COMPARISON CRITERIA FOR ESTIMATES

For the comparison of the optimal estimates we use the mean and the covariance of the estimates.

<u>Theorem</u>: (1) If X, Y random vectors and $E\{YY^T\}$ non sing. then $E\{XX^T\} - E\{XY^T\}(E\{YY^T\})^{-1}E\{YX^T\} \geq 0$

<u>Proof</u>: $E\{(X-KY)(X-KY)^T\} \geq 0$ which is equivalent to

$$E\{XX^T\} - KE\{YX^T\} - E\{XY^T\}^T K^T + \\ + KE\{YY^T\}K \geq 0 \quad \forall K$$

choosing $K = K^* = E\{XY^T\}(E\{YY^T\})^{-1}$

the theorem is proved.

Cramer - Rao lower bound

x vector to be estimated, Y measurement vector

$p_{Y|X}$ density available (differentiable w.r.t. x)

$E_{Y|X}$ expectation evaluated with $p_{Y|X}$

<u>Theorem</u>: $\forall$ estimate $\hat{x} = f(Y)$

$R(x) = E_{Y|X}\{(f(Y)-x)(f(Y)-x)^T$ satisfies

$$R(x) \geq \underbrace{\left(I + \frac{\partial S(x)}{\partial x}\right) \Lambda^{-1}(x) \left(I + \frac{\partial S(x)}{\partial x}\right)^T}_{L(x)}$$

where $\Lambda(x) = E_{Y|X}\left\{\left(\frac{\partial}{\partial x} \ln p_{Y|X}(y,x)\right)^T \cdot \left(\frac{\partial}{\partial x} \ln p_{Y|X}(y,x)\right)\right\}$

and $S(x) = E_{Y|X}\{f(Y) - x\}$, with Λ non sing. $\forall x$

$L(x)$ is the lower-bound

$S(x)$ is the polarization of the estimate

$\Lambda(x)$ is the Fisher information matrix

\# If x deterministic $p_{Y|X}(y,x) = p_Y(y,x)$

Proof: Since $\underbrace{\int_{\mathbb{R}^n} p_{Y|X}(y,x)\,dy = 1}_{\frac{\partial}{\partial x} = 0}$

It follows under regularity assumptions on $p_{Y|X}$ that

$$\int_{\mathbb{R}^n} \underbrace{\frac{\partial}{\partial x} p_{Y|X}(y,x)}\, dy = 0$$

$$= \left[\frac{\partial}{\partial x} \ln p_{Y|X}(y,x)\right] p_{Y|X}(y,x) = \gamma \rightarrow \text{for simplicity}$$

premultiplying by $x \in \mathbb{R}^n$:

$$\int_{\mathbb{R}^n} x \cdot \gamma = 0$$

$$\Rightarrow \boxed{E_{Y|X}\left\{x \frac{\partial}{\partial x} \ln p_{Y|X}(Y,x)\right) = 0} \quad (*)$$

By definition of $S(x)$:

$$\frac{\partial}{\partial x} E_{Y|X}\{f(Y)\} = \frac{\partial}{\partial x}(x + S(x)) = \boxed{I + \frac{\partial S(x)}{\partial x}} \quad (**)$$

But since $E_{Y|X}\{f(Y)\} = \int_{\mathbb{R}^n} f(y)\, p_{Y|X}(y,x)\, dy$

differentiating w.r.t. x:

$$\frac{\partial}{\partial x} E_{Y|X}\{f(y)\} = \int_{\mathbb{R}^n} f(y) \cdot \gamma \, dy =$$

$$= E_{Y|X}\left\{f(Y) \frac{\partial}{\partial x} \ln p_{Y|X}(Y,x)\right\} \quad (***)$$

Using $(*)$, $(**) = (***)$

$$I + \frac{\partial S(x)}{\partial x} = E_{Y|X}\left\{f(Y)\frac{\partial}{\partial x} \ln p_{Y|X}(Y,x)\right\} = E_{Y|X}\left\{f(Y)\frac{\partial}{\partial x} \ln p_{Y|X}(y,x)\right\}$$

$(****)$

Define $v = f(Y) - x$

$w = \left[\partial/\partial x \ln p_{Y|X}(Y,x)\right]^T$

From theorem (1):

$$E_{Y|X}\{v v^T\} \geq E_{Y|X}\{v W^T\}\big(\underbrace{E_{Y|X}\{W W^T\}}_{\Lambda(x) \text{ non sing.}}\big)^{-1} E_{Y|X}\{W V^T\}$$

