Utilisation d'approches probabilistes basées sur les critères entropiques pour la recherche d'informations sur supports multimédia.

Guilhem Coq

Sous la direction de :
M. Arnaudon (LMA), C. Olivier (XLIM-SIC) et O. Alata (XLIM-SIC).

5 décembre 2008
Université de Poitiers

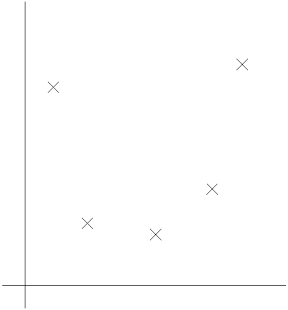
Modélisation de données par une expérience probabiliste.

- Données :
  - $\longrightarrow$ Signaux unidimensionnels : son, mesure physiologique, télécommunication.
  - $\longrightarrow$ Signaux bidimensionnels : images.
- Modélisation
  - $\longrightarrow$ Description du phénomène.
  - $\longrightarrow$ Famille de modèles, paramétriques ou non.
  - $\longrightarrow$ Sélection de modèles.
- Probabilités :
  - $\longrightarrow$ Mesurent la pertinence du modèle face à l'observation par le maximum de vraisemblance.

Surparamétrisation

Exemple de la régression gaussienne polynomiale

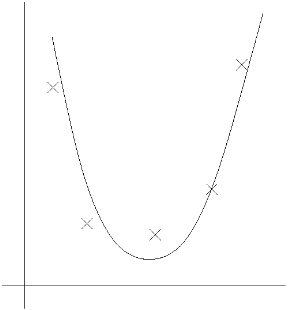
$\longrightarrow$ pertinence $\sim$ erreur quadratique par rapport aux données.



Surparamétrisation

Exemple de la régression gaussienne polynomiale

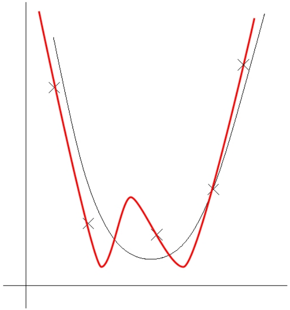
$\longrightarrow$ pertinence $\sim$ erreur quadratique par rapport aux données.

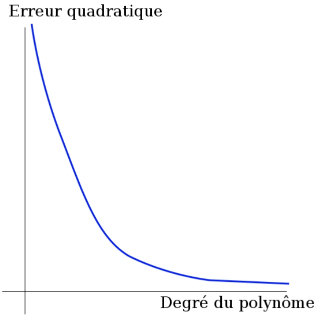


Surparamétrisation

Exemple de la régression gaussienne polynomiale

$\longrightarrow$ pertinence $\sim$ erreur quadratique par rapport aux données.





Critères d'information

- $(\Theta_i)_{i \in I}$: les modèles en compétition.
- $-\log \widehat{\mathbb{P}}_{\Theta_i}(x)$: mesure d'adéquation, à minimiser.
- Critère d'information :

$$\mathrm{IC}(\Theta_i) = -2 \log \widehat{\mathbb{P}}_{\Theta_i}(x) + |\Theta_i| \alpha(n).$$

$$\text{vraisemblance} + \text{pénalité.}$$

- Sélection du modèle par :

$$\widehat{i} = \mathrm{Argmin}(\mathrm{IC}(\Theta_i),\ i \in I).$$

$\longrightarrow$ Comportement inverse, recherche de compromis.

Critères d'information

- $(\Theta_i)_{i\in I}$: les modèles en compétition.
- $-\log \widehat{\mathbb{P}}_{\Theta_i}(x)$: mesure d'adéquation, à minimiser.
- Critère d'information :

$$\begin{array}{rcccc} \mathrm{IC}(\Theta_i) & = & -2\log \widehat{\mathbb{P}}_{\Theta_i}(x) & + & |\Theta_i|\alpha(n). \\ & & \searrow & & \nearrow \\ & & \text{vraisemblance} & + & \text{pénalité.} \end{array}$$

- Sélection du modèle par :

$$\widehat{i} = \mathrm{Argmin}(\mathrm{IC}(\Theta_i),\ i \in I).$$

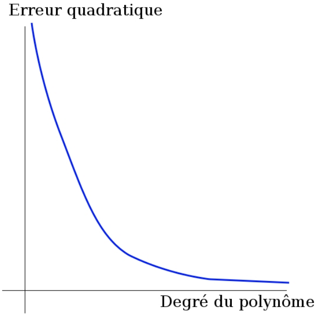
$\longrightarrow$ Comportement inverse, recherche de compromis.

