1. Montrons que OLS est l'estimateur avec la plus faible variance.

- $\mathbb{E}(\tilde{\beta}) = \mathbb{E}[Cy] = Cx\beta + C\mathbb{E}(\varepsilon)$

$$= Hx\beta + Dx\beta$$

or β non biaisé donc $Dx\beta = 0 \Rightarrow Dx = 0$

- $Var(\tilde{\beta}) = Var(Cy) = \sigma^2 CC^T$

$$= \sigma^2 (H+D)(H^T + D^T)$$

$$= \sigma^2 (HH^T + DD^T + \underbrace{HD^T}_{=0} + \underbrace{DH^T}_{=0})$$

→ d'après $Dx = 0$

et $Var(\beta^*) = \sigma^2 (X^TX)^{-1} X^TX (X^TX)^{-1}$

$$= \sigma^2 HH^T$$

ainsi $\boxed{Var(\tilde{\beta}) = Var(\beta^*) + \sigma^2 DD^T}$

d'où le résultat.

$\boxed{\text{OLS}}$

Ridge Regression

- Ridge estimateur est biaisé ?

$$\mathbb{E}(\beta^*_{ridge}) = \mathbb{E}\left[(x_c^T x_c + \lambda I)^{-1} x_c^T y_c\right]$$

$$= \left[(x_c^T x_c + \lambda I)^{-1} x_c^T\right] \underbrace{\mathbb{E}(y_c)}_{= \beta}$$

$$\neq \beta \text{ sauf si } \lambda = 0$$

Oui il est biaisé.

- SVD décomposition :

$$\beta^*_{ridge} = (x_c^T x_c + \lambda I)^{-1} x_c^T y_c$$

$$= (V D^T D V^T + \lambda V V^T)^{-1} x_c^T y_c$$

V orthogonale

$$= V \operatorname{diag}(\mu_i^2 + \lambda)^{-1} V^T x_c^T y_c$$

$$= V \operatorname{diag}\left(\frac{1}{\mu_i^2 + \lambda}\right) V^T x_c^T y_c$$

- Mq $Var(\beta^*_{OLS}) \geq Var(\beta^*_{ridge})$

à partir de l'expression ci dessus on a

$$Var(\beta^*_{ridge}) = \sigma^2 V \operatorname{diag}\left(\frac{\mu_i^2}{(\mu_i^2 + \lambda)^2}\right) V^T$$

or $$Var(\beta^*_{OLS}) = \sigma^2 V \operatorname{diag}\left(\frac{1}{\mu_i^2}\right) V^T$$

ainsi $Var(\beta^*_{OLS}) - Var(\beta^*_{ridge}) = \sigma^2 V diag\left(\frac{1}{\mu_i^2} - \frac{\mu_i^2}{(\mu_i^2+\lambda)^2}\right)V^T$

$\geqslant 0$ ■

- Si $\lambda \nearrow$ alors $Var(\beta^*_{ridge}) \searrow$ et inversement.

- $\beta^*_{ridge} = ((1+\lambda)I_d)^{-1} x_c^T y_c = \frac{\beta^*_{OLS}}{1+\lambda}$