Projet 1 : Estimation d'une fonction de régression.

En statistiques, la prevision ou l'explication d'une variable Y par une variable X est fondamental. On sait que la meilleure prevision de Y en fonction de $X = x$ (au sens L_2) est l'esperance conditionnelle ($\mathbb{E}[Y/X=x] = f(x)$). On a alors un modele $Y = f(X) + \varepsilon$ avec $\mathbb{E}[\varepsilon/X] = 0$. Il faut alors pouvoir estimer la fonction de regression f à partir de n copies independantes $(X_i, Y_i)_{i=1..n}$. Cette année, nous avons d'abord vu le modèle linéaire où on suppose que f est de la forme $f(x) = x^T\beta$ $(x \in \mathbb{R}^d)$ et alors il suffit d'estimer le paramètre fini dimensionnel β pour estimer f. Cependant, en réalité souvent on ne peut pas faire cette hypothèse sur f au même on ne peut pas faire l'hypothèse que f a une forme paramétrique où il suffirait juste d'estimer un paramètre fini - dimensionnel. On se retrouve alors dans un problème non paramétrique.

Le but de ce projet va être d'étudier un problème de ce type. Nous allons étudier l'estimateur par projection. On se place dans le cadre où X suit une loi uniforme sur $[0,1]^d$ et on suppose que $f \in L^2([0,1]^d)$. On va alors se ramener au cas du modèle linéaire avec un paramètre fini dimensionnel à estimer. En effet, avec une base orthonormée de $L^2([0,1]^d)$ $(\varphi_i)_{i\geq 1}$ on peut décomposer f dans une base c'est à dire $f = \sum_{j\geq 1} \theta_j \varphi_j$

avec $\sum_{j=1}^{N} v_j \varphi_j \xrightarrow[N \to +\infty]{L^2} f$, $v_j = \langle f, \varphi_j \rangle_{L^2([0,1]^d)}$ vérifiant

$\sum_{j=1}^{+\infty} v_j^2 < +\infty$ (et donc $v_j \xrightarrow[j \to +\infty]{} 0$).

Ainsi on va approcher f par $f_N = \sum_{j=1}^{N} v_j \varphi_j$ et obtenir un modèle semblable au modèle linéaire :

$Y = f_N(X) + \varepsilon$ (modèle approché) avec le paramètre $(v_1, \dots v_N)^T = v$ qui joue le rôle de β à estimer.

Cependant, il se rajoute maintenant le problème du choix de l'hyperparamètre N (le niveau de troncature).

Nous allons donc dans ce projet voir comment on pourrait faire ce choix de façon à avoir le meilleur estimateur, c'est à dire un N qui donnera un bon équilibre biais - variance

I) Estimateur par projection.

a) l'estimateur des moindres carrés $\hat{\vartheta}_{m,N}$ du paramètre ϑ dans le modèle approché $Y_i = f_{N,\vartheta}(x_i) + \upsilon_i$

avec $f_{N,\vartheta}(x) = \sum_{j=1}^{N} \vartheta_j \varphi_j(x)$ $\forall x \in \mathbb{R}^d$, respond

$$\hat{\vartheta}_{m,N} = \underset{\vartheta \in \mathbb{R}^N}{\operatorname{argmin}} \sum_{i=1}^{N} \left(Y_i - \sum_{j=1}^{N} \vartheta_j \varphi_j(x_i)\right)^2$$

on peut écrire Y sous la forme matricielle suivante

$Y = \Phi_N \vartheta + \Sigma_i$. On retrouve donc un modèle classique de régression linéaire.

Dans ce cas $\hat{\vartheta}_{m,N} = \underset{\vartheta \in \mathbb{R}^N}{\operatorname{argmin}} (Y - \Phi_N \vartheta)^T (Y - \Phi_N \vartheta)$

avec $\vartheta = (\vartheta_1 \dots \vartheta_N)^T$.

On appelle $S(\vartheta) = (Y - \Phi_N \vartheta)^T (Y - \Phi_N \vartheta)$

$\nabla S(\vartheta) = -2\Phi_N^T (Y - \Phi_N \vartheta)$

$\nabla^2 S(\vartheta) = 2\Phi_N^T \Phi_N$ définie positive par l'énoncé d'où S est convexe par rapport à ϑ. Dans ce cas là

$\hat{\vartheta}$ vérifie $\nabla_\vartheta S(\hat{\vartheta}) = 0$

$\Rightarrow \Phi_N^T Y = \Phi_N^T \Phi_N \hat{\vartheta}$.

