第五章　信号参量的估计

5-1 概述

例：设有一个主动式声呐，假定目标是存在的，但不知道距离和速度，发射机发射的一个单频脉冲波，设为

$$S(t) = V_o \sin(\omega_o t + \varphi_o) \qquad o \le t \le T$$

那么其接收机收到的接收波形可表示为

$$x(t) = V_r \sin\left[\left(\omega_o + \omega_d\right)(t-\tau) + \phi_r\right] + n(t)$$

式中，$\mathbf{V_o}$ 为发射振幅，ω_0 为发射信号频率，φ_o 为初相位，经过双程传播和海中吸收的损失，其接收波形的振幅变为 V_r，初相位变为 ϕ_r（它们都是随机变量），ω_d 代表由于目标径向速度所引起的多卜勒频偏，τ 代表信号双程传播所需要的时间，**n(t)**为干扰噪声的总和。

显然，声呐系统必须根据 **x(t)**来获得参数ω_d与τ的估计$\hat{\omega}_d(x)$和$\hat{\tau}(x)$。

如果令参数$\omega_d=\theta_1,\tau=\theta_2$则接收信号 **x(t)**可改写为

$$
\begin{aligned}
x(t) &= V_r\sin\left[(\omega_o+\theta_1)(t-\theta_2)+\phi_r\right]+n(t)\\
&= S(t;\theta_1,\theta_2)+n(t)
\end{aligned}
$$

相应地，参量估计为$\hat{\theta}_1(x)$和$\hat{\theta}_2(x)$。

[注：$s(t;\theta_1,\theta_2)$实际上还应该是V_r与ϕ_r的函数，因为它们不是估计参量，所以省略不写。]

推广到多个参量问题，就可得到参量估计的一般概念：设在（**0**，**T**）时间内，观测波形 **x(t)**为

$$x(t)=s(t;\theta_1,\theta_2,\cdots,\theta_M)+n(t)$$

式中 $s(t;\theta_1,\theta_2,\cdots,\theta_M)$ 代表信号， $\theta_1,\theta_2,\cdots,\theta_M$ 是信号参量，它们可能是随机变量或未知的确定量，在**(0**，**T)**内不随时间变化（否则就是波形估计问题），这些参量可用矢量表示，并记为 $\theta=(\theta_1,\theta_2,\cdots,\theta_M)$ 。**n(t)**代表噪声，对其统计特性有一定先验知识。参量估计就是根据**(0**，**T)**内的观测波形 **x(t)**或观测数据 $\{x_k;k=1,2,\cdots,N\}$ ，按照某种“最佳”估计准则，对信号参量作出估计 $\hat{\theta}(k)$ ，它是观测值 **x** 的函数。

实际上，这种“最佳”是比较含糊的，因为观测值总是随机的。

1、 可以从出现概率的大小来作为最佳性的判决准则；
2、 同样也可以从数字特征如均值方差等作为最佳性的判决准则。

不同的准则，将会得到不同的“最佳”估计。

也就是说，对于同一组观测数据，从不同的出发点可以构造出许多不同的观测值的函数作为参量的估计。所以，哪一种估计才是真正的“最佳”估计，即对估计质量的评价是重要的问题。

由于观测值的随机性，所以，作为信号参量估计的观测值的函数必定也是个随机变量，对于随机变量的评价，当然还是只能用统计特性来描述。

-2 估计量的性质

一、估计量的性能指标

<1>无偏性

当观测重复地进行时，我们总希望算出的估计量都分布在真值附近，而不是在偏离真值的地方。由此，对于随机参量，如果估计量的均值等于参量自身的均值，即

$$E(\hat{\theta})=E(\theta)=\int_{-\infty}^{\infty}\theta p(\theta)d\theta$$

则称估计是无条件无偏的，若 $E(\hat{\theta})$ 不等于 $E(\theta)$，则称两均值之[redacted]

对于非随机参量，如果估计量的均值等于真值，即：

$$E\left[\hat{\theta}(x)/\theta\right]=\theta$$

则称估计量是条件无偏的，若估计量的均值不等于真值，则称均值与真值之差为估计的偏量。有偏量的估计称为有偏估计。

<2>有效性

对于无偏的估计量，它在真值附近的摆动程度常用估计误差方差来度量，我们总希望方差：

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

越小越好。对于两种估计 $\hat{\theta}_1$和$\hat{\theta}_2$，如果 $D(\hat{\theta}_1) < D(\hat{\theta}_2)$，

则称估计 $\hat{\theta}_1$比$\hat{\theta}_2$ 有效，能达到最小下限方差即 **-R** 界的估计量称为有效估计量。关于 **-R** 界，将在下面作专门讨论。

< >一致性

按照一般的直观概念，当观测时间加长或者观测数据增加时，信息量总是随之变得更靠近真值。换句话说，参数估计 $\overset{\frown}{\theta}(x)$ 的概率分布随观测样本的增加而密集于真值附近。

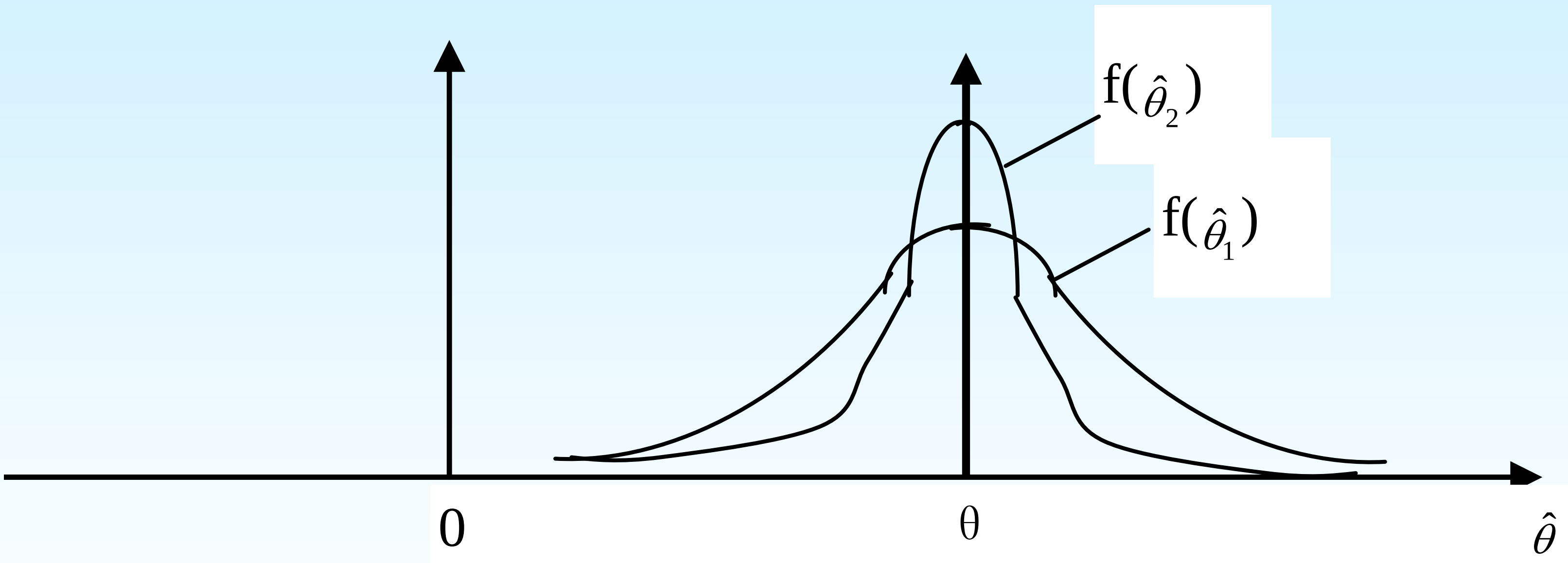
如果，估计误差的绝对值超过任意指定正数 ε 的概率趋于零；

$$\lim_{N\to\infty} P\left[\left|\theta-\hat{\theta}(x_1,x_2,\cdots,x_N)\right|>\varepsilon\right]=0$$

