

Pénalisation minimale ℓ_1
pour l'estimation de densité.
(Dantzig et lasso adaptatif)

1

Cadre théorique

• $X_1, \dots, X_m$ iid $\sim f(x) dx$

Hyp: $f \in L^2$ (mais pas $f \in L^\infty$ et surtout pas $\|f\|_\infty$ connu)

$\Rightarrow$ on cherche à estimer f comme combinaison linéaire d'atomes d'un dictionnaire $\{\varphi_k\}$

$$f_\lambda = \sum_{k=1}^p \lambda_k \varphi_k$$

Hyp: $\varphi_k \in L^2$ (souvent $\|\varphi_k\|_2^2 = 1$ mais pas nécessaire)

• $\varphi_k \in L^\infty$ et $\|\varphi_k\|_\infty$ connu.

• $p \gg m$ possible.

Outil de base

• $\hat{\beta}_k = \frac{1}{m} \sum_{i=1}^m \varphi_k(x_i)$

$\rightarrow$ équivalent empirique de $E(\varphi_k(x)) = \langle f, \varphi_k \rangle = \beta_k$

• $\forall \lambda \in \mathbb{R}^p, \quad \langle f_\lambda, \varphi_k \rangle = \sum \lambda_{k'} \langle \varphi_{k'}, \varphi_k \rangle$
 $= (G\lambda)_k$

avec G la matrice de Gram des φ_k $G_{k,k'} = \langle \varphi_k, \varphi_{k'} \rangle$

$\Rightarrow$ Pb en λ : $(G\lambda)_k \approx \hat{\beta}_k$
 $G\lambda \approx \hat{\beta}$

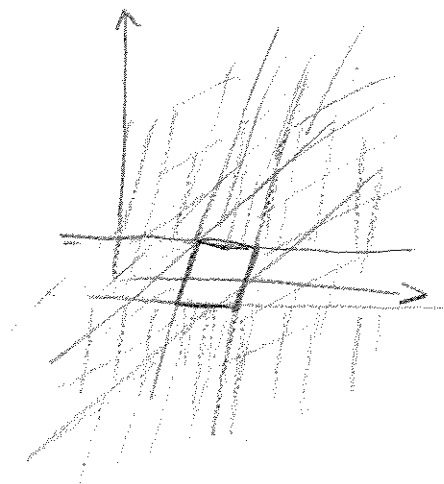
Contraintes ℓ_∞

L2

■ On recherche λ tel que $\|G\lambda - \hat{\beta}\|_\infty \leq \eta$

→ Contrainte sur λ donné par des hyperplans

$$\hat{\beta}_e - \eta \leq (G\lambda)_e \leq \hat{\beta}_e + \eta$$



■ "Plein" de solution → régularisation

$\|\lambda\|_0$ minimale $\Rightarrow$ • solution parcimonieuse
• Pb de calculabilité $\rightarrow$ algorithme d'exhaustif

$\|\lambda\|_1$ minimale $\Rightarrow$ • solution parcimonieuse
sous de bonnes hypothèses
• Pb de calculabilité $\rightarrow$ analyse aréce
linéaire

$\|\lambda\|_2$ minimale $\Rightarrow$ • solution non parcimonieuse...

Santraig

$$\hat{\lambda}^0 = \arg \min \|\lambda\|_1 \text{ sous } \|G\lambda - \hat{\beta}\|_\infty \leq \eta$$

$$\hat{\beta}^0 = G\hat{\lambda}^0$$

$$\hat{\lambda} = \argmin \frac{1}{2} \|G\lambda - \hat{\beta}\|_2^2 + \eta \|\lambda\|_1$$

$$\Rightarrow \text{Karush-Kuhn-Tucker} : \|G^2 \lambda - G\hat{\beta}\|_\infty \leq \eta$$

$$\hat{\lambda} = \argmin \frac{1}{2} (\lambda^* G \lambda - 2 \lambda^* \hat{\beta}) + \eta \|\lambda\|_1$$

$$\Rightarrow KKT \quad \|G\lambda - \hat{\beta}\|_\infty \leq \eta$$

Moindre carré avec une pénalisation ℓ_1 !

$$\|\beta - \beta_\lambda\|_2^2 = \|\beta\|_2^2 + \|\beta_\lambda\|_2^2 - 2 \langle \beta, \beta_\lambda \rangle$$

$$\text{et } \|\beta_\lambda\|_2^2 = \lambda^* G \lambda (= \lambda^* \phi^* \phi \lambda = \|\phi \lambda\|_2^2)$$

avec $\phi = (\varphi_1, \dots, \varphi_p)$ des vec l.i.

$$\text{et } \langle \beta, \beta_\lambda \rangle = \sum \lambda_e \langle \beta, \varphi_e \rangle = \sum \lambda_e \beta_e = \lambda^* \beta = E(\lambda^* \hat{\beta})$$

