匀分布 $U[-A_0, A_0]$

$$p(A)=\frac{1}{2A_0},\quad |A|\le A_0$$

$$p(\mathbf{x}\mid A)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}$$

$$p(A\mid \mathbf{x})=\frac{p(\mathbf{x}\mid A)\,p(A)}{\int p(\mathbf{x}\mid A)\,p(A)\,dA}$$

$$=\begin{cases}\dfrac{\dfrac{1}{2A_0\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}}{\displaystyle\int_{-A_0}^{A_0}\dfrac{1}{2A_0\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\dfrac{1}{2\sigma^2}\displaystyle\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}dA}, & |A|\le A_0\\ 0, & |A|>A_0\end{cases}$$

$$p(A|\boldsymbol{x})=\begin{cases}\dfrac{\exp\left\{-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}}{\int_{-A_0}^{A_0}\exp\left\{-\dfrac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}dA}, & |A|\le A_0\\ 0, & |A|>A_0\end{cases}$$

$$\sum_{n=0}^{N-1}\left(x[n]-A\right)^2=\sum_{n=0}^{N-1}x^2[n]-2NA\bar{x}+NA^2=N\left(A-\bar{x}\right)^2+\sum_{n=0}^{N-1}x^2[n]-N\bar{x}^2$$

$$p(A|\boldsymbol{x})=\begin{cases}\dfrac{\exp\left\{-\dfrac{1}{2\sigma^2}\left(N\left(A-\bar{x}\right)^2+\sum_{n=0}^{N-1}x^2[n]-N\bar{x}^2\right)\right\}}{\int_{-A_0}^{A_0}\exp\left\{-\dfrac{1}{2\sigma^2}\left(N\left(A-\bar{x}\right)^2+\sum_{n=0}^{N-1}x^2[n]-N\bar{x}^2\right)\right\}dA}, & |A|\le A_0\\ 0, & |A|>A_0\end{cases}$$

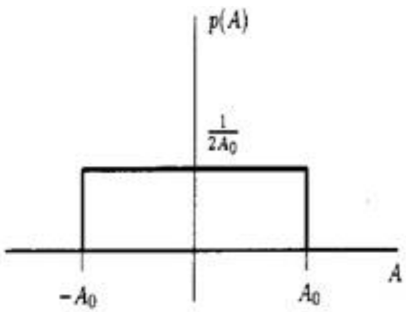
$$=\begin{cases}\dfrac{\exp\left\{-\dfrac{1}{2\sigma^2}N\left(A-\bar{x}\right)^2\right\}}{\int_{-A_0}^{A_0}\exp\left\{-\dfrac{1}{2\sigma^2}N\left(A-\bar{x}\right)^2\right\}dA}, & |A|\le A_0\\ 0, & |A|>A_0\end{cases}$$

截断高斯分布!

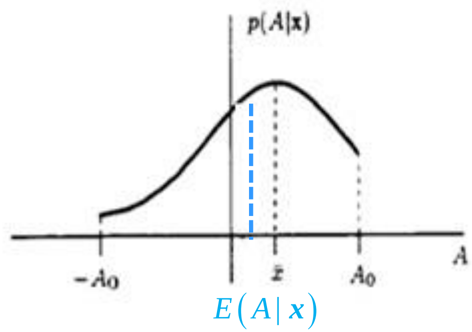
$$\text{MMSE:}\quad \hat{A}=E(A|\mathbf{x})$$

$$=\int A p(A|\mathbf{x})\,dA$$

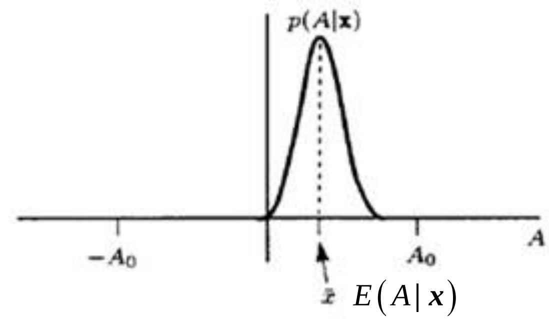
$$=\frac{\displaystyle\int_{-A_0}^{A_0} A\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left\{-\frac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right\}dA}{\displaystyle\int_{-A_0}^{A_0}\frac{1}{\sqrt{2\pi\frac{\sigma^2}{N}}}\exp\left\{-\frac{1}{2\frac{\sigma^2}{N}}(A-\bar{x})^2\right\}dA}$$