或者 $\lim\limits_{N\to\infty} P\left[\left|\theta-\hat{\theta}\right|<\varepsilon\right]=1$

则称估计量 $\hat{\theta}$ 为一致估计量。

实际应用中，总是希望采用一致估计，因为当观测值增加时，一致估计 $\hat{\theta}$ 的概率密度函数将在真值 θ 上形成又高又窄的单峰，以致出现大的估计误差的可能性极小。



二、 ramer-Rao 下限（ R 界）

对于任何无偏估计，总希望具有尽可能小的估计误差方差。

但是要找一个最小方差的无偏估计一般是比较困难的，这就给判断估计的优劣带来了困难。

经过推导可以发现，任何无偏估计量的估计方差均不能低于某个界限，即 **ramer-Rao** 下限，常称 **R** 界。

估计器无论做得多好，估计方差总不会低于这个限，这无疑给估计量的性能评价提出了一种参考，所以，**ramer-Rao** 界在参数估计中是一个很重要的量。

先讨论非随机参量的情况。

设估计是无偏的，即

$$E(\hat{\theta}-\theta)=\int_{-\infty}^{\infty}\left[\hat{\theta}(x)-\theta\right]p(x/\theta)dx=0$$

积分式对 θ 求导得

$$-\int_{-\infty}^{\infty}p(x/\theta)dx+\int_{-\infty}^{\infty}\frac{\partial p(x/\theta)}{\partial\theta}\left[\hat{\theta}(x)-\theta\right]dx=0$$

所以

$$\int_{-\infty}^{\infty}\frac{\partial p(x/\theta)}{\partial\theta}\left[\hat{\theta}(x)-\theta\right]dx=\int_{-\infty}^{\infty}p(x/\theta)dx$$

因为 $\int_{-\infty}^{\infty}p(x/\theta)dx=1$

所以有

$$\int_{-\infty}^{\infty}\frac{\partial\left[\ln p(x/\theta)\right]}{\partial\theta}p(x/\theta)\left[\hat{\theta}(x)-\theta\right]dx=1$$

由许瓦兹不等式知

$$\int_{-\infty}^{\infty}\left[\frac{\partial\left[\ln p(x/\theta)\ \right]}{\partial\theta}\sqrt{p(x/\theta)}\right]^2 dx\int_{-\infty}^{\infty}\left[\sqrt{p(x/\theta)}(\hat{\theta}-\theta)\right]^2 dx$$

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

只有当 $\dfrac{\partial[\ln p(x/\theta)]}{\partial\theta}=K(\theta)(\hat{\theta}-\theta)$ 时成立。

而

$$\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(x/\theta)}{\partial\theta}\sqrt{p(x/\theta)}\right]^2 dx$$

$$=\int_{-\infty}^{\infty}\left[\frac{\partial \ln p(x/\theta)}{\partial\theta}\right]^2 p(x/\theta)dx$$

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

$$\int_{-\infty}^{\infty}\left[\sqrt{p(x/\theta)}(\hat{\theta}-\theta)\right]^2 dx=\int_{-\infty}^{\infty}(\hat{\theta}-\theta)^2 p(x/\theta)dx=D(\hat{\theta})$$

$$\therefore D(\tilde{\theta}) = D(\hat{\theta}) \geq \frac{1}{E\left\{\left[\frac{\partial \ln p(x/\theta)}{\partial \theta}\right]^2\right\}}$$

这就是非随机参量估计的最小方差界，即 -R 界。
当条件密度满足

$$\frac{\partial \ln P(x/\theta)}{\partial \theta} = K(\theta)(\hat{\theta} - \theta)$$ 时

无偏估计 $\hat{\theta}$ 的方差才能达到 -R 界。

对式

$$\int_{-\infty}^{\infty} p(x/\theta)dx = 1$$

两边对 **θ** 求偏导，得

$$\frac{\partial}{\partial\theta}\int_{-\infty}^{\infty} p(x/\theta)dx = \int_{-\infty}^{\infty} \frac{\partial p(x/\theta)}{\partial\theta}dx = 0$$

而

$$\int_{-\infty}^{\infty} \frac{\partial p(x/\theta)}{\partial\theta}dx = \int_{-\infty}^{\infty} \frac{\partial[\ln p(x/\theta)]}{\partial\theta} p(x/\theta)dx$$

所以

$$\int_{-\infty}^{\infty} \frac{\partial[\ln p(x/\theta)]}{\partial\theta} p(x/\theta)dx = 0$$

上式再对 θ 求偏导，得

$$\int_{-\infty}^{\infty} \frac{\partial^2[\ln p(x/\theta)]}{\partial\theta^2} p(x/\theta)dx + \int_{-\infty}^{\infty} \frac{\partial[\ln p(x/\theta)]}{\partial\theta} \bullet \frac{\partial p(x/\theta)}{\partial\theta} dx = 0$$

而

$$\frac{\partial[\ln p(x/\theta)]}{\partial\theta} \bullet \frac{\partial p(x/\theta)}{\partial\theta} = \left[\frac{\partial[\ln p(x/\theta)]}{\partial\theta}\right]^2 p(x/\theta)$$

所以有

$$\int_{-\infty}^{\infty} \left[\frac{\partial[\ln p(x/\theta)]}{\partial\theta}\right]^2 p(x/\theta)dx = -\int_{-\infty}^{\infty} \frac{\partial^2[\ln p(x/\theta)]}{\partial\theta^2} p(x/\theta)dx$$

因此，可得到最小方差界（ -R 界）的另一种形式

$$D(\tilde{\theta}) = D(\hat{\theta}) \geq \frac{1}{-\int_{-\infty}^{\infty} \frac{\partial^2[\ln p(x/\theta)]}{\partial \theta^2} p(x/\theta)dx}$$

或者

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

对于随机参量情况，通过类似的证明可得 **-R** 界为

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

或者

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

当满足

$$\frac{\partial \ln p(x,\theta)}{\partial \theta} = K(\hat{\theta} - \theta) \qquad \text{（}\mathbf{K}\text{ 为常数）}$$

时，上面等式成立，即估计的方差能达到 **-R** 界。与非随机参量的差异只是将条件概率密度换成联合概率密度。

可以看出，不是所有估计的方差都能达 -R 界的，只有在 $p(x/\theta)$ 或 $p(x,\theta)$ 满足一定条件下，估计误差的方差才能达到这个下限，否则总比它大。这样，只要看估计误差方差与 -R 界的接近程度，即可评价估计的优劣。严格的有效估计是指对任意样本一致达到 -R 界的估计。所以，不是任何参数估计都存在有效估计的。

- 随机参量的估计——**Bayes** 估计

在讨论最佳估计问题时，常常是以“使估计的性能指标达到极值”作为最佳估计准则的。因此，所谓选择合适的最佳估计准则的问题，实质上就是选择合适的性能指标的问题。性能指标的选择在很大程度上取决于对被估计量性质的了解和对估计的要求，广义上说，性能指标应具备两个特点：

<1>性能指标应尽可能反映出估计的效果；（合理性）

<2>性能指标应该是一个用现有方法可以计算得出的量。（可实现性）。

例如，对于一个通讯系统，一个合适的性能指标可以是在标准信号输入时，在输出端的信号与噪声的功率比即通常所说的输出信噪比。显然，当其它条件相同时，高信噪比要比低信噪比性能好，即选择它作为性能指标是合理的。

