## Statistiques

- modèle statistique : famille de lois $\mathcal{M} = \{\mathbb{P}_\theta : \theta \in \Theta\}$
- Biais $(\hat{g}, g) = E_\theta(\hat{g}(Y_1 \dots Y_n)) - g(\theta)$
- $Var(Y) = E(Y^2) - [E(Y)]^2$
- risque quadratique : $R(\hat{g}) = \mathbb{E}[(\hat{g} - g)^2]$ $= Var(\hat{g}) + \theta^2$ (biais)
- $V = \langle u, y \rangle u$, $u$ unitaire
- $Cov(X) = E[(X - E(X))(X - E(X))^T]$
- matrice positive : $\forall u, u^T X u \geq 0$
- décomposition spectrale : $S = U \Lambda U^T = \sum_{i=1}^{p} \lambda_i u_i u_i^T$
$|S| = |U \Lambda U^T| = |\Lambda| = \prod_{i=1}^{p} \lambda_i$

- Théorème de projection : soit $\mathcal{F} \subset \mathbb{R}^p$, $y \in \mathbb{R}^p$, et $z^* : \inf_{z \in \mathcal{F}} \|y - z\|^2$ le point $z$ est atteint de façon unique ; il est caractérisé par les équations normales : $\langle y - z, f \rangle = 0 \; \forall f \in \mathcal{F}$
- MCO : $y = Y \in \mathbb{R}^n$, $\mathcal{F} = vect(X)$, $z^* = X\hat{\theta}_n$, $\langle Y, X\hat{\theta}_n, f \rangle = 0 \; \forall f \in \mathcal{F}$
- on cherche $\min_{\theta \in \mathbb{R}^p} \|Y - X\theta\|_2^2$
$(Y - X\hat{\theta}_n)^T X = 0 \Leftrightarrow X^T(Y - X\hat{\theta}_n) = 0$
$\Leftrightarrow (X^T X)\hat{\theta}_n = X^T Y$
- biais $(\hat{\theta}_n) = E(\hat{\theta}_n) - \theta^* = 0$
$Var(\hat{\theta}_n) = \sigma^2 (X^T X)^{-1} = \sigma^2/n \; \hat{G}^{-1}$
propriété : $\hat{\theta}_n - \theta^* = (X^T X)^{-1} X^T \varepsilon = A_n \varepsilon$
- $R_{quad}(\hat{\theta}_n) = \mathbb{E}[\|\hat{\theta}_n - \theta^*\|_2^2]$
$= tr(Cov(\hat{\theta}_n)) + \|Biais(\hat{\theta}_n)\|_2^2$
$= \sigma^2 tr((X^T X)^{-1})$
- NB : $\|u\|_2^2 = \sum_{k}^{p} u_k^2 = tr(u u^T)$
$tr(X) = \sum_{k=1}^{p} \lambda_k$
- $f(x)$ est convexe si sa Hessienne $\frac{\partial^2 f}{\partial x_i x_j}$ est positive
- $M$ est positive $\Leftrightarrow$ ses valeurs propres $\lambda_i$ sont $\geq 0$
- $M$ concave $\Leftrightarrow$ Hessienne négative i.e $x M x \leq 0 \; \forall x$
- OLS normale : $\min \|Y - X\theta\|_2^2$
centré : $\min \|Y_c - X_c \theta\|_2^2$
$= \min \|Y - (\tilde{X}_c \tilde{\theta} + \mathbb{1}\bar{Y})\|_2^2$
- risque prédictif : $R_{pred}(\hat{\theta}_n) = \mathbb{E}[\|Y^* - \hat{Y}\|_2^2]$
$= \mathbb{E}[\|X(\theta^* - \hat{\theta})\|_2^2]$
- biais variance résidus :
$\mathbb{E}(\sum \hat{\varepsilon}_i^2) = \mathbb{E}[\|Y - \hat{Y}\|_2^2] = \mathbb{E}[\|(I - H_n)Y\|_2^2]$
$= \mathbb{E}[\|(I - H_n)\varepsilon\|_2^2] = \sigma^2 tr(I - H_n) = \sigma^2(n - p - 1)$
- sous-gaussien : si $\mathbb{E}(e^{\lambda\varepsilon}) \leq exp(\lambda^2 \sigma^2 / 2)$
- $Z_1 \dots Z_n$ gaussiennes $N(0,1)$ ; alors $\sum z_i^2 \sim \chi^2(n)$
- $Z \sim N(0,1)$, $Y \sim \chi^2(n)$, $Z \perp\!\!\!\perp Y$ : $X = Z / \sqrt{Y/n} \sim t_n$
- lemme de Cochran : si $\varepsilon \sim N(0, \sigma^2)$
i) $\hat{\theta}_{n k} \perp\!\!\!\perp \hat{\sigma}_n^2$
ii) $\sqrt{n}(\hat{\theta}_n - \theta^*) \sim N(0, n\sigma^2 (X^T X)^{-1})$
iii) $\hat{\sigma}_n^2 \frac{(n-p-1)}{\sigma^2} \sim \chi^2(n-p-1)$
iv) $n^{1/2} / (\hat{s}_{nk} \hat{\sigma}_n)(\hat{\theta}_{nk} - \theta_k^*) \sim t_{n-p-1}$
$\hat{s}_{nk} = e_k^T \hat{G}_n^{-1} e_k$

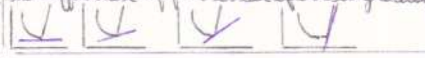
## Ridge

- soit $M$ symétrique ; alors $M = UDU^T$ avec : $D = diag(\lambda_1 \dots \lambda_p)$ et $U \in \mathbb{R}^{p \times p}$, $UU^T = I_p$
- OLS : $\hat{G} = X^T X / n$ ; $\hat{\theta}_n$ unique si $\lambda_1 \leq \dots \leq \lambda_n > 0$
si $\lambda_{p-k} = \dots = \lambda_p = 0$, sélectionne $u_1 \dots u_k$ tels que : $\lambda_1 \dots \lambda_k > 0$ et on transforme $\tilde{X} = X(u_1 \dots u_k)$
$X \in \mathbb{R}^{n \times p} \rightarrow \tilde{X} \in \mathbb{R}^{n \times k}$
- approche équivalente : $\hat{\theta}_n = (X^T X)^+ X^T Y$ avec $(X^T X)^+$ inverse généralisé de $(X^T X)$
- note : $(X^T X) = UDU^T = U \begin{pmatrix} \lambda_1 & & \\ & \lambda_k & \\ & & 0 \end{pmatrix}$
$= \sum_{k=1}^{k} \lambda_k u_k u_k^T$, implique : $(X^T X)^+ = \sum_{k=1}^{k} \lambda_k^{-1} u_k u_k^T$

