3 静态参数估计理论

蔡远利

✧ 静态参数估计

- 点估计
- 区间估计

✧ 经典点估计理论

- 最小二乘估计(K.F. Gauss)
- 极大似然估计（R.A. Fisher）
- 贝叶斯估计(T. Bayes)

✧ 工程应用

- 系统参数辨识

根据测量到的输入-输出数据来评估已知数学模型中的系数

- 飞行器气动参数辨识

从飞行测量数据中提取飞行器的气动导数（飞行器气动导数包含在飞行器动态方程系数中）

3.0 问题描述

✧ 设X为一随机矢量，Z为另外一个与X统计相关随机矢量，已知随机矢量Z的一个样本（实现）z；

✧ 推断随机矢量X最可能的取值 $\hat{x}$，或称想办法获得随机矢量X的最优估计值$\hat{x}$。

【Remarks】

- 显然，估计值$\hat{x}$应该是z的函数，可以记为 $\hat{x}(z)$ 。
- 更一般地，函数 $\hat{x}(Z)$ 称为由Z构成的统计量。
- 在不至混淆的情况下，以后我们将对随机量及其样本采用相同的符号。
- 符号惯例说明见表 3-1。

表 3-1 符号惯例

符号	含义
$\boldsymbol{x}$	被估计量
$\boldsymbol{z}$	测量（量测）量
$\hat{\boldsymbol{x}}$	被估计量的估计值，是量测量的向量函数
$\tilde{\boldsymbol{x}}$	估计误差，$\boldsymbol{x}-\hat{\boldsymbol{x}}$

3.1 贝叶斯估计(Bayes Estimation)

[定义3-1] (代价函数) 若标量函数$L[x-\hat{x}(z)]=L(\tilde{x})$满足:

1) 当$\|\tilde{x}_2\| \geq \|\tilde{x}_1\|$时,有$L(\tilde{x}_2) \geq L(\tilde{x}_1)$;

2) 当$\tilde{x}=0$ 时, $L(\tilde{x})=0$;

3) $L(\tilde{x})=L(-\tilde{x})$.

则称$L(\tilde{x})$为用$\hat{x}(z)$对x进行估计时的**代价函数**(或**损失函数**).

[定义3-2] (贝叶斯风险) $B(\tilde{x}) = E[L(\tilde{x})]$称为估计$\hat{x}(z)$的贝叶斯风险（Bayes' Risk）。

[定义3-3] (贝叶斯估计) 若$\hat{x}(z)$是使$B(\tilde{x}) \Rightarrow \min$的估计，那么称$\hat{x}(z)$为$x$的贝叶斯估计。

3.1.1 最小方差估计

[定义3-4] (**最小方差估计**) 取$L(\tilde{x}) = [x - \hat{x}(z)]^T[x - \hat{x}(z)]$的贝叶斯估计，称为最小方差估计，记为$\hat{x}_{MV}(z)$.

[定理 3-1] (**条件均值**)

$$\hat{x}_{MV}(z) = E[x|z] \tag{3.1.1}$$

[证明] 根据最小方差估计的定义，可知

$$
\begin{aligned}
B(\tilde{x}) &= E[x-\hat{x}(z)]^T[x-\hat{x}(z)] \\
&= \iint_{-\infty}^{+\infty}[x-\hat{x}(z)]^T[x-\hat{x}(z)]f(x,z)dxdz \\
&= \int_{-\infty}^{+\infty} f_z(z)\left\{\int_{-\infty}^{+\infty}[x-\hat{x}(z)]^T[x-\hat{x}(z)]f(x|z)dx\right\}dz \\
&= \int_{-\infty}^{+\infty} f_z(z)B_z(\tilde{x})dz
\end{aligned}
$$

不难发现

$$B(\tilde{x}) \Rightarrow \min \Leftrightarrow B_Z(\tilde{x}) \Rightarrow \min$$

由极值必要条件，可得

$$\frac{\partial B_Z(\tilde{x})}{\partial \hat{x}} = -2\int_{-\infty}^{+\infty} (x-\hat{x}) f(x|z)dx = 0$$

$$\Rightarrow \hat{x}_{MV}(z) = \int_{-\infty}^{+\infty} x f(x|z)dx = E[x|z]$$

验证充分条件

$$\frac{\partial^2 B_Z(\tilde{x})}{\partial^2 \hat{x}} = 2I > 0 \text{ (Positive)}$$

证毕!

[推论 3-1] 最小方差估计的均值及估计误差协方差分别为

$$(1)\, E\hat{x}_{MV}(z) = Ex \quad (E\tilde{x}_{MV} = 0)$$

$$(2)\, P_{\tilde{x}_{MV}} = \int_{-\infty}^{+\infty} P_{x|z} f_z(z)dz = E_z P_{x|z}$$

[证明] 根据均值的定义

$$E\overline{x|z} = \int_{-\infty}^{+\infty} \overline{x|z} f_z(z)dz$$

$$= \iint_{-\infty}^{+\infty} x f_{x|z}(x) f_z(z) dx dz$$

$$= \iint_{-\infty}^{+\infty} x f_{xz}(x, z) dx dz = \bar{x}$$

因此$E\hat{x}_{MV}(z) = Ex$，即$E\tilde{x}_{MV} = 0$. 再由协方差的定义

$$P_{\tilde{x}_{MV}} = E\tilde{x}_{MV}\tilde{x}_{MV}^T$$

$$= E\{(x - \overline{x|z})(x - \overline{x|z})^T\}$$

$$= \int_{-\infty}^{+\infty} f_z(z)dz \int_{-\infty}^{+\infty} (x - \overline{x|z})(x - \overline{x|z})^T f_{x|z}(x)dx$$

$$= E_z P_{x|z}$$

证毕!

[推论 3-2] 若x与z相互独立，则

$$\hat{x}_{MV} = E[x|z] = E\,x = \bar{x}.$$

[证明] 由贝叶斯公式可知

$$E[x|z]=\int_{-\infty}^{+\infty} x f_{x|z}(x)dx=\int_{-\infty}^{+\infty} x\frac{f_{xz}(x,z)}{f_z(z)}dx$$

$$=\int_{-\infty}^{+\infty} x\frac{f_x(x)f_z(z)}{f_z(z)}dx=\int_{-\infty}^{+\infty} x f_x(x)dx$$

所以，x与z相互独立时$E[x|z]=E\,x=\bar{x}$，得证!

[推论 3-3] (最小方差估计一般求解公式)

$$\hat{x}_{MV}(z)=\frac{\int_{-\infty}^{+\infty}x\,f_{z|x}(z|x)f_x(x)\mathrm{d}x}{\int_{-\infty}^{+\infty}f_{z|x}(z|x)f_x(x)\mathrm{d}x} \qquad (3.1.2)$$

[证明] 因为

$$\hat{x}_{MV}(z)=\overline{x|z}=\int_{-\infty}^{+\infty}x\,f_{x|z}(x|z)\mathrm{d}x$$

$$= \int_{-\infty}^{+\infty} x \frac{f_{xz}(x,z)}{f_z(z)} \mathrm{d}x = \frac{\int_{-\infty}^{+\infty} x\, f_{xz}(x,z)\mathrm{d}x}{\int_{-\infty}^{+\infty} f_{xz}(x,z)\mathrm{d}x}$$

分子、分母分别再一次应用贝叶斯公式

$$f_{xz}(x,z) = f_{z|x}(z|x) f_x(x),$$

即得证。

[定理 3-2] (最优估计的不变性) 假设

(1) $L(\tilde{x})$ 关于$\tilde{x}=0$对称且凹;

(2) $B_z(\tilde{x})=\int_{-\infty}^{+\infty}L(x-\hat{x}_B)f(x|z)dx$存在;

(3) $f(x|z)$ 关于条件均值$\hat{x}_{MV}=E[x|z]$对称且凸。

那么，$\hat{x}_B=\hat{x}_{MV}$.