Étant donnée que $\Phi_N^T \Phi_N$ est une matrice définie strictement positive, alors $\Phi_N^T \Phi_N$ est inversible d'où

$\hat{\vartheta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y$.

On sait par ailleurs que $f_{N,\vartheta}(x) = \varphi'(x) \vartheta$

avec $\vartheta = (\vartheta_1 \dots \vartheta_N)^T$ et $\varphi'(x) = (\varphi_1(x), \dots \varphi_N(x))$

alors $\hat{f}_{m,N}(x) = \varphi'(x)\,\hat{v} = \varphi'(x)(\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y$

$$= \sum_{j=1}^{N'} \varphi_j(x)\,\hat{v}_j$$

b) Comme $\hat{f}_{m,N}(X_1) = \varphi'(X_1)\,\hat{v}$

$$= (\varphi_1(X_1), \dots \varphi_N(X_1))(\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y$$

alors $\hat{f}_{m,N}(X_i) = (\varphi_1(X_i), \dots \varphi_N(X_i))(\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y.$

$$(\hat{f}_{m,N}(X_1), \dots \hat{f}_{m,N}(X_n))' = \begin{pmatrix} \varphi_1(X_1) & \dots & \varphi_N(X_1) \\ & & \\ \varphi_1(X_n) & \dots & \varphi_N(X_n) \end{pmatrix}(\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y$$

d'où $(\hat{f}_{m,N}(X_1), \dots \hat{f}_{m,N}(X_n))' = \Phi_N(\Phi_N^T\Phi_N)^{-1}\Phi_N^T Y.$

$$= A_N Y.$$

Montrons maintenant que $A_N = \Phi_N(\Phi_N^T\Phi_N)^{-1}\Phi_N^T$ est un projecteur orthogonal sur le sous espace vectoriel de $\mathbb{R}^n$ engendré par les colonnes de la matrice Φ_N.

Premierement verifions que $A_N^T = A_N$.

$(A_N)^T = (\Phi_N(\Phi_N^T\Phi_N)^{-1}\Phi_N^T)^T$

$$= \Phi_N^T((\Phi_N^T\Phi_N)^{-1})^T\Phi_N^T$$

Comme $\Phi_N^T\Phi_N$ est symetrique $(\cdot\,(\Phi_N^T\Phi_N)^T = \Phi_N^T\Phi_N)$ son inverse l'est aussi donc $((\Phi_N^T\Phi_N)^{-1})^T = (\Phi_N^T\Phi_N)^{-1}$

d'où le fait que $(A_N)^T = A_N$.

Par la suite vérifions que l'on a $A_N^2 = A_N$.

En effet $A_N^2 = A_N \times A_N$

$$= \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T \times \Phi_N \times (\Phi_N^T \Phi_N)^{-1} \times \Phi_N^T$$

$$= \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T = A_N.$$

On vient donc de montrer que A_N est un projecteur orthogonal.

Montrons maintenant qu'il projette bien sur ~~le sous~~ espace vectoriel de $\mathbb{R}^n$ engendré par les colonnes de Φ_N.

Soit $Z \in \mathbb{R}^n$, $A_N Z = \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Z$

on appelle $(\Phi_N^T \Phi_N)^{-1} \Phi_N^T Z = \tilde{Z}$ avec $\tilde{Z} \in \mathbb{R}^N$. On a donc

$A_N Z = \Phi_N \tilde{Z} \in \mathrm{Im}(\Phi_N)$ le sous espace ~~vectoriel~~

de $\mathbb{R}^n$ engendré par les colonnes de la matrice Φ_N.

Ainsi A_N projette orthogonalement sur un sous espace

de $\mathrm{Im}(\Phi_N)$.

Si $Z \in \mathrm{Im}(\Phi_N)$, $Z = \Phi_N W$ avec $W \in \mathbb{R}^N$ alors

$A_N Z = A_N \Phi_N W = \Phi_N (\Phi_N^T \Phi_N)^{-1} \Phi_N^T \Phi_N W = \Phi_N W$.

