Maurice Diesendruck

LinMod M1 : 1A

Generally, for M p.p.m onto $C(X)$, $M = X(X'X)^{-}X'$

M_0: orthogonal projection onto $\mathbf{1} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ $\Rightarrow M_0 = \mathbf{1}(\mathbf{1}'\mathbf{1})^{-}\mathbf{1}' = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}(3)^{-}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}$

$$= \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \frac{1}{3} = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

M_1: orthog proj onto $X_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}$ $\Rightarrow M_1 = X_1(X_1'X_1)^{-}X_1' = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\left(\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\right)^{-}\begin{bmatrix} 1 & 0 & -1 \end{bmatrix}$

$$= \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}(2)^{-}\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & 0 & -1 \end{bmatrix} \cdot \frac{1}{2} = \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}$$

M_2: orthog. proj. onto $X_2 = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}$ $\Rightarrow M_2 = X_2(X_2'X_2)^{-}X_2' = \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}\left(\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}\begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}\right)^{-}\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix}$

$$= \begin{bmatrix} 1/3 \\ -2/3 \\ 1/3 \end{bmatrix}\left(\frac{6}{9}\right)^{-}\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} = \begin{bmatrix} \frac{1}{3} \\ -\frac{2}{3} \\ 1/3 \end{bmatrix}\begin{bmatrix} \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \cdot \frac{9}{6} = \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{bmatrix}$$

$$M_0 + M_1 + M_2 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{bmatrix} + \begin{bmatrix} \frac{1}{2} & 0 & -\frac{1}{2} \\ 0 & 0 & 0 \\ -\frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} + \begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I_3 \quad \blacksquare$$

Lin Mod M1 : 1B

MB

Given $Y = 1 \cdot \beta_0 + X_1 \cdot \beta_1 + X_2 \cdot \beta_2 + \varepsilon$ with $\varepsilon \sim N(\underline{0}, \sigma^2 I_3)$, if only 1 and X_1 are considered, are the OLS estimates of intercept (β_0) and slope (β_1) unbiased estimates of β_0 and β_1 in the true model?

Yes.

The OLS estimator for the simpler model, $Y = 1 \cdot \beta_0 + X_1\beta_1 + \varepsilon$ is $\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = (X'X)^{-1}X'Y$

Here, $X = \begin{bmatrix} | & | \\ 1 & X_1 \\ | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix}$

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -1 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$= \frac{1}{6} \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 2 & 2 & 2 \\ 3 & 0 & -3 \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \frac{1}{6} \begin{bmatrix} 2Y_1 + 2Y_2 + 2Y_3 \\ 3Y_1 - 3Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3}(Y_1+Y_2+Y_3) \\ \frac{1}{2}(Y_1 - Y_3) \end{bmatrix}$$

The OLS estimator for the true model, $Y = 1 \cdot \beta_0 + X_1 \cdot \beta_1 + X_2 \cdot \beta_2 + \varepsilon$ is $\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = (X'X)^{-1}X'Y$

Here, $X = \begin{bmatrix} | & | & | \\ 1 & X_1 & X_2 \\ | & | & | \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1/3 \\ 1 & 0 & -2/3 \\ 1 & -1 & 1/3 \end{bmatrix}$

$$\hat{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = \left(\begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ 1/3 & -2/3 & 1/3 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1/3 \\ 1 & 0 & 2/3 \\ 1 & -1 & 1/3 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$= \left(\begin{bmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \frac{2}{3} \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{3}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 \\ 1 & 0 & -1 \\ \frac{1}{3} & -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{2} & 0 & -\frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{1}{3}(Y_1 + Y_2 + Y_3) \\ \frac{1}{2}(Y_1 - Y_3) \\ \frac{1}{2}(Y_1 - 2Y_2 + Y_3) \end{bmatrix}$$

In both cases, $\beta_0 = \frac{1}{3}(Y_1 + Y_2 + Y_3)$ and $\beta_1 = \frac{1}{2}(Y_1 - Y_3)$. ∎

MD

Lin Mod M1 : 1C

Let $\hat{\varepsilon}_s$ denote residuals from fitting simpler model, $Y = 1\cdot\beta_0 + X_1\beta_1 + \varepsilon_s$.

$\hat{\varepsilon}_s = Y - (1\hat{\beta}_0 + X_1\hat{\beta}_1)$. Find $E[\hat{\varepsilon}_t]$ where $\hat{\varepsilon}_t = Y - (1\cdot\hat{\beta}_0 + X_1\hat{\beta}_1 + X_2\hat{\beta}_2)$

$$E[\hat{\varepsilon}_t] = E\left[Y - 1\cdot\hat{\beta}_0 - X_1\hat{\beta}_1 - X_2\hat{\beta}_2\right]$$

$$\hat{\varepsilon}_s = \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} - \left(\tfrac{1}{3}(Y_1+Y_2+Y_3)\right)\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} - \left(\tfrac{1}{2}(Y_1-Y_3)\right)\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} = \underset{(3\times1)}{\begin{bmatrix} Y_1 - \frac{1}{3}(Y_1+Y_2+Y_3) - \frac{1}{2}(Y_1-Y_3) \\ Y_2 - \frac{1}{3}(Y_1+Y_2+Y_3) \\ Y_3 - \frac{1}{3}(Y_1+Y_2+Y_3) + \frac{1}{2}(Y_1-Y_3) \end{bmatrix}} = \underset{(3\times1)}{\begin{bmatrix} \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ -\frac{1}{3}Y_1 + \frac{2}{3}Y_2 - \frac{1}{3}Y_3 \\ \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \end{bmatrix}}$$

Note: $\hat{\varepsilon}_s$ is the residual after projecting Y onto 1 and X_1.
$\hat{\varepsilon}_t$ is the residual after projecting Y onto 1, X_1, and X_2.

