1. $\mathbb{E}(\hat{\beta}_\tau) = S_\tau^{-1} S \beta^*$, $\operatorname{cov}(\hat{\beta}_\tau) = S_\tau^{-1} S S_\tau^{-1} \sigma^2$

proof:

$$\hat{\beta}_\tau = \underset{\beta}{\operatorname{argmin}} (Y - X\beta)^T (Y - X\beta) + \tau \beta^T \beta .$$

The solution is $\hat{\beta}_\tau = (X^T X + \tau \mathbb{1})^{-1} X^T Y$

Now, by hypothesis $Y = X\beta + \varepsilon$, where $\varepsilon \sim N(\underline{0}, \mathbb{1}\sigma^2)$
so Y is just a shifted normal vector $Y \sim N(X\beta, \mathbb{1} \cdot \sigma^2)$

$\hat{\beta}_\tau$ is again a linear combination of Y, let's say $\hat{\beta}_\tau = A \cdot Y$
where $A = (X^T X + \tau \mathbb{1})^{-1} X^T$
So, since Y it's a Gaussian vector, $\hat{\beta}_\tau \sim N(AX\beta, A \mathbb{1} \sigma^2 A^T)$

Thus, $\mathbb{E}(\hat{\beta}_\tau) = (X^T X + \tau \mathbb{1})^{-1} X^T X \beta = S_\tau^{-1} S \beta$

$$\operatorname{cov}(\hat{\beta}_\tau) = \left[(X^T X + \tau \mathbb{1})^{-1} X^T \right] \sigma^2 \cdot \mathbb{1} \left[(X^T X + \tau \mathbb{1})^{-1} X^T \right]^T =$$

$$= \{ \mathbb{1} X = X \} = \sigma^2 \left[(X^T X + \tau \mathbb{1})^{-1} X^T X \underbrace{(X^T X + \tau \mathbb{1})^{-T}}_{\text{symmetric } -T = -1} \right]$$

$$= \sigma^2 \left[(X^T X + \tau \mathbb{1})^{-1} X^T X (X^T X + \tau \mathbb{1})^{-1} \right] =$$

$$= \sigma^2 S_\tau^{-1} S S_\tau^{-1}$$

$\square$

2. Hp: • $\frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y_i^* - X_i \beta)^2 = 0$

• $\mu_1 = \frac{1}{N_1} \sum_{i: y_i^* = 1} X_i \qquad \mu_{-1} = \frac{1}{N_{-1}} \sum_{i: y^* = -1} X_i$

• $\Sigma = \frac{1}{N} \left[\sum_{i: y_i^* = 1} (X_i - \mu_1)^T (X_i - \mu_1) + \sum_{i: y_i^* = 1} (X_i - \mu_{-1})^T (X_i - \mu_{-1}) \right]$

• $\left.\begin{array}{l} N_1 = N_{-1} = N/2 \ , \ \sum_{i=1}^{N} X_i = 0 , \ \mu_1 + \mu_{-1} = 0. \end{array}\right\}$ balanced class

Th: $\Sigma \beta + \frac{1}{4} (\mu_1 - \mu_{-1})^T (\mu_1 - \mu_{-1}) \beta = \frac{1}{2} (\mu_1 - \mu_{-1})^T$

proof:

let's call $X_1 := \{ X_i : y_i^* = 1 \}$ $\quad X_{-1} := \{ X_i : y_i^* = -1 \}$

$Y_1 := \{ y_i^* = 1 \}$ $\quad Y_{-1} = \{ y_i^* = -1 \}$

$$\frac{\partial}{\partial \beta} \sum_{i=1}^{N} (y_i^* - X_i \beta)^2 = 0$$

$$\Updownarrow$$

$$\frac{\partial}{\partial \beta} \left(\sum_{i: y_i^* = 1} (y_i^* - X_i \beta)^2 + \sum_{i: y_i^* = -1} (y_i^* - X_i \beta)^2 \right) = 0$$

