

## Mémoire de Master 1 Mathématiques Appliquées de l'Université de Paris Dauphine

Titre : Estimation non paramétrique pour le modèle de régression

Par : Killian MANCHERON et Camille MICHAUX

Directeur de Mémoire : Professeur Vincent RIVOIRARD

Numéro de groupe : 11

Confidentialité : ☒ Non    ☐ Oui    (Durée : ☐ 1 an    ☐ 2 ans)



# Remerciements

Nous souhaitons remercier toutes les personnes qui nous ont aidés lors de la mise en place et la rédaction de ce mémoire.

En premier, nous souhaitons remercier Monsieur Vincent RIVOIRARD pour nous avoir proposé le sujet et pour nous avoir apporté son aide tout au long de nos recherches et de la rédaction de ce mémoire. Ses conseils ont alimentés nos investigations et nous ont donné un avant goût de la recherche.

Nous souhaitons également remercier Monsieur Julien STOEHR d'avoir supervisé l'organisation de cette UE de recherche pour la majeure statistique et Monsieur Pierre CARDALIAGUET pour sa création. Cette expérience nous a permis de découvrir une infime partie de la recherche dans le domaine des mathématiques. Nous avons pu, au cours de cette année, faire un lien entre nos cours parfois très théoriques et leurs mise en pratique dans des problèmes que nous pourrions être amenés à résoudre dans notre future vie professionnelle.

Nous désirons enfin remercier le corps professoral de l'Université Paris Dauphine. Nos professeurs nous ont fourni les outils nécessaires à la réussite de nos études universitaires et nous ont apporté de solides connaissances que nous avons pu mettre à profit lors de la rédaction de ce mémoire.



# Table des matières

<b>Remerciements</b>	<b>3</b>
<b>Table des matières</b>	<b>5</b>
<b>Introduction</b>	<b>7</b>
<b>1 Estimateur par projection</b>	<b>9</b>
1.1 Modèle utilisé . . . . .	9
1.2 Etude du risque quadratique . . . . .	11
1.2.1 Contrôle du terme de variance . . . . .	12
1.2.2 Contrôle du terme de biais . . . . .	13
1.3 Un exemple de base : la base de Fourier réelle . . . . .	14
1.3.1 Contrôle du terme de biais et de variance . . . . .	15
1.3.2 Majoration du risque quadratique . . . . .	17
1.3.3 Choix de $N$ . . . . .	17
<b>2 Estimateur de Nadaraya - Watson</b>	<b>23</b>
2.1 Exemple de tableau . . . . .	23
2.2 Exemples . . . . .	25
2.3 Lien avec estimateur par projection sur la base de Fourier . . . . .	27
2.4 Risque quadratique de l'estimateur . . . . .	29
<b>3 Applications numériques</b>	<b>35</b>
3.1 Étude des moyennes mensuelles de température . . . . .	35
3.2 Étude des données prélevées par la WMAP . . . . .	38
<b>Conclusion</b>	<b>41</b>
<b>Bibliographie</b>	<b>43</b>
<b>Annexes</b>	<b>45</b>



# Introduction

Ce mémoire est basé sur le chapitre 13 extrait du livre *Statistique mathématique en action* écrit par les professeurs Vincent RIVOIRARD et Gilles STOLTZ ([RIV09]), ainsi que le livre *Introduction à l'estimation non paramétrique* écrit par le professeur Alexandre TSYBAKOV ([TSY00]). Nous cherchons à estimer une fonction  $f$  à partir de  $n$  observations bruitées provenant de l'échantillonnage de notre fonction  $f$ . En effet, à partir de notre fonction continue  $f$ , nous obtenons  $n$  observations équidistantes,  $Y_i$ ,  $i \in \{1, \dots, n\}$ . Certaines hypothèses sur  $f$  seront précisées selon les estimateurs choisis néanmoins nous étudierons  $f$  à support dans  $[0,1]$ .

Ces observations sont dites bruitées puisque celles-ci sont de la forme :

$$Y_i = f(X_i) + \epsilon_i \text{ où les } X_i \text{ peuvent être aléatoires ou déterministes selon le modèle} \\ \text{et } \epsilon_i \sim \mathcal{N}(0, \sigma^2), \text{ pour } i \in \{1, \dots, n\} \quad (1)$$

Nous supposons ces observations indépendantes et on remarque que leur forme nous rappelle celle du modèle de régression linéaire. Ce problème d'estimation est non paramétrique car le paramètre à estimer est une fonction qui n'appartient donc pas à un sous-ensemble d'espace fini.

Dans un premier temps nous avons étudié la méthode d'estimation par projection dans une base  $(\phi_j, 1 \leq j)$  orthonormée de l'ensemble des fonctions de carré intégrable. En utilisant cette méthode,  $f$  admettra la représentation :

$$f(x) = \sum_{j=1}^{+\infty} \theta_j \phi_j(x) \quad (2)$$

Il s'agit alors d'estimer les coefficients  $\theta_j$  que nous pouvons construire à partir de nos observations  $Y_k$ . Nous prendrons pour exemple la base de Fourier dans  $\mathbb{R}$ .

Dans un second temps nous étudierons les estimateurs de Nadaraya-Watson pour le modèle de régression qui utilisent une fenêtre  $h$  et un noyau  $K$ . Dans le cas d'un modèle à effets fixes nous définirons alors  $f$  par :

$$f(x) = \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) \quad (3)$$

Pour illustrer ce mémoire, nous allons utiliser les 4 fonctions données en exemple théorique dans le sujet, mais aussi des données réelles correspondantes à la moyenne mensuelle des températures à Nottingham entre 1920 et 1939 ainsi que des données de WMAP de la première année de la NASA. Les premières données sont stockées par défaut dans Rstudio et possède l'avantage de bien vérifier l'hypothèse  $f(0) = f(1)$  tandis que les secondes sont récoltées sur le site <https://lambda.gsfc.nasa.gov/>.

