CHOIX DU NOMBRE DE NOEUDS EN RÉGRESSION SPLINE PAR L'HEURISTIQUE DES PENTES

Marie Denis

Université Montpellier I

marie.denis@inserm.fr

Montpellier, le 23 Mars 2009

plan

plan

1. Introduction
   - Le modèle de régression spline
   - La sélection de modèles via une procédure de pénalisation
2. L'heuristique des pentes
   - Existence de pénalités minimales
   - Pénalités guidées par les données : Data-driven penalties
3. Applications
   - Introduction
   - Application : Simulations
4. Perspectives

La régression spline

- On observe un vecteur Gaussien $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ dont les coordonnées vérifient.

$$Y_i = s(x_i) + \epsilon_i \text{ avec } \epsilon_1 \dots, \epsilon_n \text{ i.d.d. } \mathcal{N}(0, \sigma^2), \tag{1}$$

avec s une fonction spline et $(x_1, \dots, x_n)$, n points fixés.

- Fonctions splines définies par le degré d, le nombre k de nœuds intérieurs et leur position $r = (r_1, \dots, r_k)'$.
- On utilise la base de B-splines $\{B_1^d(., r), \dots, B_m^d(., r)\}$

$$Y_i = \sum_{l=1}^{m} \beta_l B_l^d(x_i, r) + \epsilon_i. \tag{2}$$

avec $\beta = (\beta_1, \dots, \beta_m)'$ le vecteur des cœfficients spline.

- À degré fixé, on associe à chaque séquence de nœuds l'espace linéaire $\mathcal{S}_m$ engendré par les vecteurs $\{B_j^d(x, r)\}_{1 \le j \le m}$ de dimension $D_m = m$. On appelle $\mathcal{S}_m$ un modèle.

La régression spline

- On observe un vecteur Gaussien $Y = (Y_1, \dots, Y_n) \in \mathbb{R}^n$ dont les coordonnées vérifient.

$$Y_i = s(x_i) + \epsilon_i \text{ avec } \epsilon_1 \dots, \epsilon_n \text{ i.d.d. } \mathcal{N}(0, \sigma^2), \tag{1}$$

avec s une fonction spline et $(x_1, \dots, x_n)$, n points fixés.

- Fonctions splines définies par le degré d, le nombre k de nœuds intérieurs et leur position $r = (r_1, \dots, r_k)'$.
- On utilise la base de B-splines $\{B_1^d(., r), \dots, B_m^d(., r)\}$

$$Y_i = \sum_{l=1}^{m} \beta_l \, B_l^d(x_i, r) + \epsilon_i. \tag{2}$$

avec $\beta = (\beta_1, \dots, \beta_m)'$ le vecteur des cœfficients spline.

- À degré fixé, on associe à chaque séquence de nœuds l'espace linéaire $\mathcal{S}_m$ engendré par les vecteurs $\{B_j^d(x, r)\}_{1\leq j\leq m}$ de dimension $D_m = m$. On appelle $\mathcal{S}_m$ un modèle.

BUT : Sélectionner le "bon" nombre de nœuds $k \Longleftrightarrow$ Problème de choix de modèles parmis une collection finie de modèles $\{\mathcal{S}_m,\ m \in \mathcal{M}\}$.

- Deux points de vue différents sur la sélection de modèles :
  1. Les procédures **efficaces** basées sur l'idée de l'estimation non biaisée du risque, le but est de choisir un modèle minimisant ce risque $\longrightarrow$ AIC, C_p de Mallows.
  2. Les procédures **consistantes** qui supposent l'existence d'un vrai modèle de taille minimale, le but est de le trouver $\longrightarrow$ BIC.

- ICI, on s'intéresse aux procédures efficaces et à la sélection de modèles via une procédure de pénalisation.

- On considère le critère des moindres carrés :

$$\gamma_n(t) = \frac{1}{n}\sum_{i=1}^{n}(y_i - t(x_i))^2 \ = \ \|y - t\|_n^2,$$

avec $\|.\|_n$ la norme euclidienne normalisée sur $\mathbb{R}^n$.

- $\{\hat{s}_m,\ m \in \mathcal{M}\}$: estimateurs des moindres carrés associés à $\{\mathcal{S}_m,\ m \in \mathcal{M}\}$:

$$\hat{s}_m = \arg\min_{t \in \mathcal{S}_m} \{\gamma_n(t)\}.$$

- Soit $l(s,\ t) = \|s - t\|_n^2$ la fonction de perte associée.

- La qualité d'un estimateur (ou d'un modèle) est donnée par le risque quadratique correspondant :

$$\begin{aligned} \mathbb{E}[l(s, \hat{s}_m)] &= l(s, \bar{s}_m) + \mathbb{E}[l(\bar{s}_m, \hat{s}_m)] \\ &= l(s, \bar{s}_m) + \tfrac{\sigma^2}{n} D_m \end{aligned}$$

avec $\bar{s}_m$ la projection de s sur $\mathcal{S}_m$.

