

Estimation adaptative par sélection de partitions en rectangles dyadiques

Nathalie AKAKPO

Journées MAS 2010
Vendredi 3 septembre 2010

Université Paris 6, LPMA

- 1 Principe de sélection d'histogrammes
- 2 Choix de la forme de pénalité
- 3 Calcul de l'estimateur pénalisé
- 4 Adaptation au sens minimax
- 5 Extensions et autres cadres d'étude

Principe de sélection d'histogrammes

Cadre : $n \in \mathbb{N}^*$ fixé, X_1, \dots, X_n i.i.d. $\sim s$ inconnue sur $[0, 1]^d$

But : Estimer $s \in \mathbb{L}_2([0, 1]^d)$ à partir de l'observation de X_1, \dots, X_n

Contraste des moindres carrés : s minimise en $t \in \mathbb{L}_2([0, 1]^d)$

$$\begin{aligned} \|s - t\|^2 - \|s\|^2 &= \|t\|^2 - 2 \int_{[0,1]^d} t s \, d\mu \\ &= \|t\|^2 - 2\mathbb{E}_s \left[\frac{1}{n} \sum_{i=1}^n t(X_i) \right] \end{aligned}$$

d'où

$$\gamma(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(X_i).$$

Histogramme construit sur la partition m :

$$\hat{s}_m = \operatorname{argmin}_{t \in S_m} \gamma(t),$$

où $S_m = \{t : [0, 1]^d \rightarrow \mathbb{R} \text{ constante par morceaux sur } m\}$.

$$\hat{s}_m \mathbb{1}_l = \frac{1}{n\mu(l)} \sum_{i=1}^n \mathbb{1}_l(X_i) \text{ sur } l \in m.$$

Soit \mathcal{M} une famille finie de partitions de $[0, 1]^d$ et $(\hat{s}_m)_{m \in \mathcal{M}}$ les histogrammes associés.

Risque d'un histogramme

Soit m une partition de $[0, 1]^d$ en D_m rectangles,

- $S_m = \{t : [0, 1]^d \rightarrow \mathbb{R} \text{ constante sur chaque rectangle } l \in m\}$,
- s_m projeté orthogonal de s sur S_m .

$$\mathbb{E}_s [\|s - \hat{s}_m\|^2] = \underbrace{\|s - s_m\|^2}_{\text{biais/erreur d'approximation}} + \underbrace{\mathbb{E}_s [\|s_m - \hat{s}_m\|^2]}_{\text{variance/ erreur d'estimation}}$$

$$\|s - s_m\|^2 + \left(\inf_{[0,1]^d} s \right) \frac{D_m - 1}{n} \leq \mathbb{E}_s [\|s - \hat{s}_m\|^2] \leq \underbrace{\|s - s_m\|^2}_{\substack{\text{↘ quand } D_m \nearrow}} + \underbrace{\|s\|_\infty \frac{D_m - 1}{n}}_{\substack{\text{↗ quand } D_m \nearrow}}.$$

Choix d'un meilleur histogramme

Dans l'idéal : choisir la partition $m^{oracle} \in \mathcal{M}$ qui minimise $\mathbb{E}_s [\|s - \hat{s}_m\|^2]$, i.e. qui réalise le meilleur compromis biais-variance, MAIS $m^{oracle} = m^{oracle}(s)$.

Sélection de modèle d'après Birgé et Massart [BM97] : La meilleure partition est choisie en fonction des observations :

$$\begin{aligned}\hat{m}((X_i, Y_i)_{1 \leq i \leq n}) &= \operatorname{argmin}_{m \in \mathcal{M}} \{ \gamma(\hat{s}_m) + \text{pen}(D_m) \} \\ &= \operatorname{argmin}_{m \in \mathcal{M}} \{ -\|\hat{s}_m\|^2 + \text{pen}(D_m) \}\end{aligned}$$

et s est estimée par

$$\tilde{s} = \hat{s}_{\hat{m}} \text{ (estimateur pénalisé de } s\text{),}$$

où $\text{pen} : \mathbb{N}^* \rightarrow \mathbb{R}^+$, si possible telle que

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C \min_{m \in \mathcal{M}} \mathbb{E}_s [\|s - \hat{s}_m\|^2].$$

