BLUE (Best Linear Unbiased Estimator)

○ Non-Bayesian Estimation Problems

- In Bayesian case,
  - $\underline{\Theta}$ random,
  - $\underline{Y}$ random,
  - $\underline{\Theta}$ & $\underline{Y}$ are jointly random. ($f_{\underline{\Theta}\underline{Y}}(\underline{\theta},\underline{y})$ given)

In non-Bayesian case,
- $\underline{\theta}$ not random or $f_{\underline{\Theta}}(\theta)$ not available
- $\underline{Y}$ random
- $\{f_{\underline{Y};\theta}(\underline{y}) \mid \underline{\theta} \in I\}$ given

I: parameter/index set. uncountable set. (if countable, detection problems!)

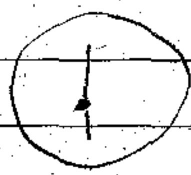
(The use of $f_{\underline{Y}|\theta}(\underline{y})$ is O.K. but it's not a conditional density.)

- ex/ Given a biased coin w/ unknown $Pr(\text{Head}) = \theta$. We want to estimate θ based on observation $\underline{Y} = \underline{y}$

- What about using Bayes-risk minimization?
  → Does not work. We cannot average w.r.t. $\underline{\theta}$.
  Similar situation to non-Bayesian detection problems.

- We only have $f_{\underline{Y};\underline{\theta}}(\underline{y})$. Hence, $E\{\ \}$ means $\int_{\underline{y}} \cdot\ f_{\underline{Y};\theta}(\underline{y})\,d\underline{y}$ which is dependent on $\underline{\theta}$. So $E_{\underline{\theta}}\{\cdot\}$.

○ A broken clock & its analogue

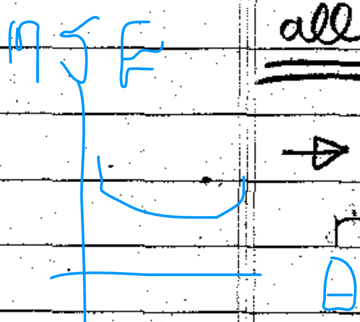
-  Suppose that you have a broken clock stopped at 6 o'clock.

  Perfect twice a day. So what?

- When $\theta \in [0, 1]$, what about using
  $\hat{\theta}(\underline{Y}) = 0.5$?
  $\rightarrow$ Perfect only when $\theta = 0.5$.

- We want $\underline{\hat{\theta}(Y)}$ that works good for <u>all</u> $\underline{\theta} \in I$.

  $\rightarrow$ "works good for all $\underline{\theta}$" needs to be refined.



○ Accuracy vs Precision for measurement devices

- Accuracy (정확) : a measure of reliability,
  accurate $\Leftrightarrow$ $\theta - \hat{\underline{\theta}}(\underline{Y})$ is small in some sense for all $\underline{\theta}$

- Precision (정밀) : a measure of repeatability
  precise $\Leftrightarrow$ $\hat{\underline{\theta}}(\underline{Y})$ does not vary a lot given the same $\underline{\theta}$. $\forall \theta$

- A choice for accuracy & precision pair

$\hat{\underline{\theta}}(\underline{y})$ is accurate : /ex/ $\underbrace{E_{\underline{\theta}}\left[\hat{\underline{\theta}}(\underline{Y}) - \underline{\theta}\right]}_{\triangleq \text{ bias}} = 0 \quad \forall \underline{\theta} \in I$

$= 0.$

$\hat{\underline{\theta}}(\underline{y})$ is precise : /ex/ $\underbrace{E_{\underline{\theta}}\{\|\hat{\underline{\theta}}(\underline{Y}) - \underline{\theta}\|^2\}}_{\text{given bias}=0,\ \text{this becomes variance}}$ small $\forall \underline{\theta}$ ← MSE

MSE w/ non-random θ

different environments

○ **BLUE** (Best Linear Unbiased Estimator)

- $\hat{\underline{\theta}}_{BLUE}(\underline{y}) \triangleq \underset{\hat{\underline{\theta}}(\underline{y})}{\operatorname{argmin}} \; E_{\underline{\theta}}\{\|\hat{\underline{\theta}}(\underline{Y}) - \underline{\theta}\|^2\}, \forall \underline{\theta}$

cf) $E_X\{\ \}$, $E_\theta\{\ \}$
average over X . average parameterized by θ.

subject to $E_{\underline{\theta}}\{\hat{\underline{\theta}}(\underline{Y})\} = \underline{\theta}, \forall \underline{\theta}$

