Let $\beta = (\beta_1, \cdots, \beta_p)^T$ and $\eta = (\eta_1, \cdots, \eta_n)^T$ be two jointly dist. random vectors with joint PDF $p_{(\beta,\eta)}(t,s)$

Def: The <u>conditional PDF</u> of β given $\eta = s$ is $p_{\beta|\eta}(t|s) \stackrel{def}{=} \dfrac{p_{(\beta,\eta)}(t,s)}{p_\eta(s)}$ where $p_\eta(s)$ is the marginal PDF of η

The <u>conditional PDF</u> of η given $\beta = t$ is $p_{\eta|\beta}(s|t) \stackrel{def}{=} \dfrac{p_{(\beta,\eta)}(t,s)}{p_\beta(t)}$ where $p_\beta(t)$ is the marginal PDF of β.

Bayes' Theorem:

$$\left(p_{(\beta,\eta)}(t,s) = \right) p_{\beta|\eta}(t|s)\, p_\eta(s) = p_{\eta|\beta}(s|t)\, p_\beta(t)$$

5.3.2 Theory

In this section, we study the Bayersian model:

$$\eta = X^T \beta + \epsilon$$

where $X \in \mathbb{R}^{p\times n}$,

$\beta = (\beta_1, \cdots, \beta_p)^T$ with i.i.d $\beta_i \sim N(0, \sigma_\beta^2)$

$\epsilon = (\epsilon_1, \cdots, \epsilon_n)^T$ with i.i.d $\epsilon_i \sim N(0, \sigma_\epsilon^2)$

Remark: $p_\beta(t)$ is known as the <u>prior PDF</u>

$p_{\beta|\eta}(t|s)$ - - - - - - - <u>a posterior PDF</u>

$p_{\eta|\beta}(s|t)$ - - - - - - - <u>the likelihood PDF</u>

Def: The *maximum a posteriori* (MAP) *estimator* of β is defined as

$$\beta^{map} \overset{def}{=} \arg\max_{t} p_{\beta|\eta}(t|s)$$

Prop: $\beta^{map} = (XX^T + \lambda I)^{-1} X s$ where $\lambda = \frac{\sigma_\epsilon^2}{\sigma_\beta^2}$

Proof: The PDFs of β and ϵ are (from previous lecture)

$$p_\beta(t) = \frac{1}{\sqrt{(2\pi)^p}\,\sigma_\beta^p} e^{-\frac{\|t\|^2}{2\sigma_\beta^2}}$$

$$p_\epsilon(r) = \frac{1}{\sqrt{(2\pi)^n}\,\sigma_\epsilon^n} e^{-\frac{\|r\|^2}{2\sigma_\epsilon^2}}$$

To apply Beyers' Thm, we compute $p_{\eta|\beta}(s|t)$:

notice that if $\beta = t$ is realized, then

$$\eta = \underbrace{X^T t}_{\text{deterministic}} + \underbrace{\epsilon}_{\text{random}} \quad \text{so} \quad \eta_i \sim N(X^T t, \sigma_\epsilon^2), \text{ thus}$$

$$p_{\eta|\beta}(s|t) = \frac{1}{\sqrt{(2\pi)^n}\,\sigma_\epsilon^n} e^{-\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2}}$$

Take "ln" in the Beyers' Thm $p_{\beta|\eta}(t|s) = \frac{p_{\eta|\beta}(s|t)\, p_\beta(t)}{p_\eta(s)}$

$$\ln p_{\beta|\eta}(t|s) = \ln p_{\eta|\beta}(s|t) + \ln p_\beta(t) - \ln p_\eta(s)$$

$$= \underbrace{\ln \frac{1}{\sqrt{(2\pi)^n}\,\sigma_\epsilon^n}}_{\text{const in } t} - \underbrace{\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2}}_{\text{quadratic in } t}$$

$$+ \underbrace{\ln \frac{1}{\sqrt{(2\pi)^p}\,\sigma_\beta^p}}_{\text{const in } t} - \underbrace{\frac{\|t\|^2}{2\sigma_\beta^2}}_{\text{quadratic in } t}$$

$$- \underbrace{\ln p_\eta(s)}_{\text{const in } t}$$

$$= -\frac{\|s - X^T t\|^2}{2\sigma_\epsilon^2} - \frac{\|t\|^2}{2\sigma_\beta^2} + \text{const in } t$$

$$\|s - X^T t\|^2 = \|s\|^2 - 2s^T X^T t + \|X^T t\|^2$$

$$= -\frac{\|s\|^2 - 2s^T X^T t + \|X^T t\|^2}{2\sigma_\epsilon^2} - \frac{\|t\|^2}{2\sigma_\beta^2} + \text{const in } t$$

$$= -\frac{1}{2\sigma_\epsilon^2} t^T X X^T t - \frac{1}{2\sigma_\beta^2} t^T t + \frac{1}{\sigma_\epsilon^2} s^T X^T t$$

$$+ \text{const in } t \left(\text{including } -\frac{\|s\|^2}{2\sigma_\epsilon^2}\right)$$

Set $\frac{\partial}{\partial t} \ln p_{\beta|y}(t|s) = 0$ to find the critical pts:

$$0 = -\frac{1}{\sigma_\epsilon^2} X X^T t - \frac{1}{\sigma_\beta^2} t + \frac{1}{\sigma_\epsilon^2} X s$$

$$\Rightarrow \quad t = \left(X X^T + \frac{\sigma_\epsilon^2}{\sigma_\beta^2} I\right)^{-1} X s$$

As $\frac{\partial^2}{\partial t^2} \ln p_{\beta|y}(t|s) = -\frac{1}{\sigma_\epsilon^2} X X^T - \frac{1}{\sigma_\beta^2} I$ is negative definite, the critical pt is the unique maximizer. Thus

$$\beta^{map} = \left(XX^T + \frac{\sigma_\epsilon^2}{\sigma_\beta^2} I\right)^{-1} Xs \qquad \blacksquare$$

Remark: $\beta^{map} = \beta^{ridge}$ (if we denote s by y).

with i.i.d Gaussian prior.