## Modèle linéaire

- Singular value decomposition : $X = V \Sigma U^T$
$n \left( x_1 \dots x_p \right) = \begin{pmatrix} v_1 & \dots & v_r \end{pmatrix} \begin{pmatrix} s_1 & & \\ & \ddots & \\ & & s_r \end{pmatrix} \begin{pmatrix} - u_1^T - \\ \vdots \\ - u_r^T - \end{pmatrix}$
$r = rang(X)$, $V^T V = I_r$, $U^T U = I_r$
- SVD pour MCO : $\hat{\theta} = (X^T X)^+ X^T Y$
$= [(U \Sigma V^T)(V \Sigma U^T)]^+ (U \Sigma V^T) Y = \dots$
$= X^+ Y = \sum_{k=1}^{r} s_k^{-1} u_k v_k^T$ (...) par $S_k^{-1}$
- en général : $M^+ = U D^+ U = \sum_{k=1}^{r} \lambda_k^{-1} u_k u_k^T$
- PCA before OLS : $\hat{\Sigma}_n = \sum_{i=1}^{n} (X_i - \bar{X})(X_i - \bar{X})^T / n$
on conserve $l$ directions (valeurs propres $> 0$) et fait OLS sur $Y_c$ et $X_c (u_1 \dots u_l)$ - alors :
$\hat{\Sigma}_n = [X_c (u_1 \dots u_l)]^T [X_c (u_1 \dots u_l)] / n$
$= \begin{pmatrix} u_1^T \\ \vdots \\ u_l^T \end{pmatrix} \underbrace{X_c^T X_c}_{UDU^T} (u_1 \dots u_l) = D$
- Ridge : $\hat{\theta}_n^{Ridge} \in \underset{\theta \in \mathbb{R}^p}{argmin} \{\|Y - X\theta\|_2^2 + \lambda \|\theta\|_2^2\}$
CPO : $\nabla_\theta \{\|Y - X\theta\|_2^2 + \lambda\|\theta\|_2^2\} = 0 \Rightarrow$
$[(X^T X) + \lambda I]\theta = X^T Y \Leftrightarrow \hat{\theta}_n^{Ridge} = (X^T X + \lambda I)^{-1} X^T Y$
- Ridge est biaisé : $\mathbb{E}[\hat{\theta}_n^{Ridge}] = \mathbb{E}[(X^T X + \lambda I)^{-1} X^T Y]$
$= \mathbb{E}[(X^T X + \lambda I)^{-1} X^T (X\theta^* + \varepsilon)] = (X^T X + \lambda I)^{-1} (X^T X)\theta^*$
$= (X^T X + \lambda I)^{-1} (X^T X + \lambda I - \lambda I)\theta^* = \theta^* - \lambda (X^T X + \lambda I)\theta^*$
- Variance de Ridge : $\sigma^2 (X^T X + \lambda I)^{-1} (X^T X)(X^T X + \lambda I)^{-1}$
- $cov$ Ridge $<$ $cov$ OLS : preuve par ordre matriciel : montre que $u^T(cov OLS - cov R)u > 0$
$u^T[(X^T X)^{-1}\sigma^2 - \sigma^2 (X^T X + \lambda I)^{-1} (X^T X)(X^T X + \lambda I)^{-1}] u$
$= \sigma^2 \left[ \begin{pmatrix} \lambda_1^{-1} & & \\ & \ddots & \\ & & \lambda_p^{-1} \end{pmatrix} - \begin{pmatrix} (\lambda_1 + \lambda)^{-1} & & \\ & \ddots & \\ & & (\lambda_p + \lambda)^{-1} \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \begin{pmatrix} (\lambda_1 + \lambda)^{-1} & & \\ & \ddots & \\ & & (\lambda_p + \lambda)^{-1} \end{pmatrix} \right]$
$= \sigma^2 \begin{pmatrix} \lambda_1^{-1} - \frac{\lambda_1}{(\lambda_1 + \lambda)^2} & & \\ & \ddots & \\ & & \lambda_p^{-1} - \frac{\lambda_p}{(\lambda_p + \lambda)^2} \end{pmatrix}$
donc $U$ : vecteurs propres de ($cov OLS - cov R$) et les valeurs propres sont $\lambda_i^{-1} - \lambda_i / (\lambda_i + \lambda)^2 > 0$ si $\lambda_i^{-1} \geq \lambda_i / (\lambda_i + \lambda)^2 \Leftrightarrow (\lambda_i + \lambda)^2 > \lambda_i^2$ : toujours vrai
- or : ordre matriciel : si $M$ symétrique $\in \mathbb{R}^{p \times p}$, $M \succeq 0$ (semi positive) si $u^T M u \geq 0 \; \forall u$, ou si toutes les valeurs propres de $M$ sont positives

*normes*
- $\|u\|_2^2 = \sqrt{\sum_{k=1}^{n} u_k^2}^{\,2} = \sum_{l=1}^{n} u_l^2 = u^T u = tr(u u^T)$
- $\|u\|_1 = \sum_{k=1}^{n} |\theta_k|$
- let $f : \mathbb{R}^p \rightarrow \mathbb{R}$ be convex. subgradient of $f$ at $x$ : any $u \in \mathbb{R}^p$ with : $\forall y \in \mathbb{R}^p$, $f(y) - f(x) > u^T(y - x)$
subdifferential of $f$ at $x$ : set of all subgradients



## Lasso

$\hat{\theta}_{LASSO} \in \underset{\theta \in \mathbb{R}^p}{argmin} \left\{ \frac{1}{2}\|Y - X\theta\|_2^2 + \lambda \|\theta\|_1 \right\}$