- Un modèle idéal est celui minimisant ce risque :

$$m^* = \inf_{m \in \mathcal{M}} \mathbb{E}[l(s, \hat{s}_m)] = \inf_{m \in \mathcal{M}} \{l(s, \bar{s}_m) + \frac{\sigma^2}{n} D_m\},$$

- L'estimateur $\hat{s}_{m^*}$ correspond à l'oracle. MAIS impossible de le choisir car s inconnue.

- La qualité d'un estimateur (ou d'un modèle) est donnée par le risque quadratique correspondant :

$$\begin{aligned}\mathbb{E}[l(s, \hat{s}_m)] &= l(s, \bar{s}_m) + \mathbb{E}[l(\bar{s}_m, \hat{s}_m)] \\ &= l(s, \bar{s}_m) + \frac{\sigma^2}{n} D_m\end{aligned}$$

avec $\bar{s}_m$ la projection de s sur $\mathcal{S}_m$.

- Un modèle idéal est celui minimisant ce risque :

$$m^* = \inf_{m \in \mathcal{M}} \mathbb{E}[l(s, \hat{s}_m)] = \inf_{m \in \mathcal{M}} \{l(s, \bar{s}_m) + \frac{\sigma^2}{n} D_m\},$$

- L'estimateur $\hat{s}_{m^*}$ correspond à l'oracle. MAIS impossible de le choisir car s inconnue.

- Le but de la sélection de modèles est de construire une procédure basée sur les observations permettant de sélectionner un modèle $\hat{m}$ d'estimateur $\hat{s}_{\hat{m}}$ de risque aussi proche que possible du risque de l'oracle i.e. satisfaisant

$$\mathbb{E}[l(s, \hat{s}_{\hat{m}})] \leq C \inf_{m \in \mathcal{M}} \{\mathbb{E}[l(s, \hat{s}_m)]\} \quad (3)$$

avec $C > 1$ une constante indépendante de s.

- L'approche via une procédure de pénalisation consiste à choisir le modèle minimisant la somme d'un risque empirique (le critère des moindres carrés, la log-vraisemblance) et d'une fonction de pénalité. Dans le cadre de la régression spline, on a :

$$crit(m) = \gamma_n(\hat{s}_m) + pen(m)$$

avec $pen : \mathcal{M} \to \mathbb{R}^+$.

- Le but de la sélection de modèles est de construire une procédure basée sur les observations permettant de sélectionner un modèle $\hat{m}$ d'estimateur $\hat{s}_{\hat{m}}$ de risque aussi proche que possible du risque de l'oracle i.e. satisfaisant

$$\mathbb{E}[l(s, \hat{s}_{\hat{m}})] \leq C \inf_{m \in \mathcal{M}} \{\mathbb{E}[l(s, \hat{s}_m)]\} \tag{3}$$

avec $C > 1$ une constante indépendante de s.

- L'approche via une procédure de pénalisation consiste à choisir le modèle minimisant la somme d'un risque empirique (le critère des moindres carrés, la log-vraisemblance) et d'une fonction de pénalité. Dans le cadre de la régression spline, on a :

$$crit(m) = \gamma_n(\hat{s}_m) + pen(m)$$

avec $pen : \mathcal{M} \to \mathbb{R}^+$.

- Le problème final est de chercher une fonction de pénalité appelée **pénalité optimale** proposant un estimateur vérifiant (3).
- La méthode de l'heuristique des pentes de Birgé et Massart (2001, 2007) propose une estimation de cette pénalité, ou une calibration en utilisant les données. Ils se placent d'un point de vue non asymptotique.

plan

Introduction

- Méthode basée sur un mélange de théorie et d'idées heuristiques :
  - Existence de **pénalités minimales** dans le cadre général des processus linéaires Gaussiens (Birgé et Massart, 2001, 2007). Généralisation au cadre de la régression hétéroscédastique sur un support aléatoire (Arlot et Massart, 2008).
  - Mise en place d'une procédure pratique pour estimer cette pénalité à partir des données (Birgé et Massart, 2007) $\longrightarrow$ Algorithme de calibration des pénalités (Arlot et Massart, 2008).