获得数据前



数据较少时



大数据量时

2. 若信号幅度服从高斯分布 $N\left(\mu_A,\sigma_A^2\right)$

$$p(A)=\frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\left\{-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right\}$$

$$p(\mathbf{x}\mid A)=\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}$$

$$p(A\mid\mathbf{x})=\frac{p(\mathbf{x}\mid A)p(A)}{\int p(\mathbf{x}\mid A)p(A)dA}=\frac{\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}\frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\left\{-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right\}}{\int_{-\infty}^{\infty}\frac{1}{\left(2\pi\sigma^2\right)^{N/2}}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2\right\}\frac{1}{\sqrt{2\pi\sigma_A^2}}\exp\left\{-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right\}dA}$$

$$=\frac{\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right\}}{\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}\left(x[n]-A\right)^2-\frac{1}{2\sigma_A^2}(A-\mu_A)^2\right\}dA}$$

$$\sum_{n=0}^{N-1}\left(x[n]-A\right)^2=\sum_{n=0}^{N-1}x^2[n]-2NA\bar{x}+NA^2$$

$$p(A|\boldsymbol{x})=\frac{\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\left(NA^2-2NA\bar{x}\right)+\frac{1}{\sigma_A^2}(A-\mu_A)^2\right)\right\}}{\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}\left(NA^2-2NA\bar{x}\right)+\frac{1}{\sigma_A^2}(A-\mu_A)^2\right)\right\}dA}$$

令 $Q(A)=\frac{1}{\sigma^2}\left(NA^2-2NA\bar{x}\right)+\frac{1}{\sigma_A^2}(A-\mu_A)^2$

$$=\left(\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}\right)A^2-2\left(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}\right)A+\frac{\mu_A^2}{\sigma_A^2}$$

令 $\sigma_{A|x}^2=\frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$， $\mu_{A|x}=\left(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}\right)\sigma_{A|x}^2$

$$\Rightarrow Q(A)=\frac{1}{\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2-\frac{\mu_{A|x}^2}{\sigma_{A|x}^2}+\frac{\mu_A^2}{\sigma_A^2}$$

$$\Rightarrow p(A|\boldsymbol{x})=\frac{\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\}}{\int_{-\infty}^{\infty}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\}dA}=\frac{\frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\}}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\}dA}$$

$$=\frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\}$$

$$p(A|\mathbf{x})=\frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\},\quad \mu_{A|x}=\left(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}\right)\sigma_{A|x}^2,\quad \sigma_{A|x}^2=\frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

$$\text{MMSE:}\quad \hat{A}=E(A|\mathbf{x})=\mu_{A|x}=\left(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}\right)\sigma_{A|x}^2$$

$$=\frac{\frac{N}{\sigma^2}\bar{x}+\frac{1}{\sigma_A^2}\mu_A}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

$$=\frac{\frac{N}{\sigma^2}}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}\bar{x}+\frac{\frac{1}{\sigma_A^2}}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}\mu_A$$

$$=\underbrace{\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}}_{\alpha}\bar{x}+\frac{\frac{\sigma^2}{N}}{\sigma_A^2+\frac{\sigma^2}{N}}\mu_A=\mu_A+\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}\left(\bar{x}-\mu_A\right)$$

$$=\mu_A+\frac{\frac{1}{\sigma^2/N}}{\frac{1}{\sigma_A^2}+\frac{1}{\sigma^2/N}}\left(\bar{x}-\mu_A\right)$$

$$=\alpha\bar{x}+(1-\alpha)\mu_A$$

先验信息对估计量的影响：

- MMSE估计量在**先验信息**与**观测数据**间加权折中
- 观测数据较少时，倚重于先验知识，否则倚重观测数据
- 先验知识越准确，越趋于先验均值，否则趋于观测数据
- 贝叶斯原理充分体现了“**先验信息**”和“**观测数据**”间的**融合**