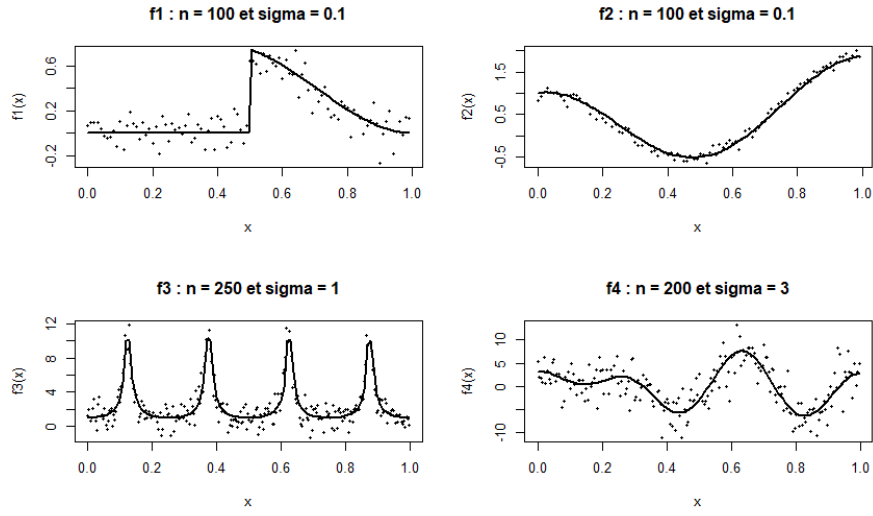


FIGURE 1: Représentation des fonctions simulées avec  $f_1$  (haut à gauche),  $f_2$  (haut à droite),  $f_3$  (bas à gauche),  $f_4$  (bas à droite).

Afin de visualiser le comportement des estimateurs selon le type de fonction étudiée, nous utiliserons les 4 fonctions représentées sur la figure 1 définies par :

- $f_1(x) = (1 - x^2) \sin\left(\frac{\pi x}{3}\right) \mathbb{1}_{\{x \geq 0.5\}}$  pour  $\sigma = 0.1$  et  $n = 100$
- $f_2(x) = \sin\left(\frac{x\pi}{3}\right) + \cos(x\pi 2)$  pour  $\sigma = 0.1$  et  $n = 100$
- $f_3(x) = \min\{10, 1/|\cos(4\pi x)|\}$  pour  $\sigma = 1$  et  $n = 250$
- $f_4(x) = 4 \sin(4\pi x) + 3 \cos(6\pi x) - 2 \sin(6\pi x)$  pour  $\sigma = 3$  et  $n = 200$

Le choix de ces fonctions est justifié par l'idée que  $f_1$  est discontinue,  $f_2$  ne vérifie pas  $f(0) = f(1)$ ,  $f_3$  n'est pas combinaison linéaire de sinus et cosinus et  $f_4$  nous servira de fonction idéale pour le choix de la base de Fourier.



# Chapitre 1

## Estimateur par projection

Nous avons dans un premier temps étudié l'estimation par projection. Cette estimation consiste à projeter dans une base orthonormée de  $\mathbb{L}^2([0, 1])$ . Dans cette partie nous posons  $X_i = \frac{i}{n} \forall i \in \{1, \dots, n\}$ .

### 1.1 Modèle utilisé

Pour l'estimateur par projection nous supposons en plus des hypothèses précédentes que la fonction est 1-périodique, c'est-à-dire qu'elle vérifie  $f(0) = f(1)$ .

Soit  $E = (\phi_j)_{\{j \geq 1\}}$  une famille de fonction réelle de  $\mathbb{L}^2([0, 1])$ . On considèrera que notre famille vérifie les hypothèses suivantes :

- E est une base de Hilbert, ce qui veut dire qu'elle est orthonormale

$$\begin{aligned} \forall (i, j) \in I^2 \quad i \neq j &\Rightarrow \langle \phi_i | \phi_j \rangle = 0 \\ \forall i \in I \quad \langle \phi_i | \phi_i \rangle &= |\phi_i|^2 = 1 \end{aligned}$$

et qu'elle est complète au sens :

$$\forall f \in L^2([0, 1]), \exists \lambda_i, \quad \sum_{i \in I} \lambda_i \phi_i = f$$

Ainsi la base permet de vérifier l'égalité de Parseval que nous utiliserons ici.

- Les fonctions  $(\phi_j)_{\{j \geq 1\}}$  sont indéfiniment dérivables en tout point de  $[0, 1]$ , il en résulte de cette hypothèse que ces fonctions sont bornées sur  $[0, 1]$ .
- La famille  $(\phi_j)_{\{j \geq 1\}}$  est une famille de fonctions périodiques.

La dernière hypothèse est nécessaire pour que la fonction soit périodique. En effet pour le démontrer nous ferons appel aux propriétés usuelles d'une base de Hilbert :

Définissons le produit scalaire dans  $\mathbb{L}^2([0, 1])$  par

$$\langle f, g \rangle_{L^2} = \int_0^1 f(x) \overline{g(x)} dx$$

Le théorème de Dirichlet nous permet d'écrire  $f = \sum_{j=1}^{+\infty} \langle f(x_i), \phi_j(x_i) \rangle_{L^2} \phi_j(x_i)$ , au sens d'une série convergente dans  $\mathbb{L}^2([0, 1])$ . Posons  $\theta_j = \langle f, \phi_j \rangle_{L^2}$ , la projection de  $f$  sur la coordonnée  $j$  de notre base

E. Nous avons donc :

$$\begin{aligned}\theta_j &= \int_0^1 f(x) \overline{\phi_j(x)} dx \\ &= \int_0^1 f(x) \phi_j(x) dx\end{aligned}$$

Donc :

$$f = \sum_{j=1}^{+\infty} \theta_j \phi_j \quad (1.1)$$

Comme la base est supposée bornée, cette série existe si les coefficients vérifient  $\sum_{i=1}^n |\theta_j| < \infty$ .