$\hat{\underline{\theta}}(\underline{y}) = W\underline{y}$

- "Best" : $\underbrace{\text{uniformly}}_{\forall \underline{\theta} \in I}$ $\underbrace{\text{minimum variance}}_{\min E_{\underline{\theta}}\{\|\hat{\underline{\theta}}(\underline{Y}) - \underline{\theta}\|^2\}}$ (if unbiased) MSE

- "Linear" : $\hat{\underline{\theta}}(\underline{y}) = W\underline{y}$

- "Unbiased" : $E_{\underline{\theta}}\{\hat{\underline{\theta}}(\underline{Y})\} = \underline{\theta}, \ \forall \underline{\theta} \in I$

- If we do not have the linearity constraint, the estimator, if it exists, is called the Uniformly Minimum Variance Unbiased (**UMVU**)

· Existence of unbiased estimator : Example

$N(\theta \mathbf{1}, \sigma^2 I)$

$\uparrow$ unknown. $\theta \in (-\infty, \infty)$

The estimator $\dfrac{Y^T \mathbf{1}}{N} = \dfrac{1}{N} \sum_{i=1}^{N} y_i$

is unbiased.

(proof) $E_\theta \left\{ \frac{1}{N} \sum_{i=1}^{N} Y_i \right\} = \frac{1}{N} \sum_{i=1}^{N} \theta = \theta, \ \forall \theta$

estimator, or Minimum Variance Unbiased Estimator **(MVUE)**.

- Existence of BLUE, UMVUE) ← not guaranteed!

- Recall the definition of UMP detector.

  U ↑ for all θ ↑ nuisance parameter; MP ↑ maximum P_D.

- Recall the definition of MMSE estimator

  ↑ for Bayesian problems. average w.r.t. $\underline{Y}$ & $\underline{\Theta}$ of $\| \hat{\underline{\theta}}(\underline{Y}) - \underline{\Theta} \|^2$

Ref. Kay's book. (K1)

○ BLUE for **scalar** parameter $\theta \,(\in \mathbb{R}^{1\times 1})$

- Definition

$$\hat{\theta}_{BLUE}(\underline{y}) = \underset{\hat{\theta}(\underline{y})}{\operatorname{argmin}} \; E_\theta\{(\hat{\theta}(\underline{Y})-\theta)^2\} \quad \forall \theta$$

s.t. $E_\theta\{\hat{\theta}(\underline{Y})\} = \theta, \; \forall \theta$

& $\underbrace{\hat{\theta}(\underline{y})}_{\mathbb{R}^{1\times 1}} = \underbrace{\underline{w}^T}_{\mathbb{R}^{1\times N}} \underbrace{\underline{y}}_{\mathbb{R}^{N\times 1}}$.

on average its very accurate.

- Existence (Example)

$\underline{Y} \sim N(\underline{0}, \sigma^2 I)$, $\sigma^2 > 0$ is unknown.

$\theta := \sigma^2$.

$\hat{\theta}_{BLUE}(\underline{y}) = ?$

A. $E_\theta\{\underline{w}^T \underline{Y}\} = \underline{w}^T E_\theta\{\underline{Y}\} = 0, \forall \theta$

$\therefore$ BLUE does not exist ($\because$ No w can have $E_\theta\{\underline{w}^T \underline{Y}\} = \theta, \forall \theta$)

- Existence of BLUE & linear observation model

$$\underset{\mathbb{R}^{N\times 1}}{\underline{Y}} = \underset{\mathbb{R}^{N\times 1}}{\underline{x}}\ \theta + \underset{\text{known}}{\underline{N}}$$

Assume (i) $\mathbb{R}^{N\times 1}$ known (ii) $E_\theta\{\underline{N}\} = \underline{0}, \forall \theta$

$\Rightarrow E_\theta\{\underline{w}^T \underline{Y}\} = E_\theta\{\underline{w}^T(\underline{x}\theta + \underline{N})\} = (\underline{w}^T \underline{x})\theta$.