[证明] **（略）**

[定理 3-3] (随机正交原理) 设x与z是两随机矢量，那么对任意的函数$g(z)$，均有

$$E\, g(z)(x - E[x|z])^T = 0 \tag{3.1.3}$$

[证明] $E\, g(z)(x - E[x|z])^T = E\, g(z)x^T - E\, g(z)\overline{x|z}^T$

$$= E\, g(z)x^T - \int g(z)\,[\int x f_{x|z}(x)dx]^T f_z(z)dz$$

$$= \overline{g(z)x^T} - \iint g(z)x^T f(x,z)dxdz = 0$$

证毕!

[推论 3-4] (随机投影定理) 设$g(\bullet)$为任一函数，那么

$$E\|x - E[x|z]\|^2 \le E\|x - g(z)\|^2 \qquad (3.1.4)$$

[证明] 根据范数的定义

$$\|x-g(z)\|^2=\|x-E[x|z]+E[x|z]-g(z)\|^2$$

$$=\|x-E[x|z]\|^2+\|E[x|z]-g(z)\|^2$$

$$+(x-E[x|z])^T(E[x|z)-g(z))+\left(E[x|z]-g(z)\right)^T(x-E[x|z])$$

根据随机正交定理(3.1.3)，可知

$$E\|x-g(z)\|^2 = E\|x-E[x|z]\|^2 + E\|E[x|z]-g(z)\|^2$$

因此 $E\|x-E[x|z]\|^2 \le E\|x-g(z)\|^2$。证毕!

[定理 3-4] (高斯条件均值与协方差) 设

$$x\sim N(\bar{x},P_x), z\sim N(\bar{z},P_z),\quad y=[x^T,z^T]^T\sim N(\bar{y},P_y),$$

其中，$P_y = \begin{bmatrix} P_x & P_{xz} \\ P_{zx} & P_z \end{bmatrix}$, $P_{xz} = E(x-\bar{x})(z-\bar{z})^T$. 那么

$$E[x|z] = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z}) \tag{3.1.5}$$

$$P_{x|z} = P_x - P_{xz}P_z^{-1}P_{zx} \tag{3.1.6}$$

【证明】因为

$$f_{x|z}(x|z) = \frac{f_{xz}(x,z)}{f_z(z)} = \frac{f_y(y)}{f_z(z)}$$

$$= \frac{\sqrt{(2\pi)^m|P_z|}}{\sqrt{(2\pi)^{n+m}|P_y|}} \exp\{-\frac{1}{2}[(y-\bar{y})^T P_y^{-1}(*) - (z-\bar{z})^T P_z^{-1}(*)]\}$$

考虑到

$$P_y^{-1}=\begin{bmatrix} D^{-1} & -D^{-1}P_{xz}P_z^{-1} \\ -P_z^{-1}P_{zx}D^{-1} & P_z^{-1}+P_z^{-1}P_{zx}D^{-1}P_{xz}P_z^{-1} \end{bmatrix} \tag{3.1.7}$$

其中，$D=P_x-P_{xz}P_z^{-1}P_{zx}$.

从而有

$$\begin{bmatrix} x-\bar{x} \\ z-\bar{z} \end{bmatrix}^T P_y^{-1}\begin{bmatrix} x-\bar{x} \\ z-\bar{z} \end{bmatrix}-(z-\bar{z})^T P_z^{-1}(z-\bar{z})=[x-E(x|z)]^T P_{x|z}^{-1}[*]$$

其中，$E(x|z)=\bar{x}+P_{xz}P_z^{-1}(z-\bar{z}), P_{x|z}=P_x-P_{xz}P_z^{-1}P_{zx}$.

注意到

$$\frac{|P_y|}{|P_z|} = |P_{x|z}| \qquad (3.1.8)$$

于是可导出

$$f_{x|z}(x|z) = \frac{1}{\sqrt{(2\pi)^2|P_{x|z}|}} \exp\{-\frac{1}{2}[x - E(x|z)]^T P_{x|z}^{-1}[**]\}$$

证毕!

[推论 3-5] 在[定理 3-4]条件下，$P_{\tilde{x}_{MV}} = E_z P_{x|z} = P_{x|z}$，即

$\tilde{x}_{MV} \sim N(0, P_x - P_{xz} P_z^{-1} P_{zx})$。

[证明] 由[推论 3-1]可知

$$P_{\tilde{x}_{MV}} = \int_{-\infty}^{+\infty} P_{x|z} f_z(z) dz = E_z P_{x|z}$$

而[定理 3-4]表明$P_{x|z} = P_x - P_{xz} P_z^{-1} P_{zx}$，是常值矩阵，因此

$E_z P_{x|z} = P_{x|z}$。结合最小方差估计是无偏估计的结论，所以[定理 3-4]条件下$\tilde{x}_{MV} \sim N(0, P_x - P_{xz} P_z^{-1} P_{zx})$。 证毕!

[定理 3-5] 若$z = Hx + v, v \sim N(0, R), x \sim N(\bar{x}, P_x)$，而且$Exv^T = 0$（正交），则有

$$E[x|z] = \bar{x} + P_{x|z} H^T R^{-1} (z - H\bar{x}) \tag{3.1.9}$$

$$P_{x|z} = P_{\tilde{x}_{MV}} = (P_x^{-1} + H^T R^{-1} H)^{-1} \tag{3.1.10}$$

[证明] 根据给定的量测方程$z = Hx + v$，可导出

$$\bar{z} = H\bar{x}, \qquad \breve{z} = z - \bar{z} = Hx + v - H\bar{x} = H\breve{x} + v$$

$$P_z = E\breve{z}\breve{z}^T = E(H\breve{x} + v)(H\breve{x} + v)^T = HP_xH^T + R$$

$$P_{xz} = E\breve{x}\breve{z}^T = E\breve{x}(H\breve{x} + v)^T = P_xH^T$$

根据[定理 3-4]可得

$$E[x|z] = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z})$$

$$= \bar{x} + P_x H^T (HP_x H^T + R)^{-1}(z - H\bar{x})$$

$$P_{x|z} = P_{\tilde{x}_{MV}} = P_x - P_{xz}P_z^{-1}P_{zx} = P_x - P_x H^T (HP_x H^T + R)HP_x$$

由矩阵求逆引理

$$[(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1} = A_{22}^{-1} + A_{22}^{-1}A_{21}(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}A_{12}A_{22}^{-1}]$$

$$P_x - P_x H^T (HP_x H^T + R)HP_x = (P_x^{-1} + H^T R^{-1} H)^{-1}$$

所以在线性量测高斯分布条件下，最小方差估计为

$$E[x|z] = \bar{x} + P_{x|z}H^T R^{-1}(z - H\bar{x})$$

$$P_{x|z} = P_{\tilde{x}_{MV}} = (P_x^{-1} + H^T R^{-1} H)^{-1}$$

证毕!

上述估计误差协方差关系可以表示为

$$P_{x|z}^{-1} = P_x^{-1} + H^T R^{-1} H \qquad (3.1.11)$$

具有重要的物理含义。另外，注意到

$$I - P_{x|z}H^TR^{-1}H = P_{x|z}\left(P_{x|z}^{-1} - H^TR^{-1}H\right) = P_{x|z}P_x^{-1} \quad (3.1.12)$$

我们可得[定理 3-5]的另外一种形式。

[定理 3-6] (线性量测高斯分布最小方差估计) 若 $z = Hx + v, v \sim N(0,R), x \sim N(\bar{x}, P_x)$，且 $Exv^T = 0$，则

$$\hat{x}_L = P_{\tilde{x}_L}(P_x^{-1}\bar{x} + H^TR^{-1}z) \quad (3.1.13)$$

$$P_{\tilde{x}_L}^{-1} = P_x^{-1} + H^TR^{-1}H \quad (3.1.14)$$

[例3-1] 设$x \sim U[0,1]$，即$f_x(x) = \begin{cases} 1, & 0 \le x \le 1 \\ 0, & \text{otherwise} \end{cases}$。

（1） 求x的验前估计$\bar{x}$;

