Bayesian Linear Model summarized

$$X = H\theta + w$$

H known $N \times p$ matrix

θ unknown parameter vector $p \times 1$

PDF $\mathcal{N}(\mu_\theta, C_\theta)$ $\leftarrow$

w = WGN with σ^2 and independent of θ

* It can be shown that X and θ are jointly Gaussian ; hence :

(1) $E[\theta \mid X] = E[\theta] + C_{\theta X} C_{XX}^{-1} (X - E(X))$

(2) $C_{\theta \mid X} = C_{\theta\theta} - C_{\theta X} C_{XX}^{-1} C_{X\theta}$

Let $z = \begin{pmatrix} \theta \\ X \end{pmatrix}$ then covariance matrix for z is

$$C_z = \begin{bmatrix} C_{\theta\theta} & C_{\theta X} \\ C_{X\theta} & C_{XX} \end{bmatrix}$$

* It can be shown that

$$\boxed{C_{\theta\theta} = C_\theta} \quad \boxed{C_{\theta X} = C_\theta H^T}$$

$$\boxed{C_{XX} = H C_\theta H^T + C_w}$$

* Substituting those into (1) & (2) we get

(1a) $E[\underline{\theta} \mid X] = \overset{\mu_\theta}{E[\theta]} + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (X - \underbrace{E(X)}_{H\mu_\theta})$

(2a) $C_{\theta \mid X} = C_\theta - C_\theta H^T (H C_\theta H^T + C_w)^{-1} H C_\theta$

Finally, using the matrix inversion lemma (1a) and (2a) can be rewritten as:

$$(1b) \quad E[\theta | x] = \mu_\theta + \left(C_\theta^{-1} + H^T C_w^{-1} H\right)^{-1} H^T C_w^{-1} (x - H\mu_\theta)$$

$$(2b) \quad C_{\theta/x} = \left(C_\theta^{-1} + H^T C_w^{-1} H \right)^{-1}$$

These forms are ~~no~~ generally more useful.

Question: what happens to (1b) as $C_\theta^{-1} \rightarrow 0$ (no prior knowledge, or uninformative prior case)

$$\hat{\theta} \rightarrow \mu_\theta + \left(H^T C_w H \right)^{-1} H^T C_w^{-1} (x - H\mu_\theta)$$

$$= \cancel{\mu_\theta} - \cancel{\mu_\theta} + (H^T C_w H)^{-1} H^T C_w^{-1} x$$

This is the MVU estimator for linear model.

MMSE estimator - vector parameter

Nuisance parameters: Suppose we have a set of unknown parameters which can be broken into 2 subsets: $\underline{\theta} = \begin{pmatrix} \underline{\beta} \\ \underline{\alpha} \end{pmatrix}$
$\underline{\beta}$: parameters to be estimated and $\underline{\alpha}$: parameters we don't care about (nuisance parameters). In classical estimation, we have no choice but to estimate both because of interdependencies.

Bayesian case: $p(\underline{\beta} \mid x) = \int p(\underline{\beta}, \underline{\alpha} \mid x)\, d\underline{\alpha}$

Integrate $\underline{\alpha}$ out.

Vector MMSE: $\underline{\theta}$: $p \times 1$ parameter vector.

Choose θ_i to minimize

$$E[(\theta_i - \hat{\theta}_i)^2] = \int (\theta_i - \hat{\theta}_i)^2 p(x, \theta_i)\, dx\, d\theta_i$$

Notice, this is the same as the scalar case; hence

$$\hat{\theta}_i = E[\theta_i \mid x] = \int \theta_i\, p(\theta_i \mid x)\, d\theta_i$$

Now take $i = 1$, we have

$$\hat{\theta}_1 = \int \theta_1\, p(\theta_1 \mid x)\, d\theta_1$$

$$= \int \theta_1 \left[\int \cdots \int p(\underline{\theta} \mid x)\, d\theta_2 \cdots d\theta_p \right] d\theta_1$$

same trick as in nuisance param.