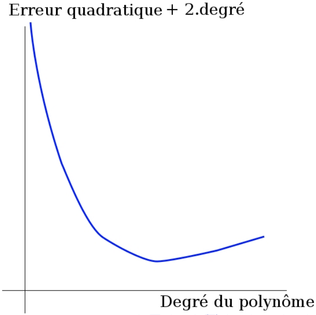
Critères d'information

- Critère d'information :

$$\begin{array}{ccccc} \mathrm{IC}(\Theta_i) & = & -2\log \widehat{\mathbb{P}}_{\Theta_i}(x) & + & |\Theta_i|\alpha(n). \\ & & \searrow & & \nearrow \\ & & \text{vraisemblance} & + & \text{pénalité.} \end{array}$$

$\longrightarrow$ Comportement inverse, recherche de compromis.





Motivation principale

Comment et *pourquoi* choisir $\alpha(n)$?

- Akaike 74 AIC : $\alpha(n) = 2$, minimisation d'un risque.
- Schwarz 78 BIC : $\alpha(n) = \log n$, bayésien.
- Hannan-Quinn 79 φ : $\alpha(n) = \log\log n$, consistance.
- Rissanen 86 MDL : $\alpha(n) = \log n$, théorie de l'information
  $\rightarrow$ Complexité stochastique, inégalité entropique
- Nishii 88 : conditions sur $\alpha(n)$ entrainant consistance.
  $\rightarrow$ El-Matouat, Hallin 96 φ_β : $\alpha(n) = n^\beta \log\log n$, $\beta \in]0,1[$
- Rissanen 92 : cadre non-paramétrique, histogrammes.
- Broersen 00 : intérêt pour le non-asymptotique.
- Baraud, Barron, Birgé, Castellan, Massart récemment : $\alpha(n) =$constante, contrôle non-asymptotique du risque, inégalités oracle.

Motivation principale

Comment et *pourquoi* choisir $\alpha(n)$?

- Akaike 74 AIC : $\alpha(n) = 2$, minimisation d'un risque.
- Schwarz 78 BIC : $\alpha(n) = \log n$, bayésien.
- Hannan-Quinn 79 φ : $\alpha(n) = \log \log n$, consistance.
- Rissanen 86 MDL : $\alpha(n) = \log n$, théorie de l'information
  $\longrightarrow$ Complexité stochastique, inégalité entropique
- Nishii 88 : conditions sur $\alpha(n)$ entrainant consistance.
  $\longrightarrow$ El-Matouat, Hallin 96 φ_β : $\alpha(n) = n^\beta \log \log n$, $\beta \in]0, 1[$
- Rissanen 92 : cadre non-paramétrique, histogrammes.
- Broersen 00 : intérêt pour le non-asymptotique.
- Baraud, Barron, Birgé, Castellan, Massart récemment : $\alpha(n) =$constante, contrôle non-asymptotique du risque, inégalités oracle.

Codages et inégalité entropique de Rissanen

- Shannon 48 (Jensen) :

$$-\sum_{x \in E} P(x) \log P(x) = H(P) \leq H(P,Q) = -\sum_{x \in E} P(x) \log Q(x)$$

$\longrightarrow$ le terme de vraisemblance dans IC estime $H(P)$.

- Rissanen : complexité stochastique, notion d'*adaptation* :

$$C_\Theta(x) = -\sum_{i=1}^{n} \log \widehat{P}_\Theta^{(i-1)}(x_i),$$

$$H_n(P) + |\Theta| \frac{\log n}{2} \leq \mathbb{E} C_\Theta(x).$$

$\longrightarrow$ $\alpha(n) = \log n$, IC(Θ) estime $C_\Theta(x)$

$\longrightarrow$ Adaptation $\sim$ pénalisation.

Codages et inégalité entropique de Rissanen

- Shannon 48 (Jensen) :

$$-\sum_{x\in E} P(x)\log P(x) = H(P) \leq H(P,Q) = -\sum_{x\in E} P(x)\log Q(x)$$

$\longrightarrow$ le terme de vraisemblance dans IC estime $H(P)$.