（2） x的量测为$z = \ln\frac{1}{x} + v$，v服从指数分布，即

$$f_v(v) = \begin{cases} e^{-v}, v \ge 0 \\ 0, v < 0 \end{cases}$$

此外假设x与v相互独立，试求x的最小方差估计。

[解] (1) $\bar{x} = \int_{-\infty}^{+\infty} x f_x(x) dx = \frac{1}{2}$;

(2)因为x与v相互独立，所以

$$f_{z|x}(z|x) = f_v\left(z - \ln\frac{1}{x}\right)$$

$$= \begin{cases} e^{-z-\ln\frac{1}{x}}, & z - \ln\frac{1}{x} \geq 0 \\ 0, & z - \ln\frac{1}{x} < 0 \end{cases} = \begin{cases} \frac{1}{x}e^{-z}, & x \geq e^{-z} \\ 0, & x < e^{-z} \end{cases}$$

由[推论 3-3]可知

$$\hat{x}_{MV}(z) = \frac{\int_{e^{-z}}^{1} e^{-z} dx}{\int_{e^{-z}}^{1} \frac{1}{x} e^{-z} dx} = \frac{e^{-z}(1-e^{-z})}{e^{-z} \ln x \,|_{e^{-z}}^{1}} = \frac{1-e^{-z}}{z}.$$

3.1.2 线性最小方差估计

[定义3-5] 取$L(\tilde{x}) = [x - \hat{x}(z)]^T[x - \hat{x}(z)]$且形式上限制为$\hat{x} = a + Bz$的 Bayes 估计，其中$a$和$B$分别是常系数向量和矩阵，则称为线性最小方差估计，记为$\hat{x}_L(z)$.

[引理3-1] 设 A,B,C 是合适维数的矩阵，关于矩阵迹有如下公式：

$$\mathrm{tr}(AB) = \mathrm{tr}(BA)$$

$$\frac{\partial \mathrm{tr}\left(ABA^{T}\right)}{\partial A} = AB+AB^{T}$$

$$\frac{\partial \mathrm{tr}(BAC)}{\partial A} = B^{T}C^{T}$$

[引理3-2] 令$L=[x-a-Bz]^T[x-a-Bz]$，有

$$\frac{\partial L}{\partial a}=2(a+Bz-x) \tag{3.1.15}$$

$$\frac{\partial L}{\partial B}=2(Bz+a-x)z^T \tag{3.1.16}$$

[证明] 展开L得

$$L=x^Tx+a^Ta+z^TB^TBz-2x^Ta+2a^TBz-2x^TBz$$

直接可得式(3.1.15)。另外，对于任意相同维数向量u和v，有

$$u^T v = \mathrm{tr}(vu^T)$$

因此

$$\frac{\partial L}{\partial B} = \frac{\partial}{\partial B}(z^T B^T Bz + 2a^T Bz - 2x^T Bz)$$

$$= \frac{\partial}{\partial B}\mathrm{tr}(Bzz^T B^T + 2Bza^T - 2Bzx^T)$$

根据[引理 3-1]，可导出

$$\frac{\partial L}{\partial B} = 2Bzz^T + 2az^T - 2xz^T$$

$$= 2(Bz + a - x)z^T$$

证毕!

[定理 3-7]　(线性最小方差估计的充要条件)

$$E\{[x-\hat{x}_L(z)]z^T\}=0 \tag{3.1.17}$$

$$\mathrm{E}[\,x-\hat{x}_L(z)]=0 \tag{3.1.18}$$

[证明] (1)充分性：设以上两式成立，对任意的$\hat{x}$，有

$$E\|x-\hat{x}\|^2=E\|x-\hat{x}_L(z)+\hat{x}_L(z)-\hat{x}\|^2$$

$$=E\|x-\hat{x}_L(z)\|^2+E\|\hat{x}_L(z)-\hat{x}\|^2$$

$$+E[x-\hat{x}_L(z)]^T[\hat{x}_L(z)-\hat{x}]+E[\hat{x}_L(z)-\hat{x}]^T[x-\hat{x}_L(z)]$$

记 $\hat{x}_L(z)=a_o+B_o z$, $\hat{x}=a+Bz$，那么 $\hat{x}_L(z)-\hat{x}=(a_o-a)+(B_o-B)z$, 考虑到

$$E[\hat{x}_L(z)-\hat{x}][x-\hat{x}_L(z)]^T$$

$$=E[a_o-a][x-\hat{x}_L(z)]^T$$

$$+E\{(B_o-B)z[x-\hat{x}_L(z)]^T\}=0$$

说明$E[\hat{x}_L(z)-\hat{x}]^T[x-\hat{x}_L(z)]=0$。类似地，可知$E[x-\hat{x}_L(z)]^T[\hat{x}_L(z)-\hat{x}]=0$。因而有

$$E\|x-\hat{x}\|^2=E\|x-\hat{x}_L(z)\|^2+E\|\hat{x}_L(z)-\hat{x}\|^2$$

所以

$$E\|x-\hat{x}_L(z)\|^2\leq E\|x-\hat{x}\|^2$$

说明$\hat{x}_L(z)$是线性最小方差估计。

(2)必要性：设$\hat{x}_L(z)=a_o+B_oz$是线性最小方差估计，由式(3.1.17)可得

$$\frac{\partial}{\partial a}E\|x-a-Bz\|^2\Big|_{a=a_o,B=B_o}=-2E[x-a_o-B_oz]=0$$

即$E[x-\hat{x}_L(z)]=0$. 而由式(3.1.18)可得

$$\frac{\partial}{\partial B}E\|x-a-Bz\|^2\Big|_{a=a_o,B=B_o}=-2E\{[x-a_o-B_oz]z^T\}=0$$

因此 $E\{[x-\hat{x}_L(z)]z^T\}=0$。证毕!

[推论 3-6] 对任意合适维数的矩阵A和向量b，都有

$$E\|x-\hat{x}_L(z)\|^2 \le E\|x-Az-b\|^2 \tag{3.1.19}$$

[定理 3-8] (线性最小方差估计及其协方差)

$$\hat{x}_L(z)=\bar{x}+P_{xz}P_z^{-1}(z-\bar{z}) \tag{3.1.20}$$

$$P_{\tilde{x}_L}=E\tilde{x}_L\tilde{x}_L^T=P_x-P_{xz}P_z^{-1}P_{zx} \tag{3.1.21}$$

[证明]设$\hat{x}_L(z) = a_o + B_o z$，根据线性最小方差估计充要条件(3.1.18)可知

$$\bar{x} = a_o + B_o \bar{z} \tag{3.1.22a}$$

而由(3.1.17)可得

$$E\{[x - \bar{x} + \bar{x} - \hat{x}_L(z)]z^T\} = 0$$

$$\Rightarrow \; E\{[x - \bar{x}]z^T\} = E\{[\hat{x}_L(z) - \bar{x}]z^T\}$$

因此

$$P_{xz} = E\{[\hat{x}_L(z) - \bar{x}][z - \bar{z}]^T\} \qquad (3.1.22b)$$

注意到 $\hat{x}_L(z) - \bar{x} = B_o(z - \bar{z})$

上式两边右乘$[z - \bar{z}]^T$并取数学期望，并带入$(3.1.22b)$，得

$$P_{xz} = B_o P_z \Rightarrow B_0 = P_{xz} P_z^{-1}$$

将上式带入$(3.1.22a)$，可得

$$a_o = \bar{x} - P_{xz}P_z^{-1}\bar{z}$$

将a_o, B_o带入$\hat{x}_L(z) = a_o + B_o z$，即得

$$\hat{x}_L(z) = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z})$$

由上式可知

$$P_{\hat{x}_L} = E\{[\hat{x}_L(z) - x][\hat{x}_L(z) - x]^T\}$$

$$= E\{[\bar{x} - x + P_{xz}P_z^{-1}(z - \bar{z})][\bar{x} - x + P_{xz}P_z^{-1}(z - \bar{z})]^T\}$$

$$= P_x - P_{xz}P_z^{-1}P_{zx}$$

上式中用到了$P_z^T = P_z, P_{zx} = P_{xz}^T$。证毕!