$$= \int \theta_1\, p(\underline{\theta} \mid x)\, d\underline{\theta}$$

In general, $\hat{\theta}_i = \int \theta_i\, p(\underline{\theta} \mid x)\, d\underline{\theta}$

so $\hat{\underline{\theta}} = \int \underline{\theta}\, p(\underline{\theta} \mid x)\, d\underline{\theta}$

$$\boxed{\hat{\underline{\theta}} = E[\underline{\theta} \mid x]}$$

$$\mathrm{Bmse}(\hat{\theta}_1) = \iint \left(\theta_1 - E[\theta_1 | x]\right)^2 p(\theta_1 | x)\, p(x)\, d\theta_1\, dx$$

$$= \int \Big[\underbrace{\int \left(\theta_1 - E[\theta_1|x]\right)^2 p(\underline{\theta} | x)\, d\underline{\theta}}_{\substack{\text{variance of } \theta_1 \text{ for posterior } p(\theta/x) \\ [C_{\theta|x}]_{ii}}} \Big] p(x)\, dx$$

where $C_{\theta|x} = \underset{\theta|x}{E}\left[(\theta - E(\theta|x))(\theta - E(\theta/x))^T\right]$

<u>Example</u>: Bayesian Fourier analysis

$$x[n] = a \cos 2\pi f_0 n + b \sin 2\pi f_0 n + w[n]$$
$$n = 0, \ldots, N-1$$

* f_0: known frequency, multiple of $1/N$
  $f_0 = K/N$ but $f_0 \neq 0$ or $1/2$
* a, b: unknown parameters. Let $\theta = \begin{pmatrix} a \\ b \end{pmatrix}$
* Prior PDF: $\theta \sim N(0, \sigma_\theta^2 I)$
* θ independent of w, w WGN with σ^2

We can cast this problem in the form of the Bayesian Linear Model: $x = H\theta + w$

$$H = \begin{bmatrix} 1 & 0 \\ \cos 2\pi f_0 & \sin 2\pi f_0 \\ \vdots & \vdots \\ \cos 2\pi f_0 (N-1) & \sin 2\pi f_0 (N-1) \end{bmatrix}$$

Notice $\mu_\theta = \underline{0} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$, $C_\theta = \sigma_\theta^2 I$, $C_w = \sigma^2 I$

~~Then~~ We also just showed that the vector MMSE estimator is: $\hat{\theta} = E[\theta \mid x]$. Now, using our results from jointly Gaussian random vars.:

$$\hat{\theta} = E[\theta \mid x] = \underbrace{E[\theta]}_{\mu_\theta = 0} + C_\theta H^T (H C_\theta H^T + C_w)^{-1} (x - \underbrace{E(x)}_{H\mu_\theta + E[w] = 0})$$

From summary (1a)

$$\hat{\theta} = \sigma_\theta^2 H^T (\sigma_\theta^2 H H^T + \sigma^2 I)^{-1} x$$

or alternate form

$$\hat{\theta} = \left(\frac{1}{\sigma_\theta^2} I + \frac{1}{\sigma^2} H^T H \right)^{-1} H^T \frac{1}{\sigma^2} x$$

and

$$C_{\theta|x} = \left(\frac{1}{\sigma_\theta^2} I + \frac{1}{\sigma^2} H^T H \right)^{-1}$$

Substituting $H^T H = I \times \frac{N}{2}$ (Orthogonal basis) we get

$$\hat{\theta} = \frac{1/\sigma^2}{1/\sigma_\theta^2 + \frac{N}{2\sigma^2}} H^T x = \underbrace{\frac{1}{1 + \frac{2\sigma^2/N}{\sigma_\theta^2}}}_{\substack{\text{scale} \\ \text{factor} \leq 1}} \underbrace{\begin{pmatrix} \frac{2}{N} \sum_{n=0}^{N-1} x[n] \cos 2\pi f_0 n \\ \frac{2}{N} \sum_{n=0}^{N-1} x[n] \sin 2\pi f_0 n \end{pmatrix}}_{\substack{\text{Classical} \\ \text{estimator}}}$$

Notice, result biased towards $\underline{0}$ because of prior. If $\sigma_\theta^2 >> 2\sigma^2/N$ (low confidence in prior) then we get the classical estimator.