Théorème (Birgé et Massart, 2007)

Soit la collection de modèles $\{\mathcal{S}_m\}_{m\in\mathcal{M}}$, *on considère une famille de poids positifs* $\{L_m\}_{m\in\mathcal{M}}$ *satisfaisant la condition*

$$\Sigma = \sum_{\{m\in\mathcal{M}|D_m>0\}} \exp^{-L_m D_m} < +\infty.$$

Deux nombres, $\theta \in (0,1)$ *et* $\kappa > 2-\theta$ *et supposons qu'il existe un sous ensemble fini (ou vide)* $\overline{\mathcal{M}}$ *de* $\mathcal{M}$ *tel que la fonction de pénalité vérifie*

$$pen(m) \geq Q_m \; pour \, m \in \mathcal{M}\backslash\overline{\mathcal{M}},$$

avec $Q_m = (\sigma^2/n)\, D_m \,(\kappa + 2\,(2-\theta)\,\sqrt{L_m} + 2\,\theta^{-1}\,L_m)$ *pour tout* $m\in\mathcal{M}$. *Ainsi l'estimateur pénalisé correspondant* $\hat{s}_{\hat{m}}$ *existe presque sûrement et satisfait*

$$(1-\theta)\mathbb{E}[\|s-\hat{s}_{\hat{m}}\|_n^2] \leq \inf_{m\in\mathcal{M}} \left\{\|s-\overline{s}_m\|_n^2 + pen(m) - (\sigma^2/n)\,D_m\right\}$$

$$+ \sup_{m\in\overline{\mathcal{M}}} \{Q_m - pen(m)\} + (\sigma^2/n)\,\Sigma\left[(2-\theta)^2\,(k+\theta-2)^{-1} + 2\,\theta^{-1}\right].$$

Résultats

1. Permet de définir la forme des pénalité $pen(m) = \alpha \frac{\sigma^2}{n} D_m \left(1 + a\sqrt{L_m} + b\, L_m\right)$ où $\alpha > 1$, $a > 2$ et $b > 2$, avec la condition limite

$$pen(m) > \frac{\sigma^2}{n} D_m \left(1 + 2\sqrt{L_m} + 2\, L_m\right).$$

2. On pose $pen_{min}(m) = \frac{\sigma^2}{n} pen_{shape}(m)$ avec

$$pen_{shape}(m) = D_m \left(1 + 2\sqrt{L_m} + 2\, L_m\right).$$

Concept de pénalité minimale

- Concept de **pénalité minimale** permet de valider et de comprendre l'heuristique des pentes.
- Si l'on choisit une pénalité de la forme $pen(m) = \alpha\, pen_{min}(m)$ alors :
  - Si $\alpha < 1$: la dimension $D_{\hat{m}}$ tend à être proche de la dimension des plus grands modèles. Phénomène émergeant des résultats théoriques et visibles sur des données simulées.
  - Si $\alpha > 1$: la dimension $D_{\hat{m}}$ est substantiellement plus petite que la dimension des plus grands modèles
- Dans le cas de la régression spline, on prend $L_m = L$ arbitrairement petit pour tout $m \in \mathcal{M}$ et on pose $pen(m) = 2\,\frac{\sigma^2}{n}\,D_m$, on a que

$$\mathbb{E}[\|s - \hat{s}_{\hat{m}}\|_n^2] \leq C \inf_{m \in \mathcal{M}} \left\{ \|s - \bar{s}_m\|_n^2 + (\sigma^2/n)\,(D_m + 1) \right\}$$

avec C une constante adéquate.

$\Rightarrow$ Dans la suite on prendra $pen_{shape}(m) = D_m$ pour la régression spline.

Concept de pénalité minimale

- Concept de **pénalité minimale** permet de valider et de comprendre l'heuristique des pentes.
- Si l'on choisit une pénalité de la forme $pen(m) = \alpha\, pen_{min}(m)$ alors :
  - Si $\alpha < 1$: la dimension $D_{\hat{m}}$ tend à être proche de la dimension des plus grands modèles. Phénomène émergeant des résultats théoriques et visibles sur des données simulées.
  - Si $\alpha > 1$: la dimension $D_{\hat{m}}$ est substantiellement plus petite que la dimension des plus grands modèles
- Dans le cas de la régression spline, on prend $L_m = L$ arbitrairement petit pour tout $m \in \mathcal{M}$ et on pose $pen(m) = 2\,\frac{\sigma^2}{n}\, D_m$, on a que

$$\mathbb{E}[\|s - \hat{s}_{\hat{m}}\|_n^2] \leq C \inf_{m \in \mathcal{M}} \left\{ \|s - \bar{s}_m\|_n^2 + (\sigma^2/n)\,(D_m + 1) \right\}$$

avec C une constante adéquate.

$\Rightarrow$ Dans la suite on prendra $pen_{shape}(m) = D_m$ pour la régression spline.

- Le choix de $\alpha = 2$ (ou proche de 2) conduit à des stratégies optimales dans de nombreux travaux.