[定理 3-9] 当(x,z) 服从高斯分布时，$\hat{x}_L(z) = \hat{x}_{MV}(z)$.

[证明]在高斯分布假设下，根据[定理 3-4]可知

$$\hat{x}_{MV}(z) = E[x|z] = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z})$$

比较(3.1.20)即得证. 此外，还有$P_{x|z} = P_{\tilde{x}_L}$. 即在正态分布情况下，线性最小方差估计即为最小方差估计. 证毕!

[推论 3-7] 设z为测量值，x_1, x_2为未知的随机量，A, B, r是确定性量，此外，$y = Ax_1 + Bx_2 + r + w$。w为噪声，与z不相关，而且$Ew = 0$，那么

(1) $\hat{y}_L(z) = A\hat{x}_{1L}(z) + B\hat{x}_{2L}(z) + r$;

(2) $E\tilde{y}_L(z) = 0.$

[证明]由[定理 3-8]可知

$$\hat{y}_L(z) = \bar{y} + P_{yz}P_z^{-1}(z - \bar{z})$$

而
$$\bar{y} = A\bar{x}_1 + B\bar{x}_2 + r$$

$$P_{yz} = E[y - \bar{y}][z - \bar{z}]^T = AP_{x_1z} + BP_{x_2z}$$

所以

$$\hat{y}_L(z) = A\bar{x}_1 + B\bar{x}_2 + r + \left(AP_{x_1z} + BP_{x_2z}\right)P_z^{-1}(z - \bar{z})$$

$$= A\bar{x}_1 + AP_{x_1z}P_z^{-1}(z - \bar{z}) + B\bar{x}_2 + BP_{x_2z}P_z^{-1}(z - \bar{z}) + r$$

$$= A\hat{x}_{1L}(z) + B\hat{x}_{2L}(z) + r$$

另外

$$E\hat{y}_L(z) = A\bar{x}_1 + B\bar{x}_2 + r = \bar{y}$$

所以

$$E\tilde{y}_L(z) = E[\hat{y}_L(z) - E\hat{y}_L(z)] = E[\hat{y}_L(z) - \bar{y}] = 0$$

证毕!

[推论 3-8] 若$z_1, z_2, \ldots, z_k$ 互不相关,那么

$$\hat{x}_L(z_1, z_2, \ldots, z_k)$$

$$= \hat{x}_L(z_1) + \hat{x}_L(z_2) + \cdots + \hat{x}_L(z_k) - (k-1)\bar{x}$$

[证明] 令

$$z = [z_1^T, z_2^T, \cdots, z_k^T]^T$$

那么

$$\bar{z} = [z_1^T, z_2^T, \cdots, \bar{z}_k^T]^T$$

$$P_{xz} = E(x - \bar{x})(z - \bar{z})^T = [P_{xz_1}, P_{xz_2}, \cdots, P_{xz_k}]$$

$$P_z = \begin{bmatrix} P_{z_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & P_{z_k} \end{bmatrix}$$

根据[定理 3-8]，可知

$$\hat{x}_L(z) = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z}) = \bar{x} + \sum_{i=1}^{k} P_{xz_i} P_{z_i}^{-1}$$

考虑到

$$\hat{x}_L(z_i) = \bar{x} + P_{xz_i}P_{z_i}^{-1}$$

所以

$$\hat{x}_L(z) = \sum_{i=1}^{k} \hat{x}_L(z_i) - (k-1)\bar{x}$$ 证毕!

[例3-2] 设 $x \sim N(\bar{x}, P_x)$，$v \sim N(0, P_v)$，两者不相关，如果

$$z = Hx + v \tag{3.1.23}$$

试求x的线性最小方差估计。

[解] 根据线性最小方差估计([定理 3-8])

$$\hat{x}_L(z) = \bar{x} + P_{xz} P_z^{-1} (z - \bar{z})$$

$$P_{\tilde{x}_L} = P_x - P_{xz} P_z^{-1} P_{zx}$$

由(3.1.23)可知$\bar{z} = H\bar{x}$， 令 $\breve{x} = x - \bar{x}$, $\breve{z} = z - \bar{z} = H\breve{x} + v$,

可得

$$P_z = E[\breve{z}\breve{z}^T] = HP_xH^T + P_v$$

$$P_{xz} = E(\breve{x})(\breve{z})^T = E(\breve{x})(H\breve{x} + v)^T = P_xH$$

因此

$$\hat{x}_L(z) = \bar{x} + P_xH(HP_xH^T + P_v)^{-1}(Hz - \bar{z}) \qquad (3.1.24)$$

$$P_{\tilde{x}_L} = P_x - P_x H(HP_x H^T + P_v)^{-1} H^T P_x \qquad (3.1.25)$$

考虑到

$$I - P_x H(HP_x H^T + P_v)^{-1} H = P_{\tilde{x}_L} P_x^{-1}$$

$$P_x H(HP_x H^T + P_v)^{-1} H = I - P_{\tilde{x}_L} P_x^{-1} = P_{\tilde{x}_L}(P_{\tilde{x}_L}^{-1} - P_x^{-1})$$

将以上两式代入(3.1.24)，可得

$$\hat{x}_L(z) = P_{\tilde{x}_L} P_x^{-1} \bar{x} + P_{\tilde{x}_L}\left(P_{\tilde{x}_L}^{-1} - P_x^{-1}\right) z \qquad (3.1.26)$$

另外应用矩阵求逆引理，由(3.1.25)可知

$$P_{\tilde{x}_L}^{-1} = P_x^{-1} + H^T P_v^{-1} H \tag{3.1.27}$$

代入(3.1.26)，最后导出

$$\hat{x}_L(z) = P_{\tilde{x}_L} P_x^{-1} \bar{x} + P_{\tilde{x}_L} H^T P_v^{-1} H z \tag{3.1.28}$$

(3.1.27)和(3.1.28)即为简化表达的结果。总结起来，x的线性最小方差估计及协方差矩阵为

$$\hat{x}_L(z) = P_{\tilde{x}_L}(P_x^{-1} \bar{x} + H^T P_v^{-1} H z) \tag{3.1.29}$$

$$P_{\tilde{x}_L}^{-1} = P_x^{-1} + H^T P_v^{-1} H \tag{3.1.30}$$

以上两式和(3.1.24)与(3.1.25)数学上是等价的。

[例3-3] 已知 $z=x+v$，$x \sim N(\bar{x},\sigma_x^2)$，$v \sim N(0,\sigma_v^2)$，试求$x$的线性最小方差估计。

[解] 由于$H=1$，$P_v=\sigma_v^2$，所以

$$H^T P_v^{-1} H=\sigma_v^{-2}$$

又$P_x=\sigma_x^2$，由(3.1.30)可知

$$P_{\tilde{x}_L}^{-1}=\sigma_{\tilde{x}_L}^{-2}=\sigma_x^{-2}+\sigma_v^{-2}$$

由(3.1.29)可以求得

$$\hat{x}_L(z)=\frac{\sigma_x^2\sigma_v^2}{\sigma_x^2+\sigma_v^2}\left(\frac{1}{\sigma_x^2}\bar{x}+\frac{1}{\sigma_v^2}z\right)$$

- 如果没有任何先验信息，意味着$\sigma_x^{-2}=0$，此时$\hat{x}_L(z)=z$.
- 如果量测误差特别大，即$\sigma_v^{-2}\rightarrow 0$，则$\hat{x}_L(z)=\bar{x}$（称为先验估计）。