Montrons que  $f$  est 1-périodique si et seulement si la famille  $(\phi_j)_{\{j \geq 1\}}$  est 1-périodique. Soit  $x \in [0, 1]$  on a :

$$f(x) = f(x+1) = \sum_{j=1}^{+\infty} \theta_j \phi_j(x+1)$$

par 1.1, par identification (puisque nous avons unicité des coefficients de  $f$ ) nous avons bien :

$$\forall j \geq 1, \forall x \in [0, 1], \phi_j(x+1) = \phi_j(x)$$

ce qui prouve le résultat par équivalence.

Désormais, estimer  $f$  revient à estimer les coefficients  $\theta_j$ . Nous devons tout d'abord fournir un estimateur  $\hat{\theta}_j$  de  $\theta_j$ . Afin d'écrire une approximation de  $\theta_j$  nous utiliserons le lemme issue de [LEJ85] :

**Lemme :**

Soit  $f$  une fonction réelle, continue et indéfiniment dérivable sur  $[0, 1]$  telle que  $f'(0) = f'(1)$ . Pour une suite croissante de points  $x_1, x_2, \dots, x_n$  telle que  $\Delta_i = x_{i+1} - x_i$  soit constant et  $0 = x_1 - \Delta_i/2$ ,  $1 = x_n + \Delta_i/2$ , nous avons :

$$\sum_{i=1}^n f(x_i) = n \int_0^1 f(x) dx + o\left(\frac{1}{n^3}\right)$$

**Démonstration :**

Supposons  $f$  dérivable sur  $[0, 1]$ . Par un développement de  $f$  en série de Taylor au voisinage de  $x_i$  puis en intégrant sur l'intervalle  $\left[x_i - \frac{\Delta_i}{2}, x_i + \frac{\Delta_i}{2}\right]$  et en sommant sur  $i$ , nous obtenons :

$$\sum_{i=1}^n f(x_i) = \frac{1}{\Delta_i} \int_0^1 f(x) dx - \frac{\Delta_i^2}{24} \sum_{i=1}^n f''(x_i) + o(\Delta_i^3)$$

En réitérant l'égalité avec la fonction  $f''$  et en rappelant que  $f'(0) = f'(1)$ , nous avons alors

$$\sum_{i=1}^n f''(x_i) = -\frac{\Delta_i^2}{24} \sum_{i=1}^n f^{(4)}(x_i) + o(\Delta_i^3)$$

En substituant cette égalité dans 1.1 nous obtenons :

$$\sum_{i=1}^n f(x_i) = \frac{1}{\Delta_i} \int_0^1 f(x) dx + o(\Delta_i^3)$$

Comme  $\Delta_i = \frac{1}{n}$  il en résulte l'égalité 1.1. □

Néanmoins il est possible de se dispenser des hypothèses fortes sur la régularité de  $f$  en considérant la base de Fourier sur laquelle nous reviendrons dans une autre partie.

Par le lemme 1.1, nous pouvons alors approcher les coefficients de notre fonction sous certaines hypothèses par :

$$\theta_j \approx \frac{1}{n} \sum_{i=1}^n f(x_i) \phi_j(x_i)$$

En utilisant la définition de nos observations, nous pouvons voir que  $Y_i$  peut être vu comme un estimateur sans biais de  $f(x_i)$ . Cela nous permet alors de poser :

$$\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^n Y_i \phi_j(x_i)$$

Nous pourrions dans ce cas associer à nos coefficients, un estimateur de  $f$  donné par  $\hat{f} = \sum_{j=1}^{+\infty} \hat{\theta}_j \phi_j$  mais nous rencontrerions un problème au niveau de l'application car cela requerrait une infinité d'estimateur de  $\hat{\theta}_j$ . Nous introduisons,  $N$ , appelé paramètre de lissage, pour définir un nouvel estimateur par projection de  $f$  :

$$\hat{f}_N = \sum_{j=1}^N \hat{\theta}_j \phi_j \text{ pour } N \in 1, 2, \dots, n-1$$

Nous prenons ainsi la projection de notre fonction  $f$  sur les  $N$  première coordonnées de notre base. Le choix de  $N$  sera discuté dans la suite de notre mémoire.

## 1.2 Etude du risque quadratique

Dans cette partie nous allons procéder à l'étude du risque quadratique, aussi appelé erreur quadratique moyenne, associé à notre estimateur de  $\hat{f}$ . Le risque quadratique est un bon indicateur de performance de notre estimateur. Nous souhaitons avoir le risque le plus faible possible. Nous définissons ce risque quadratique par :

$$R_n(\hat{f}_N) = \mathbb{E} \left[ \int_0^1 (\hat{f}_N(t) - f(t))^2 dt \right]$$

Pour étudier le risque quadratique nous allons dans un premier temps le décomposer comme une somme d'un terme de biais et d'un terme de variance. Cette décomposition biais-variance est très souvent utilisée dans des algorithmes d'apprentissage supervisé (régression, classification et prédiction).

