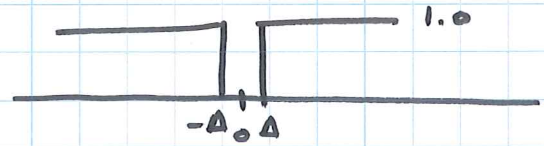
Wednesday March 10

Lect. 21

* No Live Class Monday 3/15 → Owe you 2 recorded classes.

Last Time : MAP Estimation



⇒ A reas. approach is to find max of posterior dist of θ given $Y=y$.

$$w(\theta|y)$$

$$\Rightarrow \quad \underset{\theta}{\operatorname{argmax}}\, w(\theta|y) \triangleq \hat{\theta}_{MAP}(y)$$

$$\Rightarrow \quad \frac{\partial}{\partial\theta} w(\theta|y) = 0 \quad \text{solve for } \theta = \hat{\theta}_{MAP}(y).$$

In all the cases : MMSE, MMAE, MAP our approach in all cases is find cond. dist. of Θ given $Y=y$.

$$w(\theta|y)$$

Our model is

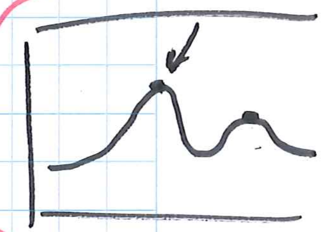
- $\{f_\theta : \theta \in \Lambda\}$ of the cond. pdfs. $Y \mid \Theta = \theta$ $\Big\{ f_{Y|\Theta}(y|\theta)$
- prior dist. on Θ.

To get $w(\theta|y)$ from the given model we use Bayes.

$$w(\theta|y) = \frac{f_\theta(y)\, w(\theta)}{\underbrace{\int_\Lambda f_\theta(y)\, w(\theta)\, d\theta}_{f_Y(y)}}$$

If interested in MAP don't need to comp. denom.

$$\hat{\theta}_{MAP}(y) = \operatorname*{argmax}_\theta w(\theta|y) = \operatorname*{argmax}_\theta f_\theta(y)\, w(\theta)$$



$$\text{ln} \qquad = \operatorname*{argmax}_\theta \log\left(f_\theta(y)\, w(\theta)\right)$$

$$= \operatorname*{argmax}_\theta \left\{ \log f_\theta(y) + \log w(\theta). \right\}$$

Subject to smoothness, a <u>necc.</u> condition for max is

$$\frac{\partial}{\partial\theta} \log f_\theta(y) \Big|_{\theta = \hat{\theta}_{MAP}(y)} = -\frac{\partial}{\partial\theta} \log w(\theta) \Big|_{\theta = \hat{\theta}_{MAP}(y)}.$$

Called the MAP Equation

Estimation of Signal Amplitude

$\Gamma = \mathbb{R}^n \quad \Lambda = \mathbb{R}$ $\qquad N \sim N(0, \Sigma) \qquad N \perp\!\!\!\perp \Theta$

$Y_k = N_k + \Theta s_k \quad 1 \le k \le n$ $\qquad \Theta \sim N(\mu, \upsilon^2)$

s is known.

For all cases want $w(\theta | y)$.

Model $\quad \Theta \sim N(\mu, \upsilon^2)$

$\qquad Y \sim N(\theta s, \Sigma)$ conditioned on $\Theta = \theta$.

$$w(\theta|y) = \frac{f_\theta(y)\, w(\theta)}{\int_\Lambda f_\theta(y)\, w(\theta)\, d\theta} = \frac{\frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(y-\theta s)^T \Sigma^{-1} (y - \theta s)\right\} \times \frac{1}{\sqrt{2\pi}\upsilon} \exp\left\{-(\theta-\mu)^2/2\upsilon^2\right\}}{\text{denom.}}$$

$$= K(y) \exp\left\{-\frac{1}{2} y^T \Sigma^{-1} y + \frac{\theta}{2} s^T \Sigma^{-1} y + \frac{\theta}{2} y^T \Sigma^{-1} s - \frac{1}{2}\theta^2 s^T \Sigma^{-1} s\right\} \times$$
$$\exp\left\{-\frac{1}{2\upsilon^2}\left[\theta^2 - 2\theta\mu + \mu^2\right]\right\}.$$

With a new def. for $K(y)$ ...

$$w(\theta|y) = K'(y) \exp\left\{ -\frac{\theta^2}{2}\left(d^2 + \frac{1}{\upsilon^2}\right) + \theta\left(s^T \Sigma^{-1} y + \frac{\mu}{\upsilon^2}\right)\right\}$$

$$d^2 = s^T \Sigma^{-1} s$$

Know $w(\theta|y)$ is Gaussian .... exponential of a quadratic in θ so

$$w(\theta|y) \sim N(m, q^2)$$

$$= \frac{1}{\sqrt{2\pi}\, q} e^{-(\theta - m)^2/2q^2}$$

$$= \frac{e^{-m^2/2q^2}}{\sqrt{2\pi}\, q} e^{-\frac{\theta^2 q^{-2}}{2} + \theta m q^{-2}}$$

Equate terms

$$q^2 = \left(d^2 + \frac{1}{\upsilon^2}\right)^{-1} \qquad m = \left(d^2 + \frac{1}{\upsilon^2}\right)^{-1}\left(s^T \Sigma^{-1} y + \frac{\mu}{\upsilon^2}\right)$$

$$K'(y) = e^{-m^2/2q^2} / \sqrt{2\pi}\, q$$

Immediately ...

$$\hat{\Theta}_{MMSE}(y) = \frac{s^T \Sigma^{-1} y + \frac{\mu}{\upsilon^2}}{d^2 + \frac{1}{\upsilon^2}} = \frac{\upsilon^2 d^2 \hat{\Theta}_1(y) + \mu}{\upsilon^2 d^2 + 1}$$

having defined $\hat{\Theta}_1(y) = s^T \Sigma^{-1} y / d^2$.

Also follows ...

$$MMSE = E\{Var\{\Theta \mid Y\}\} = \frac{1}{d^2 + 1/\upsilon^2} = \frac{\upsilon^2}{\upsilon^2 d^2 + 1}.$$

where note $Var\{\Theta \mid Y\}$ does not dep. on Y.

Turns out .... also have $\hat{\Theta}_{ABS}(y) = \hat{\Theta}_{MAP}(y) = \hat{\Theta}_{MMSE}(y)$.

Interpretation: $\hat{\Theta}_{MMSE}(y) = \dfrac{v^2 d^2\, \hat{\Theta}_1(y) + \mu}{v^2 d^2 + 1}$

v^2 determines accuracy of prior knowledge about Θ

d^2 is a meas. of quality with which s can be distinguished from $N(0, \Sigma)$ noise.

(1) $v^2 d^2$ is small. $\Longrightarrow$ $\hat{\Theta}_{MMSE}(y) \approx \mu$

(2) $v^2 d^2$ is large $\Longrightarrow$ $\hat{\Theta}_{MMSE}(y) \approx \hat{\Theta}_1(y)$ estimate driven by obs. & not prior model.

$MSE \approx 1/d^2$

Special Case:

$\Sigma = \sigma^2 I \qquad s = [1, 1, 1 \ldots 1] \qquad v^2 d^2 = n v^2/\sigma^2$

$$\hat{\Theta}_1(y) = \frac{1}{n} \sum_{k=1}^{n} y_k.$$