

# Inégalités d'Oracle pour l'Estimation de la Régression

Yun CAO

Sous la direction de Yuri Golubev

le 3 avril 2008

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines
- 4 Idées des démonstrations
- 5 Conclusion

# Plan

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines
- 4 Idées des démonstrations
- 5 Conclusion

# Problème d'estimation de la régression

**Modèle statistique :**

$$Y_i = f(X_i) + \epsilon_i \quad i = 1, \dots, n$$

où :

- $Y_i \in \mathbb{R}^1$  sont les observations
- $f : [0, 1] \rightarrow \mathbb{R}^1$  est la fonction de régression à estimer
- $X_i \in [0, 1]$  sont des points d'observation distincts
- Le bruit blanc gaussien :  $\epsilon_1, \dots, \epsilon_n$  sont *i.i.d.*  $\mathcal{N}(0, \sigma^2)$

On va considérer deux cas

- $\sigma^2$  est connue
- $\sigma^2$  est inconnue

# Problème d'estimation de la régression

**Problématique** : Reconstruire la fonction de régression  $f$  à partir des observations  $\mathbf{Y} = (Y_1, \dots, Y_n)^T$ .

**Risque d'un estimateur**  $\hat{f}(x, \mathbf{Y})$  :

$$R(\hat{f}, f) = \frac{1}{n} \mathbf{E}_f \sum_{i=1}^n [f(X_i) - \hat{f}(X_i, \mathbf{Y})]^2 \triangleq \mathbf{E}_f \|f - \hat{f}\|^2$$

**But** : Trouver l'estimateur qui minimise le risque  $R(\hat{f}, f)$ .

# Approche minimax

- Approche minimax

L'idée se base sur l'hypothèse qu'il y a une information *a priori* sur  $f : f \in \mathcal{F}$ . On cherche l'estimateur minimax

$$\hat{f}^* = \arg \min_{\hat{f}} \{ \sup_{f \in \mathcal{F}} R(\hat{f}, f) \}.$$

Par exemple

- $\mathcal{F} = \mathcal{H}(\beta, L)$ —Espace d'Hölder
- $\mathcal{F} = \mathbb{W}_2^m(L)$ —Espace de Sobolev

# Approche d'oracle

- Approche d'oracle

Le but est de trouver un estimateur parmi une famille d'estimateurs  $\mathcal{M}$  qui fonctionne presque comme l'oracle

$$\hat{f}_{or} = \arg \min_{\hat{f} \in \mathcal{M}} R(\hat{f}, f)$$

Par exemple

- En estimation par polynômes :  $\mathcal{M} = \{\hat{f}_q, q = 1, \dots, n\}$

$$\hat{f}_q = \arg \min_{p \in \mathcal{P}^q} \sum_{i=1}^n [Y_i - p(X_i)]^2$$

- En estimation par splines :  $\mathcal{M} = \{\hat{f}_\alpha, \alpha > 0\}$

$$\hat{f}_\alpha = \arg \min_g \left\{ \|Y - g\|^2 + \alpha \int_0^1 [g^{(m)}(u)]^2 du \right\}.$$

# Estimation du risque sans biais

Dans cette thèse, l'approche d'oracle se base sur l'estimation du risque sans biais.

Supposons que l'on a une famille d'estimateurs

$$\mathcal{M} = \{\hat{\mathbf{f}}_\alpha : \hat{\mathbf{f}}_\alpha = S_\alpha \mathbf{Y}\},$$

où  $\hat{\mathbf{f}}_\alpha = (\hat{f}(X_1), \dots, \hat{f}(X_n))^T$  et  $S_\alpha$  est une  $n \times n$ -matrice.

L'estimation du risque sans biais se base sur la formule suivante :

$$\mathbf{E}\|\mathbf{f} - S_\alpha \mathbf{Y}\|^2 = \underbrace{\mathbf{E}\|\mathbf{Y} - S_\alpha \mathbf{Y}\|^2}_{\text{risque empirique}} + \underbrace{\frac{2\sigma^2 \text{tr}(S_\alpha)}{n}}_{\text{pénalité}} - \sigma^2.$$



## Estimation du risque sans biais

Choix d'estimateur est équivalent à la minimisation du risque empirique pénalisé.