D'après l'égalité de Parseval nous avons :

$$\int_0^1 (\hat{f}_N(t) - f(t))^2 dt = \sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2 + \sum_{j=N+1}^{+\infty} \theta_j^2$$

Nous avons donc :

$$\begin{aligned}
 R_n(\hat{f}_N) &= \mathbb{E} \left[ \int_0^1 (\hat{f}_N(t) - f(t))^2 dt \right] \\
 &= \mathbb{E} \left[ \sum_{j=1}^N (\hat{\theta}_j - \theta_j)^2 + \sum_{j=N+1}^{+\infty} \theta_j^2 \right] \\
 &= \sum_{j=1}^N \mathbb{E} \left[ (\hat{\theta}_j - \theta_j)^2 \right] + \sum_{j=N+1}^{+\infty} \theta_j^2
 \end{aligned}$$

La dernière égalité s'obtient grâce à la linéarité de l'espérance. Le but est d'effectuer une décomposition biais-variance de notre risque quadratique pour faciliter le contrôle de celui-ci.

Nous pouvons donc écrire :

$$\begin{aligned}
 \mathbb{E} \left[ (\hat{\theta}_j - \theta_j)^2 \right] &= \mathbb{E} \left[ (\hat{\theta}_j - \mathbb{E}[\hat{\theta}_j] + \mathbb{E}[\hat{\theta}_j] - \theta_j)^2 \right] \\
 &= \mathbb{E} \left[ (\hat{\theta}_j - \mathbb{E}[\hat{\theta}_j])^2 \right] + \left[ \mathbb{E}[\hat{\theta}_j] - \theta_j \right]^2
 \end{aligned}$$

Nous pouvons à partir de là, par identification écrire :

$$R_n(\hat{f}_N) = \sum_{j=1}^N \left[ \mathbb{E} \left[ (\hat{\theta}_j - \mathbb{E}[\hat{\theta}_j])^2 \right] + \left[ \mathbb{E}[\hat{\theta}_j] - \theta_j \right]^2 \right] + \sum_{j=N+1}^{+\infty} \theta_j^2 \stackrel{not}{=} b_{n,N} + v_{n,N}$$

Avec :

$$\begin{aligned}
 b_{n,N} &\stackrel{not}{=} \sum_{j=1}^N \left( \mathbb{E}[\hat{\theta}_j] - \theta_j \right)^2 + \sum_{j=N+1}^{+\infty} \theta_j^2 \\
 v_{n,N} &\stackrel{not}{=} \sum_{j=1}^N \mathbb{E} \left[ \left( \hat{\theta}_j - \mathbb{E}[\hat{\theta}_j] \right)^2 \right] = \sum_{j=1}^N Var \hat{\theta}_j
 \end{aligned}$$

### 1.2.1 Contrôle du terme de variance

Nous commençons d'abord par contrôler le terme de variance. Nous noterons  $(b_{mj})$  les constantes telles que pour tout  $m$  et  $j$  dans  $\mathbb{N}$  Nous avons :

$$b_{mj} = \frac{1}{n} \sum_{i=1}^n \phi_m(x_i) \phi_j(x_i)$$

.

Nous pouvons écrire :

$$\begin{aligned}
\text{Var}(\hat{\theta}_j) &= \text{Var} \left( \frac{1}{n} \sum_{k=1}^n Y_i \phi_j(x_i) \right) \\
&= \sum_{i=1}^n \frac{(\phi_j(x_i))^2}{n^2} \text{Var}(Y_i) \\
&= \frac{\sigma^2}{n} \left( \frac{1}{n} \sum_{i=1}^n (\phi_j(x_i))^2 \right) \quad \text{car } \forall i \in [1; n], \text{Var}(Y_i) = \sigma^2 \\
&= \frac{\sigma^2}{n} b_{jj}
\end{aligned}$$

Il est important de noter que ce terme de variance dépend de notre paramètre de lissage  $N$  et de notre base  $(\phi_j, j \in [0; 1])$ . Ainsi nous concluons par le terme général de la variance :

$$v_{n,N} = \sum_{j=1}^N \text{Var} \hat{\theta}_j = \frac{\sigma^2}{n} \sum_{j=1}^N b_{jj} \quad (1.2)$$

Par le lemme 1.1 et par définition des  $b_{mj}$  nous pouvons dire que si la base est suffisamment dérivable, alors  $b_{jj} = 1 + o(1/n^3)$  et donc

$$v_{n,N} = \frac{N\sigma^2}{n} + o\left(\frac{1}{n^3}\right) \quad (1.3)$$

### 1.2.2 Contrôle du terme de biais

Le terme de biais en revanche ne peut être calculer exactement. Nous allons devoir faire des supposition supplémentaire sur  $f$  afin de pouvoir majorer ce terme de biais. Posons  $\alpha_j$  le biais pour notre estimateur  $\hat{\theta}_j$ . Nous avons donc  $\alpha_j = \mathbb{E}(\hat{\theta}_j) - \theta_j$ . Nous pouvons insérer cette expression dans notre formule du terme de biais. Nous obtenons :

$$b_{n,N} = \sum_{j=1}^N \alpha_j^2 + \sum_{j=N+1}^{+\infty} \theta_j^2 \quad (1.4)$$

Nous pouvons remarquer que le biais se divise en deux termes, le premier croît avec  $N$  tandis que le second décroît. Pour le contrôler nous devons majorer (1.4), pour cela nous allons dans un premier temps détailler les termes  $\alpha_j$ .

Supposons que  $\sum_{m \geq 1} |\theta_m| < \infty$ .