- standardize $X$ by using $XD^{-1/2}$ were $D = diag(X^T X)$ + recentrage
- active set : $S^* \subset \{1 \dots p\}$ tel que $S^* = \{j = 1 \dots p : \theta_j^* \neq 0\}$
- propriété : $\|\hat{\theta}_{LASSO} - \theta^*\|_2 \leq \frac{6}{\gamma} \sqrt{\frac{2\sigma^2 S \log(2p)}{n}}$ avec probabilité $1 - \delta$
- calcul : utilise : "si $f$ est convexe, $x^*$ est un minimum si et seulement si $0 \in \partial f(x^*)$, où $\partial f$ est le sous-différentiel de $f$.
- détail du calcul : $\hat{\theta}_{LASSO} \in \underset{\theta \in \mathbb{R}^p}{argmin} \{\|Y - X\theta\|_2^2 + 2\lambda\|\theta\|_1\}$

on a : $\|\theta\|_1 = \sum_{k=1}^{p} |\theta_k| = |\theta_k| + \sum_{j=1}^{j \neq k} |\theta_j|$
également : $X\theta = X^{(1)}\theta_1 + X^{(2)}\theta_2 + \dots + X^{(p)}\theta_p = Z_k + X^{(k)}\theta_k$
donc on définissant $\varepsilon_k = Y - Z_k \Rightarrow \|Y - X\theta\|_2^2 = \|\varepsilon_k - X^{(k)}\theta_k\|_2^2$
$= \|\varepsilon_k\|_2^2 + \|X^{(k)}\theta_k\|_2^2 - 2\langle \varepsilon_k, X^{(k)}\theta_k \rangle = \|\varepsilon_k\|^2 + \theta_k^2 \|X^{(k)}\|_2^2 - 2\theta_k \langle \varepsilon_k, X^{(k)} \rangle$
$= \|\varepsilon_k\|_2^2 + \|X^{(k)}\|_2^2 \left( \theta_k^2 - 2\theta_k \frac{\langle \varepsilon_k, X^{(k)}\theta_k \rangle}{\|X^{(k)}\|_2^2} \right)$

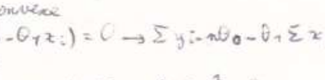
$\eta_\lambda(z) = \begin{cases} z + \lambda & \text{if } z < -\lambda \\ 0 & \text{if } z \in [-\lambda, \lambda] \\ z - \lambda & \text{if } z > \lambda \end{cases}$
$= sign(z)(|z| - \lambda)_+$

$= \|\varepsilon_k\|_2^2 + \|X^{(k)}\|_2^2 \left[ \left( \frac{\langle \varepsilon_k, X^{(k)}\theta_k \rangle}{\|X^{(k)}\|_2^2} - \theta_k \right)^2 - \left( \frac{\langle \varepsilon_k, X^{(k)} \rangle}{\|X^{(k)}\|_2^2} \right)^2 \right]$

En combinant, on obtient :
$argmin \{\|Y - X\theta\|_2^2 + 2\lambda\|\theta\|_1\} = \|X^{(k)}\|_2^2 \left( \frac{\langle \varepsilon_k, X^{(k)}\theta_k \rangle}{\|X^{(k)}\|_2^2} - \theta_k \right)^2 + 2\lambda|\theta_k| + t.i.\theta_k$
$= \left( \frac{\langle \varepsilon_k, X^{(k)}\theta_k \rangle}{\|X^{(k)}\|_2^2} - \theta_k \right)^2 + \frac{2\lambda}{\|X^{(k)}\|_2^2}|\theta_k| = (z - x)^2 + 2\tilde{\lambda}|x|$

- 3 cas : $x > 0 \rightarrow |x| = x$, $f$ devient : $(z - x)^2 + 2\tilde{\lambda} x$
CPO : $f'(x) = 0 \rightarrow -2z + 2x + 2\tilde{\lambda} = 0 \rightarrow x = z - \tilde{\lambda} > 0 \rightarrow z > \tilde{\lambda}$
$x < 0 \rightarrow |x| = -x$, $f$ devient : $(z - x)^2 - 2\tilde{\lambda} x$
CPO : $f'(x) = 0 \rightarrow -2z + 2x - 2\tilde{\lambda} = 0 \rightarrow x = z + \tilde{\lambda} < 0 \rightarrow z < -\tilde{\lambda}$
$x = 0$ : on note que la sous différentielle à $x = 0$ est $[-1, 1]$.
CPO : $f'(x) = 0 \rightarrow -2z + 2x + 2\tilde{\lambda}\partial|x| = 0 \rightarrow x = z - \tilde{\lambda}\partial|x| = 0$
impliquant $z - \tilde{\lambda} \leq 0 \leq z + \tilde{\lambda} \Leftrightarrow -\tilde{\lambda} \leq -z \leq \tilde{\lambda} \Leftrightarrow \tilde{\lambda} \geq z \geq -\lambda \rightarrow |z| < \lambda$

- Général : biais et risque quadratique de $\hat{\sigma}^2 = \frac{1}{n}\sum(y_i - \bar{y}_n)^2 \Leftrightarrow \frac{n}{\sigma^2}\hat{\sigma}^2 = \frac{\sum(y_i - \bar{y})^2}{\sigma^2}$ est distribué $\chi^2(n-1)$, donc $\mathbb{E}(\frac{n}{\sigma^2}\hat{\sigma}^2) = n - 1 \rightarrow E(\hat{\sigma}^2) = (\frac{n-1}{n})\sigma^2$ et donc le biais est : $\mathbb{E}(\hat{\sigma}^2) - \sigma^2 = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$.
risque quadratique : on ajoute $var(\frac{n}{\sigma^2}\hat{\sigma}^2) = 2(n-1) \rightarrow \frac{n^2}{\sigma^4} var(\hat{\sigma}^2) = 2(n-1)$
$\rightarrow var(\hat{\sigma}^2) = \frac{2\sigma^4(n-1)}{n^2}$ donc risque quadratique $= var(\hat{\sigma}^2) + (biais\,\hat{\sigma}^2)^2$
$= \frac{2\sigma^4(n-1)}{n^2} + \frac{\sigma^4}{n^2} = \frac{\sigma^4(2n-1)}{n^2}$