Si  $\sigma^2$  est connue, le choix de  $\alpha$  préconisé par la méthode d'Akaike (1973) est

$$\alpha^A = \arg \min_{\alpha} \left\{ \|\mathbf{Y} - S_{\alpha} \mathbf{Y}\|^2 + \frac{2\sigma^2}{n} \text{tr}(S_{\alpha}) \right\}.$$

Dans le cas où  $\sigma^2$  n'est pas connue, on l'estime par :

$$\hat{\sigma}^2 = \|\mathbf{Y} - S_{\alpha} \mathbf{Y}\|^2,$$

d'où on obtient :

$$\tilde{\alpha} = \arg \min_{\alpha} \left\{ \|\mathbf{Y} - S_{\alpha} \mathbf{Y}\|^2 [1 + 2\text{tr}(S_{\alpha})/n] \right\}.$$

# Estimation du risque sans biais

On considère deux autres choix de  $\alpha$  dans cette thèse :

- 
- **Generalized Cross Validation** (P.J.Green & B.W.Silverman (1994)) :

$$\alpha^{GCV} = \arg \min_{\alpha} \frac{\|\mathbf{Y} - S_{\alpha} \mathbf{Y}\|^2}{[1 - \text{tr}(S_{\alpha})/n]^2}$$

en remplaçant le terme  $1 + 2\text{tr}(S_{\alpha})/n$  dans l'expression de  $\tilde{\alpha}$  à l'aide du développement de Taylor.

# Plan

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines
- 4 Idées des démonstrations
- 5 Conclusion

# Estimation par polynômes

Famille d'estimateurs  $\mathcal{M} = \{\hat{f}_q, q = 1, \dots, n, \}$  où :

$$\hat{f}_q = \arg \min_{p \in \mathcal{P}^q} \sum_{i=1}^n [Y_i - p(X_i)]^2,$$

où  $\mathcal{P}^q$  est l'espace des polynômes de degré  $q$  :

$$\mathcal{P}^q = \left\{ p(x) : p(x) = \sum_{k=0}^q a_k x^k, \quad a_k \in \mathbb{R}^1 \right\}.$$

Oracle

$$\hat{f}_{or}^* = \arg \min_{\hat{f}_q \in \mathcal{M}} \mathbf{E}_f \|f(X_i) - \hat{f}_q(X_i)\|^2$$

Si  $\sigma^2$  est connue

La méthode d'Akaike nous préconise le choix suivant :

$$q^A = \arg \min_q \left\{ \|Y_i - \hat{f}_q(X_i)\|^2 + \frac{2\sigma^2 q}{n} \right\}.$$

Pour décrire la performance de cette méthode, on définit le risque d'oracle :

$$R(f) = \min_q R(f, \hat{f}_q).$$

Soit  $U(\cdot)$

$$U(x) = -1 - \frac{\log(1 - 2x)}{2x}, \quad x \in (0, 0.5).$$

Il est facile de voir que  $U(x) = x + O(x^2)$ ,  $x \rightarrow 0$

## Si $\sigma^2$ est connue

### Théorème

On a pour tout  $\gamma \in (0, 1)$  et pour tout  $f$

$$\mathbf{E}_f \|\hat{f}_{q^A} - f\|^2 \leq \frac{R(f)}{1 - \gamma} + \frac{2\sigma^2}{n(1 - \gamma)} \left[ \frac{1}{\gamma} + \frac{1}{U^{-1}(\gamma/2)} \right],$$

### Exemples :

- $R(f) \asymp \sigma^2/n$  Estimation paramétrique.  
 Avec  $\gamma = 1/2$ , on obtient :

$$\mathbf{E}_f \|\hat{f}_{q^A} - f\|^2 \leq \frac{C\sigma^2}{n};$$

où  $C$  est une constante.