Note. $\underline{w} \in \mathbb{R}^{N\times 1}$

Hence, $\underline{w}^T \underline{x} = 1 \Rightarrow$ an Unbiased estimator exists

- In the above linear case, & $E_\theta\{\underline{N}\underline{N}^T\} = R, \forall \theta$

$$\underline{w}_{BLUE} = \underset{\underline{w}}{\operatorname{argmin}}\ E_\theta\{(\underline{w}^T \underline{Y} - \theta)^2\} \ \forall \theta$$

$$\text{s.t.} \quad \underline{w}^T \underline{x} = 1$$

Since

$$E_\theta\{(\underline{w}^T\underline{Y}-\theta)^2\} = E_\theta\{(\cancel{\underline{w}^T\underline{x}\theta} + \underline{w}^T\underline{N} - \cancel{\theta})^2\}$$

$$= E_\theta\{(\underline{w}^T\underline{N})^2\} = E_\theta\{(\underline{w}^T\underline{N})(\underline{w}^T\underline{N})\}$$

$$= \underline{w}^T E\{\underline{N}\underline{N}^T\}\underline{w} = \underline{w}^T R \underline{w}, \quad \forall\theta$$

we need to solve

$$\underline{w}_{BLUE} = \arg\min_{\underline{w}} \underline{w}^T R \underline{w} \quad (\forall\underline{\theta})$$

$$\text{s.t.} \quad \underline{w}^T\underline{x} = 1. \quad (\Leftrightarrow \underline{w}^T\underline{x} - 1 = 0)$$

This is a quadratic minimization problem w/ an equality constraint. Hence, we use **the Lagrange multiplier technique**. Let $L(\underline{w},\lambda)$ be **the Lagrangian function** defined as

$$L(\underline{w},\lambda) \triangleq \underline{w}^T R \underline{w} + \lambda(\underline{w}^T\underline{x} - 1).$$

R is symmetric

Then, FONC implies

$$\left[\begin{array}{l} \left.\frac{\partial L}{\partial \underline{w}}\right|_{\underline{w}^*,\lambda^*} = \underline{0} \Rightarrow 2R\underline{w}^* + \lambda^*\underline{x} = \underline{0} \\ \qquad\qquad \Rightarrow \underline{w}^* = -\frac{\lambda^*}{2}R^{-1}\underline{x} \quad \cdots (*) \\ \left.\frac{\partial L}{\partial \lambda}\right|_{\underline{w}^*,\lambda^*} = 0 \Rightarrow \underline{w}^{*T}\underline{x} = 1 \quad \cdots (**) \end{array}\right.$$

Note that (*) and (**) lead to

$$-\frac{\lambda^*}{2}\underline{x}^T R^{-1}\underline{x} = 1 \Rightarrow \lambda^* = \frac{-2}{\underline{x}^T R^{-1}\underline{x}}$$

$$\Rightarrow \underline{w}^* = \frac{R^{-1}\underline{x}}{\underline{x}^T R^{-1}\underline{x}}$$

Therefore,

$$\hat{\Theta}_{BLUE}(\underline{y}) = \overbrace{\frac{\underline{x}^T R^{-1} \underline{y}}{\underline{x}^T R^{-1} \underline{x}}}^{\underline{w}^{*T}} \quad \text{\} matched filter?!}$$

&

$$E_\theta\{(\hat{\Theta}_{BLUE}(\underline{Y}) - \underline{\theta})^2\} = \underline{w}^{*T} R \, \underline{w}^*$$

$$= \frac{1}{\underline{x}^T R^{-1} \underline{x}}, \quad \forall \theta$$

not a function of θ.

" The BLUE exists for $\underline{Y} = \underline{x}\theta + \underline{N}$ w/ $E_\theta\{\underline{N}\} = \underline{0}$, $E_\theta\{\underline{N}\underline{N}^T\} = R$ $\forall \theta$. "

What if $E_\theta\{\underline{N}\} = 0$ $\forall \theta$ but $E_\theta\{\underline{N}\underline{N}^T\} = R_\theta$?

→ The BLUE does not exist in general.

Proof?

○ BLUE for **vector** parameter $\underline{\theta} \in \mathbb{R}^{M \times 1}$.

- Def.

$$\hat{\underline{\theta}}_{BLUE}(\underline{y}) = \underset{\hat{\underline{\theta}}(\underline{y})}{\operatorname{argmin}} \; E_{\underline{\theta}}\{\|\hat{\underline{\theta}}(\underline{Y}) - \underline{\theta}\|^2\}, \; \forall \underline{\theta}$$

s.t. $\hat{\underline{\theta}}(\underline{y}) = W \underline{y}$ &

$E_{\underline{\theta}}\{W\underline{Y}\} = \underline{\theta}$, $\forall \underline{\theta}$

where $W \in \mathbb{R}^{M \times N}$.

- BLUE for linear observation model

$$\underline{Y} = X\underline{\theta} + \underline{N}$$

where $\underline{Y} \in \mathbb{R}^{N\times 1}$

$\underline{\theta} \in \mathbb{R}^{M\times 1}$ unknown non-random parameter

$X \in \mathbb{R}^{N\times M}$ known **$M \le N$** Rank$(X)=M$.