- On obtient donc l'égalité

$$pen_{opt}(m) = 2 * pen_{min}(m), \tag{4}$$

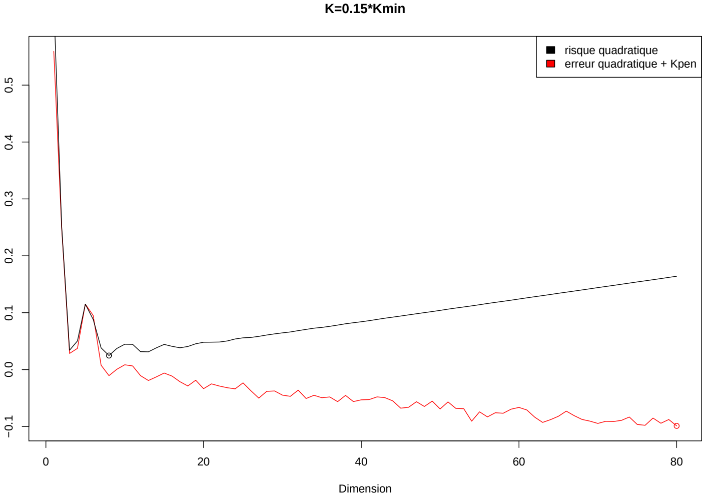
- C'est la relation caractérisant l'heuristique de Birgé et Massart. Le but final est d'estimer la pénalité minimale à partir des données.

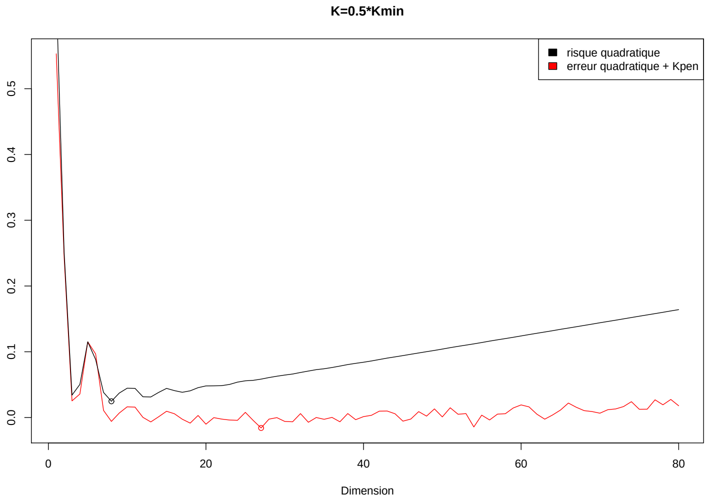
Première approche

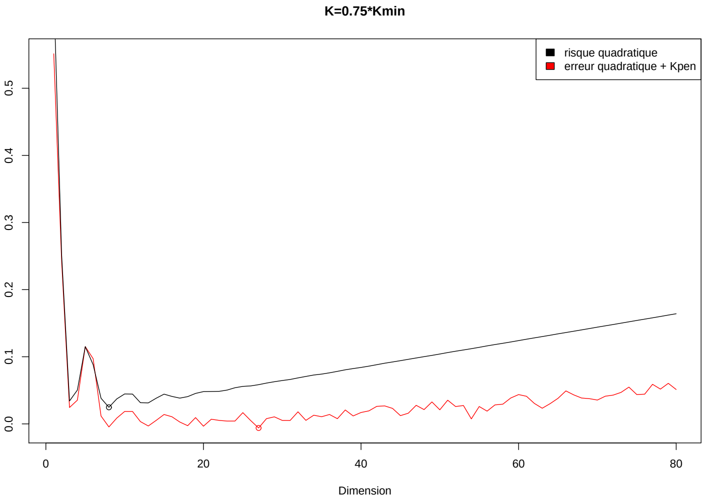
Birgé et Massart proposent une procédure pratique pour choisir une pénalité efficace quand σ^2 est inconnue.

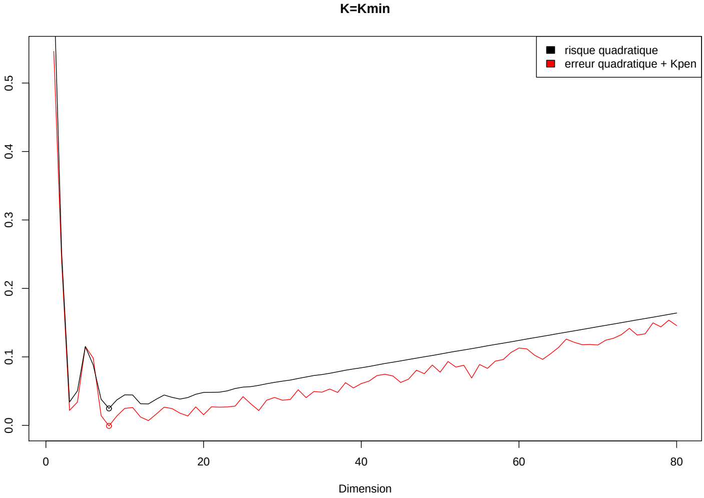
- Soit la famille continue de pénalités $pen_K(m) = K\, pen_{shape}(m)$ pour $K > 0$. On observe que pour de petites valeurs de K, les valeurs de $D_{\hat{m}}$ restent très grandes. À partir d'une valeur seuil K_{min} les valeurs chutent. $K_{min} \Leftrightarrow$ un saut de dimension (ou saut de compléxité)