- projection orthogonale de $y \in \mathbb{R}^n$ sur $1_n$ :
$u$ : vecteur unitaire de $1_n$ : $u = \frac{1}{\sqrt{n}} 1_n$ donc $V = \langle u, y \rangle . u = \langle \frac{1}{\sqrt{n}} 1_n, y \rangle \frac{1}{\sqrt{n}} 1_n$
$= \frac{1}{n}(\sum y_i) 1_n = (\bar{y}, \bar{y}, \dots, \bar{y})$
- fonction $(\theta_0, \theta_1) \rightarrow \frac{1}{i}\sum(y_i - \theta_0 - \theta_1 x_i)^2$ : Hessienne : $\begin{pmatrix} 2 & 2x_i \\ 2x_i & 2x_i^2 \end{pmatrix}$
matrice singulière $\rightarrow \lambda_1 = 0$ ; puis $\lambda_2 = 2(1 + x_i^2)$ ($tr = sum(eig)$) donc H est semi-définie positive, fonction est convexe
- estimateurs OLS : CPO : $-\sum(y_i - \theta_0 - \theta_1 x_i) = 0 \rightarrow \sum y_i - n\theta_0 - \theta_1 \sum x_i = 0$
$\rightarrow \bar{y} - \theta_0 - \theta_1 \bar{x} = 0 \rightarrow \theta_0 = \bar{y} - \theta_1 \bar{x}$
CPO : $-\sum(y_i - \theta_0 - \theta_1 x_i)x_i = 0 \rightarrow \sum y_i x_i - \theta_0 \sum x_i - \theta_1 \sum x_i^2 = 0$
$\rightarrow \sum y_i x_i - (\bar{y} - \theta_1 \bar{x})\sum x_i - \theta_1 \sum x_i^2 = 0 \rightarrow \theta_1(\sum x_i^2 - \bar{x}\sum x_i) = \sum y_i x_i - \bar{y}\sum x_i \rightarrow$
$\theta_1 = \frac{\sum y_i x_i - \bar{y}\sum x_i}{\sum x_i^2 - \bar{x}\sum x_i} = \frac{\frac{1}{n}\sum y_i x_i - (\frac{1}{n}\sum y_i)(\frac{1}{n}\sum x_i)}{\frac{1}{n}\sum x_i^2 - (\frac{1}{n}\sum x_i)(\frac{1}{n}\sum x_i)} = \frac{\frac{1}{n}\sum y_i x_i - \bar{x}\bar{y}}{\frac{1}{n}\sum x_i^2 - \bar{x}^2} = \frac{cov(x,y)}{var(x)}$
- pour $X$, que vaut $ker(X^T X)$ ?
$\rightarrow$ soit $u \in ker(X)$ : $Xu = 0 \rightarrow (X'X)u = 0 \rightarrow u \in ker(X'X) \rightarrow ker(X'X) \subset ker(X)$
$\rightarrow$ soit $u \in ker(X'X)$ : $X'Xu = 0 \rightarrow u'X'Xu = 0 \rightarrow (Xu)'(Xu) = 0 \rightarrow \|Xu\|_2^2 = 0 \rightarrow Xu = 0$
donc $u \in ker(X) \rightarrow ker(X) \subset ker(X'X) \Rightarrow ker(X'X) = ker(X)$

## Tests et intervalles

si $\varepsilon \sim N(0, \sigma^2)$ avec $\hat{\sigma}^2 = \|Y - X\hat{\theta}\|_2^2 / n - rang X$

- $T_j = \frac{\hat{\theta}_j - \theta_j^*}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}} \sim t_{n - rang(X)}$
région de rejet $= [-t_{1-\alpha/2}, t_{1-\alpha/2}]^c$ avec $t_{1-\alpha/2}$ quantile d'ordre $1 - \alpha/2$
On rejette si $|T_j| > t_{1-\alpha/2}$
- intervalle de confiance : $\hat{I} = [\underline{c}, \overline{c}]$ tel que $\mathbb{P}(\underline{c} \leq \theta$ et $\overline{c} \geq \theta) = 1 - \alpha$
généralement : $\hat{I} : (y_1 \dots y_n) \rightarrow \hat{I}$
$\hat{I} = [\underline{c}(y_1 \dots y_n), \overline{c}(y_1 \dots y_n)]$ tel que $\mathbb{P}[y \in \hat{I}(y_1 \dots y_n)] \geq 1 - \alpha$
- inégalité de Markov : $\forall \delta > 0$ $\mathbb{P}(Y > \delta) \leq \mathbb{E}(Y)/\delta$
- inégalité de Tchebychev : $\forall \delta > 0$ $\mathbb{P}(|X - \mathbb{E}(X)| > \delta) \leq \frac{Var(X)}{\delta^2}$
ex : Bernoulli : $\hat{p} = 1/n \sum y_i$
$\mathbb{E}(\hat{p}) = p$, $Var(\hat{p}) = p(1-p)/n$ donc :
$\mathbb{P}(|\hat{p} - p| > \delta) \leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4n\delta^2}$
$\Leftrightarrow \mathbb{P}(-\delta \leq p - \hat{p} \leq \delta) \geq 1 - \frac{1}{4n\delta^2}$
$\Leftrightarrow \mathbb{P}(p \in I = [\hat{p} - \delta, \hat{p} + \delta]) \geq 1 - \alpha$

- TLC : soit $y_1 \dots y_n$ i.i.d de carré intégrable soit $\mu$ et $\sigma$ la moyenne / écart-type alors $\sqrt{n}\left(\frac{\bar{y}_n - \mu}{\sigma}\right) \xrightarrow{d} \mathcal{N}(0,1)$
- lemme de Slutsky : si $\hat{\sigma} \rightarrow \sigma$, alors $\sqrt{n}\left(\frac{\bar{y}_n - \mu}{\hat{\sigma}}\right) \xrightarrow{d} \mathcal{N}(0,1)$
- ex : Bernoulli : $\hat{\sigma}^2 = \sum(y_i - \hat{p})/n = \hat{p} - \hat{p}^2$, $\alpha = 0.05$
Slutsky : $\sqrt{n}(\hat{p} - p)/\hat{\sigma} \sim \mathcal{N}(0,1)$ ; $n = 1000$, $\hat{p} = 0.55$
$\mathbb{P}\{-1.96 < \sqrt{n}\frac{\hat{p} - p}{\hat{\sigma}} < 1.96\} \approx 0.95$
$\Leftrightarrow \mathbb{P}\{-1.96\,\hat{\sigma}/\sqrt{n} + \hat{p} < p < 1.96\,\hat{\sigma}/\sqrt{n} + \hat{p}\} \approx 0.95$
$\Leftrightarrow \mathbb{P}\{-1.96 \times 0.5/31.63 + 0.55 < p < 1.96 \times 0.5/31.63\} \approx 0.95$
$\Leftrightarrow \mathbb{P}\{0.52 < p \leq 0.58\} \approx 0.95$
- Regression : $T_j = (\hat{\theta}_j - \theta_j^*)/\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}} \sim t_{n - rang(X)}$
$\mathbb{P}\{\hat{\theta}_j - t_{1-\alpha/2}\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}} < \theta_j \leq \hat{\theta}_j + t_{1-\alpha/2} \dots\} = 0.95$
- test d'hypothèse : on rejette $H_0$ : $\hat{\theta}_j = 0$ si
$\frac{|\hat{\theta}_j|}{\hat{\sigma}\sqrt{(X^T X)^{-1}_{jj}}} \leq t_{1-\alpha/2}$
note : 1.96 car si $\alpha = 0.05$ $\Phi^{-1}(1 - \alpha/2) = 1.96$