- $R(f) \gg \sigma^2/n$  Estimation non-paramétrique. Dans ce cas :

$$\mathbf{E}_f \|\hat{f}_{q^A} - f\|^2 \leq R(f)[1 + o(1)].$$

## Si $\sigma^2$ est inconnue

En pratique,  $\sigma^2$  est rarement connue. C'est la raison pour laquelle on considère les choix de  $q$  suivants :

$$q^B = \arg \min_q \frac{\|Y_i - \hat{f}_q(X_i)\|^2}{[1 - 2q/n]}$$

et

$$q^C = \arg \min_q \frac{\|Y_i - \hat{f}_q(X_i)\|^2}{[1 - q/n]^2} \quad (\text{GCV}).$$

### Théorème

Pour tout  $\gamma \in (0, 1)$  et pour tout  $f$ , soit  $\tilde{q} = q^B$ , soit  $\tilde{q} = q^C$ , on a :

$$\mathbf{E}_f \|\hat{f}_{\tilde{q}} - f\|^2 \leq \frac{1}{1 - \gamma} \left\{ R(f) \left[ 1 + \frac{C_2 R(f)}{\sigma^2} \right] + \frac{2\sigma^2}{n\gamma} + \frac{2\sigma^2}{nU^{-1}(\gamma/2)} + \frac{\sigma^2}{n} \right\}.$$

\*

## Simulation

On mesure la performance de l'estimation du risque sans biais par l'efficacité d'oracle :

$$e_{or}(f, \tilde{q}) = \frac{\inf_q \mathbf{E}_f \|\hat{f}_q - f\|^2}{\mathbf{E}_f \|\hat{f}_{\tilde{q}} - f\|^2},$$

Il est évident que l'on ne peut pas la calculer pour toutes les fonctions  $f$ , c'est pourquoi on utilise une sous famille paramétrisée

$$f_a(X_i) = a \sum_{k=0}^{n-1} \frac{\mu_k \pi_k(X_i)}{1 + (k/W)^m}$$

- $a$  est le rapport signal/bruit ;
- $\mu_k$  sont des v.a. *i.i.d.* qui prennent  $-1$  et  $1$  avec la même probabilité ;
- $\{\pi_k\}$  est un système de polynômes orthogonaux ;
- $W$  est la fenêtre ;
- $m$  est le paramètre de lissage.



# Simulation

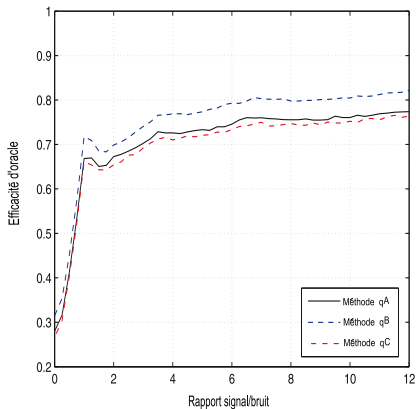


Figure: Efficacité d'oracle de l'estimation du risque sans biais.

## Méthode de la sélection du modèle

En fait, la méthode d'Akaike peut être considérée comme un cas particulier de la méthode de la sélection du modèle (Y.Golubev (2004)).

Cette méthode nous propose le choix de  $q$  suivant

$$q^K = \arg \min_q \left\{ \|Y_i - \hat{f}_q(X_i)\|^2 + \frac{Kq\sigma^2}{n} \right\}, \quad K > 1.$$

Le risque de cette méthode est contrôlé par

$$R^K(f) = \min_q \left\{ \mathbf{E}_f \|f(X_i) - \hat{f}_q(X_i)\|^2 + \frac{(K-2)q\sigma^2}{n} \right\}.$$

## Méthode de la sélection du modèle

### Théorème

Uniformément pour tout  $f$ , on a

- si  $1 < K < 2$  et pour tout  $\gamma \in (0, K - 1)$

$$\mathbf{E}_f \|\hat{f}_{q^K} - f\|^2 \leq \frac{1 + \gamma}{K - 1 - \gamma} \left[ R^K(f) + \frac{2(1 + \gamma)\sigma^2}{n(2 - K + 2\gamma)} + \frac{K\sigma^2}{nU^{-1}(\gamma)} \right].$$

- si  $K \geq 2$  et pour tout  $\gamma \in (0, 0.5)$

$$\mathbf{E}_f \|\hat{f}_{q^K} - f\|^2 \leq (1 + 2\gamma) \left[ R^K(f) + \frac{(1 + \gamma)\sigma^2}{n\gamma} + \frac{2\sigma^2}{nU^{-1}(\gamma)} \right],$$