(inégalité de type oracle non asymptotique)

Choix de la forme de pénalité

Hypothèse (P) Soient $J_* \in \mathbb{N}$ tel que 2^{J_*} de l'ordre de $(n/\ln^2(n))^{1/d}$ et m_* la partition régulière de $[0, 1]^d$ en 2^{J_*} rectangles. Toutes les partitions de \mathcal{M} sont construites sur m_* .

Hypothèse (B) s est bornée et $\inf_{[0,1]^d} s > 0$.

Théorème 1 (d'après N.A. et C. Durot [AD10])

Pour $1 \leq D \leq 2^{J^*}$, on note $\mathcal{M}_D = \{m \in \mathcal{M} / D_m = D\}$. Soit $\{L_D\}_{1 \leq D \leq 2^{J^*}}$ une famille de réels positifs, qui peuvent dépendre de n , tels que

$$\Sigma := \sum_{D=1}^{2^{J^*}} \exp\left(-D \left(L_D - \frac{\ln |\mathcal{M}_D|}{D}\right)\right) \leq 1.$$

Si pen vérifie, pour $1 \leq D \leq 2^{J^*}$,

$$\text{pen}(D) = c(1 + L_D) \frac{\|\hat{s}_{m^*}\|_\infty D}{n},$$

où $c \geq 0$ constante suffisamment grande, alors

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1 \left(1 + \max_{1 \leq D \leq 2^{J^*}} L_D\right) \min_{m \in \mathcal{M}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}.$$

- \mathcal{M}^{reg} partitions **régulières** de $[0, 1]$ en 2^j intervalles, $j = 0, \dots, J_*$:

$$|\mathcal{M}_D^{reg}| \in \{0, 1\} \text{ d'où } L_D = \ln(2)$$

$$\text{pen}(D) = c \|\hat{s}_{m_*}\|_\infty \frac{D}{n}$$

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1 \min_{m \in \mathcal{M}^{reg}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}$$

- \mathcal{M}^{irreg} partitions de $[0, 1]$ en intervalles de longueur $\geq 2^{-J_*}$:

$$|\mathcal{M}_D^{irreg}| \leq \left(\frac{en}{D}\right)^D \text{ d'où } L_D = \ln(2e) + \ln\left(\frac{n}{D}\right)$$

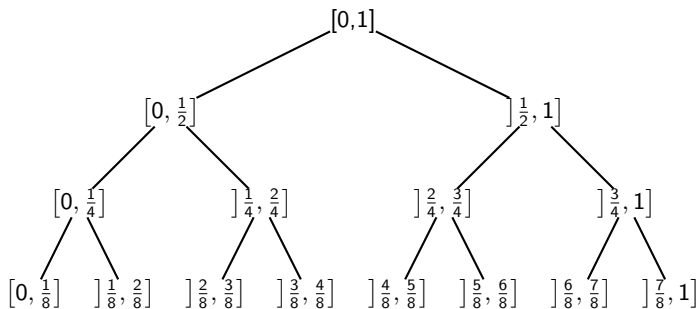
$$\text{pen}(D) = \left(c_1 + c_2 \ln\left(\frac{n}{D}\right)\right) \|\hat{s}_{m_*}\|_\infty \frac{D}{n}$$

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1(1 + \ln(n)) \min_{m \in \mathcal{M}^{irreg}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}$$

(voir aussi Barron, Birgé et Massart [BBM99], Castellan [Cas00], Lebarbier [Leb02], Sauvé [Sau09])