[例3-4] 已知$z=\ln\frac{1}{x}+v, x\sim U[0,1]$，$v$服从指数分布，即

$$f_v(v)=\begin{cases} e^{-v}, & v\geq 0 \\ 0, & v<0 \end{cases}$$

假设x与v相互独立，试求x的线性最小方差估计。

[解] 因为

$$\bar{x}=\frac{1}{2}, \qquad f_{xz}(x,z)=f_{z|x}(z|x)f_x(x),$$

$$f_{z|x}(z|x) = f_v(z - \ln\frac{1}{x}) = \begin{cases} e^{-(z-\ln\frac{1}{x})}, & z - \ln\frac{1}{x} \geq 0 \\ 0, & z - \ln\frac{1}{x} < 0 \end{cases}$$

$$= \begin{cases} \frac{1}{x}e^{-z}, & x \geq e^{-z} \\ 0, & x < e^{-z} \end{cases}$$

$$f_{xz}(x,z) = f_{z|x}(z|x)f_x(x) = \begin{cases} \frac{1}{x}e^{-z}, & 1 \geq x \geq e^{-z} \\ 0, & 0 \leq x < e^{-z} \end{cases}$$

所以

$$f_z(z) = \int_{-\infty}^{+\infty} f_{xz}(x,z)dx = \begin{cases} ze^{-z}, & z \geq 0 \\ 0, & z < 0 \end{cases}.$$

因此

$$Exz = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} xzf_{xz}(x,z)dxd\,z = \frac{3}{4},$$

$$P_{xz} = E(x-\bar{x})(z-\bar{z}) = \ Exz - \bar{x}\bar{z} = -\frac{1}{4},$$

$$P_z = E(z-\bar{z})^2 = Ez^2 - \bar{z}^2$$

$$= \int_{-\infty}^{+\infty} z^2 f_z(z)dz - [\int_{-\infty}^{+\infty} zf_z(z)dz]^2 = 6 - 2^2 = 2.$$

根据[定理 3-8]，可得

$$x_L = \bar{x} + P_{xz}P_z^{-1}(z - \bar{z})$$

$$= \frac{1}{2} - \frac{1}{4} \cdot \frac{1}{2}(z - 2) = \frac{3}{4} - \frac{1}{8}z$$

注:

$$P(Z \le z|X = x) = P\left(\frac{1}{\ln x} + V \le z\right) = P\left(V \le z - \frac{1}{\ln x}\right)$$

$$\Rightarrow \ F_{z|x}(z|x) = F_v(z - \frac{1}{\ln x})$$

$$\Rightarrow f_{z|x}(z|x) = f_v(z - \frac{1}{\ln x})$$

3.1.3 极大验后估计

[定义3-6] 使$f(x|z) \Rightarrow \max$ 的估计称为x的极大验后估计，记为$\hat{x}_{MA}(z)$。称

$$\left.\frac{\partial \ln f(x|z)}{\partial x}\right|_{x=\hat{x}_{MA}} = 0 \tag{3.1.31}$$

为验后方程。

[定理 3-10] 若(x,z)服从联合高斯分布，那么

$$\hat{x}_{MA}(z)=\hat{x}_{MV}(z)\ 。$$

[证明]$f_{x|z}(x|z)=\dfrac{1}{\sqrt{(2\pi)^n|P_{x|z}|}}\exp\{-\frac{1}{2}(x-\overline{x|z})^T P_{x|z}(x-\overline{x|z})\}$

$$\frac{\partial \ln f_{x|z}(x|z)}{\partial x}=0 \Leftrightarrow \frac{1}{2}(x-\overline{x|z})^T P_{x|z}(x-\overline{x|z}) \Rightarrow \min$$

$$\Rightarrow \hat{x}_{MA}(z)=\hat{x}_{MV}(z)=\overline{x|z}.$$

[定理 3-11] 取代价函数为

$$L(\tilde{x})=\begin{cases}0, & \|\tilde{x}\|<\dfrac{\varepsilon}{2}\\ \dfrac{1}{\varepsilon}, & \|\tilde{x}\|\geq\dfrac{\varepsilon}{2}\end{cases}\quad(\varepsilon>0\text{且足够小})$$

由此获得的 Bayes 估计$\hat{x}_B(z)$即为$\hat{x}_{MA}(z)$.

[证明] （参考资料）

3.2 极大似然估计(Maximum-Likelihood Estimation)

[定义3-7] 使似然函数

$$\Lambda(x) = f(z|x) \tag{3.2.1}$$

最大的估计称为x的极大似然估计，记为$\hat{x}_{ML}(z)$. 称

$$\left.\frac{\partial \ln \Lambda(x)}{\partial x}\right|_{x=\hat{x}_{ML}} = \left.\frac{\partial \ln f(z|x)}{\partial x}\right|_{x=\hat{x}_{ML}} = 0 \tag{3.2.2}$$

为似然方程。

[例3-5] 已知$z=\ln\frac{1}{x}+v, x\sim U(0,1)$, v服从指数分布，即

$$f_v(v)=\begin{cases}e^{-v}, & v\geq 0\\ 0, & v<0\end{cases}$$

假设x与v相互独立，试求x的极大似然估计$\hat{x}_{ML}$。

【解】因为

$$f_{z|x}(z|x)=f_v\left(z-\ln\frac{1}{x}\right)$$

$$
= \begin{cases} e^{-\left(z-\ln\frac{1}{x}\right)}, & z-\ln\dfrac{1}{x} \geq 0 \\ 0, & z-\ln\dfrac{1}{x} < 0 \end{cases}
$$

$$
= \begin{cases} \dfrac{1}{x}e^{-z}, & x \geq e^{-z} \\ 0, & x < e^{-z} \end{cases}
$$

所以 $\hat{x}_{ML} = e^{-z}$.

[定理 3-12] 当对x的验前信息一无所知时,

$$\hat{x}_{ML}(z)=\hat{x}_{MA}(z).$$

【证明】由

$$f_{z|x}(z|x)=\frac{f_{zx}(z,x)}{f_x(x)}=\frac{f_{x|z}(x|z)f_z(z)}{f_x(x)}$$

可知

$$\ln f_{z|x}(z|x) = \ln f_{z|x}(x|z) + \ln f_z(z) - \ln f_x(x)$$

当对x的验前信息一无所知时

$$\frac{\partial \ln f_x(x)}{\partial x} = 0$$

因此$\frac{\partial \ln f_{z|x}(z|x)}{\partial x} = \frac{\partial \ln f_{x|z}(x|z)}{\partial x}$,

$$\Rightarrow \hat{x}_{ML}(z) = \hat{x}_{MA}(z).$$

[定理 3-13] (Linear Gaussian Measurement)

若

$$z = Hx + v, v \sim N(0, R)$$

则

$$\hat{x}_{ML}(z) = (H^T R^{-1} H)^{-1} H^T R^{-1} z \tag{3.2.3}$$

【证明】

$$f_{z|x}(z|x) = f_v(z - Hx)$$

$$= \frac{1}{\sqrt{(2\pi)^n |R|}} \exp\{-\frac{1}{2} v^T R^{-1} v\}\Big|_{v=z-Hx}$$

$$= \frac{1}{\sqrt{(2\pi)^n |R|}} \exp\{-\frac{1}{2} (z - Hx)^T R^{-1} (z - Hx)\}$$

显然，极大化似然函数相当于

$$J = \frac{1}{2}(z - Hx)^T R^{-1}(z - Hx) \Rightarrow \min$$

由此，有

$$\frac{\partial J}{\partial x} = -H^T R^{-1}(z - Hx) = 0.$$

当$H^T R^{-1} H$非奇异时，即有

$$\hat{x}_{ML}(z) = (H^T R^{-1} H)^{-1} H^T R^{-1} z.$$

[推论 3-9] 线性量测高斯噪声条件下

$$(1)\, E\tilde{x}_{ML} = 0(\text{无偏估计})$$

$$(2)\, P_{\tilde{x}_{ML}} = (H^T R^{-1} H)^{-1}$$

$$(3)\, \hat{x}_{ML} = P_{\tilde{x}_{ML}} H^T R^{-1} z$$

Remarks:

(1) [推论 3-9]中(2)、(3)合称为 Gauss-Markov 估计器;