另外，对于一个线性系统，在已知了输入信号和噪声的频谱以及系统参数后，这个信噪比是可写出表达式的，即选择它作为性能指标是可实现的。在以前讲到的匹配滤波器实际上就是在这一准则下导出的。

在估计中，用得较多的性能指标是估计误差的非负对称函数，也称代价函数，用$C(\theta-\hat{\theta})$表示，它具有如下特性：

<1> （**0**）**=0** 估计误差为零的代价为零

<2>$\left|\theta_1-\hat{\theta}\right| \geq \left|\theta_2-\hat{\theta}\right|$时$C(\theta_1-\hat{\theta}_1) \geq C(\theta_2-\hat{\theta}_2)$，估计误差越大，所花代价也越大。

< >$C(\theta_1-\hat{\theta}_1)=C(\hat{\theta}_1-\theta_1)$原点对称性

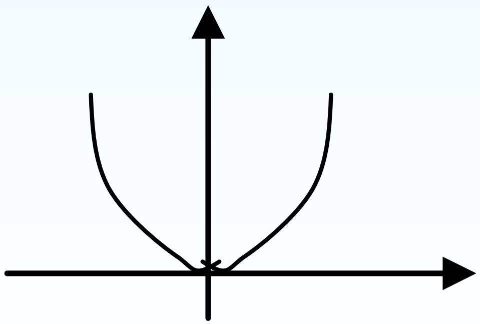
我们这里介绍两种满足上述特性的常用代价函数，即

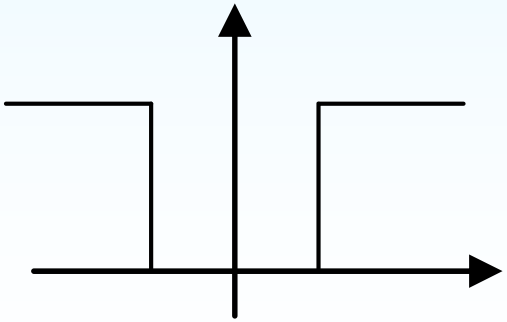
平方误差代价函数，令 $\tilde{\theta} = \theta - \hat{\theta}$

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

和均匀代价函数

$$C(\tilde{\theta}) = \begin{cases} 1 & \left|\theta - \hat{\theta}\right| \geq \dfrac{\Delta}{2} \\ 0 & \left|\theta - \hat{\theta}\right| < \dfrac{\Delta}{2} \end{cases}$$





由图可见，平方代价函数特别强调大误差的代价，以平方关系增长，均匀代价则将误差分为二截，落在区内的代价为 **0**，落在区外的恒定代价而不管误差的大小变化。

一旦规定了代价函数并已知先验概率密度函数后，就不难写出统计平均代价的表达式，用 $\overline{C}$ 表示：

$$\overline{C} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} C(\theta - \hat{\theta})\, p(x,\theta)\, dx\, d\theta$$

有时称为平均风险，贝页斯估计是使平均代价为最小的一种估计。

由于代价函数的不同，以致贝叶斯估计有许多形式。结合上述二种代价函数，我们来导出下面二种重要的估计。

一、最小均方估计$\hat{\theta}_{ms}$——最小方差估计

即$\hat{\theta}$的均方误差最小。由概率乘法知

$$p(x,\theta) = p(\theta / x)\,p(x)$$

∴平均代价可以分解为

$$\overline{C} = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} C(\theta-\hat{\theta})\,p(\theta / x)\,d\theta\right]p(x)\,dx$$

因为方括号内的值和 **p(x)**都是非负的，所以，使对 **x** 的积分值最小与使积分内函数最小是等价的，又由于 **p(x)** 不是$\hat{\theta}$的函数，所以，使平均代价$\overline{C}$最小与使方括号内积分最小是等价的，我们称此内积分为条件平均代价或条件风险，用$\overline{C}(\hat{\theta} / x)$表示。

$$\overline{C}(\hat{\theta}/x)=\int_{-\infty}^{\infty}C(\theta-\hat{\theta})p(\theta/x)d\theta$$

将 $C(\theta-\hat{\theta})=(\theta-\hat{\theta})^2$ 代入上式有

$$\overline{C}_{ms}(\hat{\theta}/x)=\int_{-\infty}^{\infty}(\theta-\hat{\theta})p(\theta/x)d\theta$$

将此条件风险对 $\hat{\theta}$ 求导，并令导数为零，所得解即为估计 $\hat{\theta}_{ms}$

$$\partial[\overline{C}_{ms}(\hat{\theta}/x)]\Big/\partial\hat{\theta}=-\int_{-\infty}^{\infty}2(\theta-\hat{\theta})p(\theta/x)d\theta\overset{令}{=}0$$

$$\therefore 得\quad \hat{\theta}_{ms}=\int_{-\infty}^{\infty}\theta p(\theta/x)d\theta=E(\theta/x)$$

现在来求它的均值：

$$E(\hat{\theta}_{ms}) = E[E(\theta / x)] = \int_{-\infty}^{\infty} E(\theta / x) p(x) dx$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \theta p(\theta / x) p(x) d\theta dx = E(\theta)$$

所以，这是一个无偏估计。

再来求估计误差的方差：

$$
\begin{aligned}
D(\theta-\hat{\theta}_{ms}) &= E[(\theta-\hat{\theta}_{ms})^2]-E^2(\theta-\hat{\theta}_{ms}) \\
&= E[(\theta-\hat{\theta}_{ms})^2] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\theta-\hat{\theta}_{ms})^2 p(\theta,x)d\theta dx \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta-E(\theta/x)\right]^2 p(\theta/x)p(x)d\theta dx \\
&= \int_{-\infty}^{\infty}D(\theta/x)p(x)dx
\end{aligned}
$$

如果设θ的任一其它无偏估计为$\hat{\theta}$，则其相应的估计误差的方差为

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

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta-E(\theta/x)+E(\theta/x)-\hat{\theta}\right]^2 p(\theta/x)p(x)d\theta dx$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left\{\left[\theta-E(\theta/x)\right]^2+2\left[\theta-E(\theta/x)\right]\left[E(\theta/x)-\hat{\theta}\right]\right.$$

$$\left.+\left[E(\theta/x)-\hat{\theta}\right]^2\right\}\bullet p(\theta/x)p(x)d\theta dx$$

$$\because \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta - E(\theta / x)\right]\left[E(\theta / x) - \hat{\theta}\right] p(\theta / x) p(x) d\theta dx$$

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

$$= 0$$

$$\therefore D(\theta - \hat{\theta}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\theta - E(\theta / x)\right]^2 p(\theta / x) p(x) d\theta dx$$

$$+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[E(\theta / x) - \hat{\theta}\right]^2 p(\theta / x) p(x) d\theta dx$$

$$= D(\theta - \hat{\theta}_{ms}) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[E(\theta / x) - \hat{\theta}\right]^2 p(\theta / x) p(x) d\theta dx$$

$$\geq D(\theta - \hat{\theta}_{ms})$$

只有当$\hat{\theta} = E(\theta / x) = \hat{\theta}_{ms}$时，等式成立。

不等式表明最佳估计$\hat{\theta}_{ms}=E(\theta/x)$的估计误差方差将小于任何其它无偏估计的方差。这就是为什么称它为最小均方估计的原因。另外由于$\hat{\theta}_{ms}$等于条件均值，所以，最小均方估计又称条件均值估计。（注：最小方差估计不一定是有效估计,有效估计误差的方差必须达到 -R 下限）。