Donc A_N projette orthogonalement sur $\mathrm{Im}(\Phi_N)$ tout entier.

c)

Soit $C = \Phi_N^T \Phi_N$ avec $\Phi_N = (\Phi_{ij})_{\substack{1 \le i \le n \\ 1 \le j \le N}}$

avec $\Phi_{ij} = \varphi_j(X_i)$.

Dans ce cas $C = (C_{ij})_{\substack{1 \le i \le N \\ 1 \le j \le N}}$ avec

$$C_{ij} = \sum_{k=1}^{n} \Phi_{ki} \Phi_{kj} = \sum_{k=1}^{n} \varphi_i(X_k) \varphi_j(X_k)$$

alors si $i \neq j$ on a

$$\frac{C_{ij}}{n} = \sum_{q=1}^{n} \frac{\varphi_j(X_q)\,\varphi_i(X_q)}{n}.$$

Par la loi forte des grands nombres ($(X_i)_{i=1\dots n}$ iid de loi uniforme] avec $\mathbb{E}[X_i] = \frac{1}{2} < +\infty$)

$$\frac{C_{ij}}{n} \xrightarrow[n \to +\infty]{\mathbb{P}\text{-p.s}} \mathbb{E}[\varphi_i(X_q)\,\varphi_j(X_q)]$$

avec $\mathbb{E}[\varphi_i(X_q)\varphi_j(X_q)] = \int_0^1 \varphi_i(x)\varphi_j(x)\,dx = \langle \varphi_i, \varphi_j \rangle$

et $\langle \varphi_i, \varphi_j \rangle = 0$ car $i \neq j$ et $(\varphi_i)_{i \geq 1}$ est une base orthonormée.

d'où $\frac{1}{n} C_{ij} \xrightarrow[n \to +\infty]{\mathbb{P}\text{-p.s}} 0$ si $i \neq j$.

si $i = j$ $\frac{C_{ij}}{n} = \sum_{q=1}^{n} \frac{\varphi_i^2(X_q)}{n}$

$\frac{C_{ij}}{n} \xrightarrow[n \to +\infty]{\mathbb{P}\text{-p.s}} \mathbb{E}[\varphi_i(X_q)^2]$ d'après la loi forte des grands nombres et on a

$$\mathbb{E}[(\varphi_i(X_q))^2] = \int_0^1 (\varphi_i(x))^2\,dx = \|\varphi_i\|^2 = 1.$$

car $(\varphi_i)_{i \geq 1}$ est une base orthonormée.

Dans ce cas là $\frac{C_{ij}}{n} \xrightarrow[n \to +\infty]{\mathbb{P}\text{-p.s}} 1$

Comme on a convergence des coordonnées diagonaux vers 1 et des termes croisés vers 0, la matrice $\frac{\Phi_N^T \Phi_N}{n}$ converge vers la matrice identité de taille $N \times N$.

De plus comme $\frac{\Phi_N^T \Phi_N}{n} \xrightarrow[n \to +\infty]{} I_{N \times N}$

alors $\Phi_N^T \Phi_N \overset{+\infty}{\sim} n I_{N \times N}$.

Etant donnée que $\hat{f}_{n,N}(x) = \varphi'(x)\hat{\vartheta}$

$$= \sum_{j=1}^{N} \varphi_j(x) \hat{\vartheta}_j$$

et $\hat{\vartheta} = (\Phi_N^T \Phi_N)^{-1} \Phi_N^T Y$

$$\tilde{\vartheta} \overset{(*)}{=} (n I_{N \times N})^{-1} \Phi_N^T Y$$

$$\tilde{\vartheta} = \frac{1}{n} \Phi_N^T Y$$

avec $\tilde{\vartheta}_j = \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) Y_i$

On obtient l'estimateur $\tilde{f}_{n,N}(x) = \varphi'(x)\tilde{\vartheta} = \sum_{j=1}^{N} \varphi_j(x) \tilde{\vartheta}_j$ et on appelle $\tilde{\vartheta}$ la nouvelle expression

(*) on remplace $\Phi_N^T \Phi_N$ par son approximation

d) Montrons que $\tilde{\vartheta}_j = \frac{1}{n} \sum_{i=1}^{n} \varphi_j(X_i) Y_i$ est l'estimateur par la methode des moments du paramètre ϑ_j.

$$Y_i = \sum_{\ell=1}^{N} \vartheta_\ell \varphi_\ell(X_i) + U_i$$

donc $\mathbb{E}[Y_i \varphi_j(X_i)] = \sum_{\ell=1}^{N} \mathbb{E}[\varphi_\ell(X_i) \vartheta_\ell \varphi_j(X_i)]$

$+ \mathbb{E}[\varphi_j(X_i) U_i]$.