Choix de η

Contrôle de $|\beta_e - \hat{\beta}_e|$

$\rightarrow$ lien entre le vrai coeff $\beta_e = E(\varphi_e(x)) = \langle \beta, \varphi_e \rangle$

et le coeff empirique $\hat{\beta}_e = \frac{1}{n} \sum_{i=1}^n \varphi_e(x_i)$

$\rightarrow$ On souhaite que la zone de confiance contienne le vrai paramètre avec grande proba

Inégalités de concentration

$$P(|\hat{\beta}_e - \beta_e| > \eta_e) \leq 2e^{-\eta_e^2/2}$$

$$\rightarrow \text{Hoeffding} \quad \eta_e = \sqrt{\frac{\|\varphi_e\|_\infty}{\sqrt{n}}}$$

$$\rightarrow \text{Bernstein} \quad \eta_e = \sqrt{\frac{\sigma_e}{\sqrt{n}}} + \frac{\sqrt{2}}{3} \frac{\|\varphi_e\|_\infty}{n}$$

$$\text{Pg. } \sigma_\epsilon \leq \|f_\epsilon\|_\infty \rightarrow \text{pas intéressant}$$

$$\leq \|f\|_\infty^{1/2} \|f_\epsilon\|_2^2 \rightarrow \|f\|_\infty \text{ pas connu...}$$

→ Adaptatif

$$\hat{\sigma}_\epsilon^2 = \frac{1}{2n(n-1)} \sum_{i=2}^n \sum_{j=1}^{i-1} (f_\epsilon(x_i) - f_\epsilon(x_j))^2$$

$$(E[(f_\epsilon(x_i) - f_\epsilon(x_j))^2] = 2\sigma_\epsilon^2 + U\text{-stat})$$

$$\hat{\sigma}_\epsilon^2 = \left(\hat{\sigma}_\epsilon^2 + U \frac{\|f_\epsilon\|_\infty}{\sqrt{n}} \right)^2 + 3 U^2 \frac{\|f_\epsilon\|_\infty^2}{n}$$

$$\eta_\epsilon = U \frac{\hat{\sigma}_\epsilon}{\sqrt{n}} + \frac{U^2}{3} \frac{\|f_\epsilon\|_\infty}{n}$$

■ $U = \sqrt{2\delta \log p} \Rightarrow$ contrôle simultané pour tous les ϵ ..

$$P(\forall \epsilon, |\hat{\beta}_\epsilon - \beta_\epsilon| \leq \eta_\epsilon) \geq 1 - 2 \left(\frac{1}{p}\right)^{\delta-1}$$

$$\rightarrow \eta_\epsilon = \sqrt{2\delta} \|f_\epsilon\|_\infty \sqrt{\frac{\log p}{n}}$$

$$\text{ou } \eta_\epsilon = \sqrt{2\delta} \hat{\sigma}_\epsilon \sqrt{\frac{\log p}{n}} + \frac{2\delta}{3} \|f_\epsilon\|_\infty \frac{\log p}{n}$$

Pg η_ϵ aléatoire mais contrôle avec grande proba du type

$$\sqrt{\delta} \sigma_\epsilon \sqrt{\frac{\log p}{n}} + \frac{2\delta}{3} \frac{\log p}{n} \leq \eta_\epsilon \leq 4\sqrt{\delta} \sigma_\epsilon \sqrt{\frac{\log p}{n}} + 10\delta \|f_\epsilon\|_\infty \frac{\log p}{n}$$

$$\leadsto \eta_\epsilon^2 \geq 2\delta \sigma_\epsilon^2 \frac{\log p}{n}$$

Dantzig et Lasso adaptatif:

15

on pose $D_g = \begin{pmatrix} \eta_1(x) \\ \vdots \\ \eta_p(x) \end{pmatrix}$

Dantzig : $\hat{\lambda}^D = \argmin \|\lambda\|_1$ sous $\|D_g^{-1}(G\lambda - Y)\|_\infty \leq 1$

Lasso : $\hat{\lambda}^L = \argmin \frac{1}{2} [\lambda^* G \lambda - 2 \lambda^* \hat{\beta}] + \|D_g \lambda\|_1$
+ $\sum \eta_e |\lambda_e|$

$\Rightarrow$ + η_e est grand - la contrainte est forte
+ on régularise.

Rq : Algo de minimisation de type chemin d'homotopie (LARS) existe pour les deux méthodes.

Hypothèses sur le dictionnaire

■ Structure de $\{p_e\} \cong$ Structure d'une base orthonormée.