## Bootstrap

- principe plug-in : remplacer $\mathbb{P}$ par $\mathbb{P}_n = \frac{1}{n}\sum \delta_{x_i}$
- racine statistique : une quantité qui converge en loi (e.g : TLC)
- Bootstrap : méthode pour approcimer $\sqrt{n}(T(\mathbb{P}_n) - T(\mathbb{P}))$
- Monde réel : $X_1 \dots X_n \sim \mathbb{P}$(inconnu) $\Downarrow$ inférence $T(\mathbb{P}_n) \Rightarrow \sqrt{n}(T(\mathbb{P}_n) - T(\mathbb{P}))$
Monde bootstrap : $X_1^* \dots X_n^* \sim \mathbb{P}_n^* = \sum \delta_{X_i}^*$ $\Downarrow$ $T(\mathbb{P}_n^*) \Rightarrow \sqrt{n}(T(\mathbb{P}_n^*) - T(\mathbb{P}_n)) \approx \sqrt{n}(T(\mathbb{P}_n) - T(\mathbb{P})) \approx N(0,1)$
- algorithme : for $b = 1 \dots B \rightarrow (X_i^*)_{i=1 \dots n} \sim \mathbb{P}_n$ (échantillon bootstrap)
$\rightarrow \theta_b^* = \hat{\theta}(X_1^* \dots X_n^*)$
on obtient $(\rightarrow R_b^* = \sqrt{n}(T(\mathbb{P}_n^*) - T(\mathbb{P}_n)) = \sqrt{n}(\theta_b^* - \hat{\theta}_n))$

$\{\theta_b^*\}$ ou $\{R_b^*\}$ : échantillon bootstrap qui simule la loi de $\sqrt{n}(T(\mathbb{P}_n) - T(\mathbb{P}))$. Idéalement, $B \rightarrow \infty$ $\hat{R}^* = n^{1/2}(\hat{\theta}^* - \hat{\theta})$
- ex : biais : $\mathbb{E}[\hat{\theta}] - \theta_0 \rightarrow B^{-1}\sum_{b=1}^{B} \theta_b^* - \hat{\theta}$
variance : $\mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] \rightarrow B^{-1}\sum_{b=1}^{B} (\hat{\theta}_b^* - B^{-1}\sum \hat{\theta}_b^*)^2$
- racine statistique : cas régulier $n^{1/2}(\hat{\theta} - \theta_0) \xrightarrow{L} \mathcal{N}(0, \sigma^2)$
Bootstrap : but, reproduire le comportement d'une racine statistique
- on note $\xi_\alpha$ le $\alpha$-quantile de $n^{1/2}(\hat{\theta} - \theta_0)$. Alors :
ic : $\mathbb{P}\{\theta_0 \in [\hat{\theta} - \xi_{1-\alpha/2}/n^{1/2}, \hat{\theta} - \xi_{\alpha/2}/n^{1/2}]\} = 1 - \alpha$
test : $\mathbb{P}(n^{1/2}(\hat{\theta} - 1) \leq \xi_{\alpha/2}$ or $n^{1/2}(\hat{\theta} - 1) \geq \xi_{1-\alpha/2}) = \alpha$
- asymptotique : $\left[\hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}}\xi_{1-\alpha/2}^{\infty}, \hat{\theta} - \frac{\hat{\sigma}}{\sqrt{n}}\xi_{\alpha/2}^{\infty}\right]$
bootstrap : $\left[\hat{\theta}_n - \frac{1}{\sqrt{n}}\hat{\xi}_{B,1-\alpha/2}, \hat{\theta}_n - \frac{1}{\sqrt{n}}\hat{\xi}_{B,\alpha/2}\right]$ $\hat{\xi}_{B,\alpha}$ : estimateur bootstrap du $\alpha$-quantile de $n^{1/2}(\hat{\theta} - \theta_0)$ basé sur B échantillons bootstrap

- montrer que $(X'X)$ non inversible $\Leftrightarrow \Sigma(x_i-\mu)(x_i-\mu)'$ non inversible $\begin{pmatrix} 1 & \tilde{X}_1 \\ \vdots & \\ 1 & \tilde{X}_n \end{pmatrix}\begin{pmatrix} u_1 \\ u_{-1} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}$

$(X'X)$ non inversible $\Leftrightarrow \exists u$ s.t $(X'X)u = 0 \Leftrightarrow Xu = 0$ ($\ker X'X = \ker X$) $\Rightarrow$

$\Rightarrow \tilde{X}_i^T u_{-1} + u_1 = 0\ \forall i = 1 \dots n \Leftrightarrow \tilde{X}_i^T u_{-1} = -u_1 \Leftrightarrow (\Sigma \tilde{X}_i^T) u_{-1} = -nu_1 \Leftrightarrow \bar{\tilde{X}}^T u_{-1} = -u_1$ donc

$\tilde{X}_i^T u_{-1} = \bar{\tilde{X}}^T u_{-1} \Leftrightarrow (X_i - \bar{\tilde{X}})' u_{-1} = 0 \Leftrightarrow (X_i - \tilde{X})(X_i - \bar{\tilde{X}})' u_{-1} = 0 \Leftrightarrow \Sigma(X_i - \bar{\tilde{X}})(X_i - \bar{\tilde{X}})' = 0$

réciproque : $\frac{1}{n}\left[\Sigma(x_i - \hat{\mu}_n)(x_i - \hat{\mu}_n)^T\right] u = 0 \Leftrightarrow \frac{1}{n}(\tilde{X} - 1_n \hat{\mu}_n)'(\tilde{X} - 1_n \hat{\mu}_n) u = 0$

$\Leftrightarrow (\tilde{X} - 1_n \hat{\mu}_n) u = 0$ ($\ker A = \ker A'A$) $\Leftrightarrow \tilde{X} - 1_n(\hat{\mu}_n u) = 0 \Leftrightarrow \tilde{X} - 1_n c = 0$

$\Leftrightarrow (1_n\ \tilde{X})\begin{pmatrix} -c \\ u \end{pmatrix} = 0 \Leftrightarrow Xw = 0,\ u \neq 0$ donc $\ker X \neq \{\emptyset\}$ et $\ker(X'X) \neq \{\emptyset\}$

donc $(X'X)$ est non inversible.