This gives exactly the result of the theorem since:

from $****$, $E_{Y|X}\{v w^T\} = I + \frac{\partial S(x)}{\partial x}$

III Properties of estimates

- **Efficiency**

Efficient estimate if $R(x) = \left(I + \frac{\partial S(x)}{\partial x}\right) \Lambda^{-1}(x) \left(I + \frac{\partial S(x)}{\partial x}\right)^T$
$\forall$ admissible x

- **Centering**

Centered estimate if $S(x) = 0 \quad \forall$ admissible x

<u>Theorem</u> Nec. cond to be efficient is that is also centered

<u>Proof</u>: If an estimate is efficient, and considering

$$V = g(Y) - x \quad , \quad W = \left[\partial/\partial x \ \ln p_{Y|X}(Y, x) \right]^T$$

one has: $E_{Y|X}\{VV^T\} = E_{Y|X}\{VW^T\} \left(E_{Y|X}\{WW^T\}\right)^{-1} E_{Y|X}\{WV^T\}$

Given $K^* = E_{Y|X}\{VW^T\}\left(E_{Y|X}\{WW^T\}\right)^{-1}$:

$E_{Y|X}\{(V - K^* W)(V - K^* W)^T\} \geq 0$ positive semidef matrix

$= 0$ iff $V - K^* W = 0$

which is $V = g(Y) - x = E_{Y|X}\{VW^T\}\left(E_{Y|X}\{WW^T\}\right)^{-1} \cdot W$

Therefore by applying $E_{Y|X}$ and since $E_{Y|X}\{W\} = 0$:

$$E_{Y|X}\{V\} = E_{Y|X}\{g(Y) - x\} = 0 \Rightarrow E_{Y|X}\{g(Y)\} = x$$

- **Consistency**

Vector Y can be larger in dimension (more measurements lead to a more precise precision)

A sequence of estimates $\{\hat{x}_N\}$ $N = 1 \dots$ of x is consistent if $\hat{x}_N = g_N(Y)$ converges in probability $P_{Y|X}$:

$\hat{x}_N = g_N(Y) \xrightarrow{P_{Y|X}} x$, or in alternative

$$\lim_{N \to +\infty} P_{Y|X}\{\| g_N(Y(\omega)) - X(\omega) \| > \varepsilon\} = 0 \quad \forall \varepsilon > 0$$

N: temporal index or discretization index

- Example 1: (Maximum likelihood)

Estimate the mean m and the variance σ^2 of a gaussian random variable.

N experiments Y_i $i=1\dots N$ independent.

Vector to be estimated: $x = \begin{pmatrix} m \\ \sigma \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ (deterministic)

Maximum likelihood estimate:

Density (gaussian) $p_{Y_i}(y_i, x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-(y_i - m)^2/2\sigma^2}$

$= \frac{1}{\sqrt{2\pi}\, x_2} e^{-(y_i - x_1)^2/2x_2^2}$

$$p_{Y_1\dots Y_N}(y_1 \dots y_N; x) = \prod_{i=1}^{N} p_{Y_i}(y_i, x) \quad \text{(independent experiments)}$$

$$= \frac{1}{(2\pi)^{N/2}} \cdot \frac{1}{x_2^N} e^{-\frac{1}{2x_2^2}\sum_{i=1}^{N}(y_i - x_1)^2}$$

Estimates:

$$\hat{x} = \underset{x:\, x_2 > 0}{\operatorname{argmax}}\; p_{Y_1\dots Y_N}(y_1 \dots y_N; x) \overset{\text{more convenient}}{=} \underset{x:\, x_2 > 0}{\operatorname{argmax}}\; \ln p_{Y_1\dots Y_N}(y_1 \dots y_N; x)$$

The extremals are obtained from:

$$0 = \partial/\partial x_1 \left(\ln p_{Y_1 \dots Y_N}(y_1 \dots y_N; x) \right)\Big|_{x = \hat{x}}$$

$$= \frac{1}{\hat{x}_2^2} \sum_{i=1}^{N} (y_i - \hat{x}_1) \Rightarrow \boxed{\hat{x}_1 = \frac{1}{N}\sum_{i=1}^{N} y_i} \quad (*)$$

$$0 = \partial/\partial x_2 \left(\ln p_{Y_1 \dots Y_N}(y_1 \dots y_N; x) \right)\Big|_{x = \hat{x}}$$

$$= -\frac{N}{\hat{x}_2^2} + \frac{1}{\hat{x}_2^3}\sum_{i=1}^{N}(y_i - \hat{x}_1)^2 \Rightarrow \boxed{\hat{x}_2^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{x}_1)^2} \quad (**)$$