(2) 特别地，$P_{\tilde{x}_{ML}}$与测量值无关。

(3) 当$H = I, P_{\tilde{x}_{ML}} = R$，此时$\hat{x}_{ML} = z$；

(4) 如果没有先验知识，可以认为$\bar{x} = 0, P_x \to \infty$，由线性高斯量测最小方差估计知

$$\begin{cases} E[x|z] = \bar{x} + P_{x|z}H^T R^{-1}(z - H\bar{x}) \\ P_{x|z} = P_{\tilde{x}_{MV}} = (P_x^{-1} + H^T R^{-1} H)^{-1} \end{cases}$$

可见，此时$E[x|z] \to \hat{x}_{ML}$，$P_{\tilde{x}_{MV}} = P_{\tilde{x}_{\mathrm{ML}}} = (H^T R^{-1} H)^{-1}$。

[例3-6] 设x是服从参数为(μ,σ)的正态分布的随机变量，即

$$f(x|\mu,\sigma)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\{-\frac{1}{2\sigma^2}(x-\mu)^2\}$$

假设进行了N次独立的采样，获得了相互独立的$x_1,x_2,\cdots,x_N$，试估计参数(μ,σ)。

[解]根据独立性假设，可知

$$\Lambda(\mu,\sigma)=f(x_1,x_2,\cdots,x_N|\mu,\sigma)$$

$$=\frac{1}{\sqrt{(2\pi)^N\sigma^N}}\exp\{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$$

根据似然函数极大化条件$\frac{\partial\Lambda(\mu,\sigma)}{\partial\mu}=0$, $\frac{\partial\Lambda(\mu,\sigma)}{\partial\sigma}=0$，可得

$$\sum_{i=1}^{N}(x_i-\mu)=0, \qquad \sum_{i=1}^{N}(x_i-\mu)^2-N\sigma^2=0$$

由此可得两个参数的极大似然估计如下：

$$\hat{\mu}=\frac{1}{N}\sum_{i=1}^{N}x_i\,, \qquad \hat{\sigma}^2=\frac{1}{N}\sum_{i=1}^{N}(x_i-\mu)^2$$

[例3-7] 设估计确定性电压x有两种不同的方案，其一是用两个昂贵的电压表，它们的测量噪声为$N(0,2)$；其二是用四个廉价的电压表，它们的测量噪声为$N(0,3.5)$。试问哪一种方案测量到的结果更加可靠？

[解] (a)用两个昂贵的电压表时

$$z = \begin{bmatrix} 1 \\ 1 \end{bmatrix} x + v$$

其中，$v \sim N(0, \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix})$. 因此

$$P_{\tilde{x}} = (H^T R^{-1} H)^{-1} = \left[(1,1) \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \end{bmatrix} \right]^{-1} = 1.$$

(b)用四个廉价电压表时

$$z = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} x + v$$

其中，$v \sim N(0,3.5\mathrm{I})$. 此时

$$\mathrm{P}_{\tilde{x}} = (H^T R^{-1} H)^{-1} = \frac{7}{8}$$

结论：第二个方案更可靠些。

Remarks:

- 似然函数比验后概率要容易获得。
- 极大似然估计中，被估计量x可以是随机变量，也可以是非随机参数。
- 当有先验信息时，极大似然估计的精度不如极大验后估计。

3.2.1 Cramer-Rao Lower Bound

[定理 3-14] （1）设$\hat{x}$为确定性量x基于测量z的无偏估计，那么

$$P_{\tilde{x}} \geq J_F^{-1} \tag{3.2.4}$$

其中

$$J_F = E\left\{\left[\frac{\partial \ln \Lambda(x)}{\partial x}\right]\left[\frac{\partial \ln \Lambda(x)}{\partial x}\right]^T\right\} \quad (3.2.4a)$$

$$= -E\left[\frac{\partial^2 \ln \Lambda(x)}{\partial x \partial x^T}\right] \quad (3.2.4b)$$

称为费希尔信息矩阵(Fisher information matrix)．此外，$\Lambda(x) = f(z|x)$。

（2）设$\hat{x}$为随机量x基于量测z的无偏估计，那么

$$P_{\tilde{x}} \geq L^{-1} \tag{3.2.5}$$

其中

$$L = E\left\{\left[\frac{\partial \ln f(x,z)}{\partial x}\right]\left[\frac{\partial \ln f(x,z)}{\partial x}\right]^T\right\} \tag{3.2.6a}$$

$$= -E\left[\frac{\partial^2 \ln f(x,z)}{\partial x \partial x^T}\right] \tag{3.2.6b}$$

称为信息矩阵(information matrix).

下面将针对确定性标量估计情况进行证明，要用到著名的施瓦尔兹不等式。

[引理3-3] (Schwarz Inequality) 设$f,g \in L^2$，那么

$$\begin{cases} < f,g > \leq \|f\| \cdot \|g\| \\ \int_{-\infty}^{+\infty} f(x)g(x)\mathrm{d}x \leq \sqrt{\int_{-\infty}^{+\infty} |f(x)|^2 dx} \sqrt{\int_{-\infty}^{+\infty} |g(x)|^2 \mathrm{d}x} \end{cases} \quad (3.2.7)$$

(注：在空间解析几何中，一般向量的点积公式为 $\vec{a} \cdot \vec{b} \leq |\vec{a}||\vec{b}|$)

对于确定性标量估计情况，$\Lambda(x)=f(z|x)$.

$$E[\hat{x}(z)-x]=\int_{-\infty}^{+\infty}[\hat{x}(z)-x]f(z|x)dz=0$$

$$\Rightarrow \frac{d}{dx}\int_{-\infty}^{+\infty}[\hat{x}(z)-x]f(z|x)dz=0$$

$$\Rightarrow -\int_{-\infty}^{+\infty}f(z|x)dz+\int_{-\infty}^{+\infty}[\hat{x}(z)-x]\frac{\partial}{\partial x}f(z|x)dz=0$$

$$\Rightarrow \int_{-\infty}^{+\infty} [\hat{x}(z) - x] \frac{\partial \ln f(z|x)}{\partial x} f(z|x) dz = 1$$

根据施瓦尔兹不等式

$$\Rightarrow \int_{-\infty}^{+\infty} [\hat{x}(z) - x]^2 f(z|x) dz \int_{-\infty}^{+\infty} [\frac{\partial \ln f(z|x)}{\partial x}]^2 f(z|x) dz \geq 1$$

$$\Rightarrow P_{\tilde{x}} \geq \{E[\frac{\partial \ln f(z|x)}{\partial x}]^2\}^{-1}$$

由于$\int_{-\infty}^{+\infty} f(z|x)dz = 1$，可导出

$$\int_{-\infty}^{+\infty} \frac{\partial \ln f\,(z|x)}{\partial x} f(z|x)dz = 0$$

$$\Rightarrow \int_{-\infty}^{+\infty} \frac{\partial^2 \ln f\,(z|x)}{\partial x^2} f(z|x)dz$$

$$= -\int_{-\infty}^{+\infty} [\frac{\partial \ln f\,(z|x)}{\partial x}]^2 f(z|x)dz$$

$$\Rightarrow E[\frac{\partial^2 \ln f\,(z|x)}{\partial x^2}] = -E\{[\frac{\partial \ln f\,(z|x)}{\partial x}]^2\}$$

极大似然估计随着量测样本数N增加，具有许多优良的性质：

（1） 对于确定性未知量x，当$N \to +\infty$，$\hat{x}_{ML}$依概率收敛于真值x；（任何具有该性质的估计称为**一致的**）

（2） 极大似然估计是渐进高斯的，即当$N \to +\infty$，$\hat{x}_{ML} \sim N(x, P_{\tilde{x}})$；

（3） 当$N \to +\infty$，极大似然估计$\hat{x}_{ML}$是**有效的**（估计误差协方差达到 Cramer-Rao 下界）。

3.3 最小二乘估计(Least-square Estimation)

设x是未知的常值向量，考虑线性量测

$$z = Hx + v \tag{3.3.1}$$

其中，v为测量误差或噪声. 我们的问题是基于z求x的最佳估计$\hat{x}$。

3.3.1 基本最小二乘估计

[定义3-8] 使指标函数

$$J(\hat{x}) = (z - H\hat{x})^T(z - H\hat{x}) \Rightarrow \min \qquad (3.3.2)$$

的估计$\hat{x}$称为x的最小二乘估计，记为$\hat{x}_{LS}(z)$.