- $X$ plein rang : $\hat{\theta} = (X'X)X'Y$
- $X$ pas de plein rang : $\hat{\theta} = (X'X)^+ X'Y$ ; $(X'X)^+$ : Moore-Penrose inverse
- 2 cas possibles pour estimateur OLS (existence, unicité) : existence toujours garantie (théorème de projection). unicité : $(X'X)$ inversible : unique : $\hat{\theta} = (X'X)^{-1}X'Y$

$(X'X)$ non inversible : infinité de solutions avec $\{\hat{\theta}_n + \ker(X)\}$, où $\hat{\theta}_n$ représente n'importe quelle solution particulière : $(X'X)(\hat{\theta}_n + \lambda u) = X^T Y$

- $\mathrm{cov}(\hat{\theta}_n) = \sigma^2 (X'X)$
- estimateur sans biais du bruit : $\hat{\sigma}_n^2 = \sigma_n^2\left(\frac{n}{n-p-1}\right)$ (si on part du principe rang$(X) = p+1$)
- soit $X$ ; on remplace $\tilde{X}_s$ par $\tilde{X}_b$

a) soit $X_b = (1\ X_1 \dots X_s b \dots X_n) = (1\times 1\ \ 1\times X_1 \dots b \times X_s \dots 1 \times X_n)$

$= (1\ X_1 \dots X_s \dots X_n)\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & b & \\ & & & 1 \end{pmatrix} = XD$

b) $\hat{\theta}_{b_n} = (X_b'X_b)X_b'Y = (D'X'XD)^{-1}D'X'Y = D^{-1}(X'X)^{-1}(D')^{-1}D'X'Y$
$= D^{-1}(X'X)^{-1}X'Y = D^{-1}\hat{\theta}_n$

c) $\mathrm{var}(\hat{\theta}_b) = \sigma^2(X_b'X_b)^{-1} = \sigma^2(D'X'XD)^{-1} = D^{-1}\sigma^2(X'X)^{-1}(D')^{-1} = D^{-2}\mathrm{var}(\hat{\theta}_n)$

d) prédiction : $\hat{y}_b = X_b\hat{\theta}_b = XDD^{-1}\hat{\theta}_n = X\hat{\theta}_n = \hat{y}$ : inchangée

- $\Omega = \mathrm{diag}(w_1 \dots w_n)$ : $\mathrm{argmin}_\theta \frac{1}{2}(y - X\theta)'\Omega(y - X\theta) = \frac{1}{2}[y'\Omega y + \theta'X'\Omega X\theta - 2\theta'X'\Omega y]$

$\frac{\partial f}{\partial \theta} = 0 \Leftrightarrow \frac{1}{2}[2X'\Omega X\theta - 2X'\Omega Y] \Leftrightarrow X'\Omega X\theta = X'\Omega Y \Leftrightarrow \hat{\theta} = (X'\Omega X)^{-1}X'\Omega Y$

- loi asymptotique de $\sqrt{n}(\hat{\theta} - \theta^*)$. On a : $\hat{\theta} - \theta^* = A\varepsilon = (X'X)^{-1}X'\varepsilon$ donc $\hat{\theta}_n - \theta^*$ : combinaison indépendante de $\varepsilon_i$ : TLC applicable. moyenne$(\hat{\theta} - \theta^*) = 0$ et variance empirique $\hat{\sigma}^2 \to \sigma^2$ donc lemme Slutsky applicable. Donc $\sqrt{n}(\hat{\theta} - \theta^*) \xrightarrow{d} \mathcal{N}(0, \sigma^2(X'X)^{-1})$
- loi de $\hat{\theta}_n$ dans le modèle fixed design et $\varepsilon \sim \mathcal{N}(0, \sigma^2 I)$ : puisque on a $Y = X\theta^* + \varepsilon$, $Y$ est une combinaison linéaire de gaussiennes et est gaussienne. donc $\hat{\theta}_n = (X'X)^{-1}X'Y$ est gaussien. $E(\hat{\theta}_n) = \theta^*$ et $\mathrm{var}(\hat{\theta}_n) = \sigma^2(X'X)^{-1}$ donc : $\hat{\theta}_n \sim \mathcal{N}(\theta^*, \sigma^2(X'X)^{-1})$
- risque de prédiction de $\hat{\theta}_n$ : $\mathbb{E}[\|Y^* - \hat{Y}\|_2^2] = \mathbb{E}[\|X(\theta^* - \hat{\theta})\|_2^2] = \mathbb{E}[\|X(X'X)^{-1}X'\varepsilon\|_2^2]$
$= \mathbb{E}[\|H_X\varepsilon\|_2^2] = \mathbb{E}\,\mathrm{tr}[H_X\varepsilon\varepsilon'H_X'] = \mathrm{tr}[H_X\mathbb{E}(\varepsilon\varepsilon')H_X'] = \sigma^2\mathrm{tr}[H_XH_X'] = \sigma^2\mathrm{tr}(H_X)$
$= \sigma^2\mathrm{tr}[X(X'X)^{-1}X'] = \sigma^2\mathrm{tr}[X'X(X'X)^{-1}] = \sigma^2\mathrm{tr}\, I_{p+1} = \sigma^2(p+1)$

Ridge :

- dérivation de l'estimateur Ridge : $\hat{\theta}_n^{Ridge} = \underset{\theta \in \mathbb{R}^p}{\mathrm{argmin}} \|Y - X\theta\|_2^2 + \lambda\|\theta\|^2$. On a :

$\|Y - X\theta\|_2^2 + \lambda\|\theta\|^2 = (Y - X\theta)'(Y - X\theta) + \lambda\theta'\theta = Y'Y + \theta'X'X\theta - 2\theta'X'Y + \lambda\theta'\theta$