值得注意，对于随机参量，估计误差的方差不一定等于估计的方差，它们有不同的表达式：

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

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

二、最大后验概率估计 $\hat{\theta}_{map}$

将均匀代价函数代入条件风险公式有

$$\overline{C}_{unf}(\hat{\theta}/x) = 1 - \int_{\hat{\theta}-\Delta/2}^{\hat{\theta}+\Delta/2} p(\theta/x)d\theta$$

要使这个条件风险最小，应使右边的积分值达最大，对于很小的Δ而言，$p(\theta/x)$可以近似看作为常量，所以有

$$\int_{\hat{\theta}-\frac{\Delta}{2}}^{\hat{\theta}+\frac{\Delta}{2}} p(\theta/x)d\theta \approx p(\theta/x)\Big|_{\theta=\hat{\theta}} \cdot \Delta$$

所以，使条件风险最小就近似地等价于使后验概率密度 $p(\theta/x)$ 最大，而这一使后验概率达到最大的参数取值

就称作最大后验概率估计，用 $\hat{\theta}_{map}$ 表示，有

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

或者写为

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

该方程称为最大后验概率方程。

利用概率乘法，

$$p(\theta / x) = \frac{p(x / \theta)\, p(\theta)}{p(x)}$$

则最大后验概率方程可以写为另一个形式：（**p(x)**为与θ无关的量）

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

其中第一项依赖于观测值 **x**，第二项与参量先验概率密度 $p(\theta)$ 有关。物理上，最大后验概率估计 $\hat{\theta}_{map}$ 可理解为在观测值 **x** 下，被估计参量θ出现可能性最大的就是 $\hat{\theta}_{map}$。或者说随机参量θ落在 $\hat{\theta}_{map}$ 的邻域内的概率比落在任何其它值的相同邻域内的概率要大。

如果将最大后验概率估计与随机参量估计的 **-R** 界联系起来的话，将会发现，若有效估计存在，这个估计就是最大验概率估计$\hat{\theta}_{map}$。这是因为

$$
\begin{tikzcd}
\text{有效估计} \arrow[r] \arrow[dr] & \dfrac{\partial \ln p(\theta, x)}{\partial \theta} = K(\hat{\theta} - \theta) \arrow[d, equal] \\
& \text{估计误差方差达到 } \textbf{-R} \text{ 界}
\end{tikzcd}
$$

而最大后验概率估计 $\hat{\theta}_{map}$ 应满足

$$\frac{\partial \ln p(\theta / x)}{\partial \theta}\bigg|_{\theta=\hat{\theta}map} = 0$$

$$\|$$

$$\frac{\partial \ln p(\theta, x)}{\partial \theta}\bigg|_{\theta=\hat{\theta}map} \overset{有效估计}{=} K(\hat{\theta} - \hat{\theta}_{map})$$

$$\therefore \hat{\theta} = \hat{\theta}_{map}$$

进一步推理可知，若有效估计存在的话，最小均方估计和最大后验概率估计是等价的。

例子：考虑均值为零，方差为σ_n^2的正态白噪声 **n** 中接收信号 **s**，已知信号是在$-s_M$与S_M之间均匀分布的，要求用一个观测数据

$$x = s + n$$

对 **s** 作出估计

先求最大后验概率估计 $\hat{\theta}_{map}$，后验概率 $p(s/x)$ 可写为

$$p(s/x)=\frac{p(x/s)p(s)}{p(x)}$$

因为对于参量 s，p(x)相当于一个常数，所以要使 $p(s/x)$ 最小等价于使 $p(x/s)p(s)$ 最小，由给定的条件，不难得到

$$p(s)=\begin{cases}\dfrac{1}{2s_M} & -s_M\le s<s_M\\ 0 & \text{其它}s\end{cases}$$

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

所以

$$p(s)\cdot p(x/s)=\begin{cases}\dfrac{1}{2\sqrt{2\pi}\sigma_n s_M}\exp\left[-\dfrac{(x-s)^2}{2\sigma_n^2}\right] & -S_M\le S<S_M\\ 0 & \text{其它}S\end{cases}$$

由上式知，当$-S_M \le x < S_M$时，取$\hat{s}=x$有最大的$p(s/x)$

当$|x|>S_M$时取$\hat{s}=\pm S_M$有最大的$p(s/x)$

$$\therefore \hat{S}_{MAP}=\begin{cases}S_M & x\ge s_M\\ x & -s_M\le x<s_M\\ -S_M & x<-s_M\end{cases}$$

再来求最小均方估计$\hat{s}_{ms}$

$$\hat{s}_{ms} = E(s/x) = \int_{-\infty}^{\infty} sp(s/x)ds$$

$$= \int_{-\infty}^{\infty} s \cdot \frac{p(x/s)p(s)}{p(x)} ds = \frac{1}{p(x)} \int_{-\infty}^{\infty} sp(x/s)p(s)ds$$

因为

$$p(x) = \int_{\infty}^{\infty} p(s,x)ds = \int_{-\infty}^{\infty} p(x/s)p(s)ds$$

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

所以

$$\hat{s}_{ms} = \int_{-S_M}^{S_M} s\exp[-\frac{(x-s)^2}{2\sigma_n^2}]ds \Big/ \int_{-S_M}^{S_M} \exp[-\frac{(x-s)^2}{2\sigma_n^2}]ds$$

令 **s= x -t**，则

$$\hat{s}_{ms}=\frac{\int_{-\infty}^{\infty} sp(s/x)p(s)ds}{p(x)}=\frac{\int_{x+s_M}^{x-s_M}(x-t)\exp[-\frac{t^2}{2\sigma_n^2}]dt}{\int_{x+s_M}^{x-s_M}\exp[-\frac{t^2}{2\sigma_n^2}]dt}$$

$$=x-\frac{\int_{x+s_M}^{x-s_M}t\exp[-\frac{t^2}{2\sigma_n^2}]dt}{\int_{x+s_M}^{x-s_M}\exp[-\frac{t^2}{2\sigma_n^2}]dt}\overset{令}{=}x-\varphi(s_M,x)$$

其中$\varphi(S_M,x)$与 **S_M** 和 **x** 有关。

可见最大后验概率估计$\hat{\theta}_{map}$与最小均方估计并不相等，而且都是非线性估计。

-4 最大似然估计

对于任何形式的贝叶斯估计，它必须预先知道参量 θ 的先验概率密度 $p(\theta)$。这就带来了二个问题，首先是这个先验概率密度一般不易确定，其次是如果估计参量是个未知的确定量，那么就不能规定参量的先验概率，这时，贝叶斯估计就不能使用。

如果将最大后验概率方程

$$\left[\frac{\partial \ln p(x/\theta)}{\text{[redacted]}} + \frac{\partial \ln p(\theta)}{\text{[redacted]}}\right]\Bigg|_{\theta=\hat{\theta}_{\mathrm{MAP}}} = 0$$

中先验概率 $p(\theta)$ 的作用去掉，方程就变为

$$\frac{\partial \ln p(x/\theta)}{\partial \theta}\Bigg|_{\theta=\hat{\theta}_{ml}} = 0$$

满足该方程的解$\theta = \hat{\theta}_{ml}$就称为参数$\theta$的最大似然估计，表示为$\hat{\theta}_{ml}$。实际上，$\hat{\theta}_{ml}$是$p(x/\theta)$的极值点。而从物理概念上我们知道，$p(x/\theta)$应有极大值点，$p(x/\theta)$称为似然概率密度，所以$\hat{\theta}_{ml}$是使似然概率密度$p(x/\theta)$最大的估计，故称为最大似然估计。

物理上，最大似然估计可解释为$\hat{\theta}_{ml}$是使观测值 **x** 出现的可能性最大的θ值，或者说观测值 **x** 包含了使它本身最有可能出现的那个$\hat{\theta}_{ml}$。