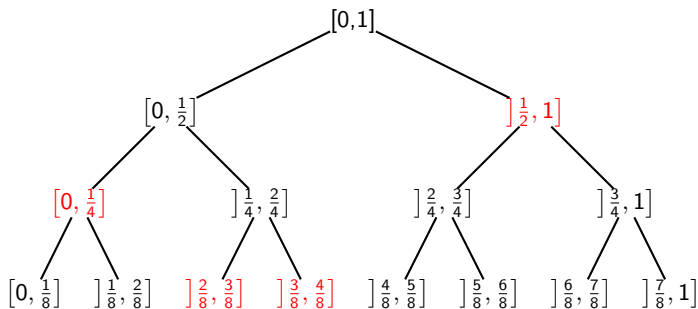
\mathcal{M}^{dyad} , partitions en intervalles dyadiques de longueur $\geq 2^{-J^*}$

Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur au moins $1/2^{J^*}$ ($J^* = 3$)



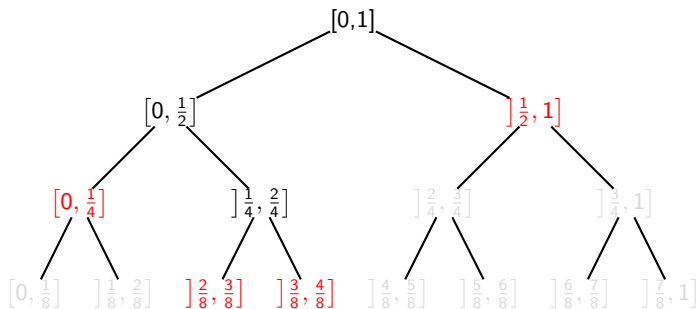
\mathcal{M}^{dyad} , partitions en intervalles dyadiques de longueur $\geq 2^{-J^*}$

Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur au moins $1/2^{J^*}$ ($J^* = 3$)



\mathcal{M}^{dyad} , partitions en intervalles dyadiques de longueur $\geq 2^{-J^*}$

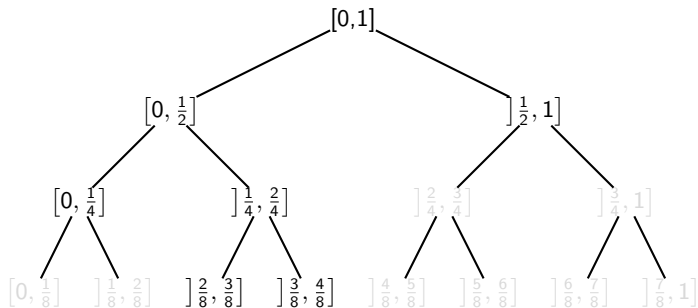
Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur au moins $1/2^{J^*}$ ($J^* = 3$)



$m \in \mathcal{M}^{dyad}$ ssi $\{I; I \in m\}$ feuilles d'un sous-arbre binaire complet de \mathcal{A}

\mathcal{M}^{dyad} , partitions en intervalles dyadiques de longueur $\geq 2^{-J^*}$

Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur au moins $1/2^{J^*}$ ($J^* = 3$)



$m \in \mathcal{M}^{dyad}$ ssi $\{I; I \in m\}$ feuilles d'un sous-arbre binaire complet de \mathcal{A}

Complexité combinatoire de \mathcal{M}^{dyad}

$$|\mathcal{M}_D^{dyad}| = \frac{1}{D} \binom{2(D-1)}{D-1} \leq 4^D, \text{ pour tout } 1 \leq D \leq 2^{J^*}$$

nombre de Catalan

d'où $L_D = \ln(8)$

- \mathcal{M}^{reg} partitions **régulières** de $[0, 1]$ en 2^j intervalles, $j = 0, \dots, J_*$:

$$\text{pen}(D) = c \|\hat{s}_{m_*}\|_\infty \frac{D}{n}$$

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1 \min_{m \in \mathcal{M}^{reg}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}$$

- \mathcal{M}^{irreg} partitions de $[0, 1]$ en intervalles de longueur $\geq 2^{-J_*}$:

$$\text{pen}(D) = \left(c_1 + c_2 \ln \left(\frac{n}{D} \right) \right) \|\hat{s}_{m_*}\|_\infty \frac{D}{n},$$