Therefore, by subtracting out $\hat{\varepsilon}_s$'s projection onto X_2, what remains is the residual after projecting onto 1, X_1, and X_2. So $\hat{\varepsilon}_s - M_2\hat{\varepsilon}_s = \hat{\varepsilon}_t$.

$$M_2\hat{\varepsilon}_s = \underset{(3\times3)}{\begin{bmatrix} \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \\ -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ \frac{1}{6} & -\frac{1}{3} & \frac{1}{6} \end{bmatrix}} \underset{(3\times1)}{\begin{bmatrix} \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ -\frac{1}{3}Y_1 + \frac{2}{3}Y_2 - \frac{1}{3}Y_3 \\ \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \end{bmatrix}} = \underbrace{\begin{bmatrix} \frac{1}{6}(\frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3) - \frac{1}{3}(-\frac{1}{3}Y_1 + \frac{2}{3}Y_2 - \frac{1}{3}Y_3) + \frac{1}{6}(\frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3) \\ -\frac{1}{3}(\quad\downarrow\quad) + \frac{2}{3}(\quad\downarrow\quad) - \frac{1}{3}(\quad\downarrow\quad) \\ \frac{1}{6}(\quad\downarrow\quad) - \frac{1}{3}(\quad\downarrow\quad) + \frac{1}{6}(\quad\downarrow\quad) \end{bmatrix}}_{(3\times1)}$$

$$\hat{\varepsilon}_t = \hat{\varepsilon}_s - M_2\hat{\varepsilon}_s = \underset{(3\times1)}{\begin{bmatrix} \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \\ -\frac{1}{3}Y_1 + \frac{2}{3}Y_2 - \frac{1}{3}Y_3 \\ \frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3 \end{bmatrix}} - \underset{(3\times1)}{\blacktriangledown}$$

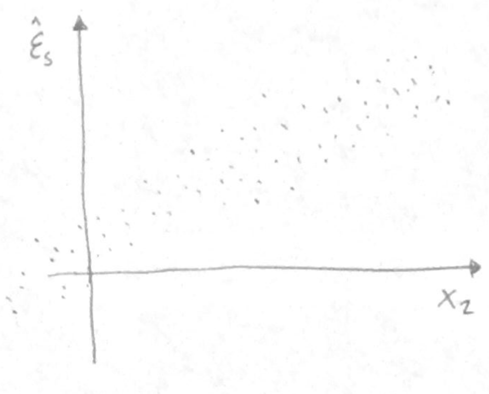
$$= \begin{bmatrix} \frac{5}{6}(\frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3) - \frac{1}{3}(-\frac{1}{3}Y_1 + \frac{2}{3}Y_2 - \frac{1}{3}Y_3) + \frac{1}{6}(\frac{1}{6}Y_1 - \frac{1}{3}Y_2 + \frac{1}{6}Y_3) \\ -\frac{1}{3}(\quad\downarrow\quad) + \frac{1}{3}(\quad\downarrow\quad) - \frac{1}{3}(\quad\downarrow\quad) \\ \frac{1}{6}(\quad\downarrow\quad) - \frac{1}{3}(\quad\downarrow\quad) + \frac{5}{6}(\quad\downarrow\quad) \end{bmatrix} = \begin{bmatrix} \frac{5}{18}Y_1 - \frac{5}{9}Y_2 + \frac{5}{18}Y_3 \\ -\frac{2}{9}Y_1 + \frac{4}{9}Y_2 - \frac{2}{9}Y_3 \\ \frac{5}{18}Y_1 - \frac{5}{9}Y_2 + \frac{5}{18}Y_3 \end{bmatrix}$$

$$\hat{\varepsilon}_t = \begin{bmatrix} \frac{5}{18}(Y_1 - 2Y_2 + Y_3) \\ -\frac{2}{9}(Y_1 - 2Y_2 + Y_3) \\ \frac{5}{18}(Y_1 - 2Y_2 + Y_3) \end{bmatrix} = \begin{bmatrix} \frac{5}{18} \\ -\frac{2}{9} \\ \frac{5}{18} \end{bmatrix} (Y_1 - 2Y_2 + Y_3)$$

(3×1)

Lin Mod M1 : 1D

Given $\hat{\varepsilon}_s$ as residuals on the simple model, $Y = 1 \cdot \beta_0 + X_1\beta_1 + \varepsilon_s$; if the data comes from the true model, $Y = 1\beta_0 + X_1\beta_1 + X_2\beta_2 + \varepsilon_t$, then plotting $\hat{\varepsilon}_s$ versus X_2 will show some correlation, since the information that "should" be expressed by X_2 is forced into the simple model's error term ε_s.



$\hat{\varepsilon}_s \not\perp X_2$ if the data comes from the true model.

MD

Lin Mod M1: 2A

Suppose you have a prior $\beta | g, \sigma^2$ on a p-dimensional vector of regression coefficients: $\beta | g, \sigma^2 \sim N(\underline{0}, \sigma^2 g (X'X)^{-1})$, $\underset{n \times p}{X}$, $r(X) = p$, $g > 0$

Given $\frac{1}{g} \sim Ga(\frac{1}{2}, \frac{n}{2})$, find marginal $p(\beta | \sigma^2)$ (and be aware of relevent normalizing constants)

To solve, integrate out g from $\beta | g, \sigma^2$ to leave marginal $\beta | \sigma^2$.