最大似然估计解决了贝叶斯估计不能解决的对于未知的确定参量的估计和未知先验概率密度的随机参量的估计。对于随机参量情况，可设想它的先验密度服从均匀分布，也就是意味着对θ一无所知，这样$\partial \ln p(\theta) \big/ \partial\theta = 0$，最大后验概率估计就转化为最大似然估计。

把这里讲到的最大似然估计与前面讲到的$c-R$界和有效估计联系起来考虑，对于确定性参量：

$$\text{有效估计} \rightarrow \text{达到}c-R\text{界} \rightarrow \frac{\partial \ln p(x/\theta)}{\partial \theta} = \left(\hat{\theta} - \theta\right) K(\theta)$$

$$\text{而由似然方程} \quad \left.\frac{\partial \ln p(x/\theta)}{\partial \theta}\right|_{\theta=\hat{\theta}_{ml}} = \left(\hat{\theta} - \hat{\theta}_{ml}\right) \kappa(\theta) = 0$$

所以$\hat{\theta} = \hat{\theta}_{ml}$

所以有结论：若存在有效估计，则这个估计就是最大似然估计。

例 2 设观测数据

$$x_k = as_k + n_k \qquad k = 1, 2, \cdots, N$$

s 为信号， $n \sim N\left(o, \sigma_n^2\right)$ 估计未知幅度 a 。（相当于信号传播衰减的估计）。

讨论最简单的模型，即信道是非随机的，这样， a 就是一个未知的非随机参量。

由随机过程的知识知，似然函数为（n维条件概率密度函数）

$$p\left(\vec{\chi}/a\right)=\left(\frac{}{\sqrt{\ \ \pi}\sigma_n}\right)^N \qquad -\sum_{k=1}^{N}\frac{\left(x_k-as_k\right)^2}{\sigma_n^2}$$

求最大似然估计：

$$\left.\frac{\partial \ln p\left(\vec{\chi}/a\right)}{\partial a}\right|_{a=\hat{a}_{ml}}=-\sum_{\kappa=I}^{N}\left.\frac{\partial\left[\dfrac{\left(x_k-as_k\right)^2}{2\sigma_n^2}\right]}{\partial a}\right|_{a=\hat{a}_{ml}}$$

$$=\frac{1}{\sigma_n^2}\sum_{\kappa=1}^{N}s_k\left.\left(x_k-as_k\right)\right|_{a=\hat{a}_{ml}}=0$$

所以 $\hat{a}_{ml} = \dfrac{\sum\limits_{\kappa=1}^{N} x_k s_k}{\sum\limits_{\kappa=1}^{N} s_k^2}$

现在来讨论其估计质量：

无偏性：

$$\mathrm{E}[\hat{a}_{ml}] = \frac{1}{\sum\limits_{\kappa=1}^{N} s_k^2} \sum_{\kappa=1}^{N} E[x_k] s_k = \frac{1}{\sum\limits_{\kappa=1}^{N} s_k^2} \sum_{\kappa=1}^{N} E[a s_k + n_k] s_k = a$$

所以估计是无偏的。

估计误差的方差为：

$$D\left(\hat{a}_{ml}-a\right)=\mathrm{E}\left[\left(\hat{a}_{ml}-a\right)^{2}\right]$$

$$=\mathrm{E}\left\{\left[\frac{\sum_{\kappa=1}^{N}(as_k+n_k)s_k}{\sum_{\kappa=1}^{N}s_k^2}-a\right]^2\right\}=\frac{1}{\left(\sum_{\kappa=1}^{N}s_k^2\right)^2}\mathrm{E}\left[\left(\sum_{\kappa=1}^{N}n_k s_k\right)^2\right]$$

考虑到 s 与 n 不相关及 n 的白色性，最终有

$$D\left(\hat{a}_{ml}-a\right)=\frac{1}{\left(\sum_{\kappa=1}^{N}s_k^2\right)^2}\sum_{\kappa=1}^{N}E[(n_k s_k)^2]=\frac{\sigma_n^2}{\sum_{\kappa=1}^{N}s_k^2}$$

将$\hat{a}_{ml}$的表达式代入$\partial \ln p(\vec{\chi}/a) / \partial a$，有

$$\frac{\partial \ln p\left(\vec{\chi}/a\right)}{\partial a}=\frac{\partial\left[-\sum_{\kappa=1}^{N}\frac{(x_k-as_k)^2}{\sigma_n^2}\right]}{\partial a}=\frac{\sum_{\kappa=1}^{N}s_k^2}{\sigma_n^2}\bullet\left(\frac{\sum_{\kappa=1}^{N}s_k x_k}{\sum_{\kappa=1}^{N}s_k^2}-a\right)$$

$$=\frac{\sum_{\kappa=1}^{N}s_k^2}{\sigma_n^2}(\hat{a}_{ml}-a)\triangleq K(a)(\hat{a}_{ml}-a)$$

满足达到$c-R$界的条件，所以$\hat{a}_{ml}$是个有效估计。

根据确定参量$c-R$界的表达式，

$$D_{\min}=\frac{1}{\mathrm{E}\left\{\left[\frac{\partial\ln p\left(\vec{\chi}/\theta\right)}{\partial\theta}\right]^2\right\}}=\left[\frac{\sigma_n^2}{\sum_{\kappa=1}^{N}s_k^2}\right]^2\frac{1}{E\left\{\left[\frac{\sum_{\kappa=1}^{N}s_k x_k}{\sum_{\kappa=1}^{N}s_k^2}-a\right]^2\right\}}$$

$$=\frac{[D(\hat{a}_{ml})]^2}{D(\hat{a}_{ml})}\overset{正好}{=}D(\hat{a}_{ml})$$

例 3：发射信号回波到达时间的估计。

解：设发射信号为 $s(t)$，回波信号为 $s(t-\tau)$，τ 为回波到达时间，背景噪声假定为正态白噪声，距离一定、声速恒定时，τ 可看成是未知的确定参量，用最大似然法估计。

其接收信号为：$x(t)=as(t-\tau)+n(t)$ ； $n(t)\sim N(o,\sigma^2)$

接收信号时间序列的联合条件概率密度函数为

$$p\left(\vec{\chi}/\tau\right)=p\left[\vec{\chi}/\vec{s}(t-\tau)\right]=\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^N \exp[-\sum_{\kappa=1}^{N}\frac{(x_k-s_k)^2}{2\sigma^2}]$$

$$\approx\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^N \exp\{-\frac{N}{2\sigma^2 \mathrm{T}}\int_o^{\mathrm{T}}\left[x(t)-s(t-\tau)\right]^2 dt\}$$

取对数，有

$$\ln p\left(\vec{\chi}/\tau\right) = \ln K - C\int_o^T \left[x(t) - s(t-\tau)\right]^2 dt$$

其中 **K**，　均为常量。所以

最大似然估计$\longrightarrow$使$\ln p\left(\vec{\chi}/\tau\right)$达到最大

$\longrightarrow$使$\int_o^T \left[x(t) - s(t-\tau)\right]^2 dt$达到最小。

而

$$\int_o^T \left[x(t)-s(t-\tau)\right]^2 dt$$

$$=\int_o^T x^2(t)dt+\int_o^T s^2(t-\tau)dt-2\int_o^T x(t)s(t-\tau)dt$$

因为

$$\int_o^T x^2(t)dt=\int_o^T [as(t-\tau)+n(t)]^2 dt$$

在假设信号与噪声不相关的条件下，有 $\int_o^T s(t-\tau)n(t)dt=0$ 近似成立，所以

$$上式\approx\int_o^T a^2s^2(t-\tau)dt+\int_o^T n^2(t)dt$$

是与 τ 无关的量，而 $\int_o^T s^2(t-\tau)dt$ 也是与 τ 无关的量。

所以，使$\int_o^T [x(t)-s(t-\tau)]^2 dt$最小与$\int_o^T x(t)s(t-\tau)dt$最大等价，也就是将接收信号与已知发射信号求互相关，相关函数最大即相关峰处的信号时延$\tau_{\max}$即为所需估计的回波到达时间