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1 (1 + \ln(n)) \min_{m \in \mathcal{M}^{irreg}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}.$$

- \mathcal{M}^{dyad} partitions de $[0, 1]$ en intervalles **dyadiques** de longueur $\geq 2^{-J_*}$:

$$\text{pen}(D) = c \|\hat{s}_{m_*}\|_\infty \frac{D}{n}$$

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C_1 \min_{m \in \mathcal{M}^{dyad}} \left\{ \|s - s_m\|^2 + \frac{\|s\|_\infty D_m}{n} \right\} + \frac{C_2}{n}$$

NB : partitions de type CART en classification ou régression (cf. Gey et Nédélec [GN05], Gey et Lebarbier [GL08])

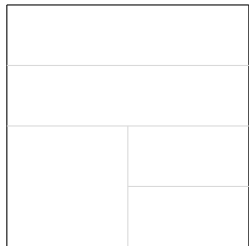
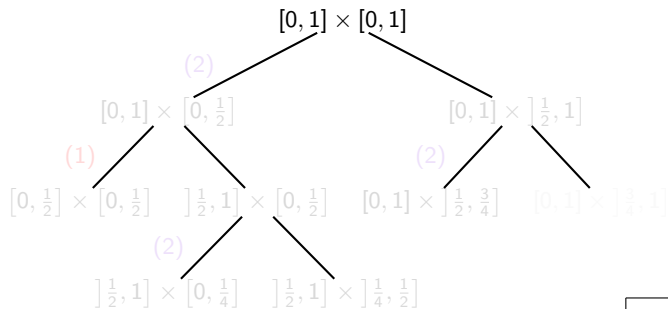
Rectangle dyadique de $[0, 1]^d$:

$$I = \left[\frac{k_1}{2^{j_1}}, \frac{k_1 + 1}{2^{j_1}} \right] \times \dots \times \left[\frac{k_d}{2^{j_d}}, \frac{k_d + 1}{2^{j_d}} \right]$$

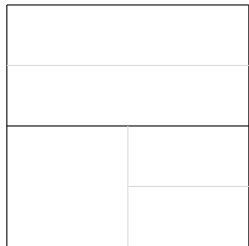
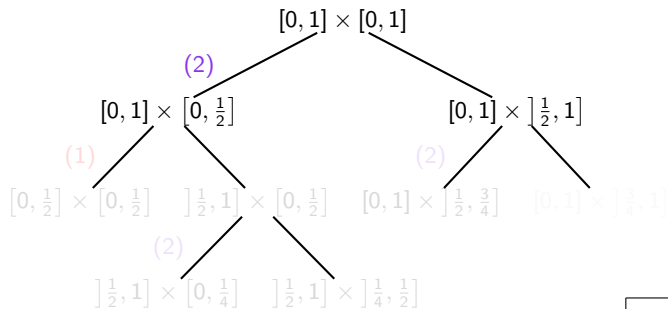
où $j_l \in \mathbb{N}$ et $k_l \in \{0, \dots, 2^{j_l} - 1\}$, pour $l = 1, \dots, d$.

\mathcal{M}^{dyad} collection des partitions de $[0, 1]^d$ en rectangles dyadiques de côtés $\geq 2^{-J_\star}$.

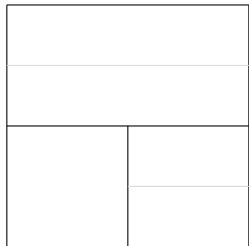
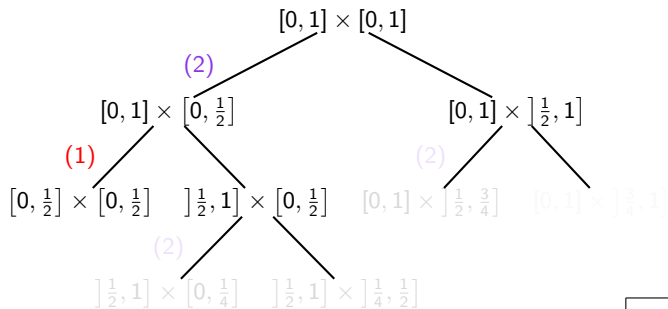
Exemple de partition de $[0, 1]^d$ en rectangles dyadiques ($d = 2$)