① $\beta | g, \sigma^2 \sim N(\underline{0}, \sigma^2 g (X'X)^{-1})$

$$= \prod_{i=1}^{p} \left[\left(2\pi \sigma^2 g [(X'X)^{-1}]_{ii}\right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \frac{1}{\sigma^2 g [(X'X)^{-1}]_{ii}} \cdot (\beta_i - 0)^2 \right\} \right]$$

② $\frac{1}{g} \sim Ga(\frac{1}{2}, \frac{n}{2})$

$$= \frac{(\frac{n}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \left(\frac{1}{g}\right)^{\frac{1}{2}-1} e^{-\frac{n}{2}(\frac{1}{g})}$$

Combine (i.e. multiply) ① and ②, and integrate g (i.e. identify kernel) w.r.t. $1/g$, regarding other variables as constant.

$$f = ① \times ② = \prod_{i=1}^{p} \left[\left(2\pi \sigma^2 g [(X'X)^{-1}]_{ii}\right)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \frac{1}{\sigma^2 g [(X'X)^{-1}]_{ii}} \cdot \beta_i^2 \right\} \right] \cdot \frac{(\frac{n}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \left(\frac{1}{g}\right)^{-\frac{1}{2}} e^{-\frac{n}{2}(\frac{1}{g})}$$

Let $[(X'X)^{-1}]_{ii} = v_i$, $f = \prod_{i=1}^{p} \left[(2\pi\sigma^2 v_i)^{-\frac{1}{2}} \cdot \left(\frac{1}{g}\right)^{\frac{1}{2}} \cdot \exp\left\{ -\frac{\beta_i^2}{2\sigma^2 v_i} \cdot \left(\frac{1}{g}\right) \right\} \right] \cdot \frac{(\frac{n}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \left(\frac{1}{g}\right)^{-\frac{1}{2}} e^{-\frac{n}{2}(\frac{1}{g})}$

$$f = \left[\prod_{i=1}^{p} (2\pi\sigma^2 v_i)^{-\frac{1}{2}} \right] \cdot \left(\frac{1}{g}\right)^{\frac{p}{2}} \cdot \exp\left\{ \left[\sum_{i=1}^{p} \frac{-\beta_i^2}{2\sigma^2 v_i} \right] \cdot \frac{1}{g} \right\} \cdot \left[\frac{(\frac{n}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \right] \cdot \left(\frac{1}{g}\right)^{-\frac{1}{2}} e^{-\frac{n}{2}(\frac{1}{g})}$$

$$\propto \left(\frac{1}{g}\right)^{\frac{p-1}{2}} \exp\left\{ \left[\left(\sum_{i=1}^{p} \frac{-\beta_i^2}{2\sigma^2 v_i} \right) - \frac{n}{2} \right] \left(\frac{1}{g}\right) \right\} \sim Ga\left(\frac{p+1}{2}, \left(\sum_{i=1}^{p} \frac{\beta_i^2}{2\sigma^2 v_i} \right) + \frac{n}{2} \right)$$

Lin Mod M1: 2B

MD

$$\underline{Y} \mid \underline{\beta}, \sigma^2, \tfrac{1}{g} \sim N(X\beta, \sigma^2 I_n)$$

$$Y = X\beta + \varepsilon, \quad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 I_n)$$

$$\underline{\beta} \mid g, \sigma^2 \sim N(\underline{0}, \sigma^2 g (X'X)^{-1})$$

$$\frac{1}{g} \sim Ga\left(\frac{1}{2}, \frac{n}{2}\right)$$

Find distribution of $\frac{1}{g} \,\big|\, \underline{\beta}, \sigma^2, \underline{Y}$. Call $f = P\left(\frac{1}{g} \,\big|\, \underline{\beta}, \sigma^2, \underline{Y}\right)$

$$f = \frac{P(\underline{Y}, \underline{\beta}, \sigma^2, \frac{1}{g})}{P(\underline{Y}, \underline{\beta}, \sigma^2)} = \frac{\text{joint}}{\text{marginal}} = \frac{P(\underline{Y} \mid \underline{\beta}, \sigma^2, \frac{1}{g}) P(\underline{\beta} \mid \sigma^2, \frac{1}{g}) P(\frac{1}{g})}{\text{marginal}}$$

$$\propto P(\underline{Y} \mid \underline{\beta}, \sigma^2, \tfrac{1}{g}) P(\underline{\beta} \mid \sigma^2, \tfrac{1}{g}) P(\tfrac{1}{g}) = N(\underline{Y} \mid X\beta, \sigma^2 I_n) \cdot N(\underline{\beta} \mid \underline{0}, \sigma^2 g (X'X)^{-1}) \cdot Ga\left(\tfrac{1}{g} \,\big|\, \tfrac{1}{2}, \tfrac{n}{2}\right)$$

$$y_i \sim N((X\beta)_i, (\sigma^2 I_n)_{ii}) \qquad \beta_i \sim N(0, (\sigma^2 g (X'X)^{-1})_{ii})$$

$$y_i \sim N(m_i, v_i) \qquad \beta_i \sim N(0, w_i)$$

$$\propto \prod_{i=1}^{n} \left[(2\pi v_i)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \cdot \frac{(y_i - m_i)^2}{v_i} \right\} \right]$$

$$\times$$

$$\prod_{i=1}^{n} \left[(2\pi w_i)^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2} \frac{\beta_i^2}{w_i} \right\} \right]$$

$$\times$$

$$\frac{\left(\frac{n}{2}\right)^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} \left(\frac{1}{g}\right)^{-\frac{1}{2}} e^{-\frac{n}{2}\left(\frac{1}{g}\right)}$$

In this step, isolate only those terms with $\frac{1}{g}$. This includes the w_i term, as $w_i = [\sigma^2 g (X'X)^{-1}]_{ii}$. Here, $w_i^{-\frac{n}{2}}$ yields $\left(\frac{1}{g}\right)^{\frac{n}{2}}$.