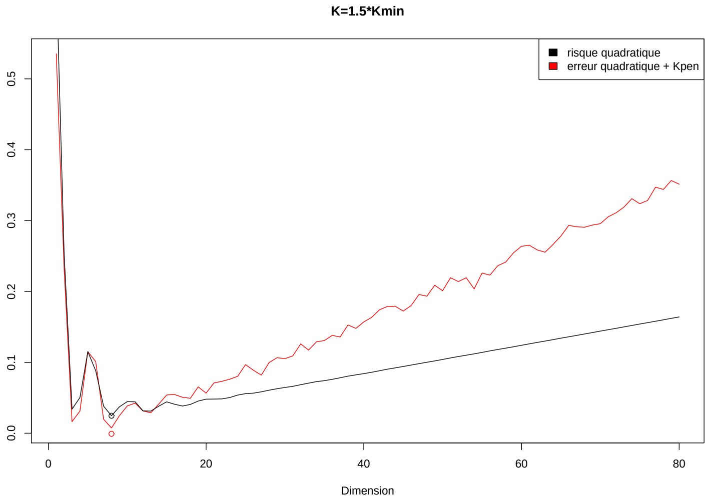
- Explosion de la procédure quand $K < K_{min}$.


K=0.15*Kmin
risque quadratique
erreur quadratique + Kpen
0.5
0.4
0.3
0.2
0.1
0.0
−0.1
0
20
40
60
80
Dimension


K=0.5*Kmin
risque quadratique
erreur quadratique + Kpen
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
Dimension


K=0.75*Kmin
risque quadratique
erreur quadratique + Kpen
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
Dimension


K=Kmin
risque quadratique
erreur quadratique + Kpen
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
Dimension


K=1.5*Kmin
risque quadratique
erreur quadratique + Kpen
0.0
0.1
0.2
0.3
0.4
0.5
0
20
40
60
80
Dimension

- On définit la pénalité optimale par

$$pen_{opt}(m) = 2\, K_{min}\, pen_{shape}(m)$$

.

IMPORTANT : K_{min} peut être estimé par les données.

$\Rightarrow$ Méthode pratique utilisée pour la détection de rupture dans la moyenne d'un signal gaussien : Lebarbier (2005).

Algorithme de calibration des pénalités (Arlot et Massart 2008)

Algorithme de calibration des pénalités

1. Pour chaque $K > 0$, calculer
$$\hat{m}(K) \in \arg\min_{m \in \mathcal{M}} \{\gamma_n(\hat{s}_m) + K\, pen_{shape}(m)\}.$$
2. Trouver $\hat{K}_{min}$ tel que $D_{\hat{m}(K)}$ est "très"grande si $K < \hat{K}_{min}$ et "raisonnable" si $K > \hat{K}_{min}$.
3. Choisir le modèle $\hat{m} = \hat{m}(2\,\hat{K}_{min})$.

Algorithme (Étape 1 de l'algorithme de calibration des pénalités)

pour tous $m \in \mathcal{M}$, on définit $f(m) = \gamma_n(\hat{s}_m)$ et $g(m) = pen_{shape}(m)$.

1. Initialisation : $K_0 = 0$, $m_0 = \arg\min_{m \in \mathcal{M}}\{f(m)\}$,
2. Étape i, $i \geq 1$: posons

$$G(m_{i-1}) = \{m \in \mathcal{M}_n \, s.t. \, f(m) > f(m_{i-1}) \; and \; g(m) < g(m_{i-1})\}.$$

Si $G(m_{i-1}) = \varnothing$, ainsi posons $K_i = +\infty$, $i_{max} = i - 1$ et stop. Sinon,

$$K_i = \inf\left\{\frac{f(m) - f(m_{i-1})}{g(m_{i-1}) - g(m)} \; s.t. \; m \in G(m_{i-1})\right\}$$

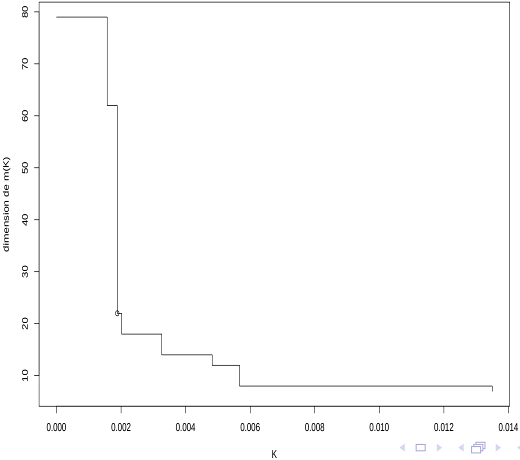
et

$$m_i = minF_i \;\; \text{avec} \;\; F_i = \arg\min_{m \in G(m_{i-1})} = \left\{\frac{f(m) - f(m_{i-1})}{g(m_{i-1}) - g(m)}\right\}$$