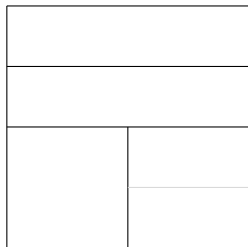
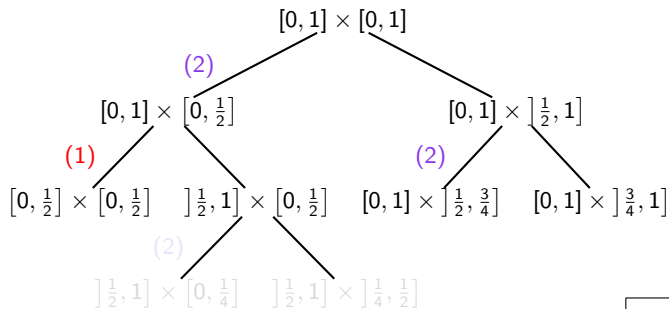
Exemple de partition de $[0, 1]^d$ en rectangles dyadiques ($d = 2$)



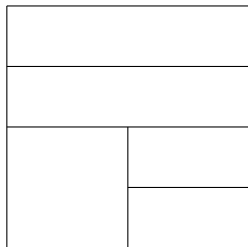
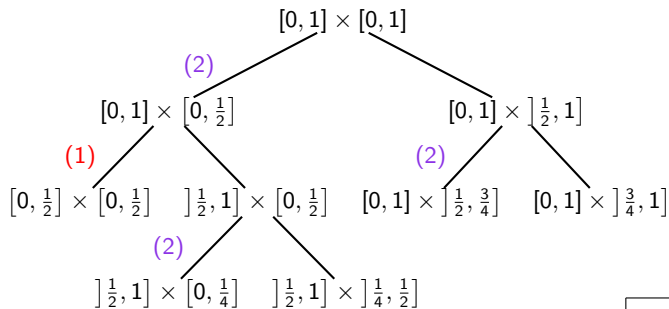
Exemple de partition de $[0, 1]^d$ en rectangles dyadiques ($d = 2$)



Exemple de partition de $[0, 1]^d$ en rectangles dyadiques ($d = 2$)



Exemple de partition de $[0, 1]^d$ en rectangles dyadiques ($d = 2$)



$$m \in \mathcal{M}_D^{dyad}$$



$(D - 1)$ -uplet d'éléments de $\{1, \dots, d\}$ (directions de coupure)
× arbre binaire complet de racine $[0, 1]^d$ à D feuilles

Complexité combinatoire de \mathcal{M}^{dyad}

$$|\mathcal{M}_D^{dyad}| \leq d^{D-1} \times \frac{1}{D} \binom{2(D-1)}{D-1} \leq (4d)^D$$

nombre de Catalan

Même forme de pénalité et même inégalité de type oracle que pour $d = 1$.

Calcul de l'estimateur pénalisé $\tilde{s} = \hat{S}_{\hat{m}}$

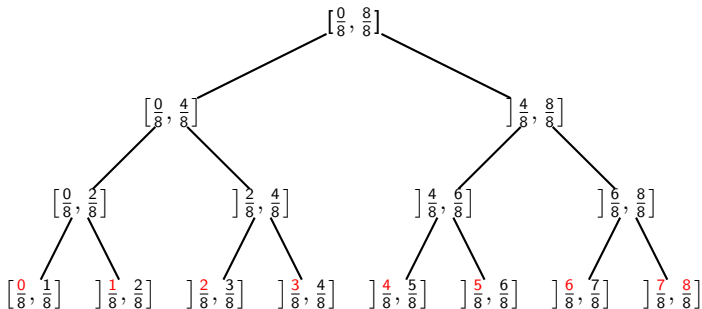
Calcul de $\tilde{\sigma}$ pour \mathcal{M}^{dyad}