Since $\ln p_{Y_1 \dots Y_N}(y_1 \dots y_N; x) = -N \log x_2 - \frac{1}{2x_2^2}\sum_{i=1}^{N}(y_i - x_1)^2$

and the limit as $\|x\| \to \infty$ is $-\infty$ then

$\hat{x} = \begin{pmatrix} \hat{x}_1 \\ \hat{x}_2 \end{pmatrix} = \begin{pmatrix} * \\ ** \end{pmatrix}$ is the maximum with admissibility condition $\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{x}_1)^2 > 0$

(*) is centered : $E\{\hat{x}_1\} = \frac{1}{N}\sum_{i=1}^{N} E\{y_i\} = \frac{1}{N} N x_1$

(**) is not centered : $E\{\hat{x}_2\} = \frac{1}{N}\sum_{i=1}^{N} E\{(y_i - \frac{1}{N}\sum_{j=1}^{N} y_j)^2\}$

$$= \frac{1}{N}\sum_{i=1}^{N} E\{(y_i - x_1 - \frac{1}{N}\sum_{j=1}^{N}(y_j - x_1))^2\}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\Big(E\{(y_i - x_1)^2\} + \frac{1}{N}E\{(\sum_{j=1}^{N}(y_j - x_1))^2\}$$

$$- \frac{2}{N}E\{(y_i - x_1)\sum_{j=1}^{N}(y_j - x_1)\}\Big)$$

But $E\{(y_i - x_1)(y_j - x_1)\} = 0 \quad \forall i \neq j$ (Y_i, Y_j are independent)

Therefore $E\{\hat{x}_2^2\} = \frac{1}{N}\sum_{i=1}^{N}(6^2 + \frac{1}{N^2} N 6^2 - \frac{2}{N}6^2) = \boxed{\frac{N-1}{N}6^2}$

The modified estimate is

$\bar{x}_2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \hat{x}_1)^2 = \frac{N}{N-1}\hat{x}_2$ will be centered but not maximum likelihood estimate

---

~ Example 2 : Fisher matrix and Cramer Rao

Consider $Y(\omega) = Ax + N(\omega)$ $\quad N \in \mathcal{N}(0, \Psi_N)$,

x deterministic unknown

$Y \in \mathcal{N}(\underbrace{Ax}_{M_Y}, \underbrace{\Psi_N}_{\Psi_Y})$

Therefore

$$p_Y(y, x) = \frac{1}{(2\pi)^{m/2}(\det \Psi_N)^{1/2}} \cdot e^{-\frac{1}{2}(y - Ax)^T \Psi_N^{-1}(y - Ax)}$$

and a maximum likelihood estimate is

$$\hat{x} = \arg\max_x \ln p_Y(y, x) = \arg\min_x \underbrace{\frac{1}{2}(y - Ax)^T \Psi_N^{-1}(y - Ax)}_{S(y,x)}$$

Extremals:

$$0 = \partial/_{\partial x} S(y, x)\big|_{x = \hat{x}} = -(y - Ax)^T \Psi_N^{-1} A\big|_{x = \hat{x}}$$

if A full column rank : $\hat{x} = A^{\#}_{\Psi_N} y$ maximum likelihood estimate

$E\{\hat{x}\} = E\{A^{\#}_{\Psi_N} y\} = A^{\#}_{\Psi_N} E\{y\} = A^{\#}_{\Psi_N} Ax = x \rightarrow \hat{x}$ is centered

The Cramer Rao lower bound is: $(I + \frac{\partial S(x)}{\partial x})^T \Lambda^{-1}(x)(I + \frac{\partial S(x)}{\partial x})$

$$\begin{cases} \Lambda(x) = E_Y\{(\partial/_{\partial x} \ln p_Y(Y, x))^T(\partial/_{\partial x} \ln p_Y(Y, x))\} = A^T \Psi_N^{-1}(Y - Ax) \cdot (Y - Ax)^T \Psi_N^{-1} A = A^T \Psi_N^{-1} A \\ S(x) = 0 \end{cases}$$

$\Rightarrow \hat{x}$ is efficient

~ Example 3 : Markov estimates (Weighted least square)