Comme $X_i \perp\!\!\!\perp U_i$ on a $\mathbb{E}[\varphi_j(X_i) U_i] = 0$ car $\mathbb{E}[U_i] = 0$.

Dans ce cas $\mathbb{E}[Y_i \varphi_j(X_i)] = \sum_{\ell=1}^{N}{}' v_\ell \, \mathbb{E}[\varphi_\ell(X_i)\varphi_j(X_i)]$

et $\mathbb{E}[\varphi_\ell(X_i)\varphi_j(X_i)] = \int_0^1 \varphi_\ell(x)\varphi_j(x) = \langle \varphi_\ell, \varphi_j \rangle$
$= \mathbb{1}_{\{\ell = j\}}$

du coup $\mathbb{E}[Y_i \varphi_j(X_i)] = v_j$.

On considère la variable $Z_i = Y_i \varphi_j(X_i)$
$Z_i = g(X_i, Y_i)$

Puisque $\{(X_i, Y_i)_{1 \le i \le n}\}$ sont n copies indépendantes, alors les $(Z_i)_{1 \le i \le n}$ sont indépendants et $\mathbb{E}[Z_i] = v_j$.

Par la méthode des moments, l'estimateur $\tilde{v}_j$ de v_j est : $\tilde{v}_j = \frac{1}{n}\sum_{i=1}^{n}{}' Z_i$
$= \frac{1}{n}\sum_{i=1}^{n}{}' Y_i \varphi_j(X_i)$

e) $b_{n,N}(x) = \mathbb{E}[\tilde{f}_{n,N}(x)] - f(x)$

$= \mathbb{E}\left[\sum_{j=1}^{N}{}' \tilde{v}_j \varphi_j(x)\right] - \sum_{j=1}^{+\infty}{}' v_j \varphi_j$

$= \sum_{j=1}^{N}{}' \mathbb{E}[\tilde{v}_j]\varphi_j(x) - \sum_{j=1}^{+\infty}{}' v_j \varphi_j$

$= \sum_{j=1}^{N}{}' v_j \varphi_j(x) - \sum_{j=1}^{+\infty}{}' v_j \varphi_j$ car $\mathbb{E}[\tilde{v}_j] = v_j$

$$b_{n,N}(x) = -\sum_{j=N+1}^{+\infty}{}' v_j \varphi_j.$$

On sait que $f = \sum_{j=1}^{+\infty}{}' v_j \varphi_j$, alors la serie $\sum_{j=1}^{+N}{}' v_j \varphi_j$

est convergente dans L^2, du coup.

$$-\sum_{j=1}^{N+1}{}' v_j \varphi_j = \sum_{j=1}^{N}{}' v_j \varphi_j - \sum_{j=1}^{+\infty}{}' v_j \varphi_j$$

$$= \sum_{j=1}^{N}{}' v_j \varphi_j - f.$$

$$\| - \sum_{j=N+1}^{+\infty}{}' v_j \varphi_j \|_{L^2} = \| \sum_{j=1}^{N}{}' v_j \varphi_j - f \|_{L^2}$$

alors si N est grand on a $\| \sum_{j=1}^{N}{}' v_j \varphi_j - f \|_{L^2} \xrightarrow[N \to +\infty]{} 0.$

En conclusion le biais est $b_{n,N}(x) = -\sum_{j=N+1}^{+\infty}{}' v_j \varphi_j$

et lorsque N est grand ce biais converge vers 0

dans L^2.