$\nabla f(\theta) = 0 \Rightarrow 2X'X\theta - 2X'Y + 2\lambda\theta = 0 \Leftrightarrow (X'X + \lambda I)\theta = X'Y \Leftrightarrow \hat{\theta} = (X'X + \lambda I)^{-1}X'Y$

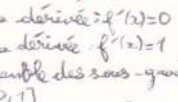
lorsque $X = I_n$, on obtient $\hat{\theta} = (I + \lambda I)^{-1}IY \Leftrightarrow \hat{\theta} = [(1+\lambda)I]^{-1}Y \Leftrightarrow \hat{\theta} = \frac{1}{1+\lambda}Y$

- formule explicite de Ridge : $\hat{\theta} = (X'X + \lambda I)^{-1}X'Y$
- variance du Ridge : $\hat{\theta} = (X'X + \lambda I)^{-1}X'Y \Rightarrow \mathrm{var}(\hat{\theta}) = (X'X + \lambda I)^{-1}X'\mathrm{var}(Y)X(X'X + \lambda I)^{-1}$
comme $Y = X\theta^* + \varepsilon$, $\mathrm{var}(Y) = \mathrm{var}(\varepsilon) = \sigma^2 I_n \Rightarrow \mathrm{var}(\hat{\theta}) = \sigma^2(X'X + \lambda I)^{-1}(X'X)(X'X + \lambda I)^{-1}$
- formule explicite de : $\hat{\theta} = \mathrm{argmin} \|Y - X\theta\|_2^2 + \lambda\|D\theta\|_2^2$. On développe $f(\theta)$ :

$f(\theta) = (Y - X\theta)'(Y - X\theta) + \lambda(\theta'D'D\theta) = (Y'Y + \theta'X'X\theta - 2\theta'X'Y) + \lambda(\theta'D'D\theta)$

$\nabla f(\theta) = 0 \Leftrightarrow 2X'X\theta - 2X'Y + 2\lambda D'D\theta = 0 \Leftrightarrow (X'X + \lambda D'D)\theta = X'Y \Leftrightarrow \hat{\theta} = (X'X + \lambda D'D)^{-1}X'Y$

Lasso :

- donner en tous points la sous-différentielle de $f(x) = \max(x, 0)$ 

$f(x)$ est continue sur $]-\infty, 0[$, sa sous différentielle est alors sa dérivée : $f'(x) = 0$
$f(x) = x$ est continue sur $]0, +\infty[$, sa sous différentielle est alors sa dérivée $f'(x) = 1$
$f(x) = 0$ est discontinue à $x = 0$ ; sa sous différentielle est l'ensemble des sous-gradients c.à.d les vecteurs $u$ tels que $f(y) - f(x) \geq u^T(y - x)$ : ici $\partial f(x) = [0, 1]$

- étape de mise à jour principale pour le problème : $\hat{\theta}_\lambda = \underset{\theta\ \mathrm{sc}\ \Omega\theta \geq 0}{\mathrm{argmin}} \frac{1}{2}\|y - X\theta\|_2^2$ ce qui en Lagrangien donne : $\underset{\theta}{\mathrm{argmin}} \frac{1}{2}\|Y - X\theta\|_2^2 + \lambda\|\Omega\theta\|_1$
- résoudre $\hat{\theta}_\lambda = \mathrm{argmin} \frac{1}{2}\|Y - X\theta\|_2^2 + \lambda\sum_{j=1}^{p} w_j|\theta_j|$ : de manière standard. On a juste un $\tilde{\lambda}_j = \lambda w_j$ spécifique à chaque entrée du vecteur $\theta$.

ACP/SVD

- pseudo inverse de $X$ : toute matrice $X^+$ telle que $XX^+X = X$. Soit $X = \sum_{i=1}^{r} s_i u_i V_i^T$ le pseudo-inverse est $X^+ = \sum_{i=1}^{r} s_i^{-1} V_i u_i^T$. En effet les définitions impliquent que : $X = VSU^T$ et $X^+ = US^{-1}V^T$ ; donc $XX^+X = VSU^TUS^{-1}V^TX = IX = X$
- que vaut : $\max f(u, v) = u^TXv$ s.c $\|u\|_2^2 = 1$ $\|v\|_2^2 = 1$

note : $\frac{\partial b^Ta}{\partial b} = a$, $\frac{\partial b^TAb}{\partial b} = 2Ab$

Test

- Soit $X_1 \dots X_n$ iid à valeur $[0,1]$. test pour $p = P(X=1) = 1/2$

On a : $\hat{p} = \frac{1}{n}\Sigma X_i$ $\mathbb{E}(\hat{p}) = \frac{1}{n}\Sigma\mathbb{E}(X_i) = \frac{1}{n}np = p$

$\mathrm{var}(\hat{p}) = \mathrm{var}(\frac{1}{n}\Sigma X_i) = \frac{1}{n^2}\Sigma\mathrm{var}(X_i)$ par indépendance $= \frac{1}{n^2}np(1-p) = \frac{p(1-p)}{n}$

Chebychev : $\mathbb{P}(|X - \mathbb{E}(X)| > \delta) \leq \frac{\mathrm{Var}(X)}{\delta^2} \Rightarrow \mathbb{P}(|\hat{p} - p| > \delta) \leq \frac{p(1-p)}{n\delta^2} \leq \frac{1}{4n\delta^2}$

$\mathbb{P}(|\hat{p} - p| \leq \delta) \geq 1 - \frac{1}{4n\delta^2} \Leftrightarrow \mathbb{P}(-\delta \leq p - \hat{p} \leq \delta) \geq 1 - \frac{1}{4n\delta^2}$

$\Leftrightarrow \mathbb{P}(\hat{p} - \delta \leq p \leq \hat{p} + \delta) \geq 1 - \frac{1}{4n\delta^2} \Leftrightarrow \mathbb{P}(\hat{p} - \delta \leq p \leq \hat{p} + \delta) \geq 1 - \alpha$

$\alpha = \frac{1}{4n\delta^2} \Leftrightarrow \delta^2 = \frac{1}{4n\alpha} \Leftrightarrow \delta = \frac{1}{2\sqrt{n\alpha}}$

test : $H_0 : p = 1/2$ $H_1 : p \neq 1/2$ détermine $\alpha$ et calcule $\delta = 1/2\sqrt{n\alpha}$ calcule $\hat{p}$ et $I = [\hat{p} - \delta, \hat{p} + \delta]$. Si $1/2 \in I$, accepte $H_0$, sinon rejette $H_0$.