Estimate deterministic x from M independent measurements

$y_i = x + n_i$, $i = 1 \dots M$, $E\{n_i\} = 0$, $E\{n_i^2\} = \sigma_{n_i}^2$

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_M \end{pmatrix} \quad N = \begin{pmatrix} n_1 \\ \vdots \\ n_M \end{pmatrix} \Rightarrow Y = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} x + N$$

$$\Psi_N = \operatorname{diag}\{\sigma_{N_1}^2, \dots, \sigma_{N_M}^2\}$$

Weighted least square estimate :

Weights: $W = \Psi_N^{-1} = \operatorname{diag}\{1/\sigma_{N_1}^2 \dots 1/\sigma_{N_M}^2\}$

$$\hat{x} = A_W^{\#} y = \frac{\sum_{i=1}^{M} \frac{y_i}{\sigma_{N_i}^2}}{\sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}} \rightarrow \text{Markov estimates}$$

Since the measurement equation is linear in x with N gaussian this estimate coincides with maximum likelihood estimate

If $\sigma_{N_i}^2 = \sigma_{N_j}^2 \ \forall i,j \Rightarrow \hat{x} = \frac{1}{M} \sum_{i=1}^{M} y_i$ (least square estimate $W = I$)

For Markov estimates:

$$x - \hat{x} = \sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}(x - y_i) \Big/ \sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}$$

$$\sigma_{\hat{x}}^2 = E\{(x - \hat{x})^2\} = \frac{1}{\sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}} \quad (*)$$

For classical least square estimate $\hat{x}$ of x:

$$\sigma_{\hat{x}}^2 = \frac{1}{M^2} \sum_{i=1}^{M} \sigma_{N_i}^2 \geq (*)$$

the variance of Markov is less then this variance

In this case :

$$\Lambda(x) = (1 \dots 1)\, \Psi_N^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \sum_{i=1}^{M} \frac{1}{\sigma_{N_i}^2}$$

~ Example 4 : estimate resistance R

R resistance , m measurements $\left(\frac{\text{voltage}}{\text{current}}\right)$, n noise

$$\underbrace{\begin{bmatrix} v_1 \\ \vdots \\ v_m \end{bmatrix}}_{\substack{\text{voltage measurements} \\ V}} = \underbrace{\begin{bmatrix} i_1 \\ \vdots \\ i_m \end{bmatrix}}_{\substack{\text{current measurements} \\ I}} R + \underbrace{\begin{bmatrix} n_1 \\ \vdots \\ n_m \end{bmatrix}}_{\substack{\text{noise} \\ N}}$$

$\Psi_N = \text{diag}\{\sigma_{n_1}^2 \dots \sigma_{n_m}^2\}$

$E\{N\} = 0$ uncorrelated noise with zero mean

assume $N \in \mathcal{N}(0, \Psi_N)$

Maximum likelihood estimate:

$$\hat{R} = \frac{\sum_{j=1}^{m} \left(i_j v_j / \sigma_{N_j}^2 \right)}{\sum_{j=1}^{m} \left(i_j^2 / \sigma_{N_j}^2 \right)}$$

If the current is affected by noise instead of voltage:

$I = V \cdot \frac{1}{R} + H$ ⤷ noise

⤷ admittance Z

In this case the estimate is

$$\hat{Z} = \frac{\sum_{i=1}^{m} i_j v_j / \sigma_{H_j}^2}{\sum_{i=1}^{m} v_j^2 / \sigma_{H_j}^2}$$

If all the variances $\sigma_{N_j}^2$ or $\sigma_{H_j}^2$ are equal

$$\hat{R} = \frac{V^T I}{I^T I} \qquad \hat{Z} = \frac{V^T I}{V^T V}$$

Cauchy - Schwarz : $\hat{R} \cdot \hat{Z} = \frac{(V^T I)^2}{(I^T I)(V^T V)} \leq 1$

~ Example 5

$Y = AX + N$, $N \in \mathcal{N}(0, \Psi_N)$, $X \in \mathcal{N}(m_x, \Psi_x)$

Y is also gaussian:

$\underbrace{E\{Y\}}_{m_y} = E\{AX + N\} = A m_x$

$\Psi_y = E\{(A(X - m_x) + N)(A(X - m_x) + N)^T\} = A \Psi_x A^T + \Psi_N$

For simplicity we suppose N and X independent

Cross covariance:

$\Psi_{XY} = E\{(X - m_x)(Y - m_y)^T\} = E\{(X - m_x)(A(X - m_x) + N)^T\} =$
$= \Psi_x A^T$