Soit  $j \in \{1, \dots, n-1\}$ , Nous avons :

$$\begin{aligned}
\alpha_j &= \mathbb{E}(\hat{\theta}_j) - \theta_j \\
&= \frac{1}{n} \sum_{i=1}^n f(x_i) \phi_j(x_i) - \theta_j \\
&= \frac{1}{n} \sum_{i=1}^n \left( \sum_{m=1}^{+\infty} \theta_m \phi_m(x_i) \right) \phi_j(x_i) - \theta_j \quad \text{d'après (1.1)}
\end{aligned}$$

Nous pouvons inverser les sommes par convergence normale de façon à obtenir :

$$\alpha_j = \sum_{m=1}^{+\infty} \theta_m \underbrace{\left( \frac{1}{n} \sum_{i=1}^n \phi_m(x_i) \phi_j(x_i) \right)}_{=b_{mj}} - \theta_j = \sum_{m=1}^{+\infty} \theta_m b_{mj} - \theta_j$$

Par le lemme 1.1 et par orthogonalité de la base, nous remarquons alors que pour tout  $m$  et  $j$  dans  $\mathbb{N}$  :

$$b_{mj} - \mathbb{1}_{m=j} = o\left(\frac{1}{n^3}\right) \quad \text{i.e } n^3(b_{mj} - \mathbb{1}_{m=j}) \xrightarrow{n \rightarrow \infty} 0$$

En effet nous remarquerons que l'ordre peut dépendre de  $m$  et  $j$ . Nous remarquerons également que  $\theta_m$  indépendant de  $n$  sous sa forme intégrale et que si pour tout  $n \in \mathbb{N}^*$  la série converge uniformément, nous avons alors

$$\lim_{n \rightarrow \infty} n^3 \alpha_j = \lim_{n \rightarrow \infty} \sum_{m=1}^{+\infty} \theta_m n^3 (b_{mj} - \mathbb{1}_{m=j}) = \sum_{m=1}^{+\infty} \theta_m \lim_{n \rightarrow +\infty} n^3 (b_{mj} - \mathbb{1}_{m=j}) = 0$$

Ainsi nous pouvons en déduire que  $\alpha_j$  est d'ordre  $1/n^3$  si la fonction  $f$  est indéfiniment dérivable, ceci nous amène à considérer que le second terme du biais est dominant sur le premier. Nous remarquerons alors que si l'ordre de la suite  $(\alpha_j)$  est  $1/n^3$ , il s'en suit que la série quadratique  $\sum (\alpha_j)^2$  est d'ordre au moins  $1/n^3$ . Il est aussi possible de majorer les termes  $\alpha_j$  par :

$$\max_{1 \leq j \leq n-1} |\alpha_j| \leq \left| \sum_{m=1}^{n-1} \theta_m b_{mj} - \theta_j \right| + \eta \sum_{m=n}^{+\infty} |\theta_m| \quad \text{où } \eta = \max_{1 \leq j \leq n-1} |\phi_j|^2 \leq +\infty \quad (1.5)$$

Dans cette expression les coefficients  $b_{mj}$  dépendent aussi de la base considérée. Nous l'utiliserons dans le cas de la base de Fourier où ces coefficients sont simples. Nous pouvons alors se ramener à la majoration du risque quadratique que nous aborderons dans la section (1.3.2)

### 1.3 Un exemple de base : la base de Fourier réelle

Comme nous l'avons vu dans les deux dernières sections, le contrôle du biais et de la variance dépendent de la base, notamment des constantes  $(b_{mj})$ . Pour cela nous introduisons une base de Hilbert, la base de Fourier définie par :

$$\begin{aligned} \phi_1 &= 1 \text{ et pour tout } j \in \mathbb{N}^* \text{ et pour tout } x \in [0,1] \\ \phi_{2j}(x) &= \cos(2\pi j x) \text{ et } \phi_{2j+1}(x) = \sqrt{2} \sin(2\pi j x) \end{aligned} \quad (1.6)$$

Nous pouvons alors calculer explicitement les coefficients  $b_{mj}$  pour  $m$  et  $j$  dans  $\{1, \dots, n-1\}$ . Il va s'agir de démontrer le Lemme 13.1 de [RIV09].

$$\forall 1 \leq j, j' \leq n-1 \quad \frac{1}{n} \sum_{i=1}^n \phi_j(x_i) \phi_{j'}'(x_i) = \begin{cases} 1 & \text{si } j = j' \\ 0 & \text{sinon} \end{cases}$$

La preuve de ce lemme s'articule autour des valeurs de  $j$  et  $j'$ . Nous distinguons quatre cas :

- $j = 1$  ou  $j' = 1$  (en effet,  $\phi_1 = 1$ )
- $j$  et  $j'$  sont de la forme  $j = 2k + 1$  et  $j' = 2l + 1 \quad \forall k, l \geq 1$

- $j$  et  $j'$  sont de la forme  $j = 2k$  et  $j' = 2l \forall k, l \geq 1$
- $j$  et  $j'$  sont de la forme  $j = 2k + 1$  et  $j' = 2l \forall k, l \geq 1$  (on peut inverser les rôles de  $j$  et  $j'$  dans ce cas)

Les quatre cas se traitent de la même manière et la preuve de chaque cas repose sur les propriétés géométriques de la base de Fourier réelle. Nous allons traiter le dernier cas où  $j$  et  $j'$  sont de parité différente et donc non égaux.