$\underline{N} \in \mathbb{R}^{N\times 1}$ $E_{\underline{\theta}}\{\underline{N}\}=0$, $E_{\underline{\theta}}\{\underline{N}\underline{N}^T\}=R$, $\forall \underline{\theta}$

symmetric positive definite.

Hence, the problem can be rewritten as,

$$W_{BLUE} = \arg\min_{W} E_{\underline{\theta}}\{\|W\underline{Y}-\underline{\theta}\|^2\}, \ \forall \underline{\theta}$$

$$\text{s.t.} \quad E_{\underline{\theta}}\{W\underline{Y}\} = \theta, \ \forall \underline{\theta}.$$

Note that the constraint is $WX = I$.
why?

the objective function is

$$E_{\underline{\theta}}\{\|W\underline{N}\|^2\} = E_{\underline{\theta}}\{(W\underline{N})^T(W\underline{N})\}$$

$$= tr[E_{\underline{\theta}}\{W\underline{N}\underline{N}^TW^T\}] = tr\{WRW^T\}$$

Now, the problem can be rewritten as

independent of $\underline{\theta}$

$$W_{BLUE} = \arg\min_{W} tr\{WRW^T\}$$

$$\text{s.t.} \quad WX = I.$$

How can we solve?

Let $W = \begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_M^T \end{bmatrix}$ Then, ...

$$WRW^T = \begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_M^T \end{bmatrix} R \begin{bmatrix} \underline{w}_1 & \underline{w}_2 & \cdots & \underline{w}_M \end{bmatrix}$$

$$= \begin{bmatrix} \underline{w}_1^T R \underline{w}_1 & \underline{w}_1^T R \underline{w}_2 & \cdots & \underline{w}_1^T R \underline{w}_M \\ \underline{w}_2^T R \underline{w}_1 & \underline{w}_2^T R \underline{w}_2 & \cdots & \underline{w}_2^T R \underline{w}_M \\ \vdots & & & \\ \underline{w}_M^T R \underline{w}_1 & \cdots & & \underline{w}_M^T R \underline{w}_M \end{bmatrix}$$

$$\Rightarrow \operatorname{tr}(WRW^T) = \sum_{i=1}^{M} \underline{w}_i^T R \underline{w}_i \quad \cdots (*)$$

In addition, $WX = I \Rightarrow \underbrace{\begin{bmatrix} \underline{w}_1^T \\ \underline{w}_2^T \\ \vdots \\ \underline{w}_M^T \end{bmatrix} \begin{bmatrix} \underline{x}_1 & \underline{x}_2 & \cdots & \underline{x}_M \end{bmatrix}}_{=I}$

Let $X \triangleq [\underline{x}_1\ \underline{x}_2 \cdots \underline{x}_M]$

$$\Rightarrow \begin{bmatrix} \underline{w}_1^T \underline{x}_1 & \underline{w}_1^T \underline{x}_2 & \cdots & \underline{w}_1^T \underline{x}_M \\ \underline{w}_2^T \underline{x}_1 & \underline{w}_2^T \underline{x}_2 & \cdots & \underline{w}_2^T \underline{x}_M \\ \vdots & & & \\ \underline{w}_M^T \underline{x}_1 & \underline{w}_M^T \underline{x}_2 & \cdots & \underline{w}_M^T \underline{x}_M \end{bmatrix} = \begin{bmatrix} 1 & & & \underline{0} \\ & 1 & & \\ & & \ddots & \\ \underline{0} & & & 1 \end{bmatrix} \quad \cdots (**)$$

By combining (i) & (ii), we obtain

$$W_{BLUE} = \underset{\{\underline{w}_i\}_i}{\operatorname{argmin}} \sum_{i=1}^{M} \underline{w}_i^T R \underline{w}_i$$

M^2 constraints

$$\text{s.t.} \quad \underline{w}_i^T \underline{x}_j = \delta_{ij} \quad \forall i,j \in \{1,2,\cdots,M\}$$