$\hat{\tau}_{ml}$。

将该问题推广两个接收信号间的时延估计，请推导估计方法。

- 线性最小均方估计

前面已经讨论了几个估计准则，由它们所求出的估计量$\hat{\theta}(\vec{\chi})$往往是观测数据的复杂函数是非线性的，求解很不容易，对于噪声是非白色的情况就更困难。而且每个估计都需要预先知道有关观测数据和被估计量的先验概率密度或条件概率密度的精确表达式，这常常也难做到。所以在工程上往往是取最佳性与可行性的折中，宁愿采用容易求得的次最佳估计，而不用虽然是最佳，但给处理带来很大难度或者根本是不可实现的方法。

线性最小均方估计是在这一愿望下导得的一种估计方法。该方法人为地规定估计量$\hat{\theta}(\vec{x})$为观测数据$\vec{x}$的线性函数，然后再以估计误差方差最小为准则寻求θ的估计。下面的讨论将会发现，在这一强制性假设下，估计量的求取将会简单得多，而且它也放松了对数据和参量的先验概率知识方面的要求，只要用到它们的一、二阶数字特征。由于正态分布的一、二阶数字特征充分代表了其概率统计特性，所以，对于正态分布情况，线性最小均方估计就是贝叶斯最小均方估计。

要注意线性最小均方估计是规定估计量为观测数据的线性函数条件下的最均方估计，线性最小均方估计与最小均方估计的差别就在于多了一个人为的前提。

讨论单参数情况，$\hat{\theta}_{lms}$ 表示θ的线性最小均方估计。

设随机参量θ与观测数据有关，且θ在整个观测过程中不变，根据N个观测数据$\{\chi_1, \chi_2 \cdots \chi_N\}$对参数$\theta$作线性最小均方估计$\hat{\theta}_{lms}$。 约束：

$$\hat{\theta}_{lms} = \sum_{i=1}^{N} a_i x_i + b$$

选择a_i和b使得估计的平方误差代价函数$\overline{C}_{lms}$ 达到最小。

$$\overline{C}_{lms} = \mathrm{E}\left[\left(\hat{\theta}_{lms} - \theta\right)^2\right] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\hat{\theta}_{lms} - \theta\right)^2 p\left(\vec{x}, \theta\right) d\vec{x} d\theta$$

$$= \mathrm{E}\left[\left(\sum_{i=1}^{N} a_i x_i + b - \theta\right)^2\right]$$

数学上，如果$\overline{C}_{lms}$的极小值存在的话，$\{a_i\}$与 **b** 的取值应满足：

$$\frac{\partial \overline{C}_{lms}}{\partial a_k} = 2\mathrm{E}\left[x_k \left(\sum_{i=1}^{N} a_i x_i + b - \theta \right) \right] = \sum_{i=1}^{N} a_i \mathrm{E}\left(x_i x_k \right) + b\mathrm{E}\left(x_k \right) - \mathrm{E}\left(\theta\, x_k \right)$$

$$= 2\mathrm{E}\left[x_k (\hat{\theta} - \theta) \right] = 0 \qquad\qquad k = 1, 2, \ldots, N$$

即估计误差与各个观测数据的相关函数为零。

$$\frac{\partial \overline{C}_{lms}}{\partial b} = 2\mathrm{E}\left[\sum_{i=1}^{N} a_i x_i + b - \theta \right] = 2\left[\sum_{i=1}^{N} a_i \mathrm{E}\left(x_i \right) + b - \mathrm{E}\left(\theta \right) \right] = 0$$

整理以上两式，得

$$\begin{cases} b = \mathrm{E}(\theta) - \sum_{i=1}^{N} a_i \mathrm{E}(x_i) \\ \sum_{i=1}^{N} a_i \mathrm{E}(x_i x_k) + b\mathrm{E}(x_k) - \mathrm{E}(\theta x_k) = 0 \end{cases}$$

k=1 2 … N

解此 $N+1$ 个联立方程组，可得 $\{a_i\}$ 和 **b**，代入估计式，即可得参数 θ 的线性最小均方估计 $\hat{\theta}_{lms}$。将上式 **b** 的表达式代入下式，并整理得

$$\sum_{i=1}^{N} a_i [\mathrm{E}(x_i x_k) - \mathrm{E}(x_i)\mathrm{E}(x_k)] = \mathrm{E}(\theta x_k) - \mathrm{E}(\theta)\mathrm{E}(x_k)$$

k=1 2 … N

令

$$c_{ik} = \mathrm{E}\left(x_i x_k\right) - \mathrm{E}\left(x_i\right)\mathrm{E}\left(x_k\right)$$

$$c_{\theta x}(k) = \mathrm{E}\left(\theta x_k\right) - \mathrm{E}\left(\theta\right)\mathrm{E}\left(x_k\right)$$

有

$$\sum_{i=1}^{N} c_{ik} a_i = c_{\theta x}(k)$$ **k=1 2 … N**

写成矩阵形式，有

$$CA = C_{\theta x}$$

$$A = C^{-1} C_{\theta x}$$

其中，

$$C = \{c_{ik}\}$$

$$A = [a_1, a_2, ..., a_N]^T$$

$$C_{\theta x} = [c_{\theta x}(1), c_{\theta x}(2), ..., c_{\theta x}(N)]^T$$

为观测 **X** 的协方差矩阵， $_{\theta x}$为观测 **X** 与参数θ的互协方差矢量，当常数 **b=0** 时，有

$$A = R^{-1}R_{\theta x}$$

其中，**R** 为观测 **X** 的自相关矩阵，**R**$_{\theta x}$为观测 **X** 与参数θ的互相关矢量。这是著名的 **Yule-Walker** 方程。

与线性预测模型相似，实际上线性预测模型就是这里的线性均方估计，只是待估计参数就是信号自己。

讨论一个单数据 **x** 的特殊情况。方程变为

$$\begin{cases} b = \mathrm{E}(\theta) - a\mathrm{E}(x) \\ \mathrm{E}(\theta x) - a\mathrm{E}\left(x^2\right) - \left[\mathrm{E}(\theta) - a\mathrm{E}(x)\right]\mathrm{E}(x) = 0 \end{cases}$$

求解得，

$$a = \frac{\mathrm{E}(\theta x) - \mathrm{E}(x)\mathrm{E}(\theta)}{\mathrm{E}\left(x^2\right) - \mathrm{E}^2(x)} = \frac{\mathrm{cov}(\theta, x)}{\sigma_x^2}$$

$$b = \mathrm{E}(\theta) - a\mathrm{E}(x)$$

所以

$$\hat{\theta}_{lms} = \frac{\mathrm{cov}(\theta, x)}{\sigma_x^2}\left[x - \mathrm{E}(x)\right] + \mathrm{E}(\theta)$$

看看估计性质。

无偏性：

$$E(\hat{\theta}_{lms})=\frac{\operatorname{cov}\left(\theta,x\right)}{\sigma_x^2}E\left[x-E(x)\right]+E\left[E(\theta)\right]=E(\theta)$$

所以是无偏估计。

估计误差的方差：

$$D(\theta-\hat{\theta}_{lms})=E[(\theta-\hat{\theta}_{lms})^2]=E[(\theta-\hat{\theta}_{lms})\theta-(\theta-\hat{\theta}_{lms})\hat{\theta}_{lms}]$$