> Norme $\|b_\lambda\|_2^2 = \|\lambda\|_2^2$ $\Leftrightarrow$ Equivalence Dantzig Lasso Seuilage doux

$$\leadsto \phi_{\min} \|\lambda\|_2^2 \leq \lambda^* G \lambda \leq \phi_{\max} \|\lambda\|_2^2$$

$\leadsto$ contrainte sur les v.p. de la matrice de Gramm

$\leadsto \Delta \neq \text{Frame} \rightarrow$ Base de Riesz

$\rightarrow$ Covariance $\langle p_q, p_{q'} \rangle = S_{qq'}$

$\leadsto \max_{q \neq q'} |\langle p_q, p_{q'} \rangle| \leq M(p)$ (Colonne mutuelle)

$\leadsto$ contrainte sur les valeurs hors-diagonales de G

$\leadsto$ résultat de reconstruction si $\|\lambda\|_0 \leq \frac{1}{2} \left(1 + \frac{1}{M(p)}\right)$

■ Restriction sur le support

6

$$\phi_{\min}(\ell) \|\lambda\|_2^2 \leq \|\beta_\lambda\|_2^2 \leq \phi_{\max}(\ell) \|\lambda\|_2^2$$

$$\text{si } \|\lambda\|_0 \leq \ell$$

$$|\langle \beta_\lambda, \beta_{\lambda'} \rangle| \leq \Theta_{\delta, \ell} \|\beta_\lambda\|_2 \|\beta_{\lambda'}\|_2$$

$\lambda^* \in \mathcal{S}_\lambda$

$$\text{si } \|\lambda\|_0 \leq \Delta, \|\lambda'\|_0 \leq \ell \text{ et } \mathcal{S}_\lambda \cap \mathcal{S}_{\lambda'} = \emptyset$$

↑ support de λ'

on type du type

$$\exists \Delta, \ell \text{ tel que } 1 \leq \Delta \leq \frac{p}{2}, \ell \geq \Delta \text{ et } \ell + \Delta \leq p$$

$$\text{H1} \quad \text{si } \phi_{\min}(\ell + \Delta) > \Theta_{\ell, \ell + \Delta}$$

$$\text{H2} \quad \text{si } \ell \phi_{\min}(\ell + \Delta) > \Delta \phi_{\max}(\ell)$$

$$\text{TR: Avec proba } \geq 1 - \left(\frac{1}{m}\right)^{\delta-1}, \text{ si H1 ou H2}$$

$$\forall \mathcal{S} \text{ tel que } |\mathcal{S}| \leq \Delta$$

$$\forall \lambda \in \mathbb{R}^p, \forall \beta > 0,$$

$$\|\hat{\beta}^0 - \beta\|_2^2 \leq \|\beta_\lambda - \beta\|_2^2 + \Delta \left(\|\eta_\delta\|_\infty + \|\eta_{\delta'}\|_\infty \right)^2 4 \left(\frac{1}{\beta} + \frac{1}{\kappa^2} \right)$$

$$+ \frac{\|\lambda_{\mathcal{S}^c}\|_1^2}{\Delta} \beta \left(1 + \frac{2\mu\sqrt{\Delta}}{\kappa} \right)$$

$$+ \frac{(\|\hat{\lambda}^0\|_1 - \|\lambda\|_1)_+}{2\Delta} \beta \left(1 + \frac{2\mu\sqrt{\Delta}}{\kappa} \right)$$

avec μ et κ comme

$$\mu = \frac{\Theta_{\ell, \ell + \Delta}}{\sqrt{\ell \phi_{\min}(\ell + \Delta)}} \text{ et } \kappa = \sqrt{\phi_{\min}} - \sqrt{\Delta} \mu = \sqrt{\phi_{\min}} \left[1 - \frac{\Theta_{\ell, \ell + \Delta}}{\phi_{\min}} \sqrt{\frac{\Delta}{\ell}} \right] \text{ sous H1}$$

et

$$\mu = \sqrt{\frac{\phi_{\max}(\ell)}{\ell}} \text{ et } \kappa = \sqrt{\phi_{\min}} \left[1 - \sqrt{\frac{\ell \phi_{\max}(\ell)}{\ell \phi_{\min}(\ell + \Delta)}} \right] \text{ sous H2}$$

■ Pas de restriction sur $\lambda \leadsto$ contrôle des problèmes...

$\rightarrow \| \beta_\lambda - \beta \|_2^2 \leadsto$ terme d'approximation classique.