$$\begin{aligned}\hat{m} &= \operatorname{argmin}_{m \in \mathcal{M}^{dyad}} \left\{ \gamma(\hat{\sigma}_m) + c \|\hat{\sigma}_{m^*}\|_\infty \frac{D_m}{n} \right\} \\ &= \operatorname{argmin}_{m \in \mathcal{M}^{dyad}} \left\{ -\|\hat{\sigma}_m\|^2 + c \|\hat{\sigma}_{m^*}\|_\infty \frac{D_m}{n} \right\} \\ &= \operatorname{argmin}_{m \in \mathcal{M}^{dyad}} \sum_{I \in m} (\mathcal{L}(I) + c \|\hat{\sigma}_{m^*}\|_\infty),\end{aligned}$$

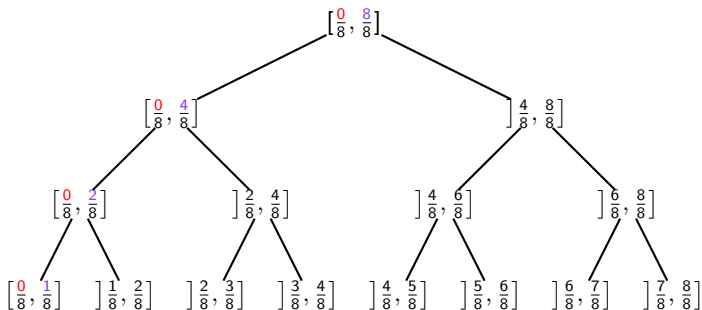
où, pour tout rectangle dyadique I ,

$$\mathcal{L}(I) = -\frac{1}{n\mu(I)} \left(\sum_{i=1}^n \mathbb{1}_I(X_i) \right)^2.$$

Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur $\geq 1/2^{J_*}$ ($J_* = 3$)



Arbre \mathcal{A} des intervalles dyadiques de $[0, 1]$ de longueur $\geq 1/2^{J_*}$ ($J_* = 3$)



Calcul de \tilde{s} pour \mathcal{M}^{dyad}

Complexité algorithmique pour $2^{J^*} \lesssim n^{1/d}$:

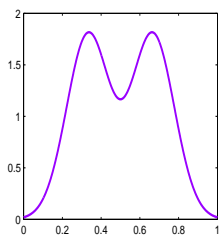
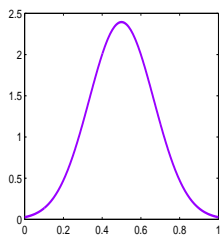
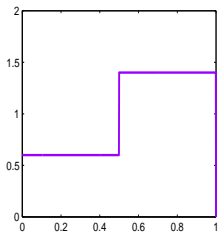
- $d = 1$: $\mathcal{O}(n)$ via un algorithme de plus court chemin [Aka10a]
- $d \geq 2$: $\mathcal{O}(n)$ via un algorithme similaire à Donoho [Don97] (voir aussi Blanchard, Schäfer, Rozenholc and Müller [BSRM07]).

NB : Complexité de calcul de \tilde{s}

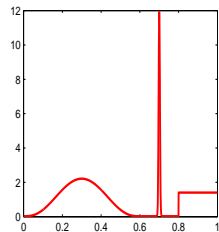
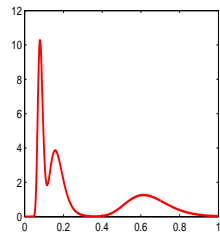
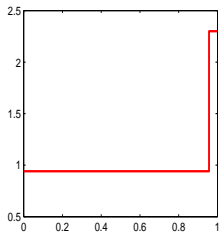
- pour \mathcal{M}^{reg} : $\mathcal{O}(n)$
- pour \mathcal{M}^{irreg} ($d = 1$) : $\mathcal{O}(n^3)$ (cf. Lebarbier [Leb02]).