f) $$\|\tilde{f}_{n,N} - f\|^2 = \| \sum_{j=1}^{N}{}' \tilde{v}_j \varphi_j - \sum_{j=1}^{+\infty}{}' v_j \varphi_j \|^2$$

$$= \| \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j) \varphi_j - \sum_{j=N+1}^{+\infty}{}' v_j \varphi_j \|^2$$

$$= \| \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j) \varphi_j \|^2 - 2 < \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j) \varphi_j , \sum_{j=N+1}^{+\infty}{}' v_j \varphi_j >$$

$$+ \| \sum_{j=N+1}^{+\infty}{}' v_j \varphi_j \|^2$$

Or $\| \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j) \varphi_j \|^2 = \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j)^2$

car $\|\varphi_j\|^2 = 1$ et que $\langle \varphi_i, \varphi_j \rangle = \delta_{ij}$

et $\| \sum_{j=N+1}^{+\infty}{}' v_j \varphi_j \|^2 = \sum_{j=N+1}^{+\infty}{}' v_j^2$ pour les mêmes raisons et par continuité de la norme

De plus $\langle \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j) \varphi_j , \sum_{j'=N+1}^{+\infty}{}' (\tilde{v}_{j'} - v_{j'}) \varphi_{j'} \rangle$

$$= \sum_{j=1}^{N}{}' \sum_{j'=N+1}^{+\infty}{}' (\tilde{v}_j - v_j)(\tilde{v}_{j'} - v_{j'}) \langle \varphi_j, \varphi_{j'} \rangle$$

NB: la somme infinie sort du p.s par continuité du p.s

et comme $\langle \varphi_i, \varphi_j \rangle = \delta_{ij}$, ce produit scalaire

est nul.

On a donc

$$\| \tilde{f}_{m,N} - f \|^2 = \sum_{j=1}^{N}{}' (\tilde{v}_j - v_j)^2 + \sum_{j=N+1}^{+\infty}{}' v_j^2$$

On a dans ce cas

$$\mathbb{E}[\| \tilde{f}_{m,N} - f \|^2] = \mathbb{E}\left(\sum_{j=1}^{N}{}' (\tilde{v}_j - v_j)^2 \right) + \sum_{j=N+1}^{+\infty}{}' v_j^2$$

$$\tilde{v}_j - v_j = \frac{1}{m} \sum_{i=1}^{N}{}' y_i \varphi_j(x_i) - \frac{v_j \times n}{m}$$

$$= \frac{1}{m} \left[\sum_{i=1}^{n}{}' (y_i \varphi_j(x_i) - v_j) \right]$$

du coup $(\tilde{v}_j - v_j)^2 = \frac{1}{m^2} \left[\sum_{i=1}^{n}{}' (y_i \varphi_j(x_i) - v_j) \right]^2$

$$= \frac{1}{n^2} \sum_{i=1}^{n}{}' (y_i \varphi_j(x_i) - v_j)^2 + \frac{2}{m^2} \sum_{\substack{i,\ell \\ i \neq \ell}}{}' (y_i \varphi_j(x_i) - v_j)(y_\ell \varphi_j(x_\ell) - v_j)$$

Dans ce cas

$$\mathbb{E}[(\tilde{v}_j - v_j)^2] = \frac{1}{n^2} \underbrace{\mathbb{E}\Big[\sum_{i=1}^{n} (Y_i\varphi_j(X_i) - v_j)^2\Big]}_{a}$$

$$+ \frac{1}{n^2} \times 2 \sum_{\substack{i,\ell \\ i\neq \ell}} \underbrace{\mathbb{E}[(Y_i\varphi_j(X_i)-v_j)(Y_\ell\varphi_j(X_\ell)-v_j)]}_{b}$$

le terme b devient :

$$\mathbb{E}[(Y_i\varphi_j(X_i) - v_j)(Y_\ell\varphi_j(X_\ell)-v_j)]$$

$$= \mathbb{E}[(Y_i\varphi_j(X_i)-v_j)]\,\mathbb{E}[(Y_\ell\varphi_j(X_\ell)-v_j)]$$

car (X_i, Y_i) est indépendant de (X_ℓ, Y_ℓ) pour $i\neq \ell$

et on sait d'après la question d) que $\mathbb{E}[Y_i\varphi_j(X_i)] = v_j$

et donc le terme b s'annule.