Therefore $Z = \begin{pmatrix} X \\ Y \end{pmatrix} \in \mathcal{N}(m_z, \Psi_z)$ $\quad m_z = \begin{pmatrix} m_x \\ A m_x \end{pmatrix}$

$$\Psi_z = \begin{bmatrix} \Psi_x & \Psi_x A^T \\ A \Psi_x & A \Psi_x A^T + \Psi_N \end{bmatrix}$$

Estimates with minimum variance of X given Y

$\hat{x} = m_x + \Psi_{XY} \Psi_Y^{-1} (Y - m_Y) = m_x + \Psi_x A^T (A \Psi_x A^T + \Psi_N)^{-1} (Y - A m_x)$

solution of $\hat{x} = \underset{x}{\operatorname{argmin}}\, S(x, Y)$

with $S(x, Y) = (x - m_x)^T \Psi_x^{-1} (x - m_x) + (Y - Ax)^T \Psi_N^{-1} (Y - Ax)$

$\hat{x}$ is close to m_x (*)

$A\hat{x}$ is close to y (**)

Since no a priori information on X (mean, covariance) available

- Maximum likelihood estimation (satisfying **):

$\hat{x} = A^{\#}(Y - A m_x) + m_x$ $\quad A^{\#} = \Psi_x A^T \underbrace{(A \Psi_x A^T + \Psi_N)^{-1}}_{\text{since } \dim(Y) > \dim(X)}$ (•)

$A^{\#} = (\Psi_x^{-1} + A^T \Psi_N^{-1} A) A^T \Psi_N^{-1}$ (••)

(•) = (••)

If $\psi_x^{-1} \to 0$ then $\hat{x} \to$ maximum likelihood estimate since $A^{\#} \to A^{\#}_{\psi_N}$

This corresponds to consider a random vector to be deterministic & unknown

---

Simultaneous estimation of random variable and params

Estimate a random vector X from the measurement vector Y with joint density $p_{X,Y}(x,y;\theta)$ depending on the parameter $\theta \in \mathbb{R}^p$

$$\hat{X} = E_\theta\{X|Y\} = f(Y;\theta)$$

we must find first $\hat{\theta} = \arg\max_\theta p_Y(y;\theta)$ (maximum likelihood)

where $p_Y(y;\theta) = \int_{\mathbb{R}^n} p_{X,Y}(x,y;\theta)\,dx$ and replace $\hat{\theta}$ in

$$\hat{x} = E_\theta\{X|Y\}\big|_{\theta=\hat{\theta}}$$

~ Example:

Consider $Y = AX + B\theta + \underbrace{N}_{\text{independent and gaussian}}$ $\quad X \in \mathbb{R}^n,\ Y \in \mathbb{R}^m,\ \theta \in \mathbb{R}^p$

$X \in N(m_x, \Psi_x) \quad N \in N(0, \Psi_N)$ (a priori information)

$$\hat{X} = m_x + \Psi_{XY}(\theta)\psi_Y^{-1}(\theta)(Y - m_x(\theta)) \quad , \quad \begin{aligned} m_Y(\theta) &= Am_x + B\theta \\ \Psi_{XY}(\theta) &= \Psi_x A^T \\ \Psi_Y(\theta) &= A\Psi_x A^T + \Psi_N \end{aligned}$$

$$= m_x + \Psi_x A^T(A\Psi_x A^T + \Psi_N)^{-1}(Y - Am_x - B\theta) \quad (*)$$

since Y is gaussian: $p_Y(y,\theta) = \dfrac{1}{(2\pi)^{m/2}(\det\Psi_Y)^{1/2}} \cdot e^{-\frac{1}{2}(y - m_Y(\theta))^T \Psi_Y^{-1}(y - m_Y(\theta))}$

$$\hat{\theta} = \arg\max_\theta p_Y(y,\theta) = \arg\max_\theta \ln p_Y(y,\theta)$$

Assuming rank $B = p$: $\hat{\theta} = (B^T\Psi_Y^{-1}B)^{-1}B^T\Psi_Y^{-1}(y - Am_x)$

replacing $\hat{\theta}$ in $(*)$:

$$\hat{x} = m_x + \Psi_x A^T \cdot \Psi_Y^{-1} \cdot (I - B(B^T\Psi_Y^{-1}B)^{-1}B^T\Psi_Y^{-1})(y - Am_x)$$