先验信息对估计性能的影响

$$\mathrm{Bmse}(\hat{A}) = E\left(\left(A-\hat{A}\right)^2\right) = \iint \left(A-\hat{A}\right)^2 p(\mathbf{x},A)\,d\mathbf{x}\,dA$$

$$= \int \left\{\int \left(A-\hat{A}\right)^2 p(A|\mathbf{x})\,dA\right\} p(\mathbf{x})\,d\mathbf{x}$$

$$= \int \left\{\underline{\int \left(A-E(A|\mathbf{x})\right)^2 p(A|\mathbf{x})\,dA}\right\} p(\mathbf{x})\,d\mathbf{x}$$

$$\boxed{= \int \mathrm{var}(A|\mathbf{x})\,p(\mathbf{x})\,d\mathbf{x}}$$

$$= \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}} < \frac{\sigma^2}{N}$$

——MVU估计量的MSE

$$\boxed{\begin{aligned} p(A|\mathbf{x}) &= \frac{1}{\sqrt{2\pi\sigma_{A|x}^2}} \exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\} \\ \mathrm{var}(A|\mathbf{x}) &= \sigma_{A|x}^2 = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}} \end{aligned}}$$

先验知识改进了估计性能!

- 另一种解释:

$$\mathrm{var}(A|\mathbf{x}) = \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}} \Rightarrow \frac{1}{\mathrm{var}(A|\mathbf{x})} = \frac{N}{\sigma^2}+\frac{1}{\sigma_A^2} \Rightarrow \frac{1}{\underline{\mathrm{var}(A|\mathbf{x})}} = \frac{1}{\frac{\sigma^2}{\underline{N}}}+\frac{1}{\underline{\sigma_A^2}}$$

后验信息=数据信息+先验信息

✓ 先验信息的引入，改善了估计性能

Bmse的含义

例：白噪声中电平估计问题：

$$x[n]=A+w[n],\ n=0,1,\ldots,N-1$$

待估计参数为信号幅度 A。$w[n]$ 为高斯白噪声，且其方差为 σ^2，即 $w[n]\sim N\left(0,\sigma^2\right)$。

1. 方法一：采用MVU方法

$$\hat{A}=\bar{x}$$

$$\operatorname{mse}\left(\hat{A}\right)=E\left(\left(\hat{A}-A\right)^2\right)=\frac{\sigma^2}{N}$$

2. 方法二：采用贝叶斯方法

MMSE： $\hat{A}=E\left(A\,|\,\mathbf{x}\right)$

假定待估计参数A 服从高斯分布 $N\left(\mu_A,\sigma_A^2\right)$，由上例有

$$p\left(A\,|\,\mathbf{x}\right)=\frac{1}{\sqrt{2\pi\sigma_{A|x}^2}}\exp\left\{-\frac{1}{2\sigma_{A|x}^2}\left(A-\mu_{A|x}\right)^2\right\},\quad \mu_{A|x}=\left(\frac{N}{\sigma^2}\bar{x}+\frac{\mu_A}{\sigma_A^2}\right)\sigma_{A|x}^2,\quad \sigma_{A|x}^2=\frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}}$$

$$\hat{A} = \mu_{A|x} = \underbrace{\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}}_{\alpha}\bar{x} + \frac{\frac{\sigma^2}{N}}{\sigma_A^2+\frac{\sigma^2}{N}}\mu_A = \alpha\bar{x}+(1-\alpha)\mu_A$$

就某次估计而言（即对某个未知 A进行估计），该估计量对应的MSE为：

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

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

$$= \mathrm{var}(\hat{A})+\left(E(\hat{A})-A\right)^2 = \alpha^2\,\mathrm{var}(\bar{x})+\left[\alpha A+(1-\alpha)\mu_A-A\right]^2$$