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^n \phi_j(x_i) \phi_{j'}'(x_i) &= \frac{2}{n} \sum_{i=1}^n \cos(\pi j x_i) \sin(\pi j' x_i) \\
&= \frac{2}{n} \sum_{i=1}^n \left( \frac{e^{\pi j x_i} + e^{-\pi j x_i}}{2} \right) \left( \frac{e^{\pi j' x_i} - e^{-\pi j' x_i}}{2i} \right) \\
&= \frac{1}{2in} \sum_{i=1}^n \left( e^{\pi x_i(j+j')} - e^{\pi x_i(j-j')} + e^{\pi x_i(-j+j')} - e^{-\pi x_i(j+j')} \right)
\end{aligned}$$

On pose  $x_i = \frac{i}{n} \forall i \in [1; n]$

$$\begin{aligned}
&= \frac{1}{2in} \sum_{i=1}^n \left( e^{\frac{\pi i(j+j')}{n}} - e^{\frac{\pi i(j-j')}{n}} + e^{\frac{\pi i(-j+j')}{n}} - e^{\frac{-\pi i(j+j')}{n}} \right) \\
&= \frac{1}{2in} \left( \frac{1 - e^{\pi(j+j')}}{1 - e^{\frac{\pi(j+j')}{n}}} - \frac{1 - e^{\pi(j-j')}}{1 - e^{\frac{\pi(j-j')}{n}}} + \frac{1 - e^{\pi(-j+j')}}{1 - e^{\frac{\pi(-j+j')}{n}}} - \frac{1 - e^{-\pi(j+j')}}{1 - e^{\frac{-\pi(j+j')}{n}}} \right) (*) \\
&= \frac{1}{2in} \times 0 \\
&= 0
\end{aligned}$$

Le passage de la l'égalité (\*) à l'égalité suivante s'obtient en regroupant et en réduisant au même dénominateur la première et la dernière fraction, ainsi que la deuxième et la troisième. Les termes s'annulent alors deux à deux et nous obtenons le résultat souhaité. Comme  $j$  et  $j'$  sont des parité différente, on a nécessairement  $j \neq j'$ . Les 3 autres cas s'obtiennent de manière équivalente.

□

### 1.3.1 Contrôle du terme de biais et de variance

Reprenons l'expression de la variance (1.2) Nous avons donc :

$$b_{jj'} = \begin{cases} 1 & \text{si } j = j' \\ 0 & \text{sinon} \end{cases}$$

Nous avons donc :

$$v_{n,N} = \frac{\sigma^2 N}{n} \quad (1.7)$$

Pour contrôler le biais nous souhaitons majorer (1.4) et pour cela référons nous à l'expression 1.5. Comme les coefficients  $(b_{mj})$  sont explicites pour  $m$  et  $j$  dans  $\{1, \dots, n-1\}$ , nous obtenons

$$\max_{1 \leq j \leq n-1} |\alpha_j| \leq 2 \sum_{m=n}^{+\infty} |\theta_m|$$

en remarquant que  $|\phi_j| \leq \sqrt{2}$ . À partir de cette majoration de la suite  $(\alpha_j)$ , majorons le premier terme de biais :

$$\begin{aligned} \sum_{j=1}^N \alpha_j^2 &\leq \sum_{j=1}^N \left( 2 \sum_{k=n}^{+\infty} |\theta_k| \right)^2 \\ &\leq 4N \left( \sum_{k=n}^{+\infty} |\theta_k| \right)^2 \end{aligned} \quad (1.8)$$

Pour majorer plus précisément (1.8) supposons que  $f$  appartient à l'espace de Sobolev  $W(\beta, L)$ . Les espaces de Sobolev sont définis dans [RIV09] par

$$W(\beta, L) = \left\{ f : [0, 1] \rightarrow \mathbb{R}, f^{(\beta)} \text{ existe et } \int_0^1 \left( f^{(\beta)}(t) \right)^2 dt \leq L^2 \text{ et } f^{(\beta)}(0) = f^{(\beta)}(1) \text{ pour } j \in [0, \beta - 1] \right\}$$

La proposition 13.3 nous permet de définir  $\beta \in \mathbb{N}^*$  et  $L > 0$  tel que si  $f$  appartient à  $W(\beta, L)$  alors les coefficients de Fourier de  $f$  vérifient

$$\sum_{j=1}^{+\infty} c_j^2 \theta_j^2 \leq \frac{L^2}{\pi^{2\beta}} \text{ où } c_j = \begin{cases} j^\beta & \text{si } j \text{ est pair} \\ (j-1)^\beta & \text{sinon} \end{cases}$$

Afin de faire apparaître ces coefficients dans notre biais, nous utilisons l'inégalité de Cauchy-Schwarz pour obtenir :

$$\left( \sum_{j=n}^{+\infty} |\theta_j| c_j \cdot c_j^{-1} \right)^2 \leq \left( \sum_{j=n}^{+\infty} c_j^2 \theta_j^2 \right) \left( \sum_{j=n}^{+\infty} c_j^{-2} \right)$$

À l'aide de ce résultat et de la proposition précédente, nous pouvons majorer (1.8) par :

$$\begin{aligned} \sum_{j=1}^N \alpha_j^2 &\leq 4N \left( \sum_{k=n}^{+\infty} c_k^2 \theta_k^2 \right) \left( \sum_{k=n}^{+\infty} c_k^{-2} \right) \\ &\leq 4N \left( \sum_{j=n}^{+\infty} c_j^2 \theta_j^2 \right) \left( \sum_{j=n}^{+\infty} c_j^{-2} \right) \\ &\leq 4N \frac{L^2}{\pi^{2\beta}} \sum_{j=n}^{+\infty} (j-1)^{-2\beta} \end{aligned}$$