先算第一项，

$$E[(\theta-\hat{\theta}_{lms})\theta]=E(\theta^2)-E(\hat{\theta}_{lms}\theta)$$

$$=E(\theta^2)-E\{\theta[\frac{\operatorname{cov}(\theta,x)}{\sigma_x{}^2}(x-E(x))+E(\theta)]\}$$

$$=E(\theta^2)-\frac{\operatorname{cov}(\theta,x)}{\sigma_x{}^2}[E(\theta x)-E(\theta)E(x)]-E^2(\theta)$$

$$=\sigma_\theta^2-\sigma_\theta^2\frac{\operatorname{cov}^2(\theta,x)}{\sigma_\theta^2\sigma_x^2}$$

$$=\sigma_\theta^2(1-\rho^2)$$

其中，$\rho=\dfrac{\operatorname{cov}(\theta,x)}{\sigma_x\sigma_\theta}$ 为 θ 与 $\mathbf{x}$ 的相关系数。

再算估计误差方差中的第二项，

$$E[(\theta-\hat{\theta}_{lms})\hat{\theta}_{lms}]=E[(\theta-\hat{\theta}_{lms})(ax+b)]$$

$$=aE[(\theta-\hat{\theta}_{lms})x]+bE(\theta-\hat{\theta}_{lms})$$

$$=0+0=0$$

其中第一项就是平方误差代价函数对系数 $\mathbf{a_k}$ 的偏导数，所以为 **0**，即估计误差与各观测数据不相关，也即相互正交。第二项因线性最小均方估计为无偏估计而等于 **0**。

$$\therefore\ E[(\theta-\hat{\theta}_{lms})^2]=E[(\theta-\hat{\theta}_{lms})\theta]=\sigma_\theta^2(1-\rho^2)$$

总结线性最小均方估计有 4 个特点：

（1）估计 $\hat{\theta}_{lms}$ 为观测 **x** 的线性组合；

（2）是无编估计；

（ ）估计误差与观测 **x** 正交，估计 $\hat{\theta}_{lms}$ 为参量 θ 在观测 **x** 组成的空间中的正交投影。

$$(\theta - \hat{\theta}_{lms}) \perp u(x)$$

（4）只需要利用过程的一、二阶矩知识就可以确定最佳线性无偏估计。

- 最小二乘估计 $\hat{\theta}_{LS}$

贝叶斯估计，最大似然估计和线性最小均方估计等都要求观测数据和待估计量在概率分布或者数字特征方面的先验知识。最小二乘估计则不需要任何先验知识，它把估计问题作为一个确定的最佳化问题来处理。由于使用条件很宽，而且估计质量也较好，所以这是一种常被使用的估计方法。

假设观测模型是线性的，即观测数据 **x** 与被估计参量 $\theta_1,\theta_2,\cdots,\theta_M$ 之间服从下面线性关系：

$$x = c_1\theta_1 + c_2\theta_2 + \cdots + c_M\theta_M + n$$

式中 c_1,c_2，…，c_M 是已知的常系数；**n** 是观测噪声。假定在时刻 $t_1,t_2,\cdots,t_N$ 对 **x** 进行了 **N** 次线性观测，则可得到 **N** 个类似的线性方程：

$$x_i = c_{i1}\theta_1 + c_{i2}\theta_2 + \cdots + c_{iM}\theta_M + n_i$$

$$i = 1,2,\cdots,N$$

用矢量及矩阵表示，可写成

$$\vec{x} = c\vec{\theta} + \vec{n}$$

其中

$$\vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} \qquad \bar{c} = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1M} \\ c_{21} & c_{22} & \cdots & c_{2M} \\ \vdots & \vdots & \cdots & \vdots \\ c_{N1} & c_{N2} & \cdots & C_{NM} \end{bmatrix} \qquad \vec{n} = \begin{bmatrix} n_1 \\ n_2 \\ \vdots \\ n_N \end{bmatrix}$$

如果没有噪声的影响，就应该有 $\vec{x} = c\vec{\theta}$

所以，我们可以这一公式为基本出发点来构造一个性能指标：

$$J(\hat{\hat{\theta}}) = (\vec{x} - c\hat{\hat{\theta}})^T (\vec{x} - c\hat{\hat{\theta}})$$

若 N→∞，则误差的平方和则变为误差的能量。
也可构成更一般的二次型性能指标

$$J_w(\hat{\hat{\theta}}) = (\vec{x} - c\hat{\hat{\theta}})^T \overline{W} (\vec{x} - c\hat{\hat{\theta}})$$

其中$\overline{W}$ 为加权矩阵，是 N×N 的对称正定阵，所要求的估计

$\hat{\hat{\theta}}$就是使得上述规定的性能指标达到极小的一种估计。这实质是一个确定性的求极小值问题。

把 $J(\hat{\hat{\theta}})$ 表达式写成展开的形式：

$$J(\hat{\hat{\theta}}) = \sum_{i=1}^{N} (x_i - \sum_{i=1}^{N} c_{ij} \hat{\theta}_j)^2$$

在极小值点应满足导数等于零

$$\frac{\partial J}{\partial \hat{\theta}_k} = -2\sum_{i=1}^{N} c_{ik} (x_i - \sum_{j=1}^{M} c_{ij} \hat{\theta}_j) \Big|_{\hat{\theta}=\hat{\theta}_{LS}} = 0$$

$k = 1, 2, \cdots, M$

写成矩阵形式，有

$$-2\overline{C}^T (\vec{x} - \overline{C}\,\hat{\hat{\theta}}) = 0$$

实际上通过对矢量$\hat{\theta}$直接求导也可得到相同的形式，

$$\frac{\partial J}{\partial \hat{\theta}} = -2\overline{C}^T(\vec{x} - \overline{C}\,\hat{\theta})\Big|_{\hat{\theta}=\hat{\theta}_{LS}} = 0$$

解此线性方程组（显然，能得到解的条件为 **N≥M**），得最小二乘估计为：

$$\hat{\theta}_{LS}(\vec{x}) = (C^T C)^{-1} C^T \vec{x}$$

相应可得加权最小二乘估计为：

$$\hat{\theta}_{LSW}(\vec{x}) = (C^T W C)^{-1} C^T W \vec{x}$$

是观测数据$\vec{x}$的线性函数，所以最小二乘估计是线性估计。

最小二乘估计是指观测 **x** 的最小二乘估计，$\hat{x}=c\hat{\vec{\theta}}$ 是 **x** 在参数 $\vec{\theta}$ 空间中的正交投影，

$$(x-\hat{x})\perp u(\theta);(x-\hat{x})\perp\hat{x}$$

可以看出，最小二乘估计，仅仅是观测数据 **x** 的线性函数，它并不象前面讨论的方法那样，要求知道先验概率或统计知识。唯一的条件是线性观测。

回顾前面线性最小均方估计也是观测数据 **x** 的线性函数，但两者是有本质区别的

一方面，最小二乘估计假设了观测数据与参数之间的线性加权模型，而线性最小均方估计没有这一假设，只是把估计用观测数据 **x** 的线性函数来近似。

另一方面，最小二乘估计只是对观测数据的最小方差拟合以求得模型参数，最小方差只是针对有限个观测数据而言，而线性最小均方估计仍然是在统计意义上的最小误差方差，最小方差是针对整个随机过程的。