$\rightarrow \Delta \left(\| \eta_\delta \|_\infty + \| \eta_{\delta'} \|_\infty \right)^2 \leadsto$ terme de variance sur le support J
 $\sim (\delta + \delta') \| \sigma_\epsilon^2 \|_\infty \frac{\log p}{m} \Delta$

$\rightarrow \frac{\| \lambda_{J^c} \|_1^2}{\Delta} \leadsto$ défaut de support $\sim$ terme d'approximation

$$\lambda(a) \sim \left(\frac{1}{e} \right)^\delta \Rightarrow \frac{\| \lambda_{J^c} \|_1^2}{\Delta} \sim \| \lambda_{J^c} \|_2^2$$

$\rightarrow \frac{(\| \hat{\lambda}^0 \|_1 - \| \lambda \|_1)_+}{2\Delta} \leadsto$ défaut de contrôle si $\| \lambda \|_1 \leq \| \hat{\lambda}^0 \|_1$

$\rightarrow$ Explique pourquoi c'est δ' qui intervient dans la pénalité et pas δ

■ Si $\beta = \beta_\lambda$ ou $\beta_\lambda = P_\phi \beta$ alors avec $P \geq 1 - 2 \left(\frac{1}{n} \right)^\delta$

$G\lambda = \beta$ satisfait le critère de Dantzig

$$\hookrightarrow \| \hat{\lambda}^0 \|_1 \leq \| \lambda \|_1$$

$\rightarrow$ Disjonction du dernier terme.

$\rightarrow$ Choix de $\delta \geq 1$ raisonnable.

Th : Equivalence Dantzig / Lasso

$$\begin{aligned} | \| \hat{\beta}^0 - \beta \|_2^2 - \| \hat{\beta}^L - \beta \|_2^2 | &\leq \Delta \left(\| \eta_\delta \|_\infty + \| \eta_{\delta'} \|_\infty \right)^2 + \left(\frac{1}{\beta} + \frac{1}{K^2} \right) \\ &\quad + \frac{\| \hat{\lambda}_{J^c}^L \|_1^2}{\Delta} \left(1 + \frac{2\sqrt{\Delta}}{K} \right) \end{aligned}$$

R₂ $\hat{\lambda}^L$ satisfait toujours le critère de Dantzig.

Localisation des hypothèses sur le dictionnaire

8

$$\|\beta_\lambda\|_2 \geq k(j) \|\lambda_j\|_2 \quad \text{si} \quad \|\lambda_{j^c}\|_1 \leq c_0 \|\lambda_j\|_1$$

$$HL(j) \quad \forall c \quad \|\beta_\lambda\|_2 \geq k(j) \|\lambda_j\|_2 - \mu(j)c \quad \text{si} \quad \|\lambda_{j^c}\|_1 \leq \|\lambda_j\|_1 + c$$

$$\underline{Pg}: H1 \text{ ou } H2 \Rightarrow HL(j) \quad \forall j, |j| = \Delta$$

$$\underline{Th}: \text{ Avec } \mu_{\text{rate}} \geq 1 - \left(\frac{1}{m}\right)^{S'-1}$$

$$\forall j \text{ tel que } HL(j),$$

$$\forall \lambda \in \mathbb{R}^p, \quad \forall \beta > 0$$

$$\begin{aligned} \|\hat{\beta}^0 - \beta\|_2^2 \leq & \|\beta_\lambda - \beta\|_2^2 + |j| (\|\eta_\delta\|_\infty + \|\eta_{j^c}\|_\infty)^2 4 \left(\frac{1}{\beta} + \frac{1}{k^2(j)} \right) \\ & + \frac{\|\lambda_{j^c}\|_1^2}{|j|} \beta \left(1 + \frac{2\mu(j)|j|}{k(j)} \right) \\ & + \frac{(\|\hat{\lambda}^0\|_1 - \|\lambda\|_1)_+}{2|j|} \beta \left(1 + \frac{2\mu(j)|j|}{k(j)} \right) \end{aligned}$$

$\rightarrow$ localisation fine $\Rightarrow$ intérêt du contrôle des $\neq$ pertes
(sauf la dernière...)

$\rightarrow$ Illustration numérique avec des mélanges Fourier/Haar...

Pénalisation minimale

19

γ petit mais $\gamma > 1$ "bonne idée" $\equiv$ inégalité oracle pour $\beta_X = P\beta$
on va montrer que $\gamma < 1$ est une "mauvaise idée"

$\rightarrow$ Il existe des cas où l'estimateur est mauvais.

TR $\beta = \mathbb{1}_{[0,1]}$ Le cas de Haar $p = m$

$$\gamma > 1 \Rightarrow E(\|\hat{\beta}^0 - \beta\|_2^2) \leq C_\gamma \left(\frac{\log m}{m} \right)$$

$$\gamma < 1 \Rightarrow \exists \delta < 1 \quad E(\|\hat{\beta}^0 - \beta\|_2^2) \geq \frac{1}{m^\delta} (1 + o_m(1))$$

$\Rightarrow$ Pénalité minimale au sens où + petit elle est mauvaise.

Illustration numérique