$$\Updownarrow$$

$$\frac{\partial}{\partial \beta} \left[(Y_1 - X_1 \beta)^T (Y_1 - X_1 \beta)^T + (Y_{-1} - X_{-1} \beta)^T (Y_{-1} - X_{-1} \beta) \right] = 0$$

$$-2(Y_1 - X_1\beta)X_1^T - 2(Y_{-1} - X_{-1}\beta)X_{-1}^T = 0$$

$$X_1^T Y_1 - X_1^T X_1 \beta + X_{-1}^T Y_{-1} - X_{-1}^T X_{-1}\beta = 0$$

$$(X_1^T X_1 + X_{-1}^T X_{-1})\beta = X_1^T Y_1 + X_{-1}^T Y_{-1}$$

Now: $X_1^T Y_1 = \sum_{i:y_i^*=1} X_i = \frac{N}{2}\mu_1^T$

$$X_{-1}^T Y_{-1} = \sum_{i:y_i^*=-1} -X_i = -\frac{N}{2}\mu_{-1}^T$$

$$\rightarrow (X_1^T X_1 + X_{-1}^T X_{-1})\beta = \frac{N}{2}(\mu_1 - \mu_{-1})^T$$

Moreover, we want to write the first term better:

$$X_1^T X_1 + X_{-1}^T X_{-1} = \sum_{i:y_i^*=1} X_i^T X_i + \sum_{i:y_i^*=-1} X_i^T X_i =$$

$$= \sum_{i:y_i^*=1} (X_i - \mu_1 + \mu_1)^T (X_i - \mu_1 + \mu_1) + \sum_{i:y_i^*=-1} (X_i - \mu_{-1} + \mu_{-1})^T (X_i - \mu_{-1} + \mu_{-1})$$

$$= \sum_{i:y_i^*=1} (X_i - \mu_1)^T (X_i - \mu_1) + \mu_1^T \mu_1 + 2\sum_{i:y_i^*=1} (X_i - \mu_1)^T \mu_1 +$$

$\Sigma =$

$$\sum_{i:y_i^*=-1} (X_i - \mu_{-1})^T (X_i - \mu_{-1}) + \mu_{-1}^T \mu_{-1} + 2\sum_{i:y_i^*=1} (X_i - \mu_{-1})^T \mu_{-1} =$$

$$= N\Sigma + \frac{N}{2}(\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}) + 0$$

since $\sum_{i:y_i^*=1} (X_i - \mu_1)^T \mu_1 = \mu_1 \sum_{i:y_i^*=1} (X_i - \mu_1)^T = \mu_1 \left[\frac{N}{2}\mu_1^T - \frac{N}{2}\mu_1^T\right] = 0$

same for $y_i^* = -1$.

$$\rightarrow \quad N \Sigma \beta + \frac{N}{2}\left(\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}\right)\beta = \frac{N}{2}\left(\mu_1 - \mu_{-1}\right)^T$$

Now $\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1} = \frac{1}{2}\left(\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}\right) + \frac{1}{2}\left(\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}\right) =$

$$= \left\{\mu_1 = -\mu_{-1}\right\} = \frac{1}{2}\left\{\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}\right\} + \frac{1}{2}\left\{-\mu_1^T \mu_{-1} - \mu_1^T \mu_{-1}\right\} =$$

$$= \frac{1}{2}\left\{\mu_1^T \mu_1 + \mu_{-1}^T \mu_{-1}\right\} - \mu_1^T \mu_{-1} = \frac{1}{2}\left(\mu_1 - \mu_{-1}\right)^T\left(\mu_1 - \mu_{-1}\right)$$

So we have:

$$\cancel{N}\Sigma \cdot \beta + \frac{\cancel{N}}{4}\left(\mu_1 - \mu_{-1}\right)^T\left(\mu_1 - \mu_{-1}\right)\beta = \frac{\cancel{N}}{2}\left(\mu_1 - \mu_{-1}\right)^T$$

$$\rightarrow \quad \Sigma \beta + \frac{1}{4}\left(\mu_1 - \mu_{-1}\right)^T\left(\mu_1 - \mu_{-1}\right)\beta = \frac{1}{2}\left(\mu_1 - \mu_{-1}\right)^T$$

$\square$