最小二乘估计的性质：

（1）无编性

$$E(\hat{\vec{\theta}}_{ls})=E\{(C^TC)^{-1}C^T\vec{x}\}=(C^TC)^{-1}C^TE(\vec{x})$$

$$=(C^TC)^{-1}C^T[CE(\vec{\theta})+E(\vec{n})]$$

同理 $E(\hat{\vec{\theta}}_{lsw})=(C^TWC)^{-1}C^TW[CE(\vec{\theta})+E(\vec{n})]$

所以当观测噪声为零均值时，即 $E(\vec{n})=0$ 时，估计是无偏的。

$$E(\hat{\vec{\theta}}_{LS})=E(\hat{\vec{\theta}}_{LSW})=E(\vec{\theta})$$

（**2**）估计误差的方差阵（设观测噪声为零均值）

因估计参量是一组量，所以要用误差方差阵来表示，

$$D(\hat{\vec{\theta}}_{LS}) = E\{[\vec{\theta} - \hat{\vec{\theta}}_{LS}][\vec{\theta} - \hat{\vec{\theta}}_{LS}]^T\}$$

化简得 $$= (C^T C)^{-1} C^T E(\vec{n} \cdot \vec{n}^T) C (C^T C)^{-1}$$

$$= (C^T C)^{-1} C^T D(\vec{n}) C (C^T C)^{-1}$$

同理得 $$D(\hat{\vec{\theta}}_{LSW}) = (C^T W C)^{-1} C^T W D(\vec{n}) W C (C^T W C)^{-1}$$

式中 $D(\vec{n})$ 是对称正定矩阵，为观测噪声的方差阵。

如果取$W=D^{-1}(\vec{n})$，则有

$$D(\hat{\bar{\theta}}_{LSW})=[C^T D^{-1}(\vec{n})C]^{-1}$$

可以证明，对任意的加权阵，有

$$D(\hat{\bar{\theta}}_{LSW})\geq[C^T D^{-1}(\vec{n})C]^{-1}$$

例 **4**：试根据两次观测

$$\begin{bmatrix}3\\1\\2\end{bmatrix}=\begin{bmatrix}1 & 1\\0 & 1\\1 & 0\end{bmatrix}\bullet\begin{bmatrix}x_1\\ \\x_2\end{bmatrix}+\begin{bmatrix}n_1\\n_2\\n_3\end{bmatrix}$$

$$5=\begin{bmatrix}1 & 2\end{bmatrix}\bullet\begin{bmatrix}x_1\\x_2\end{bmatrix}+n_4$$

求 $x=\begin{bmatrix}x_1 & x_2\end{bmatrix}$的最小二乘估计。

解：令

$$z_1 = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}, \qquad z_2 = 5$$

$$C_1 = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix} \qquad C_2 = \begin{bmatrix} 1 & 2 \end{bmatrix}$$

得到总观测矢量为

$$z = \begin{bmatrix} 3 & 1 & 2 & 5 \end{bmatrix}^T$$

常系数矩阵为

$$C = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix}$$

由最小二乘估计公式得

$$\hat{X}_{Ls} = (C^T C)^{-1} C^T Z$$

$$= \left\{ \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 2 \end{bmatrix} \right\}^{-1} \begin{bmatrix} 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 2 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \\ 5 \end{bmatrix}$$

$$= \begin{bmatrix} 3 & 3 \\ 3 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 10 \\ 14 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 10 \\ 14 \end{bmatrix} = \begin{bmatrix} 2 \\ 4/3 \end{bmatrix}$$

-7 正交原理

在线性最小均方估计的讨论中，我们已经知道，其估计误差与各个观测数据的相关函数为零，这在概率论中是一个正交的概念，即最小均方估计误差与观测数据相正交。

$$E\left[x_k(\hat{\theta}_{lms}-\theta)\right]=0 \qquad k=1,2,\cdots,N$$

所以线性最小均方估计可以直接利用上式正交条件求解系数 $\{a_i\},b$。对于[redacted]最小均方估计也是如此，这就是所谓的正交原理。

在前评我 与

设$\hat{\theta}$代表任意其它线性估计，$\hat{\theta}=a'x+b$，而$\hat{\theta}'=ax+b$满足正交条件。

$$E\left\{[\theta-a'x-b]^2\right\}$$

$$=E\left\{[\theta-ax-b+ax-a'x]^2\right\}$$

$$=E\left[(\theta-ax-b)^2\right]+2(a-a')E\left[x(\theta-ax-b)\right]$$

$$+(a-a')^2E(x^2)$$

由正交化条件

$$=E\left[(\theta-ax-b)^2\right]+(a-a')^2E\left[x^2\right]\geq E\left[\left(\theta-ax-b\right)^2\right]$$

不等式说明满足正交化条件的线性估计是所有线性估计中误差最小的。所以线性最小均方估计的充分必要条件是估计误差与观测数据正交。

正交条件实际上就是借用几何的观点把不相关性解释为正交性（互相垂直）。θ 与 **x** 本来不正交，但从 θ 中减去一个 **x** 的线性函数构成的$\hat{\theta}_{lms}$后，便与 **x** 正交。因此可以说$\hat{\theta}_{lms}$是 θ 在 **x** 上的正交投影。

另有一点要说明的是，最小方差估计同样也满足正交原理。因为

$$\mathrm{E}\left\{[\theta-\hat{\theta}_{ms}]x\right\}=\mathrm{E}\left\{\left[\theta-\mathrm{E}(\theta/x)\right]x\right\}=\mathrm{E}\left\{\theta x-\mathrm{E}(\theta/x)x\right\}$$

$$=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta\, x\, p\left(\theta,x\right)d\theta dx-\int_{-\infty}^{\infty}x\mathrm{E}(\theta/x)p\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}x[\int_{-\infty}^{\infty}\theta p(\theta/x)p\left(x\right)d\theta]dx-\int_{-\infty}^{\infty}x\mathrm{E}(\theta/x)p\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}x[\int_{-\infty}^{\infty}\theta p(\theta/x)d\theta]p\left(x\right)dx-\int_{-\infty}^{\infty}x\mathrm{E}(\theta/x)p\left(x\right)dx$$

$$=\int_{-\infty}^{\infty}x\mathrm{E}(\theta/x)p\left(x\right)dx-\int_{-\infty}^{\infty}x\mathrm{E}(\theta/x)p\left(x\right)dx=0$$

即最小方差估计误差与观测 **x** 的相关函数为 **0**，满足正交性。

习题：

习题 1

测量数据为

$$x_k = a + n_k \qquad k = 1,2,\cdots,N$$

其中 **a** 为未知的确定参量，$n_k \sim N(0,\sigma_n^2)$，求最大似然估计和最小二乘估计。

习题 2

设有观测序列 $\{z_i = x + v_i\}$ $\quad i = 1,2,\cdots,N$，其中 **x** 为正态随机变量 $N(0,\sigma_x^2)$，观测噪声为互相独立的正态随机变量 $N(0,\sigma_v^2)$，求最大后验概率估计 $\hat{x}_{MA}$

习题

设用已知时间信号$b(n)$激励一传递函数未知的信道，其输出为带畸变的信号$x(n)$，为了消除信道的畸变影响，设$x(n)$通过一个$2k+1$抽头的横向滤波器以产生信号

$$r(n)=\sum_{k=-K}^{K}c_i x(n-i)$$

试利用 **n=0 1 … N-1** 的 **r(n)**值及 **n=-K -K+1 … N-1+K** 的 **x(n)**值，求使 $\sum_{N=0}^{N-1}\left|r(n)-b(n)\right|^2$ 为最小的滤波器系数集｛$\mathbf{c_i}$｝。

习题 4

设有二元通信系统，观测序列$\{z_i=x+v_i\}$ $i=1,2,\cdots,N$，其中信号 **x** 等概率地取值**+s** 和**-s**，观测噪声 $\mathbf{v_i}$ 为正态白噪声，$\mathbf{v_i}$～$N(0,\sigma_v^2)$，求 **x** 的最小方差估计$\hat{x}_{ms}$。