\*

## Combinaisons convexes des estimateurs par projection

On a un estimateur

$$\hat{f}_q = S(q)\mathbf{Y}$$

et on définit

$$\hat{f}_T = \sum_{k=1}^M \lambda_k S(q_k)\mathbf{Y} \triangleq T\mathbf{Y}$$

où  $S(q_k)$  sont des  $n \times n$ -matrices de projection sur l'espace engendré par les polynômes de degré  $q_k$  et

$$\sum_{k=1}^M \tau_k = 1.$$

# Combinaisons convexes des estimateurs par projection

Supposons que l'on a la famille d'estimateurs linéaires

$$\mathcal{M} = \{\hat{f}_T : \hat{f}_T = T\mathbf{Y}, T \in \mathcal{T}_M\},$$

où l'ensemble  $\mathcal{T}_M$  est défini comme suit :

$$\mathcal{T}_M = \left\{ T : T = \sum_{k=1}^M \tau_k S(q_k), \text{ tels que } \right. \\ \left. \tau_k \geq 0, \sum_{k=1}^M \tau_k = 1 \text{ et } q_k \in \{1, \dots, n\} \right\}.$$

# Combinaisons convexes des estimateurs par projection

Dans ce cas, le choix de  $T$  basé sur l'estimation du risque sans biais est

$$\hat{T} = \arg \min_{T \in \mathcal{T}_M} \left\{ \|\mathbf{Y} - T\mathbf{Y}\|^2 + \frac{2\sigma^2 \text{tr}(T)}{n} \right\}.$$

Le théorème suivant contrôle le risque de cette méthode :

## Théorème

On a pour tout  $\gamma \in (0, 1)$  et pour tout  $f$  :

$$\mathbf{E}_f \|\hat{f}_{\hat{T}} - f\|^2 \leq \frac{1}{1 - \gamma} \left\{ \min_{T \in \mathcal{T}_M} R(\hat{f}_T, f) + 2\sigma^2 \left[ \frac{M}{\gamma} + \frac{M}{U^{-1}(\gamma/2)} \right] \right\}.$$

# Simulation

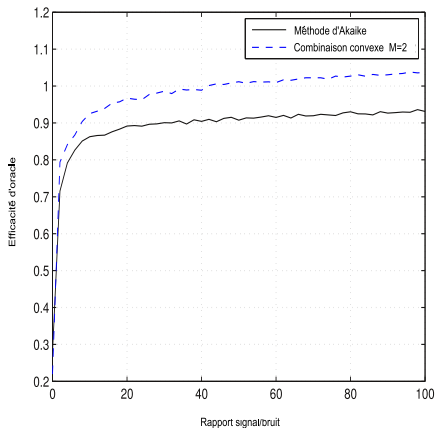


Figure: Efficacité d'oracle de combinaison convexe de 2 estimateurs par projection.

# Plan

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines**
- 4 Idées des démonstrations
- 5 Conclusion



## Estimation par splines

En estimation par splines, la famille d'estimateurs est

$$\mathcal{M} = \{\hat{f}_\alpha, \alpha > 0\}$$

où

$$\hat{f}_\alpha = \arg \min_g \left\{ \|Y - g\|^2 + \alpha \int_0^1 [g^{(m)}(u)]^2 du \right\}.$$

### Avantages des splines :

- On peut calculer les splines facilement. Grâce à l'algorithme de Reinsch, une spline se calcule avec  $m \times n$  opérations (P.J.Green & B.W.Silverman (1994))
- Les splines donnent de bonnes approximations des fonctions lisses.

Dans ce cas, l'oracle est :

$$\hat{f}_{or}^* = \arg \min_{\hat{f}_\alpha \in \mathcal{M}} \mathbf{E}_f \|f(X_i) - \hat{f}_\alpha(X_i)\|^2$$

## Si $\sigma^2$ est connue

La méthode d'Akaike nous propose le choix suivant :

$$\hat{\alpha} = \arg \min_{\alpha > 0} \left\{ \|\mathbf{Y} - S_{\alpha} \mathbf{Y}\|^2 + \frac{2\sigma^2 \text{tr}(S_{\alpha})}{n} \right\}.$$

où  $S_{\alpha}$  est la  $n \times n$ -matrice qui correspond à  $\hat{f}_{\alpha}$ .