$$\propto \prod_{i=1}^{n} \left[(2\pi v_i)^{-\frac{1}{2}} (2\pi w_i)^{-\frac{1}{2}} \right] \cdot \exp\left\{ -\frac{1}{2} \left[\sum_{i=1}^{n} \left(\frac{(y_i - m_i)^2}{v_i} + \frac{\beta_i^2}{w_i} \right) + \frac{n}{g} \right] \right\} \cdot \left(\frac{1}{g}\right)^{-\frac{1}{2}}$$

$$= C_1 \left(\frac{1}{g}\right)^{\frac{n}{2}} \cdot \exp\{\ldots\} \cdot \left(\frac{1}{g}\right)^{-\frac{1}{2}}$$

Let $C_2 = \frac{(y_i - m_i)^2}{v_i}$

Let $w_i^{-1} = \frac{1}{g} \cdot C_3$

$$= C_1 \left(\frac{1}{g}\right)^{\frac{n}{2}} \cdot \exp\left\{-\frac{1}{2}\left(\frac{n}{g}\right) - \frac{1}{2}\left(\sum_{i=1}^{n}\left(\frac{(y_i - m_i)^2}{v_i} + \frac{\beta_i^2}{w_i}\right)\right)\right\} \cdot \left(\frac{1}{g}\right)^{-\frac{1}{2}}$$

$$\propto \left(\frac{1}{g}\right)^{\frac{n-1}{2}} \exp\left\{-\frac{n}{2}\left(\frac{1}{g}\right) - \frac{1}{2}\left[\sum_{i=1}^{n}\left(C_2 + \beta_i^2 \cdot C_3 \cdot \left(\frac{1}{g}\right)\right)\right]\right\}$$

$$\propto \left(\frac{1}{g}\right)^{\frac{n-1}{g}} \exp\left\{-\frac{n}{2}\left(\frac{1}{g}\right) - \frac{1}{2}\sum_{i=1}^{n}\left(\beta_i^2 \cdot C_3\left(\frac{1}{g}\right)\right)\right\}$$

$$= \left(\frac{1}{g}\right)^{\frac{n-1}{2}} \cdot \exp\left\{-\frac{n}{2}\left(\frac{1}{g}\right) - \frac{1}{2} \cdot n\left(C_3\left(\frac{1}{g}\right)\right) \cdot \sum_{i=1}^{n} \beta_i^2\right\}$$

$$= \left(\frac{1}{g}\right)^{\frac{n-1}{2}} \exp\left\{-\left[\frac{n}{2} + \frac{n}{2} \cdot C_3 \cdot \sum_{i=1}^{n} \beta_i^2\right]\left(\frac{1}{g}\right)\right\}$$

$$\sim Ga\left(\frac{n+1}{2}, \ \frac{n}{2}\left(1 + C_3 \cdot \sum_{i=1}^{n} \beta_i^2\right)\right) \quad \blacksquare$$

where $C_3 = \left[\sigma^2 (X'X)^{-1}\right]_{ii}^{-1}$

Recall:

$w_i = \left[\sigma^2 g (X'X)^{-1}\right]_{ii}$

$= g\left[\sigma^2 (X'X)^{-1}\right]_{ii}$

$w_i^{-1} = g^{-1}\left[\sigma^2 (X'X)^{-1}\right]_{ii}^{-1}$

$= \frac{1}{g} \cdot C_3$

Lin Mod M1 : 3A

MD

Show $E\left[(X_0\hat{\gamma} - X\beta)'(X_0\hat{\gamma} - X\beta)\right] = \sigma^2 tr(M_0) + \beta' X'(I - M_0) X\beta$,

where M_0 is p.p.m. onto $C(X_0)$ and $\hat{\gamma}$ is L.S. estimator in simple model.

Here: $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I_n) \Rightarrow Y \sim N(X\beta, \sigma^2 I_n)$, $E[Y] = X\beta$

$Cov(Y) = \sigma^2 I_n$

Since $\hat{\gamma}$ is LS, $X_0\hat{\gamma} = M_0 Y$. $E[X_0\hat{\gamma}] = E[M_0 Y] = M_0 \cdot E[Y] = M_0 X\beta$

The left side, above, can be written with terms swapped: $E\left[(X\beta - X_0\hat{\gamma})'(X\beta - X_0\hat{\gamma})\right] = ⊛$

Now call $z = X\beta - X_0\hat{\gamma}$, so that $⊛ = E[z'z]$.

Using the variance identity: $E[z'z] = cov(z) + E[z]^2$.

First, simplify $z = X\beta - X_0\hat{\gamma} = X\beta - M_0 Y$

Then find $E[z] = E[X\beta - M_0 Y] = X\beta - M_0 \cdot E[Y] = X\beta - M_0 X\beta = (I - M_0) X\beta$

and $cov(z) = cov(X\beta - M_0 Y) = M_0 \cdot cov(Y) \cdot M_0' = M_0 \cdot \sigma^2 I_n \cdot M_0' = \sigma^2 I_n \cdot M_0 M_0' = \sigma^2 I_n \cdot M_0$

idempotent

Therefore $E[z'z] = cov(z) + E[z]^2$

$= \sigma^2 I_n \cdot M_0 + \left[(I - M_0) X\beta\right]'\left[(I - M_0) X\beta\right]$

$= \sigma^2 \cdot M_0 + \beta' X'(I - M_0)'(I - M_0) X\beta$

$= \sigma^2 \cdot tr(M_0) + \beta' X'(I - M_0) X\beta$ ← idempotent

∎

Lin Mod M1 : 3B

MD

Standardized estimate of prediction error is $C_p = \dfrac{SSE(X_0)}{\hat{\sigma}_F^2} - (n - 2p_0)$

Q: If the simpler model is true and we replace $\hat{\sigma}_F^2$ by σ^2, what do you expect C_p should be?