Also, remember $Bmse(\hat{\theta}_i) = [C_{\theta|x}]_{ii}$

so $C_{\theta|x} = \left(\frac{I}{\sigma_\theta^2} + \frac{H^T H}{\sigma^2} \right)^{-1}$

using $H^T H = N/2$ we get $C_{\theta|x} = \frac{1}{\frac{1}{\sigma_\theta^2} + \frac{N}{2\sigma^2}} I$

so $Bmse(\hat{a}) = Bmse(\hat{b}) = \frac{1}{1/\sigma_\theta^2 + N/2\sigma^2}$

Important Properties of MMSE estimators

① Let $\alpha = A\theta + b$ ← $r \times 1$ vector known

$r \times 1$ parameter vector — known $r \times p$ matrix — unknown param vector $p \times 1$

$$\hat{\alpha}_{MMSE} = E(\alpha|x) = E(A\theta + b|x)$$
$$= A\,E(\theta|x) + b$$
$$\boxed{\hat{\alpha}_{MMSE} = A\,\hat{\theta}_{MMSE} + b}$$

② $\underline{x}_1, \underline{x}_2$ two independent data vectors

$\theta, \underline{x}_1, \underline{x}_2$ jointly Gaussian

Let $\underline{x} = \begin{pmatrix} \underline{x}_1 \\ \underline{x}_2 \end{pmatrix}$ then $\hat{\theta} = E(\theta|x) = \mu_\theta + C_{\theta x} C_{xx}^{-1}(x - E(x))$

$$C_{xx}^{-1} = \begin{bmatrix} C_{x_1 x_1} & C_{x_1 x_2} \\ C_{x_2 x_1} & C_{x_2 x_2} \end{bmatrix}^{-1} = \begin{bmatrix} C_{x_1 x_1}^{-1} & 0 \\ 0 & C_{x_2 x_2}^{-1} \end{bmatrix}$$

0 due to $\underline{x}_1, x_2$ independence

$$C_{\theta x} = E\left[\theta \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}^T\right] = E\left[\left(\theta_1 \underline{x}_1^T \quad \theta_2 \underline{x}_2^T\right)\right]$$

$$= \left(C_{\theta x_1} \quad C_{\theta x_2}\right)$$

so $\hat{\theta} = \mu_\theta + \left[C_{\theta x_1} \; C_{\theta x_2}\right] \begin{bmatrix} C_{x_1 x_1}^{-1} & 0 \\ 0 & C_{x_2 x_2}^{-1} \end{bmatrix} \begin{bmatrix} x_1 - E(x_1) \\ x_2 - E(x_2) \end{bmatrix}$

(*) $$\hat{\theta} = \underbrace{\mu_\theta}_{\substack{\text{Prior} \\ \text{estimator}}} + \underbrace{C_{\theta x_1} C_{x_1 x_1}^{-1} \left(x_1 - E(x_1)\right)}_{\substack{\text{First data set} \\ \text{estimator}}} + \underbrace{C_{\theta x_2} C_{x_2 x_2}^{-1} \left(x - E(x_2)\right)}_{\substack{\text{Second} \\ \text{data set} \\ \text{estimator}}}$$

This is the additivity property for independent data sets.

(3) PDF of error $\varepsilon = \theta - E(\theta \mid x) = \theta - \hat{\theta}$

$$E[\varepsilon] = \iint \left(\theta - E(\theta \mid x)\right) p(\theta, x)\, d\theta\, dx$$

1st term $\int \theta\, p(\theta, x)\, d\theta$

$= \int \theta\, p(\theta \mid x) p(x)\, d\theta$

then $\iint \theta\, p(\theta \mid x)\, d\theta\, p(x)\, dx$

$= \int E(\theta \mid x)\, p(x)\, dx$

2nd term : $\iint E(\theta \mid x)\, p(\theta \mid x) p(x)\, d\theta\, dx$

$$= \int E(\theta \mid x) \underbrace{\left[\int p(\theta \mid x)\, d\theta\right]}_{1} p(x)\, dx$$

so $\boxed{E[\varepsilon] = 0}$. MMSE estimator correct on average. Notice this doesn't mean it is unbiased for a particular θ value

$\boxed{\text{also } \operatorname{var}(\varepsilon) = \operatorname{Bmse}(\hat{\theta})}$