Simplifions le terme a

$$\mathbb{E}\Big[\sum_{i=1}^{n} (Y_i\varphi_j(X_i) - v_j)^2\Big] = \sum_{i=1}^{n} \mathbb{E}[(Y_i\varphi_j(X_i) - v_j)^2]$$

$$= \sum_{i=1}^{n} \mathbb{E}[(Y_i\varphi_j(X_i))^2 - 2Y_i\varphi_j(X_i)v_j + v_j^2]$$

$$= \sum_{i=1}^{n} \mathbb{E}[(Y_i\varphi_j(X_i))^2] - 2\,\mathbb{E}[Y_i\varphi_j(X_i)]v_j + v_j^2$$

$$= \sum_{i=1}^{n} \mathbb{E}[(Y_i\varphi_j(X_i))^2] - 2\sum_{i=1}^{n} v_j^2 + \sum_{i=1}^{n} v_j^2 \text{ car}$$

$$\mathbb{E}[Y_i\varphi_j(X_i)] = v_j$$

Du coup $\mathbb{E}\left[\sum_{i=1}^{n}(Y_i \varphi_j(X_i) - v_j)^2\right]$

$$= \sum_{i=1}^{n}\left(\mathbb{E}\left[(Y_i\varphi_j(X_i))^2\right] - v_j^2\right)$$

or $\mathbb{E}\left[(Y_i\varphi_j(X_i))^2 - v_j^2\right]$

$$\leq \mathbb{E}\left[(Y_i\varphi_j(X_i))^2\right]$$

regardons l'expression $\mathbb{E}\left[(Y_i\varphi_j(X_i))^2\right]$

comme $Y_i = f(X_i) + U_i$

on a

$$\mathbb{E}\left[(Y_i\varphi_j(X_i))^2\right] = \mathbb{E}\left[\left((f(X_i)+U_i)(\varphi_j(X_i))\right)^2\right]$$

$$= \mathbb{E}\left[f(X_i)^2\varphi_j(X_i)^2\right] + \mathbb{E}\left[U_i^2\varphi_j(X_i)^2\right]$$

$$+ 2\,\mathbb{E}\left[f(X_i)\varphi_j(X_i)^2 U_i\right]$$

or $\mathbb{E}\left[f(X_i)U_i\varphi_j(X_i)^2\right] = \mathbb{E}[U_i]\,\mathbb{E}\left[f(X_i)\varphi_j(X_i)^2\right]$

car $U_i \perp\!\!\!\perp X_i$ et comme $\mathbb{E}[U_i] = 0$ cette expression est nulle

De plus $\mathbb{E}\left[U_i^2\varphi_j(X_i)^2\right] = \mathbb{E}[U_i^2]\,\mathbb{E}\left[\varphi_j(X_i)^2\right]$

$$= \sigma^2 \times \int_0^1 \varphi_j^2(x)\,dx = \sigma^2$$

car $\mathbb{E}[U_i^2] = \sigma^2$ et $\|\varphi_j\|^2 = 1$.

$$\mathbb{E}[f(X_i)^2 \varphi_j(X_i)^2] = \int_0^1 f(x)^2 \varphi_j(x)^2 dx$$

$$\leq \|f\|_\infty^2 \underbrace{\int_0^1 \varphi_j(x)^2 dx}_{= 1 \text{ car } \|\varphi\| = 1}$$

$$\leq \|f\|_\infty^2$$

du coup on a

$$\mathbb{E}[(Y_i \varphi_j(X_i))^2 - v_j^2]$$

$$\leq \mathbb{E}[(Y_i \varphi_j(X_i))^2] \leq \sigma^2 + \|f\|_\infty^2$$

dans ce cas là le terme a est majorée par :

$$\mathbb{E}\left[\sum_{i=1}^{n} (Y_i \varphi_j(X_i) - v_j)^2\right] \leq \sum_{i=1}^{n} (\sigma^2 + \|f\|_\infty^2)$$

$$\leq n(\sigma^2 + \|f\|_\infty^2)$$

et dans ce cas $\mathbb{E}[(\tilde{v}_j - v_j)^2] \leq \frac{1}{n^2} \times n(\sigma^2 + \|f\|_\infty^2)$

$$\leq \frac{\sigma^2 + \|f\|_\infty^2}{n}$$

Dans ce cas là on peut majorer l'expression du risque

$$R(\tilde{f}_{n,N}, f) = \mathbb{E}[\|\tilde{f}_{n,N} - f\|^2] = \sum_{j=1}^{N} \mathbb{E}[(\tilde{v}_j - v_j)^2]$$

$$+ \sum_{j=N+1}^{+\infty} v_j^2 \leq \sum_{j=1}^{N} \frac{\sigma^2 + \|f\|_\infty^2}{n} + \sum_{j=N+1}^{+\infty} v_j^2$$