Nous remarquerons dans la dernière ligne de calcul que la parité de  $j$  n'importe pas puisque si  $j$  est pair nous avons bien  $j^{-2\beta} \leq (j-1)^{-2\beta}$ . Nous pouvons aussi majorer notre somme par une intégrale

$$\sum_{j=n}^{+\infty} (j-1)^{-2\beta} \leq \int_{n-2}^{+\infty} t^{-2\beta} dt = \frac{1}{2\beta-1} (n-2)^{1-2\beta} \leq \frac{1}{2\beta-1} \left( \frac{n}{2} \right)^{1-2\beta}$$

la dernière inégalité est vraie si  $\beta \geq 1$  et  $1/(n-2) \leq 2/n$  donc  $n \geq 4$ . Cette majoration nous permet d'affirmer l'existence d'une constante  $C_{\beta,L}$  qui dépend de  $\beta$  et  $L$  telle que

$$\sum_{j=1}^N \alpha_j^2 \leq C_{\beta,L} N n^{1-2\beta} \quad (1.9)$$



Cette démonstration vient du livre [RIV09] dont nous nous sommes inspirés pour ce mémoire. L'intérêt de présenter cette preuve est de montrer que les résultats obtenus dans la partie 1.2.1 et 1.2.2 sont vrais avec la base de Fourier et que si  $\beta \geq 2$  alors les termes  $(\alpha_j)_{j \in \mathbb{N}}$  sont d'ordre  $1/n^3$  ce qui nous amène alors à les négliger par rapport aux deux autres termes du risque quadratique.

### 1.3.2 Majoration du risque quadratique

Dans les parties précédentes nous avons pu conclure qu'il était possible de décomposer le risque quadratique en une partie biais et une partie variance. Cette première partie admettait un terme  $\sum_{j \leq N} (\alpha_j)^2$  que nous supposons d'ordre  $1/n^3$  dès lorsque notre fonction était suffisamment dérivable et un terme  $\sum_{j \leq N} \theta_j$  qui diminuait avec  $N$ . Il en est de même pour la seconde partie de la somme, dont le premier terme augmentait avec  $N$  tandis que le reste était d'ordre  $1/n^3$ .

En négligeant les termes d'ordre  $1/n^3$ , dans (1.8) et (1.3), nous pouvons alors majorer approximativement notre risque par ces deux grandes valeurs et à l'aide de l'égalité de Parseval :

$$R_n(\widehat{f_N}) = b_{n,N} + v_{n,N} \lesssim \frac{N\sigma^2}{n} + \sum_{j=N+1}^{\infty} \theta_j^2 = \frac{N\sigma^2}{n} + \int_0^1 f^2(t)dt - \sum_{j=1}^N \theta_j^2 \quad (1.10)$$

Nous allons désormais engager une discussion sur le choix de  $N$  en utilisant ces majorations pour réduire au maximum le risque quadratique.

### 1.3.3 Choix de $N$

Afin de minimiser le risque 1.10 en fonction de  $N$ , il faut donc choisir celui qui vérifie :

$$N \in \underset{1 \leq N \leq n-1}{\operatorname{argmin}} \left\{ \frac{N\sigma^2}{n} - \sum_{j=1}^N \theta_j^2 \right\}$$

Rappelons que :

$$\alpha_j = \mathbb{E} [\widehat{\theta}_j] - \theta_j \quad \text{ou aussi} \quad \mathbb{E} [\widehat{\theta}_j] = \alpha_j + \theta_j$$

Donc d'après le théorème de König-Huygens :

$$\mathbb{E} [\widehat{\theta}_j^2] - (\theta_j + \alpha_j)^2 = \operatorname{Var} \widehat{\theta}_j = \frac{\sigma^2}{n}$$

En négligeant de nouveau les termes  $\alpha_j$ , nous obtenons alors que  $\widehat{\theta}_j - \sigma^2/n$  un estimateur sans biais de  $\theta_j$ , en substituant cette estimation dans le  $N$  précédent nous avons alors :

$$N \in \underset{1 \leq N \leq n-1}{\operatorname{argmin}} \left\{ \frac{N\sigma^2}{n} - \sum_{j=1}^N (\widehat{\theta}_j^2 - \frac{\sigma^2}{n}) \right\} = \underset{1 \leq N \leq n-1}{\operatorname{argmin}} \left\{ \frac{2N\sigma^2}{n} - \sum_{j=1}^N \widehat{\theta}_j^2 \right\}$$

On appelle ce  $N$  approximatif, le  $N_{\text{mallows}}$  car celui-ci est fondée sur l'heuristique du  $C_p$  de Mallows. Afin d'étudier le comportement de l'estimateur de nos quatre fonctions définies dans l'introduction, il nous faut regarder au préalable à quel espace de Sobolev appartient chaque fonction :

- La fonction  $f_1$  définie par

$$f_1(x) = (1 - x^2) \sin\left(\frac{\pi x}{3}\right) 1_{\{x \geq 0.5\}}$$

n'est pas continue, est 1-périodique et de carré intégrable donc il existe  $L$  tel qu'elle appartient à  $W(0, L)$

- La fonction  $f_2$  définie par

$$f_2(x) = \sin\left(\frac{\pi x}{3}\right) + \cos(x\pi 2)$$

est indéfiniment dérivable et n'est pas 1-périodique. Même si pour tout  $\beta$  il existe  $L$  tel que la fonction appartiennent aux espaces plus classique de Sobolev, celle-ci ne respecte pas la périodicité. Nous l'avons introduite pour tester la légitimité de la condition de périodicité.