Definitions: Know that $E[MSE] = \sigma^2$

$SSE[X_0]$ "residual sum of squares under simple model" $= (Y)'(M - M_0)(Y)$

$\hat{\sigma}_F^2$ "estimate of σ^2 under large, true model" $= E[MSE] = E\left[\dfrac{Y'(I-M)Y}{r(I-M)}\right]$

$$C_p = \frac{(Y)'(M-M_0)(Y)}{\underbrace{E\left[\frac{Y'(I-M)Y}{r(I-M)}\right]}_{=\sigma^2}} - (n - 2p_0)$$

$$= \frac{\sigma^2 tr(M_0) + (X\beta)'(M-M_0)(X\beta)}{\sigma^2} - (n - 2p_0)$$

Use result from (A) to replace numerator.

$$\boxed{C_p = tr(M_0) - (n - 2p_0)}$$

Here: Assuming denominator is true σ^2, and that simple model is true. The latter implies $(X\beta)'(M-M_0)(X\beta) = 0$

Lin Mod M1 : 3C

MD

If simpler model is true, is $\hat{\sigma}_F^2$ an unbiased estimate of σ^2?

Yes.

To be unbiased means $E[\hat{\sigma}_F^2] = \sigma^2$

Here, $\hat{\sigma}_F^2$ is the estimate of σ^2 under the full model; this is $E[MSE]$ under the full model. $E[MSE] = E\left[\frac{Y'(I-M)Y}{r(I-M)}\right] = \sigma^2$ by Theorem 1.3.2.

When the simpler model is true, $E[Y'(M-M_0)Y] = \sigma^2 \cdot r(M-M_0) + (X\beta)'(M-M_0)(X\beta)$

Such that $X\beta = X_0\gamma$ and

$$(M-M_0)(X_0\gamma) = 0$$
$$MX\beta - M_0X_0\gamma = 0$$
$$Y - Y = 0$$

And $E\left[\frac{Y'(M-M_0)Y}{r(M-M_0)}\right] = \sigma^2$, the same variance as in the full model,

where $\hat{\sigma}_F^2 = E[MSE] = \sigma^2$

MD

Lin Mod M1 : 3D

Is $E\left[\frac{SSE(X_0)}{\hat{\sigma}_F^2}\right] = E\left[SSE(X_0)\right] \cdot E\left[\frac{1}{\hat{\sigma}_F^2}\right]$?

Here, $SSE(X_0)$ and $\frac{1}{\hat{\sigma}_F^2}$ are independent, so expectation distributes over their product.

By Jensen's Inequality $E\left[\frac{1}{\hat{\sigma}_F^2}\right] > \frac{1}{E[\hat{\sigma}_F^2]}$ for convex function.

So two are not equal.

MD

Lin Mod M1 : 4A

Estimate β_1, β_2, and σ^2 for model $y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$, $e_i \sim N(0, \sigma^2)$, i.i.d.

Given x_{i1}, x_{i2}, y_i.

$$X = \begin{bmatrix} 10 & 15 \\ 9 & 14 \\ 9 & 13 \\ 11 & 15 \\ 11 & 14 \\ 10 & 14 \\ 10 & 16 \\ 12 & 13 \end{bmatrix} \qquad Y = \begin{bmatrix} 82 \\ 79 \\ 74 \\ 83 \\ 80 \\ 81 \\ 84 \\ 81 \end{bmatrix}$$

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = (X'X)^{-1}X'Y$$

$$\approx \boxed{\begin{pmatrix} 2.65 \\ 3.74 \end{pmatrix}}$$

$$\hat{\sigma}_2 = MSE = \frac{Y'(I-M)Y}{n-r} \quad \text{where } M = X(X'X)^{-1}X', \; n = 8, \; r = 2$$

$$= \frac{Y'Y - Y'MY}{n-r} = \frac{Y'Y - Y'(X(X'X)^{-1}X')Y}{n-r} \approx \boxed{4.70}$$

Lin Mod M1 : 4B

Give 95% C.I. for β_1 and $\beta_1 + \beta_2$

$$\frac{\lambda'\hat{\beta} - \lambda'\beta}{(MSE\,\lambda'(X'X)^{-}\lambda)^{\frac{1}{2}}} \sim t(df\,E)$$

For β_1, use $\lambda_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. For $\beta_1 + \beta_2$, use $\lambda_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$

Generally, the C.I. is defined as: $\lambda'\beta = \lambda'\hat{\beta} \pm (MSE\,\lambda'(X'X)^{-}\lambda)^{\frac{1}{2}} \cdot$ "two-sided, 95%, df=6, t-score"

Here, for β_1: $\beta_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}'\hat{\beta} \pm \left(MSE \begin{pmatrix} 1 \\ 0 \end{pmatrix}'(X'X)^{-1}\begin{pmatrix} 1 \\ 0 \end{pmatrix}\right)^{\frac{1}{2}} \cdot 2.45 \approx \boxed{[1.12, 4.17]}$

For $\beta_1 + \beta_2$: $\beta_1 + \beta_2 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}'\hat{\beta} \pm \left(MSE \begin{pmatrix} 1 \\ 1 \end{pmatrix}'(X'X)^{-1}\begin{pmatrix} 1 \\ 1 \end{pmatrix}\right)^{\frac{1}{2}} \cdot 2.45 \approx \boxed{[5.93, 6.84]}$

LinMod M1 : 4C

Perform $\alpha = 0.01$ test for $H_0 : \beta_2 = 3$

Use same approach as in (4B) :