Du coup

$$R(\tilde{f}_{m,N}, f) \leq \frac{N}{m}\left(\sigma^2 + \|f\|_\infty^2\right) + \sum_{j=N+1}^{+\infty} v_j^2$$

si on connaît f,

on prendrait N de l'ordre de $m^{1-\varepsilon}$ avec $\varepsilon > 0$

de telle sorte que $\frac{N}{m}(\sigma^2 + \|f\|_\infty^2)$ tende vers

0 pour m grand.

g) Montrons que $\sum_{j>N} v_j^2 \leq N^{-2\xi} L$

$$\sum_{j>N}' v_j^2 = \sum_{j=N+1}^{+\infty}{}' v_j^2 \times j^{2\xi} \times \frac{1}{j^{2\xi}}$$

et $\frac{1}{j^{2\xi}} \leq \frac{1}{N^{2\xi}}$ pour $j>N$.

donc
$$\sum_{j>N}' v_j^2 \leq \frac{1}{N^{2\xi}} \sum_{j>N}' v_j^2\, j^{2\xi} \leq N^{-2\xi} \sum_{j=1}^{+\infty}{}' j^{2\xi} v_j^2$$

or $\sum_{j=1}^{+\infty}{}' v_j^2 j^{2\xi} \leq L$ par l'énoncé donc

$$\sum_{j>N}' v_j^2 \leq L N^{-2\xi}.$$

Dans ce cas là comme $R(\hat{f}_{m,N}, f) = \mathbb{E}[\|\hat{f}_{m,N} - f\|^2]$

et par la question precedente on a

$$\mathbb{E}[\|\hat{f}_{m,N} - f\|^2] \leq N\left(\frac{\sigma^2 + \|f\|_\infty^2}{m}\right) + \sum_{j=N+1}^{+\infty}{}' v_j^2$$

on a $R(\hat{f}_{m,N}, f) \leq N\left(\frac{\sigma^2 + \|f\|_\infty^2}{m}\right) + N^{-2\xi} L$.

Comme f est bornée par M on peut ecrire

$\|f\|_\infty \leq M$ et dans ce cas $R(\hat{f}_{m,N}, f) \leq N\left(\frac{\sigma^2 + M^2}{m}\right) + N^{-2\xi} L$

on pose $h(N) = N^{-2k} L + N\left(\frac{M^2+\sigma^2}{n}\right)$

pour trouver la valeur de N qui minimise le majorant de $R(\hat{f}_{M,N}, f)$ il faut minimiser h.

$$\nabla h(N) = -2k\, N^{-2k-1} L + \frac{M^2+\sigma^2}{n}$$

$$\nabla^2 h(N) = 2k\,(2k+1)\, N^{-2k-2} L \geq 0.$$

h est donc convexe, ce qui implique que la fonction admet un minimum au point N^* qui vérifie

$$\nabla h(N^*) = 0$$

D'où $\quad -2k(N^*)^{-2k-1} L + \frac{M^2+\sigma^2}{n} = 0$

$$\frac{M+\sigma^2}{n} = 2k\,(N^*)^{-(2k+1)}$$

$$N^* = \left(\frac{2nk}{M+\sigma^2}\right)^{1/2k+1}$$

II) Simulations. (graphiques et codes en annexe).

Nous allons maintenant observer la qualité de l'estimateur $\tilde{f}_{m,N}$ pour différentes valeurs de N ainsi que l'efficacité d'une méthode de sélection automatique (à partir des données) du niveau de troncature N en pratique.