[定理 3-15]

(1) 当H^TH非奇异时，$\hat{x}_{LS}(z)=(H^TH)^{-1}H^Tz$.

(2) 若v为零均值噪声,那么$E\tilde{x}_{LS}(z)=0$.

[证明] 由

$$\frac{\partial J(\hat{x})}{\partial \hat{x}}=-2H^T(z-H\hat{x})=0$$

可知

$$H^T H\hat{x} = H^T z$$

$H^T H$是方阵，如果非奇异，则

$$\hat{x}_{LS}(z) = (H^T H)^{-1} H^T z \tag{3.3.3}$$

$H^T H$非奇异保证了$J(\hat{x})$关于的二阶导数矩阵非负定，说明上式的确是最优解。

将(3.3.1)代入(3.3.3)，得

$$\hat{x}_{LS}(z) = x + (H^T H)^{-1} H^T v$$

如果 $Ev = 0$，则

$$E\hat{x}_{LS}(z) = x$$

即

$$E\tilde{x}_{LS}(z) = 0.$$

[例3-8] 用万用表对未知阻值的电阻进行k次测量值，根据含噪声的k个测量$z_i(i=1,2,\cdots,k)$对电阻值x进行估计。此时，x是一个标量，k个含噪声的测量值如下：

$$z_1 = x + v_1$$

$$\vdots$$

$$z_k = x + v_k$$

写成矩阵方程形式

$$\begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} x + \begin{bmatrix} v_1 \\ \vdots \\ v_k \end{bmatrix}$$

由(3.3.3)可得电阻值x的最小二乘估计为

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

$$= \left([1 \cdots 1] \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} [1 \cdots 1] \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix} = \frac{1}{k} (z_1 + z_2 + \cdots + z_k)$$

3.3.2 加权最小二乘估计

在基本最小二乘估计中，假设了对所有的量测值具有相同的置信度。但在一定的条件下，我们可能对某些测量比其他的

更有信心。在这种情况下，可以对上一节的结果进行推广，从而获得加权最小二乘估计。

[定义3-9] 设W是正定对称方阵，使指标函数

$$J(\hat{x})=(z-H\hat{x})^TW(z-H\hat{x}) \tag{3.3.3}$$

最小的估计$\hat{x}$称为x的加权最小二乘估计，记为$\hat{x}_{WLS}(z)$.

[定理 3-16] 当H^TWH非奇异

$$\hat{x}_{WLS}(z) = (H^TWH)^{-1}H^TWz \tag{3.3.4}$$

若v为零均值噪声，$\hat{x}_{WLS}(z)$是x的无偏估计。如果进一步假设

$R = E[vv^T]$，那么

$$P_{\tilde{x}_{WLS}} = (H^TWH)^{-1}H^TWRWH(H^TWH)^{-1} \tag{3.3.5}$$

[定理 3-17] (Markov 估计)设$v \sim N(0,R)$ $(R>0)$，取$W=R^{-1}$，

那么

$$\hat{x}_{WLS}(z)=(H^T R^{-1} H)^{-1} H^T R^{-1} z \tag{3.3.6}$$

而且此时估计的均方误差最小，

$$P_{\tilde{x}_{WLS}}=(H^T R^{-1} H)^{-1} \tag{3.3.7}$$

[证明]将$W = R^{-1}$带入一般的加权最小二乘估计(3.3.4)，即可得到(3.3.6)．同时由(3.3.6)可导出(3.3.7)．此外，根据线性量测高斯分布情况下的[定理 3-6]，可知

$$\hat{x}_L = P_{\tilde{x}_L}(P_x^{-1}\bar{x} + H^T R^{-1} z) \tag{3.1.13}$$

$$P_{\tilde{x}_L}^{-1} = P_x^{-1} + H^T R^{-1} H \tag{3.1.14}$$

在最小二乘估计中，认为没有x的先验信息($P_x^{-1} = 0$)，所以此

时的最小方差估计为

$$\hat{x}_L = P_{\tilde{x}_{LMV}} H^T R^{-1} z$$

$$P_{\tilde{x}_L}^{-1} = H^T R^{-1} H$$

与(3.3.6), (3.3.7)是一致的，即说明了此时的加权最小二乘估计的均方误差最小。证毕!

[例3-9] 在[例 3-8]中，假设$v_i \sim N\left(0, \sigma_i^2\right)(i = 1, \cdots, k)$，同时认为每次测量是独立进行的，即

$$R = \mathrm{diag}(\sigma_1^2, \cdots, \sigma_k^2)$$

此时电阻值的最优估计为

$$\hat{x}_{WLS}(z) = (H^T R^{-1} H)^{-1} H^T R^{-1} z$$

$$= \left([1 \cdots 1] \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k^2 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \right)^{-1} [1 \cdots 1] \begin{bmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_k^2 \end{bmatrix}^{-1} \begin{bmatrix} z_1 \\ \vdots \\ z_k \end{bmatrix}$$

$$= \frac{1}{\sum_{i=1}^{k} 1/\sigma_i^2} \left(\frac{z_1}{\sigma_1^2} + \frac{z_2}{\sigma_2^2} + \cdots + \frac{z_k}{\sigma_k^2}\right)$$

可见每次的测量精度作为权重进入了最后的估计，当$\sigma_i^2 = \sigma^2$ (每次测量精度相同)，则退化为[例 3-8]了。

3.3.3 递推最小二乘估计

如果我们持续地进行测量，并希望随着每次新的测量值更新x的估计值，我们需要不断扩大H矩阵并完全重新计算估计值$\hat{x}$。当测量的次数变得很大时，计算量将变得非常大。例如，如果我们每秒测量一次卫星的高度，一小时就要测量 3600 次。显然，随着时间的增加，最小二乘估计的计算量可能很快超过我们的计算资源。因此，建立递推形式的最小二乘估计是非常有必要的。

假设我们已经获得了k时刻的最小二乘估计，即由

$$z_k = H_k x + v_k \tag{3.3.15}$$

得到了加权最小二乘估计

$$P_k = (H_k^T W_k H_k)^{-1} \tag{3.3.16}$$

$$\hat{x}_k = P_k H_k^T W_k z_k \tag{3.3.17}$$

如果现在又获得了$k+1$时刻的量测，即

$$z_{k+1} = H_{k+1}x + v_{k+1} \tag{3.3.18}$$

综合考虑(3.3.15)和(3.3.18)，加权最小二乘意味着

$$\left(\begin{bmatrix} z_k \\ z_{k+1} \end{bmatrix} - \begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix} \hat{x}_{k+1}\right)^T \begin{bmatrix} W_k & 0 \\ 0 & w_{k+1} \end{bmatrix} \left(\begin{bmatrix} z_k \\ z_{k+1} \end{bmatrix} - \begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix} \hat{x}_{k+1}\right)$$

$$\Rightarrow \min$$

即为

$$(z_k - H_k\hat{x}_{k+1})^T W_k (z_k - H_k\hat{x}_{k+1})$$
$$+ (z_{k+1} - H_{k+1}\hat{x}_{k+1})^T w_{k+1} (z_{k+1} - H_{k+1}\hat{x}_{k+1})$$
$$\Rightarrow \min$$

由极值必要条件，可得

$$H_k^T W_k (z_k - H_k\hat{x}_{k+1}) + H_{k+1}^T w_{k+1} (z_{k+1} - H_{k+1}\hat{x}_{k+1}) = 0$$

$$\Rightarrow (H_k^T W_k H_k + H_{k+1}^T w_{k+1} H_{k+1})\hat{x}_{k+1}$$

$$= H_k^T W_k z_k + H_{k+1}^T w_{k+1} z_{k+1}$$

$$\Rightarrow (H_k^T W_k H_k + H_{k+1}^T w_{k+1} H_{k+1})\hat{x}_{k+1}$$
$$= P_k^{-1}\hat{x}_k + H_{k+1}^T w_{k+1} z_{k+1} \qquad (3.3.19)$$