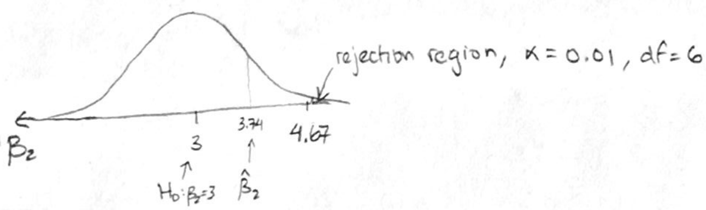
$$\frac{\lambda'\hat{\beta} - \lambda'\beta}{\left(MSE\, \lambda'(x'x)^{-}\lambda\right)^{\frac{1}{2}}} \sim t(df\,E)$$

where here, $\lambda = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$, $\lambda'\hat{\beta}$ is judged against $\lambda'\beta = 3$ and the resulting t-score is compared to the critical t-score for two-sided, 99%, df = 6 score. ($t_{critical} \approx 3.71$)

$$\frac{\begin{pmatrix} 0 \\ 1 \end{pmatrix}'\hat{\beta} - 3}{\left[4.70 \begin{pmatrix} 0 \\ 1 \end{pmatrix}'(x'x)^{-1}\begin{pmatrix} 0 \\ 1 \end{pmatrix}\right]^{\frac{1}{2}}} \approx 1.64$$

$1.64 < 3.71$

Not extreme enough to reject $H_0 : \beta_2 = 3$ at $\alpha = 0.01$.



Lin Mod M1 : 4D

Find p-value for test $H_0 : \beta_1 - \beta_2 = \emptyset$

Here, use above t-distribution formula with $\lambda = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$

$$\frac{\begin{pmatrix} 1 \\ -1 \end{pmatrix}'\hat{\beta} - \overbrace{\begin{pmatrix} 1 \\ -1 \end{pmatrix}\beta}^{\emptyset}}{\left[mse \cdot \begin{pmatrix} 1 \\ -1 \end{pmatrix}'(x'x)^{-1}\begin{pmatrix} 1 \\ -1 \end{pmatrix}\right]^{\frac{1}{2}}} \approx -1.02$$

In R: pt(q=-1.02, df=6) gives probability that a value is that far below the null or farther:

$P_{low} \approx 0.17$. Since the null is equality, the alternative is non-equality, i.e. a two-tailed test. The p-value is therefore $2 \times P_{low} \approx 0.34$.

M1.Q4 code

```
# Linear Models Midterm

library(MASS)

# QUESTION 4A

x <- matrix(c(10,9,9,11,11,10,10,12, 15,14,13,15,14,14,16,13), 8, 2)
y <- matrix(c(82,79,74,83,80,81,84,81), 8, 1)
bhat <- ginv(t(x)%*%x)%*%t(x)%*%y
yhat <- x%*%bhat

n <- 8
r <- 2
mse <- ( t(y)%*%y - (t(y)%*% x %*% ginv(t(x)%*%x) %*% t(x) %*% y) ) / (n-r)

# QUESTION 4B

L1 <- matrix(c(1,0), 2, 1)
L2 <- matrix(c(1,1), 2, 1)
t.score <- qt(p=0.975, df=n-r)
b1.conf <- c(t(L1)%*%bhat - sqrt(mse%*%t(L1)%*%ginv(t(x)%*%x)%*%L1) * t.score,
             t(L1)%*%bhat + sqrt(mse%*%t(L1)%*%ginv(t(x)%*%x)%*%L1) * t.score)
sum.conf <- c(t(L2)%*%bhat - sqrt(mse%*%t(L2)%*%ginv(t(x)%*%x)%*%L2) * t.score,
              t(L2)%*%bhat + sqrt(mse%*%t(L2)%*%ginv(t(x)%*%x)%*%L2) * t.score)

# QUESTION 4C

L3 <- matrix(c(0,1), 2, 1)
null <- 3
test.t.score3 <- (t(L3)%*%bhat - null) / sqrt(mse%*%t(L3)%*%ginv(t(x)%*%x)%*%L3)
critical.tscore <- qt(p=0.995, df=n-r)

# QUESTION 4D
L4 <- matrix(c(1,-1), 2, 1)
null <- 0
test.t.score4 <- (t(L4)%*%bhat - null) / sqrt(mse%*%t(L4)%*%ginv(t(x)%*%x)%*%L4)
```

MD

LinMod M1 : 5A

$$Y'MY = \sum_{i=1}^{p} y_{*i}^2$$

Find first p terms of Y_* : $y_{*1}, \dots, y_{*p}$

$Y = X\beta + \varepsilon$ Use: $X = ULV'$.

$Y = ULV'\beta + \varepsilon$ Multiply by $U_*' = \begin{bmatrix} U' \\ U'_{rest} \end{bmatrix}$.

$$U_* Y = U_* ULV'\beta + U_*'\varepsilon$$

$$Y_* = \begin{bmatrix} U' \\ U'_{rest} \end{bmatrix} ULV'\beta + \begin{bmatrix} U' \\ U'_{rest} \end{bmatrix} \varepsilon$$

$$\underset{n\times 1}{Y_*} = \begin{bmatrix} \underset{p\times n}{U'}\underset{n\times p}{U}\underset{p\times p}{L}\underset{p\times p}{V'}\underset{p\times 1}{\beta} + \underset{p\times n}{U'}\underset{n\times 1}{\varepsilon} \\ U'_{rest}ULV'\beta + U'_{rest}\varepsilon \end{bmatrix} = \begin{bmatrix} \underset{p\times 1}{\square} \\ \underset{n-p\times 1}{\square} \end{bmatrix}$$

Here, $U'\varepsilon = 0$ since $U' \in C(X)$, $\varepsilon \in C(\varepsilon)$ and $C(X) \perp C(\varepsilon)$.