Kronecker delta function

$$\equiv \text{(i)} \quad \underline{w}_1^* = \underset{\underline{w}_1}{\operatorname{argmin}} \ \underline{w}_1^T R \underline{w}_1$$

$$\text{s.t.} \quad \left.\begin{aligned} \underline{w}_1^T \underline{x}_1 &= 1 \\ \underline{w}_1^T \underline{x}_2 &= 0 \\ &\vdots \\ \underline{w}_1^T \underline{x}_M &= 0 \end{aligned}\right\} \ M \text{ constraints}$$

$$\text{(ii)} \quad \underline{w}_2^* = \underset{\underline{w}_2}{\operatorname{argmin}} \ \underline{w}_2^T R \underline{w}_2$$

$$\text{s.t.} \quad \left.\begin{aligned} \underline{w}_2^T \underline{x}_1 &= 0 \\ \underline{w}_2^T \underline{x}_2 &= 1 \\ &\vdots \\ \underline{w}_2^T \underline{x}_M &= 0 \end{aligned}\right\} \ M \text{ constraints}$$

$$\vdots$$

$$\text{(M)} \quad \underline{w}_M^* = \underset{\underline{w}_M}{\operatorname{argmin}} \ \underline{w}_M^T R \underline{w}_M$$

$$\text{s.t.} \quad \left.\begin{aligned} \underline{w}_M^T \underline{x}_1 &= 0 \\ &\vdots \\ \underline{w}_M^T \underline{x}_M &= 1 \end{aligned}\right\} \ M \text{ constraints}$$

Notice that each sub-problem is a quadratic minimization problem w/ M equality constraints. So we can use the Lagrange multiplier technique to find the solution.

ex/ (1)

$$L(w_1, \underline{\lambda}_1) = \underline{w}_1^T R \underline{w}_1 + \lambda_1^T \begin{bmatrix} \underline{w}_1^T \underline{x}_1 - 1 \\ \underline{w}_1^T \underline{x}_2 \\ \vdots \\ \underline{w}_1^T x_M \end{bmatrix}$$

$$= \underline{w}_1^T R \underline{w} + \underline{\lambda}_1^T \left(X^T \underline{w}_1 - \underline{e}_1 \right)$$

1st standard basis vector for $\mathbb{R}^{M \times 1}$

$$\begin{cases} \left.\dfrac{\partial L_1}{\partial \underline{w}_1}\right|_{\underline{w}_1^*, \underline{\lambda}_1^*} = \underline{0} \Rightarrow 2R\underline{w}_1^* + X\underline{\lambda}_1^* = \underline{0} \\ \left.\dfrac{\partial L_1}{\partial \underline{\lambda}_1}\right|_{\underline{w}_1^*, \underline{\lambda}_1^*} = \underline{0} \Rightarrow X^T \underline{w}_1^* = \underline{e}_1 \end{cases}$$

$$\Rightarrow \begin{cases} \underline{w}_1^* = -\frac{1}{2} R^{-1} X \underline{\lambda}_1^* \\ X^T \underline{w}_1^* = \underline{e}_1 \end{cases}$$

(R assumed invertible.)

$$\Rightarrow \quad -\frac{1}{2}\left(X^T R^{-1} X\right) \underline{\lambda}_1^* = \underline{e}_1$$

$$\Rightarrow \quad \underline{\lambda}_1^* = -2 \left(X^T R^{-1} X\right)^{-1} \underline{e}_1$$

$$\underline{w}_1^{*T} = \underline{e}_1^T \left(X^T R^{-1} X\right)^{-1} X^T R^{-1}$$

In general, $\underline{w}_i^{*T} = \underline{e}_i^T (X^T R^{-1} X)^{-1} X^T R^{-1}$

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

Hence, $$W^* = \begin{bmatrix} \underline{w}_1^{*T} \\ \underline{w}_2^{*T} \\ \vdots \\ \underline{w}_M^{*T} \end{bmatrix} = \underbrace{\begin{bmatrix} \underline{e}_1^T \\ \underline{e}_2^T \\ \vdots \\ \underline{e}_M^T \end{bmatrix}}_{I} (X^T R^{-1} X)^{-1} X^T R^{-1}$$

$$= (X^T R^{-1} X)^{-1} X^T R^{-1}$$

$$\therefore \boxed{\hat{\underline{\theta}}_{BLUE}(\underline{y}) = (X^T R^{-1} X)^{-1} X^T R^{-1} \underline{y}}$$

Now, we evaluate the performance:

$$E_\theta \{ \| W_{BLUE} \underline{Y} - \underline{\theta} \|^2 \} = E_\theta \{ \| W_{BLUE} \underline{N} \|^2 \}$$
$$= tr [E_\theta \{ W_{BLUE} \underline{N} \underline{N}^T W_{BLUE}^T \}]$$

$$= tr \{ W_{BLUE} R W_{BLUE}^T \} = tr \{ (X^T R^{-1} X)^{-1} X^T \cancel{R^{-1}} \cancel{R} R^{-1} X (X^T R^{-1} X)^{-1} \}$$
$$= \boxed{tr \{ (X^T R^{-1} X)^{-1} \}}$$