- La fonction $K \rightarrow \hat{m}(K)$ est une fonction constante par morceaux et décroissante.
- On résume la trajectoire de $(\hat{m}(K))_{K \geq 0}$ par :
  1. le nombre de sauts $i_{max} \in \{0, \ldots, Card(\mathcal{M}) - 1\}$,
  2. la position des sauts $(K_i)_{0 \leq i \leq i_{max}+1}$ suite croissante de nombre réel avec $K_0 = 0$ et $K_{i_{max}+1} = +\infty$,
  3. une séquence décroissante de modèles $(m_i)_{0 \leq i_{max}}$, avec $\forall i \in \{0, \ldots, i_{max}\},\ \forall K \in [K_i,\ K_{i+1}),\ \hat{m}(K) = m_i$.

Dans le cas où il y a clairement un saut, on définit

$$\hat{K}_{min} = K_{i_{jump}} \quad \text{avec} \quad i_{jump} = \arg \max_{i \in \{0, \ldots, i_{max}-1\}} \{D_{m_{i+1}} - D_{m_i}\}.$$



Deuxième approche

- On définit :

$$b_m = l(s, \overline{s}_m);\;\; V_m = l(\overline{s}_m, \hat{s}_m);\; \hat{b}_m = \gamma_n(\overline{s}_m) - \gamma_n(s);$$
$$\hat{V}_m = \gamma_n(\overline{s}_m) - \gamma_n(\hat{s}_m).$$

- Décomposition du risque quadratique :

$$\mathbb{E}[l(s, \hat{s}_m)] = b_m + \mathbb{E}[V_m]$$

- Hypothèses : $\mathbb{E}[V_m] \approx V_m$, $\mathbb{E}[\hat{V}_m] \approx \hat{V}_m$ et $\mathbb{E}[\hat{V}_m] \approx \mathbb{E}[V_m]$

- Minimiser $\gamma_n(\hat{s}_m) + pen(m)$ est équivalent à minimiser

$$\gamma_n(\hat{s}_m) - \gamma_n(s) + pen(m) \quad \approx \quad b_m - \mathbb{E}[\hat{V}_m] + pen(m)$$

- Sachant que notre but est de minimiser le risque quadratique :

$$pen(m) \quad \approx \quad \mathbb{E}[\hat{V}_m] + \mathbb{E}[V_m]$$

$$pen(m) \quad \approx \quad 2\,\hat{V}_m$$

- Dans le cas du C_p de Mallows, on a que $\mathbb{E}[\hat{V}_m] = \mathbb{E}[V_m]$ sont explicitement calculables (au moins si la variance est fixée et connue), on retrouve la pénalité associée $pen_{C_p}(m) = 2\,\frac{\sigma^2}{n}\,D_m$.

- Décomposition de $\hat{V}_m$:

$$\begin{aligned}\hat{V}_m = \gamma_n(\overline{s}_m) - \gamma_n(\hat{s}_m) &= \gamma_n(\overline{s}_m) - \gamma_n(s) + \gamma_n(s) - \gamma_n(\hat{s}_m) \\ &= \hat{b}_m + \gamma_n(s) - \gamma_n(\hat{s}_m).\end{aligned}$$

- Pour de "grandes" dimensions, $\hat{b}_m$ tend à se stabiliser. De plus quand la pénalité est proportionnelle à la dimension, on devine le comportement de $\hat{V}_m$ en fonction de D_m au travers du risque empirique $-\gamma_n(\hat{s}_m)$ (Maugis et Michel, 2008).
- La pénalité finale est donnée par

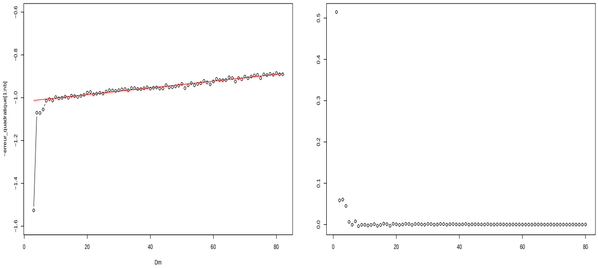
$$pen(m) \approx 2\,\alpha\,D_m$$

où α est l'estimation de la pente de la partie linéaire.

- On doit oberver un comportement linéaire de $-\gamma_n(\hat{s}_m)$ pour appliquer cette méthode.

Exemple

Graphiques de la fonction $D_m \to -\gamma_n(\hat{s}_m)$ et du terme de biais $\hat{b}_m$ en fonction de D_m.



plan

Application à la régression spline

On observe $Y = (Y_1, \ldots, Y_n)$ un vecteur Gaussien

$$Y_i = \sum_{l=1}^{m} \beta_l B_l^d(x_i, r) + \epsilon_i \text{ pour } 1 \leq i \leq n, \text{ avec } \epsilon_1, \ldots, \epsilon_n \text{ i.i.d. } \mathcal{N}(0, \sigma^2), \quad (5)$$

- On a une collection finie de modèles $\mathcal{S}_m$, $m \in \mathcal{M}$ avec $\mathcal{M} = \{d+1, \ldots, N+d\}$ où $Card(\mathcal{M}) =$nombre total de modèles et N le nombre maximal de nœuds.
- Rappel : $pen_{shape}(m) = D_m$.