First p elements of Y_* : $U'ULV'\beta = Y_{*p}$

Sum of first p elements squared : $(Y_{*p})'(Y_{*p})$

$$= [U'ULV'\beta]'[U'ULV'\beta]$$

$$= \beta'VL'U'U\underbrace{UU'}_{I}LV'\beta$$

Note: $U'U = I$, since columns of U are orthonormal,

$$= \beta'VL'U'ULV'\beta$$

$$= (ULV'\beta)'(ULV'\beta)$$

Recall : $X = ULV'$

$$= (X\beta)'(X\beta)$$

$$= (MY)'(MY)$$

$$= Y'M'MY$$

Recall: $M = M'M$ idempotence

$$= Y'MY \quad \blacksquare$$

LinMod M1 : 5B

MD

$$Y'(1-M)Y = \sum_{i=p+1}^{n} y_*^2$$

This is the sum of squares of last $(n-p)$ values of Y_*

As before :

$$\underset{n\times 1}{Y_*} = \begin{bmatrix} U'ULV'\beta + U'\varepsilon \\ U_{rest}'ULV'\beta + U_{rest}'\varepsilon \end{bmatrix} = \begin{bmatrix} \underset{p\times 1}{\square} \\ \underset{n-p\times 1}{\square} \end{bmatrix}$$

Last $(n-p)$ elements of Y_* : $U_{rest}'ULV'\beta + U_{rest}'\varepsilon = Y_{*rest}$

$$\sum_{i=p+1}^{n} y_*^2 = \left[Y_{*rest}\right]'\left[Y_{*rest}\right]$$

$$= \left[U_{rest}'ULV'\beta + U_{rest}'\varepsilon\right]'\left[U_{rest}'ULV'\beta + U_{rest}'\varepsilon\right]$$

$$= \left[U_{rest}'X\beta + U_{rest}'\varepsilon\right]'\left[U_{rest}'X\beta + U_{rest}'\varepsilon\right]$$

$$= \left[U_{rest}'\varepsilon\right]'\left[U_{rest}'\varepsilon\right]$$

$$= \varepsilon' U_{rest} U_{rest}' \varepsilon$$

$$= \varepsilon'\varepsilon$$

$$= \left[(1-M)Y\right]'\left[(1-M)Y\right]$$

$$= Y'(1-M)'(1-M)Y$$

$$= Y'(1-M)Y \quad \blacksquare$$

$X = ULV'$

$\left.\begin{array}{l} U' \in C(X), \\ U_{rest}' \in C(X)^{\perp} \\ X\beta \in C(X) \end{array}\right\} \circledast$

$\circledast$ implies $U_{rest}' \perp X\beta$

and $U_{rest}' X\beta = 0$

$\varepsilon \in C(X)^{\perp}$

$U_{rest}' \in C(X)^{\perp}$

$(1-M)$ is p.p.m onto $C(X)^{\perp}$

$\varepsilon = (1-M)Y$

U_{rest} are orthonormal,

So $U_{rest}U_{rest}' = I$

MD

Lin Mod M1 : 5c

$$\hat{\beta}'\hat{\beta} = \sum_{i=1}^{p} \frac{y_{*i}^2}{\lambda_i^2}$$

Recall: First p elements of $Y_* = U'ULV'\beta = Y_{*p}$

Recall: $\sum_{i=1}^{p} y_{*i}^2 = [U'ULV'\beta]'[U'ULV'\beta] = Y_{*p}'Y_{*p}$

$$= \begin{bmatrix} y_{*1} & \cdots & y_{*p} \end{bmatrix} \begin{bmatrix} y_{*1} \\ \vdots \\ y_{*p} \end{bmatrix}$$

Want same method but include $\frac{1}{\lambda_i}$ in each element.

Try $\begin{bmatrix} y_{*1} & \cdots & y_{*p} \end{bmatrix} \begin{bmatrix} \frac{1}{\lambda_1} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\lambda_p} \end{bmatrix} = \begin{bmatrix} \frac{y_{*1}}{\lambda_1} & \cdots & \frac{y_{*p}}{\lambda_p} \end{bmatrix}$

$$\underset{1 \times p}{Y_{*p}'} \cdot \underset{p \times p}{L^{-1}} = \underset{1 \times p}{Y_{*p}'L^{-1}}$$

Now we have $\frac{1}{\lambda_i}$ in each element:

$$\sum_{i=1}^{p} \frac{y_{*i}^2}{\lambda_i^2} = \begin{bmatrix} \frac{y_{*1}}{\lambda_1} & \cdots & \frac{y_{*p}}{\lambda_p} \end{bmatrix} \begin{bmatrix} \frac{y_{*1}}{\lambda_1} \\ \vdots \\ \frac{y_{*p}}{\lambda_p} \end{bmatrix}$$

$$= \left[Y_{*p}'L^{-1}\right]\left[Y_{*p}'L^{-1}\right]'$$

$$= \left[(U'ULV'\beta)'L^{-1}\right]\left[(U'ULV'\beta)'L^{-1}\right]'$$

$$= \beta'VL'\underbrace{U'U}_{I}L^{-1}(L^{-1})'\underbrace{U'U}_{I}LV'\beta$$

$$= \beta'VLL^{-1}L^{-1}LV'\beta$$

$$= \beta'VV'\beta$$

$$= \beta'\beta \quad \blacksquare$$

$Y_{*p} = U'ULV'\beta$

$U'U = I$ orthonormal columns of U

$L' = L$, $L^{-1} = (L^{-1})'$

since each is square, diagonal

$V'V = I$ orthonormal columns of V

$\Rightarrow V' = V^{-1}$, so $VV' = VV^{-1} = I$

Lin Mod M1 : 6A   Show $\tilde{\beta} = V\tilde{\gamma}$.   MD

Given model as in (5), let $\tilde{\beta} = (X'X + kI)^{-1}X'Y$, $\tilde{\gamma}_j = \frac{\lambda_j}{\lambda_j^2 + k} \cdot y_{*j}$ be canonical ridge regression estimates.