- Soit $X_1 \dots X_n$ iid $\mathcal{N}(\mu, \sigma^2)$. test pour $\mu = 1$. On a $\bar{X} = \frac{1}{n}\Sigma X_i$, $\mathbb{E}(\bar{X}) = \mu$
$\mathrm{var}(\bar{X}) = \mathrm{var}(\frac{1}{n}\Sigma X_i) = \frac{1}{n^2}\Sigma\mathrm{var}(X_i) = \frac{\sigma^2}{n}$. comme $\bar{X}$ combinaison linéaire de gaussiennes, $\bar{X}$ est gaussien : $\bar{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

donc $\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$ d'où : $\mathbb{P}\left(-\Phi^{-1}(1-\alpha/2) \leq \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \leq \Phi^{-1}(1-\alpha/2)\right) = 1 - \alpha$

et donc : $\mathbb{P}(\bar{X} - \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n} \leq \mu \leq \bar{X} + \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}) = 1 - \alpha$

On pose : $H_0 : \mu = 1$, $H_1 : \mu \neq 1$

calcule de $\bar{X}$, puis $I = [\bar{X} - \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}, \bar{X} + \Phi^{-1}(1-\alpha/2)\sigma/\sqrt{n}]$

si $1 \in I$, accepte $H_0$ ; sinon, rejette $H_0$.

Bootstrap

- Soit $\epsilon_1 \dots \epsilon_n$ iid $\sim \mathcal{N}(\mu, \sigma^2)$, $\mu$ inconnue. on observe $(y_i, x_i)_{i=1..n}$ avec $y_i = x_i^T\beta + \epsilon_i$, $x_i$ déterministe

a) $y_1$ et $y_2$ indépendants ? oui car $x_1^T\beta$ et $x_2^T\beta$ sont des constantes, et $\epsilon_1$ et $\epsilon_2$ sont indépendants ; lois différentes : $y_i \sim \mathcal{N}(x_i^T\beta + \mu, \sigma^2)$

On calcule $\hat{\epsilon}_i = y_i - x_i^T\hat{\beta}$ et donc $\hat{\mu} = \frac{1}{n}\Sigma\hat{\epsilon}_i$ avec $\mathbb{E}(\hat{\mu}) = \mu$

donc $(y_i^*) \sim \mathcal{N}(x_i^T\beta + \mu, \sigma^2)$

- procédure bootstrap sur les résidus pour calculer l'écart quadratique moyen des moindres carrés :

On observe $(y_1, x_1) \dots (y_n, x_n)$ ; calculer d'abord $\hat{\theta}_n = (X^TX)^{-1}X^TY$

puis estimer $\hat{\epsilon}_1 = Y_1 - X_1\hat{\theta} \dots \hat{\epsilon}_n = Y_n - X_n\hat{\theta}$

phase bootstrap : pour $b = 1 \dots B$ :

- pioche avec remise de $\epsilon_1^* \dots \epsilon_n^*$
- génère $y_i^* = X_i\hat{\theta} + \epsilon_i^*$, obtiens $y_1^* \dots y_n^*$
- calcule $\theta_b^* = (X^TX)^{-1}X^TY^*$

on obtient $\{\theta_b^*\}_{b=1\dots B}$

Statistique bootstrap : $\mathbb{E}[(\hat{\theta} - \theta)^2] \approx \frac{1}{B}\sum_{b=1}^{B}(\theta_b^* - \hat{\theta}_n)^2$

---

- dérivation elastic net : $\frac{1}{2}\|Y - X\theta\| + \lambda\left(\alpha\|\theta\|_1 + (1-\alpha)\frac{\|\theta\|_2^2}{2}\right)$

On $\|\theta\|_2^2 = \sum_{l=1}^{p}\theta_l^2 = \theta_l^2 + W$

En combinant, on obtient :

$\|X^{(l)}\|_2^2\left(\frac{\langle\varepsilon_l, X^{(l)}\theta_l\rangle}{\|X^{(l)}\|_2^2} - \theta_l\right)^2 + 2\lambda\alpha|\theta_l| + \lambda(1-\alpha)\theta_l^2$ + t.i. $\theta_l$

$= \left(\frac{\langle\varepsilon_l, X^{(l)}\theta_l\rangle}{\|X^{(l)}\|_2^2} - \theta_l\right)^2 + \frac{2\lambda\alpha}{\|X^{(l)}\|_2^2}|\theta_l| + \frac{\lambda(1-\alpha)}{\|X^{(l)}\|_2^2}\theta_l^2$

de la forme $(z - x)^2 + 2\lambda_1|x| + \lambda_2 x^2$

3 cas :

- $x > 0$ d'où $|x| = x$ et $f(x) = (z - x)^2 + 2\lambda_1 x + \lambda_2 x^2$
et $f'(x) = 0 \Leftrightarrow -2z + 2x + 2\lambda_1 + 2\lambda_2 x = 0$
$\Leftrightarrow x = \frac{z - \lambda_1}{1 + \lambda_2} > 0 \Leftrightarrow z > \lambda_1$
- $x < 0$ d'où $|x| = -x$ et $f(x) = (z - x)^2 - 2\lambda_1 x + \lambda_2 x^2$
et $f'(x) = 0 \Leftrightarrow -2z + 2x - 2\lambda_1 + 2\lambda_2 x = 0$
$\Leftrightarrow x = \frac{z + \lambda_1}{1 + \lambda_2} < 0 \Leftrightarrow z < -\lambda_1$
- $x = 0$ avec $\partial|x| = [-1, 1]$
$f'(x) = 0 \Leftrightarrow -2z + 2x + 2\lambda_1\partial|x| + 2\lambda_2 x = 0$
$\Leftrightarrow x = \frac{z - \partial|x|\lambda_1}{1 + \lambda_2} \Leftrightarrow \frac{z - \lambda_1}{1 + \lambda_2} < 0 < \frac{z + \lambda_1}{1 + \lambda_2} \Leftrightarrow z - \lambda_1 \leq 0 \leq z + \lambda_1$

$\Leftrightarrow |z| < \lambda$