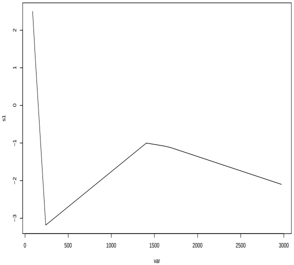
- Comparaisons avec le C_p de Mallows et le BIC :
  - $pen_{C_p}(m) = 2\,\sigma^2\,\frac{D_m}{n}$, on estime σ^2 indépendemment de la procédure de sélection de modèles.
  - $pen_{BIC}(m) = log(n)\,D_m$.
- On utilise le rapport des risques pour comparer la performance des différents critères :

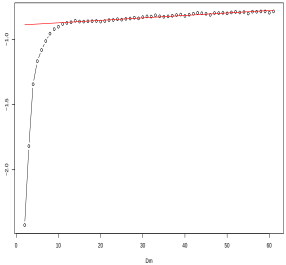
$$\mathcal{R}(\hat{s}_m) = E[l(s, \hat{s}_m)] / \inf_{m \in \mathcal{M}} E[l(s, \hat{s}_m)], \tag{6}$$

Simulations

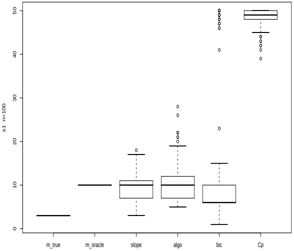
- On prend $\sigma^2 = 1$, $n = 100, 500$ et 5 fonctions différentes $(s_1, \ldots, s_5)$.
- Pour chaque fonction, on estime le rapport des risques (6) sur 500 simulations.
- On estime la fonction spline avec $d = 1$ et les k nœuds intérieurs correspondent aux k-quantiles.

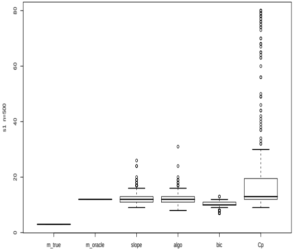
Résultats pour s_1





Résultats pour s_1

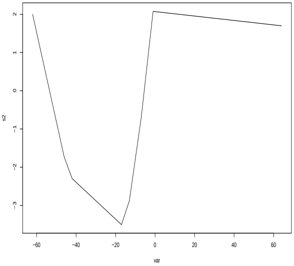


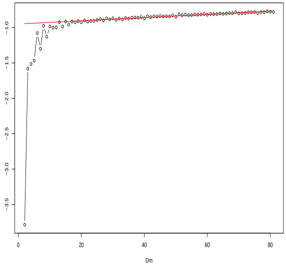


Rapports des risques correspondant :

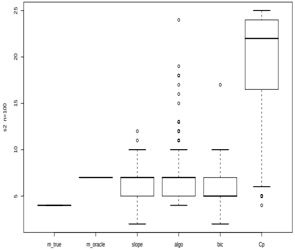
- pour $n = 100$: $BIC = 1.564838$, $Slope = 1.138617$, $Algo = 1.1374$ et $C_p = 3.474864$.
- pour $n = 500$: $BIC = 1.361813$, $Slope = 1.101046$, $Algo = 1.09442$ et $C_p = 1.851713$

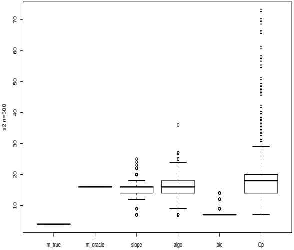
Résultats pour s_2





Résultats pour s_2

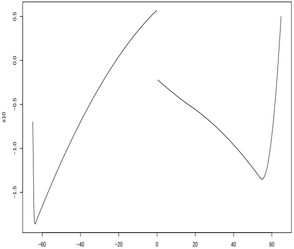


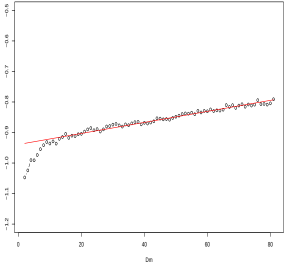


Rapports des risques correspondant :

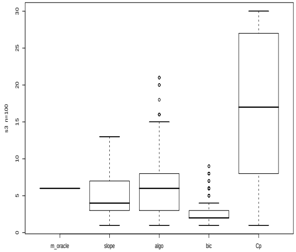
- pour $n = 100$: $BIC = 1.45197$, $Slope = 1.287763$, $Algo = 1.2747856$ et $C_p = 2.308792$.
- pour $n = 500$: $BIC = 1.460082$, $Slope = 1.106397$, $Algo = 1.118135$ et $C_p = 1.851713$