- Rissanen : complexité stochastique, notion d'*adaptation* :

$$C_\Theta(x) = -\sum_{i=1}^{n} \log \widehat{P}_\Theta^{(i-1)}(x_i),$$

$$H_n(P) + |\Theta|\frac{\log n}{2} \leq \mathbb{E}C_\Theta(x).$$

$\longrightarrow$ $\alpha(n) = \log n$, $\mathrm{IC}(\Theta)$ estime $C_\Theta(x)$

$\longrightarrow$ Adaptation $\sim$ pénalisation.

Codages et inégalité entropique de Rissanen

- Shannon 48 (Jensen) :

$$-\sum_{x\in E} P(x)\log P(x) = H(P) \leq H(P,Q) = -\sum_{x\in E} P(x)\log Q(x)$$

$\longrightarrow$ le terme de vraisemblance dans IC estime $H(P)$.

- Rissanen : complexité stochastique, notion d'*adaptation* :

$$C_\Theta(x) = -\sum_{i=1}^{n} \log \widehat{P}_\Theta^{(i-1)}(x_i),$$

$$H_n(P) + |\Theta|\frac{\log n}{2} \leq \mathbb{E}C_\Theta(x).$$

$\longrightarrow$ $\alpha(n) = \log n,$ IC(Θ) estime $C_\Theta(x)$
$\longrightarrow$ Adaptation $\sim$ pénalisation.

Application : estimation de densité par histogramme

- La minimisation d'un codage en deux temps :

  $\longrightarrow$ Codage des données discrétisées sur la partition
  $\longrightarrow$ Codage de précision sans perte

  amène le critère

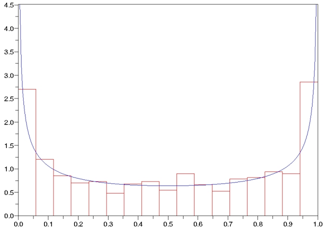
$$\mathrm{CRIT}(\Pi) = -\sum_{j=1}^{m} n_j \log \frac{n_j}{nL_j} + m\frac{\log n}{2}.$$

- On choisit la partition par :

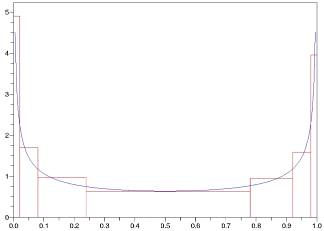
$$\widehat{\Pi} = \mathrm{Argmin}(\mathrm{CRIT}(\Pi),\ \Pi \in \mathcal{P}).$$

Application : estimation de densité par histogramme

$$\mathrm{CRIT}(\Pi) = -\sum_{j=1}^{m} n_j \log \frac{n_j}{nL_j} + m\frac{\log n}{2}.$$



$\mathrm{KL}(H, \beta) = 1,2$

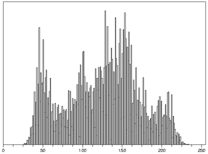


$\mathrm{KL}(H, \beta) = 0,89$

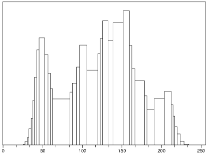
$\longrightarrow$ Application : reconnaissance de loi via Kullback-Leibler sur les histogrammes

Application : estimation de densité par histogramme

- Quantification d'image :



256 classes



39 classes

PSNR = 38,52 dB

$\longrightarrow$ Application aux domaines transformés.

Modèles paramétriques

- $\Theta \subset \mathbb{R}^d$ espace des paramètres.

$$\Theta_S = \left\{\theta \in \Theta | \theta_j = 0, j \notin S\right\}, \; S \subset [\![1, d]\!].$$

- Nishii 88 : si $\theta^\star \in \Theta_{S^\star}$ et $\widehat{S} = \text{Argmin}(\text{IC}(\Theta_S), S \subset [\![1, d]\!])$ alors

$$\frac{\log \log n}{\alpha(n)} \longrightarrow 0 \;\text{ et }\; \frac{\alpha(n)}{n} \longrightarrow 0 \Longrightarrow \widehat{S} \xrightarrow{\text{p.s.}} S^\star$$