On signale qu'il y a un algorithme qui nous permet de trouver  $\hat{\alpha}$  sans calculer  $S_{\alpha}$ . (P.J.Green & B.W.Silverman (1994))

### Théorème

On a pour tout  $\gamma \in (0, 1)$  et tout  $f$

$$\mathbf{E}_f \|\hat{f}_{\hat{\alpha}} - f\|^2 \leq \frac{1}{1 - \gamma} \left\{ R(f) + \frac{C\sigma^2}{n\gamma} \right\},$$

où le risque d'oracle  $R(f) = \min_{\alpha > 0} R(\hat{f}_{\alpha}, f)$ .

## Si $\sigma^2$ est inconnue

La méthode de **GCV modifiée** nous donne

$$\hat{\alpha}_G = \arg \min_{\alpha > 0: \text{tr}(S_\alpha) \leq \sqrt{n}} \frac{\|\mathbf{Y} - S_\alpha \mathbf{Y}\|^2}{[1 - \text{tr}(S_\alpha)/n]^2}.$$

Risque d'oracle modifié :

$$\tilde{R}(f) = \inf_{\alpha > 0: \text{tr}(S_\alpha) \leq \sqrt{n}} R(\hat{f}_\alpha, f)$$

### Théorème

On a pour tout  $\gamma \in (0, 1)$  et tout  $f$  :

$$\mathbf{E}_f \|\hat{f}_{\hat{\alpha}_G} - f\|^2 \leq \frac{1}{1 - \gamma} \left\{ \left[ 1 - \frac{1}{\sqrt{n}} \right]^{-2} \tilde{R}(f) + \frac{C_2 \sigma^2}{n\gamma} \right\}.$$

\*

# Plan

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines
- 4 Idées des démonstrations
- 5 Conclusion

## Transformation en modèle "sequence space"

Afin de démontrer nos résultats, on utilise une transformation linéaire des données en modèle **sequence space**. Soient  $\{\varphi_k\}$  une base orthonormée et

$$Z_k = \langle \mathbf{Y}, \varphi_k \rangle.$$

Il est facile de voir que (**sequence space**)

$$Z_k = \theta_k + \varepsilon \xi_k, \quad k = 1, \dots, n,$$

qui est équivalent au modèle  $Y_i = f(X_i) + \epsilon_i$ ,

où

- $\theta_k = \langle \mathbf{f}, \varphi_k \rangle$  ;
- $\xi_k$  sont supposés *i.i.d.*  $\mathcal{N}(0, 1)$  ;
- $\varepsilon = \sigma / \sqrt{n}$ .

## Transformation en modèle "sequence space"

Pour obtenir une représentation 'creuse' de  $f$ , on utilise des bases  $\{\varphi_k\}$  spéciales.

En estimation par polynômes, on utilise un système des polynômes orthogonaux  $\{\varphi_k\}$  associé au design  $X_i, i = 1, \dots, n$ .

$$\frac{1}{n} \sum_{i=1}^n \varphi_k(X_i) \varphi_l(X_i) = \delta_{lk},$$

où  $\varphi_k$  est un polynôme de degré  $k$  ;

## Transformation en modèle "sequence space"

En estimation par splines, on utilise la **base de Reinsch-Demmler** (A.Kneip (1994)) définie par :

•

$$\frac{1}{n} \sum_{i=1}^n \varphi_k(X_i) \varphi_j(X_i) = \delta_{kj},$$

•

$$\int_0^1 \varphi_i^{(m)}(x) \varphi_j^{(m)}(x) dx = \lambda_i \delta_{ij},$$

où  $0 = \lambda_1 = \dots = \lambda_{m-1} < \lambda_m < \dots < \lambda_n$ .

## Transformation en modèle "sequence space"

Cette base est l'ensemble des diagonales de l'ellipsoïde de Sobolev  $\mathbb{W}_2^m$  qui nous permet de calculer le **diamètre de Kolmogorov**  $\lambda_{N+1}^{-1}$ .