Adaptation au sens minimax

Régularité homogène



Régularité non homogène



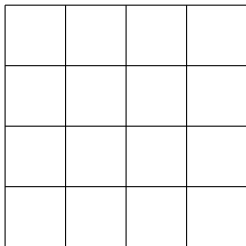


FIGURE: Homogène et isotrope

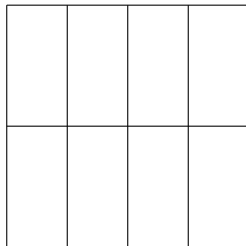


FIGURE: Homogène et anisotrope

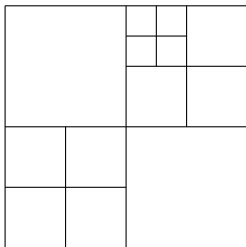


FIGURE: Non homogène et isotrope

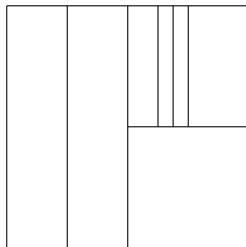


FIGURE: Non homogène et anisotrope

Adaptation (au sens minimax) à la non-homogénéité et à l'anisotropie :

- Donoho [Don97] (régression fixed design gaussienne)
- Neumann et Von Sachs [NvS97] (spectre évolutif)
- Neumann [Neu00] (modèle de bruit blanc gaussien)
- Kerkycharian, Lepski et Picard [KLP01] (modèle de bruit blanc gaussien)
- Klemelä [Kle09] (densité)

Adaptation à une puissance de $\ln(n)$ près.

Théorème 1 (bis)

Sous les hypothèses du Théorème 1. Si pen vérifie, pour $1 \leq D \leq 2^{J^*}$,

$$\text{pen}(D) = c \frac{\|\hat{s}_{m^*}\|_\infty D}{n},$$

où $c \geq 0$ constante suffisamment grande, alors

$$\mathbb{E}_s [\|s - \tilde{s}\|^2] \leq C(s) \min_{1 \leq D \leq 2^{J^*}} \left\{ \min_{m \in \mathcal{M}_D^{\text{dyad}}} \|s - s_m\|^2 + \frac{D}{n} \right\}.$$

Pour $R > 0$, $p > 0$, $\alpha = (\alpha_1, \dots, \alpha_d)$ tel que $\underline{\alpha} := \min_{1 \leq l \leq d} \alpha_l > 0$ et $\max_{1 \leq l \leq d} \alpha_l < 1$, $\rho_2 \geq \rho_1 > 0$

$$\mathcal{B}(\alpha, p, R, \rho) = \{s : [0, 1]^d \rightarrow \mathbb{R} / |s|_{B_p^\alpha} \leq R, \rho_1 \leq s \leq \rho_2\}$$

boule de Besov anisotrope de rayon R
régularité d'ordre α_l dans la l^{e} direction, mesurée dans la norme \mathbb{L}_p

En particulier :

- $\alpha_1 = \dots = \alpha_d$: boule de Besov isotrope
- $p \geq 2$: régularité homogène
- $0 < p < 2$: régularité non homogène.

Théorème 2 (N.A. [Aka09])

Pour $s \in \mathcal{B}(\alpha, p, R, \rho)$ et $D \in \mathbb{N}^*$,

$$\min_{m \in \mathcal{M}_D^{\text{dyad}}} \|s - s_m\|^2 \leq C(\alpha, p) R^2 D^{-2H(\alpha)/d}$$

où

$$H(\alpha) = \left(\frac{1}{d} \sum_{l=1}^d \frac{1}{\alpha_l} \right)^{-1} \quad (\text{moyenne harmonique}).$$

Extension du résultat d'approximation de DeVore et Yu [DY90].