- Introduction du critère φ_β par El-Matouat et Hallin 96 :

$$\alpha(n) = n^\beta \log \log n, \;\; \beta \in]0, 1[.$$

Modèles paramétriques

- $\Theta \subset \mathbb{R}^d$ espace des paramètres.

$$\Theta_S = \left\{\theta \in \Theta | \theta_j = 0, j \notin S\right\}, \ S \subset [\![1, d]\!].$$

- Nishii 88 : si $\theta^\star \in \Theta_{S^\star}$ et $\widehat{S} = \text{Argmin}(\text{IC}(\Theta_S), S \subset [\![1, d]\!])$ alors

$$\frac{\log\log n}{\alpha(n)} \longrightarrow 0 \ \text{ et } \ \frac{\alpha(n)}{n} \longrightarrow 0 \Longrightarrow \widehat{S} \xrightarrow{\text{p.s.}} S^\star$$

- Introduction du critère φ_β par El-Matouat et Hallin 96 :

$$\alpha(n) = n^\beta \log\log n, \ \beta \in]0, 1[.$$

Les méthodes comparatives

- Méthode globale : $\widehat{S} = \text{Argmin}(\text{IC}(\Theta_S), S \subset [\![1, d]\!])$:

  $\longrightarrow$ 2^d critères à calculer.
  $\longrightarrow$ Introduction des méthodes comparatives.

- Méthode comparative (Nishii 88)

$$\left\{ \begin{array}{rcl} \text{IC}_{\text{ref}} & = & \text{IC}([\![1, d]\!]) \\ \widehat{S} & = & \left\{ j \in [\![1, d]\!],\ \text{IC}_{\text{ref}} \leq \text{IC}\left([\![1, d]\!] \setminus \{j\}\right) \right\}. \end{array} \right.$$

  $\longrightarrow$ $d + 1$ critères à calculer.

- Méthode comparative inversée

  $\longrightarrow$ $d + 1$ critères à calculer.

- Méthode comparative descendante

  $\longrightarrow$ Au plus $d(d+1)/2$ critères à calculer.

Résultats concernant la régression

- Avec les méthodes
  - comparative,
  - comparative inversée dans le cas orthonormal,
  - comparative inversée adaptée,
  - comparative descendante,

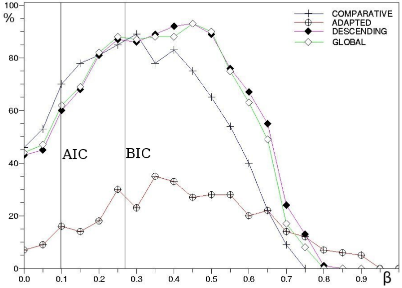
  on obtient, sous les conditions de Nishii :

  $\longrightarrow$ Consistance forte.
  $\longrightarrow$ Equivalence asymptotique du risque avec le risque oracle.

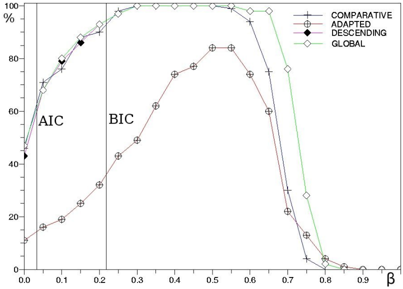
- Lorsque $\alpha(n)$ =cste, la méthode comparative descendante permet l'obtention d'une inégalité oracle non-asymptotique.

Résultats concernant la régression : illustration

- Fonction inconnue : $-x + \cos 2x - \sin 2x$.
- Régression sur Vect$(x, x^2, \cos x, \cos 2x, \sin x, \sin 2x)$.
- Utilisation de φ_β, $\alpha(n) = n^\beta \log\log n$, pour $\beta \in]0,1[$.



$n = 50$



$n = 200$

Conclusions, perspectives

- La pénalité dépend du type de propriété souhaitée
  $\longrightarrow$ Pas de "critère-miracle" mis en évidence
- Les méthodes comparatives permettent une description précise et rapide du modèle :
  $\longrightarrow$ Choix des harmoniques principales d'un signal périodique.
  $\longrightarrow$ Détermination du support d'autorégression d'une image.
- L'estimation par histogramme régulier ou non :
  $\longrightarrow$ Reconnaissance de loi : à partir des densités estimées.
  $\longrightarrow$ Images : quantification des coefficients transformés, élimination de la corrélation résiduelle.

MERCI A TOUS !