### Théorème

$\{\varphi_k(x), k = 1, \dots, n\}$  est la base de Reinsch-Demmler.  $\phi$  est une base orthonormée quelconque. Soit  $D^N(\phi)$  l'erreur maximale d'approximation des fonctions dans  $\mathbb{W}_2^m$  telle que :

$$D^N(\phi) = \sup_{f \in \mathbb{W}_2^m} \frac{1}{n} \sum_{i=1}^n \left[ f(X_i) - \sum_{k=1}^N \langle f, \phi_k \rangle \phi_k(X_i) \right]^2,$$

alors :

$$D^N(\varphi) = \inf_{\phi} D^N(\phi) = \lambda_{N+1}^{-1}.$$



## Transformation en modèle "sequence space"

Avec ces bases, les estimateurs s'écrivent comme suit :

$$\hat{\theta}_k = h_k Z_k, \quad k = 1, \dots, n,$$

où

- en estimation par polynômes

$$h_k \in \mathcal{H}_{pol} = \left\{ h_k : h_k = \mathbf{1}(k \leq q), \quad 0 < q \leq n-1 \right\}$$

- en estimation par splines

$$h_k \in \mathcal{H}_{spl} = \left\{ h_k : h_k(\alpha) = [1 + \alpha \lambda_k]^{-1}, \quad \alpha > 0 \right\}$$

Le meilleur choix de  $h$  est

$$h^* = \arg \min_{h \in \mathcal{H}} \left\{ \|hZ - Z\|^2 + 2\varepsilon^2 \|h\|_1 \right\}, \quad \text{où } \|h^*\|_1 = \sum_{k=0}^{n-1} h_k^*.$$

## Décomposition du risque

L'idée principale est de contrôler  $R(\theta, h^*)$  en se basant sur

$$R(\theta, h^*) = \mathbf{E}_\theta \|h^* Z - \theta\|^2 = \mathbf{E}_\theta \underbrace{\left\{ \|h^* Z - Z\|^2 + 2\varepsilon^2 \|h^*\|_1 - \varepsilon^2 \|\xi\|^2 \right\}}_I$$

$$+ \underbrace{2\varepsilon \mathbf{E}_\theta \langle \theta - h^* \theta, \xi \rangle}_{II} + \underbrace{2\varepsilon^2 \mathbf{E}_\theta \langle h^* \xi, \xi \rangle - 2\varepsilon^2 \|h^*\|_1}_{III}$$

La motivation de cette décomposition est la suivante

- $I$  est facile d'être contrôlé car  $h^*$  le minimise
- $II$  et  $III$  sont des termes petits

## Contrôle de $I$

Il est évident que

$$\mathbf{E}_\theta \left\{ \|h^* Z - Z\|^2 + 2\varepsilon^2 \|h^*\|_1 \right\} \leq \mathbf{E}_\theta \left\{ \|hZ - Z\|^2 + 2\varepsilon^2 \|h\|_1 \right\}$$

est vrai pour tout  $h$  qui ne dépend pas de données. D'où on obtient

$$\mathbf{E}_\theta \|h^* Z - Z\|^2 + 2\varepsilon^2 \|h^*\|_1 - \varepsilon^2 \|\xi\|^2 \leq R(\theta, h).$$

## Contrôle de $II$ et de $III$

Afin de majorer  $\mathbf{E}_\theta \langle \theta - h^*\theta, \xi \rangle$  et  $\mathbf{E}_\theta \langle h^*\xi, \xi \rangle - \|h^*\|_1$ , on utilise :

$$|\mathbf{E}_\theta \langle \theta - h\theta, \xi \rangle| \leq C \sqrt{\mathbf{E}_\theta \|\theta - h\theta\|^2},$$

$$|\mathbf{E}_\theta \langle h\xi, \xi \rangle - \|h\|_1| \leq C \sqrt{\mathbf{E}_\theta \|h\|^2},$$

qui sont vrais  $\forall h \in \mathcal{H}$  et qui peuvent être démontrées à partir de :