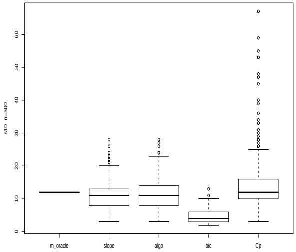
Résultats pour s_3





Résultats pour s_3

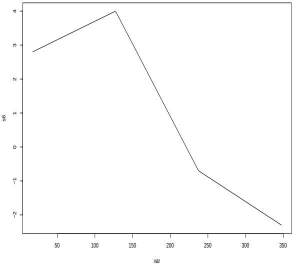


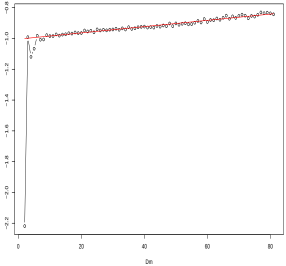


Rapports des risques correspondant :

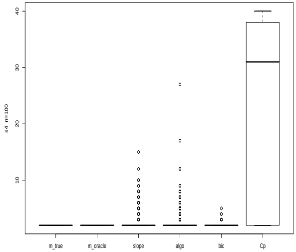
- pour $n = 100$: $BIC = 1.555163$, $Slope = 1.173173$, $Algo = 1.165988$ et $C_p = 1.689721$.
- pour $n = 500$: $BIC = 1.872631$, $Slope = 1.14936$, $Algo = 1.143538$ et $C_p = 1.229182$

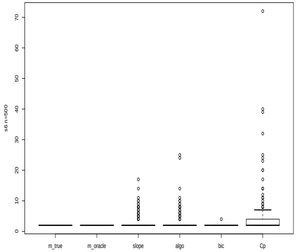
Résultats pour s_4





Résultats pour s_4

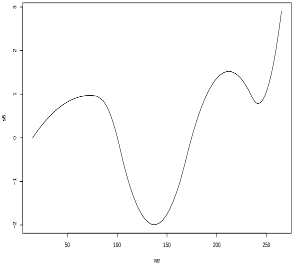


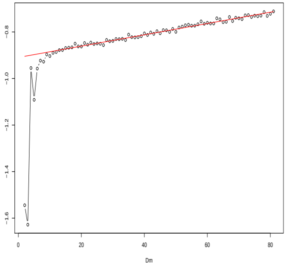


Rapports des risques correspondant :

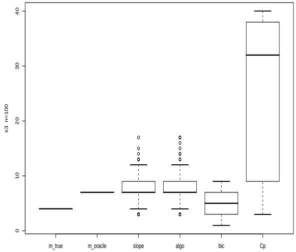
- pour $n = 100$: $BIC = 1.084681$, $Slope = 1.317911$, $Algo = 1.312262$ et $C_p = 6.021108$.
- pour $n = 500$: $BIC = 1.014020$, $Slope = 1.415023$, $Algo = 1.377729$ et $C_p = 1.581628$

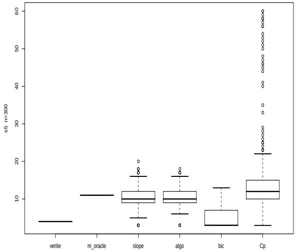
Résultats pour s_5





Résultats pour s_5





Rapports des risques correspondant :

- pour $n = 100$: $BIC = 1.627893$, $Slope = 1.230778$, $Algo = 1.21218$ et $C_p = 2.845571$.
- pour $n = 500$: $BIC = 1.924053$, $Slope = 1.15325$, $Algo = 1.141332$ et $C_p = 1.428121$

Remarques

1. Les procédures mises en place fonctionnent bien et même mieux que le C_p de Mallows (surtout pour de petits échantillons) et que le BIC en terme de risque quadratique. Cependant dans le cas où la vraie dimension coincide avec la dimension oracle, le BIC est meilleur.
2. On observe assez rapidement un comportement linéaire de la fonction $-\gamma_n(\hat{s}_m)$.
3. Le BIC semble plutôt sélectionner la vraie dimension alors que les critères de l'heuristique des pentes tendent à sélectionner la dimension oracle.

plan

- Faire varier la position des nœuds : ne plus avoir une séquence fixée de nœuds mais les considérer comme des paramètres ⇒ Problème : quelle dimension associée à chaque modèle ?
- Revoir le choix de la forme de la pénalité minimale $pen_{shape}(m) = D_m$ en utilisant une méthode de simulation.
- Une difficulté réside dans le choix et la signification du plus grand saut ⇒ Lebarbier (2005) a mis en œuvre une méthode de calibration afin de déterminer le "bon" saut dans le cadre de la détection de rupture dans la moyenne d'un signal gaussien. Objectif : adapter cette méthode à la régression spline.