Pour cela, nous nous plaçons dans le cas unidimensionnel ($d=1$) avec comme base orthonormée de $L^2([0,1])$ la base trigonométrique. Le "vrai" modèle est $Y_i = f(X_i) + \varepsilon_i$ où $f(x) = (x^2 2^{(x-1)} - (x-0,5)^3) \sin(10x)$ et les variables ε_i sont iid indépendantes de $\{X_i\}_{i=1..m}$ avec $\varepsilon_i \sim \mathcal{N}(0,\sigma^2)$, $\sigma = 0,2$. On prend $m = 100$ et on génère aléatoirement les données $\{X_i\}_{i=1..m}$ (et $(\varepsilon_i)_{i=1..m}$) par m variables iid de loi uniforme sur $[0,1]$ (et de loi $\mathcal{N}(0,\sigma^2)$).

On commence par tracer le nuage des points $(X_i, Y_i)_{i=1..m}$ avec la courbe de la fonction f (cf graphique 1).

Ensuite, pour $N = 5, 10, 15, 20, \dots 50$, on trace la courbe de $\tilde{f}_{m,N}$ avec celle de f et le nuage des points $(X_i, Y_i)_{i=1..m}$ (cf graphique 2 à graphique 11). Visuellement, le meilleur estimateur semble être pour $N=5$. On observe que plus N est petit, plus la courbe de $\tilde{f}_{m,N}$ est lisse et au contraire plus N est grand, plus il y a de variations.

Le biais de l'estimateur $\tilde{f}_{m,N}$ diminue quand N augmente mais en même temps sa variance augmente, et on a très rapidement (quand N augmente) un phénomène d' "overfitting" c'est à dire que l'estimateur colle trop aux données et ne prédit alors pas bien la valeur de Y pour de nouvelles données. On ne prédit bien que ce qui a déjà été observé ce qui est inutile.

On va maintenant comparer notre N "optimal" observé graphiquement avec le niveau sélectionné automatiquement par les données. Pour cela, on calcule d'abord la valeur de l'estimateur de σ^2, $\hat{\sigma}^2_{N_0}$ défini à la question h) pour $N_0 = 50$.
On trouve $\hat{\sigma}^2_{50} = 0,1001137$. On a $\sigma = 0,2$ donc $\sigma^2 = 0,04$
donc notre estimation n'est pas très précise.
On détermine ensuite $\hat{N} = \underset{N=1..50}{\mathrm{argmin}} \left(\| I_{m \times m} - A_N Y \|^2 - (m - 2N) \hat{\sigma}^2_{N_0} \right)$
On obtient $\hat{N} = 3$ ce qui est proche de la valeur "optimale" déterminée à la question précédente. On trace la courbe de $\tilde{f}_{n,\hat{N}}$ (avec $\hat{N} = 3$) avec celle de f (cf graphique 12).
On observe que la courbe de l'estimateur semble approcher encore légèrement mieux la courbe de f que dans le cas où N était égal à 5.

La méthode de sélection automatique du niveau de troncature semble donc donner de bons résultats.
Pour vérifier cela, nous allons déterminer 100 fois une valeur de $\hat{N}$ par cette méthode en regénérant à chaque fois un nouvel échantillon $\{X_i\}_{i=1..m}$ et $\{U_i\}_{i=1..m}$
On trace alors l'histogramme de la répartition des valeurs de $\hat{N}$ (cf annexes). On observe que ces valeurs

sont très concentrées en 3 (un peu moins de la moitié) et le reste se partage d'un coté entre 10 et 20% chacun pour 4 5 et 6 (valeurs proches de 3) et d'un autre côté à moins de 5% chacun pour 7 et 8 (valeurs toujours relativement proche de 3). Ainsi, la méthode de sélection automatique semble être robuste puisqu'elle donne souvent 3 (ou une valeur proche (voire relativement proche dans peu de cas) et précise rapport au "N optimal" (5) qu'on avait observé graphiquement.

Pour conclure cette partie, on a observé dans ce cas particulier que l'estimateur $\tilde{f}_{n,N}$ était meilleur pour de petites valeurs de N (l'overfitting arrivant très rapidement) et que la méthode de sélection automatique du niveau de troncature donnait de bons résultats par rapport au N "optimal" graphique. Le seul bémol est pour la valeur de l'estimateur de σ^2, $\hat{\sigma}_{N_0}^2$ qui n'était pas très précise.