$$\mathbf{E} \sup_{h \in \mathcal{H}} \left[ \sum_{k=0}^{n-1} (1 - h_k) \theta_k \xi_k - \gamma \sum_{k=0}^{n-1} (1 - h_k)^2 \theta_k^2 \right]_+ \leq \frac{C}{\gamma}, \quad (1)$$

$$\mathbf{E} \sup_{h \in \mathcal{H}} \left[ \sum_{k=0}^{n-1} h_k (\xi_k^2 - 1) - \gamma \sum_{k=0}^{n-1} h_k^2 \right]_+ \leq \frac{C}{\gamma}, \quad (2)$$

où  $\mathcal{H}$  soit  $\mathcal{H}_{pol}$  soit  $\mathcal{H}_{spl}$ .

Estimation par polynômes ( $\mathcal{H} = \mathcal{H}_{pol}$ )

Pour démontrer (1), notons que la distribution de  $\sum_{k=0}^{n-1} (1 - h_k) \theta_k \xi_k$  coïncide avec celle de  $W(\sum_{k=0}^{n-1} [1 - h_k]^2 \theta_k^2)$ , où  $W(\cdot)$  est un processus de Wiener. On en déduit de l'inégalité de Doob que :

$$\mathbf{E}\left\{\max_{t \geq 0} [W(t) - \gamma t]\right\} \leq \frac{1}{2\gamma}.$$

ce qui est équivalente à (1).

Pour vérifier (2), on applique l'inégalité de Doob sur la martingale

$$\eta_k = \exp \left[ \alpha \sum_{s=0}^{k-1} (\xi_s^2 - 1) - \alpha U(\gamma) k \right]$$

et on obtient :

$$\mathbf{E}\left\{\max_k \left[ \sum_{s=0}^{k-1} (\xi_s^2 - 1) - U(\gamma) k \right]\right\} \leq \frac{1}{2\gamma}$$

## Estimation par splines ( $\mathcal{H} = \mathcal{H}_{spl}$ )

En estimation par splines, on ne peut pas appliquer l'approche précédente puisque  $\sum_{k=0}^{n-1} (1 - h_k) \theta_k \xi_k$  n'est plus une martingale. C'est pourquoi on introduit **les processus ordonnés**.

### Définition

Le processus séparable  $\xi(t)$  est un *processus ordonné* si

- 

$$\mathbf{E}\xi(t_1)\xi(t_2) \geq \min\{\mathbf{E}\xi^2(t_1), \mathbf{E}\xi^2(t_2)\};$$

- il existe  $\lambda > 0$  tel que

$$\sup_{t_1, t_2} \mathbf{E} \cosh \left\{ \lambda \frac{\xi(t_1) - \xi(t_2)}{\sqrt{\mathbf{E}[\xi(t_1) - \xi(t_2)]^2}} \right\} < \infty.$$

## Estimation par splines ( $\mathcal{H} = \mathcal{H}_{spl}$ )

La propriété principale des processus ordonnés :

### Théorème

*Si  $\xi(t)$  est un processus ordonné, alors*

$$\mathbf{E} \sup_{t \geq 0} [\xi(t) - \gamma \mathbf{E} \xi(t)]_+^p \leq \frac{C(p)}{\gamma^p}.$$

Afin d'appliquer ce théorème, il suffit de vérifier que

$$\xi_1(h) = \sum_{k=0}^{n-1} (1 - h_k) \theta_k \xi_k \quad \text{et} \quad \xi_2(h) = \sum_{k=0}^{n-1} h_k (\xi_k^2 - 1), \quad h \in \mathcal{H}_{spl}$$

sont ordonnés afin de démontrer (1) et (2).

# Plan

- 1 Introduction
- 2 Estimation par polynômes
- 3 Estimation par splines
- 4 Idées des démonstrations
- 5 Conclusion



## Conclusion

- 1 Dans le cas où la variance est connue : l'expression exacte de la constante devant le risque d'oracle permet de savoir quand et pourquoi le critère d'Akaike minimise bien le risque d'oracle et donc fonctionne bien.
- 2 Dans le cas où la variance n'est pas connue : on obtient des inégalités d'oracle en changeant la constante devant le risque d'oracle.
- 3 Introduction des processus ordonnés et de leur propriété.