而由(3.3.16)可知

$$P_{k+1} = \left(\begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix}^T \begin{bmatrix} W_k & 0 \\ 0 & w_{k+1} \end{bmatrix} \begin{bmatrix} H_k \\ H_{k+1} \end{bmatrix} \right)^{-1}$$

$$= (H_k^T W_k H_k + H_{k+1}^T w_{k+1} H_{k+1})^{-1}$$

带入(3.3.16)，(3.3.19)化为出

$$\hat{x}_{k+1} = P_{k+1}(P_k^{-1}\hat{x}_k + H_{k+1}^T w_{k+1} z_{k+1}) \qquad (3.3.20)$$

其中

$$P_{k+1} = (P_k^{-1} + H_{k+1}^T w_{k+1} H_{k+1})^{-1} \tag{3.3.21}$$

(3.3.20)和(3.3.21)一起构成了递推最小二乘估计。如果知道$v_{k+1} \sim N(0, R_{k+1})$，可以取$w_{k+1} = R_{k+1}^{-1}$，此时递推最小二乘估计算法公式为

$$\hat{x}_{k+1} = P_{k+1}(P_k^{-1}\hat{x}_k + H_{k+1}^T R_{k+1}^{-1} z_{k+1}) \tag{3.3.22}$$

$$P_{k+1}^{-1} = P_k^{-1} + H_{k+1}^T R_{k+1}^{-1} H_{k+1} \tag{3.3.23}$$

递推最小二乘估计也有其他等价形式，例如由(3.3.23)可知

$$P_{k+1}P_k^{-1} = I - P_{k+1}H_{k+1}^T R_{k+1}^{-1} H_{k+1} \tag{3.3.24}$$

代入(3.3.22)，可得如下结构的递推最小二乘估计

$$\hat{x}_{k+1} = \hat{x}_k + K_{k+1}[z_{k+1} - H_{k+1}\hat{x}_k] \tag{3.3.25}$$

其中

$$K_{k+1} = P_{k+1} H_{k+1}^T R_{k+1}^{-1} \tag{3.3.26}$$

由矩阵求逆引理，(3.3.21)可以化为

$$P_{k+1} = P_k - P_k H_{k+1}^T (R_{k+1} + H_{k+1} P_k H_{k+1}^T)^{-1} H_{k+1} P_k \tag{3.3.27}$$

这样可以减小递推计算量。(3.3.25)、(3.3.26)和(3.3.27)构成了一组完整的递推算法。

由(3.3.26)，可得

$$P_{k+1}H_{k+1}^T = K_{k+1}R_{k+1} \tag{3.3.28}$$

代入(3.3.27)可知

$$\begin{aligned} P_{k+1}H_{k+1}^T &= P_kH_{k+1}^T \\ &\quad - P_kH_{k+1}^T(R_{k+1} + H_{k+1}P_kH_{k+1}^T)^{-1}H_{k+1}P_kH_{k+1}^T \\ &= P_kH_{k+1}^T(R_{k+1} + H_{k+1}P_kH_{k+1}^T)^{-1}[(R_{k+1} + H_{k+1}P_kH_{k+1}^T) \\ &\quad - H_{k+1}P_kH_{k+1}^T] \end{aligned}$$

$$= P_k H_{k+1}^T (R_{k+1} + H_{k+1} P_k H_{k+1}^T)^{-1} R_{k+1}$$

比较(3.3.24)，可得K_{k+1}的另外一个计算公式

$$K_{k+1} = P_k H_{k+1}^T (R_{k+1} + H_{k+1} P_k H_{k+1}^T)^{-1} \qquad (3.329)$$

代入(3.3.27)，可得

$$P_{k+1} = (I - K_{k+1} H_{k+1}) P_k \qquad (3.3.30)$$

(3.3.25)、(3.3.29)和(3.3.30)是另外一组常见的递推公式，

重写如下:

$$\hat{x}_{k+1} = \hat{x}_k + K_{k+1}[z_{k+1} - H_{k+1}\hat{x}_k] \tag{3.3.25}$$

$$K_{k+1} = P_k H_{k+1}^T (R_{k+1} + H_{k+1} P_k H_{k+1}^T)^{-1} \tag{3.329}$$

$$P_{k+1} = (I - K_{k+1} H_{k+1}) P_k \tag{3.3.30}$$

[例3-10] 仍然考虑[例 3-8]中的问题，即用万用表对未知阻值的电阻进行测量，这里用递推最小二乘估计来求解。k时刻的量测方程为

$$z_i = x + v_i$$

假设$v_i \sim (0, R_i)$，$R_i = R$. 本问题$H_k = 1$.

$k = 0$: （给定$\hat{x}_0, P_0$）

$$K_1 = \frac{P_0}{R + P_0}, P_1 = (1 - K_1)P_0 = \frac{RP_0}{R + P_0}$$

$$\hat{x}_1 = \hat{x}_0 + K_1(z_1 - \hat{x}_0) = (1 - K_1)\hat{x}_0 + K_1 z_1$$

$$= \frac{R}{R + P_0}\hat{x}_0 + \frac{P_0}{R + P_0} z_1$$

$k = 1$:

$$K_2 = \frac{P_1}{R + P_1} = \frac{P_0}{R + 2P_0}, P_2 = (1 - K_2)P_1 = \frac{RP_0}{R + 2P_0}$$

$$\hat{x}_2 = \hat{x}_1 + K_2(z_2 - \hat{x}_1) = \frac{R + P_0}{R + 2P_0}\hat{x}_1 + \frac{P_0}{R + 2P_0}z_2$$

一般地$k \geq 1$

$$K_k = \frac{P_{k-1}}{R + P_{k-1}} = \frac{P_0}{R + kP_0}, P_k = \frac{RP_0}{R + kP_0}$$

$$\hat{x}_k = \hat{x}_{k-1} + K_k(z_k - \hat{x}_{k-1}) = (1 - K_k)\hat{x}_{k-1} + K_k z_k$$

$$= \frac{R + (k-1)P_0}{R + kP_0}\hat{x}_{k-1} + \frac{P_0}{R + kP_0}z_k$$

我们分三种初始条件讨论:

(1) $\hat{x}_0 = \bar{x}_0, P_0 = 0.$

$$K_k = 0:\ \hat{x}_k = \hat{x}_{k-1} = \cdots = \bar{x}_0$$

说明无需量测。

(2) $R = +\infty$.

$$K_k = 0:\ \hat{x}_k = \hat{x}_{k-1} = \cdots = \bar{x}_0$$

说明量测不能带来任何有意义信息。

(3) $\hat{x}_0 = \bar{x}_0, P_0 = +\infty$.

$$K_k = \frac{1}{k}:\ \hat{x}_k = \frac{k-1}{k}\hat{x}_{k-1} + \frac{1}{k}z_k = \frac{1}{k}[(k-1)\hat{x}_{k-1} + z_k]$$

说明电阻值的最优估计就是量测值的几何平均，即

$$\hat{x}_k = \frac{1}{k}\sum_{i=1}^{k} z_i = \frac{1}{k}\left[(k-1)\frac{1}{k-1}\sum_{i=1}^{k-1} z_i + z_k\right]$$

$$= \frac{1}{k}[(k-1)\hat{x}_{k-1} + z_k]$$

Remarks:

- 最小二乘估计不需要任何统计信息，但精度较低。适用于确定量或变化规律确定的量之估计。
- 在递推最小二乘估计中，初始估计可取为零、初始P_0矩阵可取为元素足够大的对角线阵。若干步后，初值的影响就会逐渐消失。
- 滑动窗技术与遗忘因子法是工程中有效的改进算法。

3.4 本章小结（$z = Hx + v$）

(1) 最小二乘估计;

(2) 线性最小方差估计;

(3) 最小方差估计;

(4) 极大似然估计;

(5) 极大验后估计.