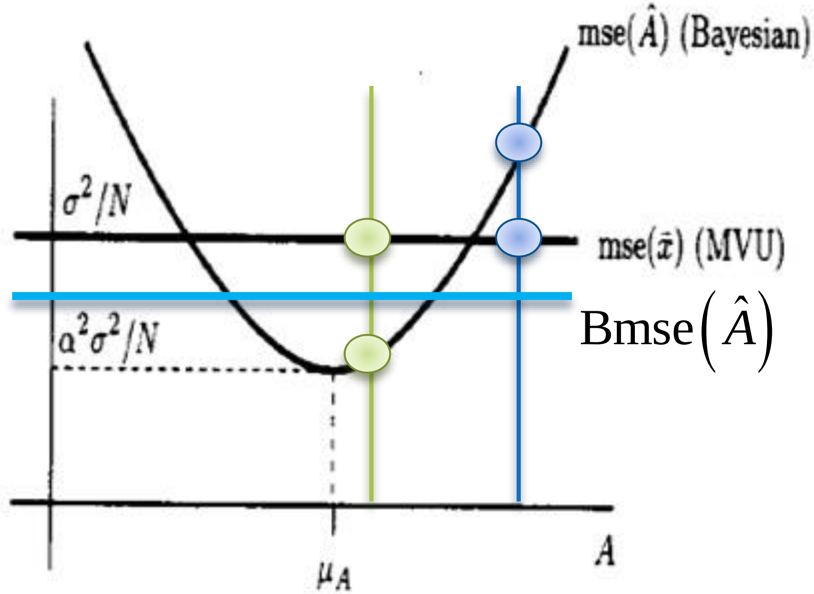
$$= \alpha^2\frac{\sigma^2}{N}+(1-\alpha)^2(A-\mu_A)^2$$ ——为A的二次函数

另一方面，在贝叶斯框架下，相应的Bmse为

$$\mathrm{Bmse}(\hat{A}) = E\left(\left(A-\hat{A}\right)^2\right) = \int\left\{\int\left(A-\hat{A}\right)^2 p(A|\mathbf{x})dA\right\}p(\mathbf{x})d\mathbf{x}$$

$$= \int\left\{\int\left(A-E(A|\mathbf{x})\right)^2 p(A|\mathbf{x})dA\right\}p(\mathbf{x})d\mathbf{x} = E\left(\sigma_{A|x}^2\right)$$

$$= \frac{1}{\frac{N}{\sigma^2}+\frac{1}{\sigma_A^2}} = \frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}\frac{\sigma^2}{N} = \alpha\frac{\sigma^2}{N}$$



$$\alpha\frac{\sigma^2}{N}$$

$$\mathrm{Bmse}\left(\hat{A}\right)$$

$$\hat{A}=\alpha\bar{x}+(1-\alpha)\mu_A$$
$$\mathrm{mse}\left(\hat{A}\right)=\alpha^2\frac{\sigma^2}{N}+(1-\alpha)^2\left(A-\mu_A\right)^2$$

$$\hat{A}=\bar{x}$$
$$\mathrm{mse}\left(\hat{A}\right)=E\left(\left(\hat{A}-A\right)^2\right)\quad=\frac{\sigma^2}{N}$$

- 就某次估计而言（即对参数的某个现实），贝叶斯估计并不见得比经典估计好
- 贝叶斯估计量对估计性能的改进，是指**平均**意义上的改进，即对大量现实进行估计时才会从**整体**上得以体现
- 这种改进，可认为源自于其利用了待估计参数**内在**的联系——由先验信息表征
- 当先验信息准确时，可以改进估计的性能，否则会起到误导作用，即恶化性能

贝叶斯MSE（Bmse）可认为是对经典MSE的“平均”：

$$E_A\left(\mathrm{mse}\left(\hat{A}\right)\right)=\alpha^2\frac{\sigma^2}{N}+(1-\alpha)^2E\left(\left(A-\mu_A\right)^2\right)$$
$$=\alpha^2\frac{\sigma^2}{N}+(1-\alpha)^2\sigma_A^2$$
$$\alpha=\frac{\sigma_A^2}{\sigma_A^2+\frac{\sigma^2}{N}}$$

$$E_A\left(\mathrm{mse}\left(\hat{A}\right)\right)=\alpha\frac{\sigma^2}{N}\ =\mathrm{Bmse}\left(\hat{A}\right)$$