Théorème 3 (N.A. [Aka09])

Soit

$$H(\alpha) = \left(\frac{1}{d} \sum_{l=1}^d \frac{1}{\alpha_l} \right)^{-1} \quad (\text{moyenne harmonique}).$$

Sous les hypothèses du Théorème 1, si $H(\alpha)/d > 1/p - 1/2$ et

$$\frac{\alpha}{H(\alpha)} \frac{d + 2H(\alpha)}{H(\alpha)} \left(\frac{H(\alpha)}{d} + \frac{1}{2} - \frac{1}{p} \right) > 1,$$

alors

$$\sup_{s \in \mathcal{B}(\alpha, \rho, R, \rho)} \mathbb{E}_s \left[\|s - \tilde{s}\|^2 \right] \leq C(\alpha, \rho, R, \rho) \left(Rn^{-H(\alpha)/d} \right)^{2d/(d+2H(\alpha))}.$$

Tous les résultats précédents (inégalité de type oracle, complexité algorithmique, approximation, adaptation) restent valables

- pour d'autres modèles : avec données censurées (N.A. et C. Durot [AD10]), en densité conditionnelle avec données dépendantes (N.A. et C. Lacour [AL10]) ;
- avec des polynômes par morceaux dyadiques (N.A. et C. Lacour [AL10], N.A. [Aka10b]) ;
- avec une pénalité additive (N.A. [Aka10b]).



N. Akakpo and C. Durot.

Histogram selection for possibly censored data.

Math. Methods Statist., A paraître, 2010.



N. Akakpo.

Estimation adaptative par sélection de partitions en rectangles dyadiques.

PhD thesis, Thèse de doctorat, Université Paris Sud XI, UFR Orsay, 2009.



N. Akakpo.

Estimating a discrete distribution via histogram selection.

ESAIM Probab. Statist., A paraître, 2010.



N. Akakpo.

From adaptive approximation to adaptive estimation for functions with anisotropic and inhomogeneous smoothness.

En cours de rédaction, 2010.



N. Akakpo and C. Lacour.

Inhomogeneous and anisotropic conditional density estimation from dependent data.

En cours de rédaction, 2010.



A. Barron, L. Birgé, and P. Massart.
Risk bounds for model selection via penalization.
Probab. Theory Related Fields, 113(3) :301–413, 1999.



L. Birgé and P. Massart.
From model selection to adaptive estimation.
In *Festschrift for Lucien Le Cam*, pages 55–87. Springer, New York, 1997.



G. Blanchard, C. Schäfer, Y. Rozenholc, and K.R. Müller.
Optimal dyadic decision trees.
Machine Learning, 66(2) :209–241, 2007.



G. Castellan.
Sélection d'histogrammes à l'aide d'un critère de type Akaike.
C. R. Acad. Sci. Paris Sér. I Math., 330(8) :729–732, 2000.



D. L. Donoho.
CART and best-ortho-basis : a connection.
Ann. Statist., 25(5) :1870–1911, 1997.



R. A. DeVore and X. M. Yu.

Degree of adaptive approximation.

Math. Comp., 55(192) :625–635, 1990.



S. Gey and E. Lebarbier.

Using cart to detect multiple change-points in the mean for large samples.

Technical report, Technical report, Preprint SSB, 2008.



S. Gey and E. Nédélec.

Model selection for CART regression trees.

IEEE Trans. Inform. Theory, 51(2) :658–670, 2005.



J. Klemelä.

Multivariate histograms with data-dependent partitions.

Statist. Sinica, 19(1) :159–176, 2009.



G. Kerkycharian, O. Lepski, and D. Picard.

Nonlinear estimation in anisotropic multi-index denoising.

Probab. Theory Related Fields, 121(2) :137–170, 2001.



E. Lebarbier.

Quelques approches pour la détection de ruptures à horizon fini.
PhD thesis, Doctoral thesis, Université Paris XI, UFR Orsay, 2002.



M. H. Neumann.

Multivariate wavelet thresholding in anisotropic function spaces.
Statist. Sinica, 10(2) :399–431, 2000.



M. H. Neumann and R. von Sachs.

Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra.
Ann. Statist., 25(1) :38–76, 1997.



M. Sauvé.

Histogram selection in non Gaussian regression.
ESAIM Probab. Stat., 13 :70–86, 2009.