Recall: $X = ULV'$

$X' = VLU'$

Note: $U_r' \perp U$ by definition.

$V'V = U'U = I$ since cols of V, U form orthonormal basis

Let $\gamma = V'\beta$

Y_* is estimated by $\begin{bmatrix} L \\ \emptyset \end{bmatrix}\hat{\gamma}$

$L'L = L^2$, since L is diagonal.

$(L'L + kI)$ is square, and diag, so inverse can be expressed element-wise

$(L^2 + kI)^{-1}$ and L^2 can commute, since each is diag. of same dimension

$X'XV = VL^2$ by definition

$X'X = VLU' \cdot ULV' = VL^2V'$

$(X'X)^{-1} = (VL^2V')^{-1}$

$V' = V^{-1}$, $(VV')^{-1} = VV'$

Setup: $Y = X\beta + \varepsilon$, $U_* = [U, U_r]$, $Y_* = U_*'Y = \begin{bmatrix} U' \\ U_r' \end{bmatrix} Y$, $\varepsilon_* = U_*'\varepsilon$

$$Y = ULV'\beta + \varepsilon$$

$$Y_* = U_*'(ULV'\beta) + \varepsilon_*$$

$$= \begin{bmatrix} U' \\ U_r' \end{bmatrix} ULV'\beta + \varepsilon_* = \begin{bmatrix} U'UL \\ \emptyset \end{bmatrix} V'\beta + \varepsilon_*$$

$$Y_* = \begin{bmatrix} L \\ \emptyset \end{bmatrix} \gamma + \varepsilon_*$$

This is the canonical regression model of Y_* with parameter γ, and design matrix $\begin{bmatrix} L \\ \emptyset \end{bmatrix}$.

Use the ridge regression estimate for the regression parameter γ:

$$\tilde{\gamma} = \left([L', \emptyset] \begin{bmatrix} L \\ \emptyset \end{bmatrix} + kI \right)^{-1} [L', \emptyset] Y_*$$

$$= (L'L + kI)^{-1} [L', \emptyset] \begin{bmatrix} L \\ \emptyset \end{bmatrix} \hat{\gamma}$$

$$= (L'L + kI)^{-1}(L'L)\hat{\gamma} = (L^2 + kI)^{-1} L^2 \hat{\gamma}.$$

$$V\tilde{\gamma} = V\left[(L^2 + kI)^{-1} L^2 \hat{\gamma} \right]$$

$$= V(L^2 + kI)^{-1} L^2 \cdot L^{-1}U'Y$$

$$= V(L^2 + kI)^{-1} LU'Y$$

$$= V(L^2 + kI)^{-1} (V'V) LU'Y$$

$$= V(L^2 + kI)^{-1} V'X'Y$$

$$= (VL^2V' + kI)^{-1} X'Y$$

$$= (X'X + kI)^{-1} X'Y = \tilde{\beta} \quad \blacksquare$$

From above: $Y_* = \begin{bmatrix} L \\ \emptyset \end{bmatrix}\gamma + \varepsilon_*$

$\Rightarrow \hat{\gamma} = [L^{-1}, 0] Y_* = [L^{-1}, 0] \begin{bmatrix} U' \\ U_r' \end{bmatrix} Y$

$\hat{\gamma} = L^{-1}U'Y$

$\Rightarrow \hat{\gamma}_j = \frac{u_j'Y}{\lambda_j} = \frac{y_{*j}}{\lambda_j}$

Lin Mod M1: 6B

Complete Linear Model to a Bayesian Inference Model

MD

Bayes: $P(Y|\beta) = N(X\beta, \sigma^2 I_p)$, $p(\beta) = N(m_0, \Sigma_0)$

Note: σ^2 is fixed.

Goal: Find m_0, Σ_0 such that $E[\beta|Y] = \tilde{\beta}$

① $Y = X\beta + \varepsilon$, $\varepsilon \sim N(0, \sigma^2 I_p)$ and want to add a prior on β. Try $m_0 = 0$, as an "uninformed" guess of the value (rather than expecting $\beta > 0$ or $\beta < 0$). So $\beta \sim N(0, \Sigma_0)$.

② Just as ① alone implies $Y \sim N(X\beta, \sigma^2 I_p)$; $0 = I\beta + \tilde{\varepsilon}$, $\tilde{\varepsilon} \sim N(\phi, \Sigma_0)$ implies $\beta \sim N(\phi, \Sigma_0)$.

$$\begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X \\ I \end{bmatrix} \beta + \begin{bmatrix} \varepsilon \\ \tilde{\varepsilon} \end{bmatrix} \text{, where } \begin{bmatrix} \varepsilon \\ \tilde{\varepsilon} \end{bmatrix} \sim N\left(\begin{bmatrix} \phi_{p\times 1} \\ \phi_{n-p\times 1} \end{bmatrix}, \sigma^2 \begin{bmatrix} I_p & 0 \\ 0 & \Sigma_0 \end{bmatrix} \right)$$

Above is L.S. model whose estimates: $P(\beta|Y) \propto P(Y|\beta) \cdot P(\beta)$ characterize the complete Bayesian model.

Setting $m_0 = 0$ and keeping Σ_0 such that variance is large puts low weight on the prior and more weight on the L.S. estimate $\tilde{\beta}$: e.g. $a I_{n-p}$, $a >> 1$.

The prior would thus be $\beta \sim N(\phi, (a I_{n-p})\sigma^2)$ for large a, and $E[\beta|Y] = \tilde{\beta}$.

$$\boxed{\begin{aligned} m_0 &= 0 \\ \Sigma_0 &= a I_{n-p} \sigma^2 , \quad a >> 1 \end{aligned}}$$