- La fonction  $f_3$  définie par

$$f_3(x) = \min \left\{ 10, \frac{1}{|\cos(4\pi x)|} \right\}$$

est continue en tout point mais non dérivable si  $x \in \{1/8, 3/8, 5/8, 7/8\}$ . Elle est bien 1-périodique et de carré intégrable donc il existe  $L$  tel que  $f_3 \in W(1, L)$

- La fonction  $f_4$  définie par

$$f_4(x) = 4 \sin(4\pi x) + 3 \cos(6\pi x) - 2 \in (6\pi x)$$

est indéfiniment dérivable, est bien 1-périodique et il existe  $L$  tel que pour tout  $\beta \geq 1$ ,  $f_4 \in W(\beta, L)$

Dans la suite, nous retiendrons la valeur  $\beta = 1$  et mettrons en application l'estimation du signal en prenant les valeurs suivantes de  $N$  :

- la partie entière de  $n^{\frac{1}{3}}$  (justifié dans l'ouvrage [RIV09] pour  $\beta = 1$ )
- la valeur maximale  $N = n - 1$
- la valeur  $N_{mallows}$  donnée précédemment
- la valeur optimale de  $N$ , notée  $N_{opt}$

La valeur optimale est donnée par la minimisation du risque quadratique discrétisé :

$$N_{opt} \in \underset{1 \leq N \leq n-1}{\operatorname{armin}} \left\{ \frac{1}{n} \sum_{k=1}^n \left( f(t_k) - \hat{f}_N(t_k) \right)^2 \right\}$$

Pour illustrer cela, reprenons les fonctions définies lors de l'introduction. Nous pouvons remarquer sur la Figure 1.1 que pour chacune des fonctions, le  $N_{opt}$  et le  $N_{mallows}$  sont assez proches. En revanche, la partie entière de  $n^{\frac{1}{3}}$  et  $N = n - 1$  sont graphiquement opposés. Cela traduit le fait qu'ils ne sont pas appropriés lors de l'application car peu précis. Ces graphiques traduisent l'enjeu de bien déterminer un  $N$  performant puisque s'il est trop faible on remarque bien le pic proche de 0. Un autre point important est que les trois premiers graphes ont en commun une stabilisation à partir du  $N_{opt}$  tandis que le dernier remonte linéairement avec le même ordre de grandeur qu'en 0. Ceci pose un problème car lors des simulations il est possible de remarquer que les valeurs de notre  $N_{mallows}$  peut varier sur tout l'intervalle où cette fonction est constante. Ceci induit fortement sur le lissage de la courbe mais qui nous permet d'estimer que la fonction n'est pas une combinaison linéaire de fonctions trigonométriques comme la fonction  $f_4$  ou ne satisfait pas une des hypothèses principales. Afin d'illustrer cela nous avons fait le choix de nous concentrer sur  $f_1$  et  $f_4$  puisque l'analyse sur les fonctions  $f_2$  et  $f_3$  est semblable en apparence. Les graphiques de ces dernières sont disponibles en annexes. Nous noterons que la propriété  $f(0) = f(1)$  n'a pas influencé l'allure de l'estimateur de  $f_2$ . Néanmoins comme celle-ci est une combinaison linéaire nous aurions dû avoir un graphique semblable à celui de  $f_4$ . Un moyen de contourner ce problème serait de couper l'intervalle d'estimation où nous avons bien des bornes égales mais d'une part l'estimateur ne serait pas pertinent en dehors de ces

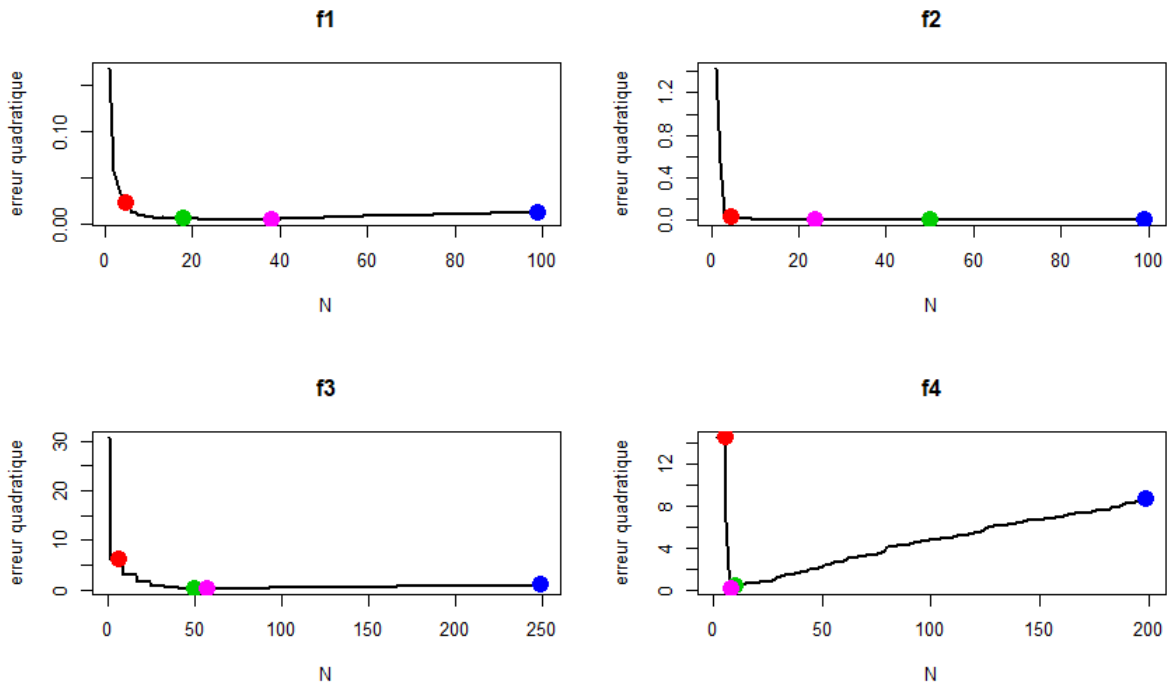


FIGURE 1.1: Erreur quadratique entre le signal et l'estimateur, en rouge la valeur de  $N = n^{1/3}$ , en vert la valeur de  $N_{opt}$ , en violet celle de  $N_{mallows}$  et en bleu de  $N = n - 1$

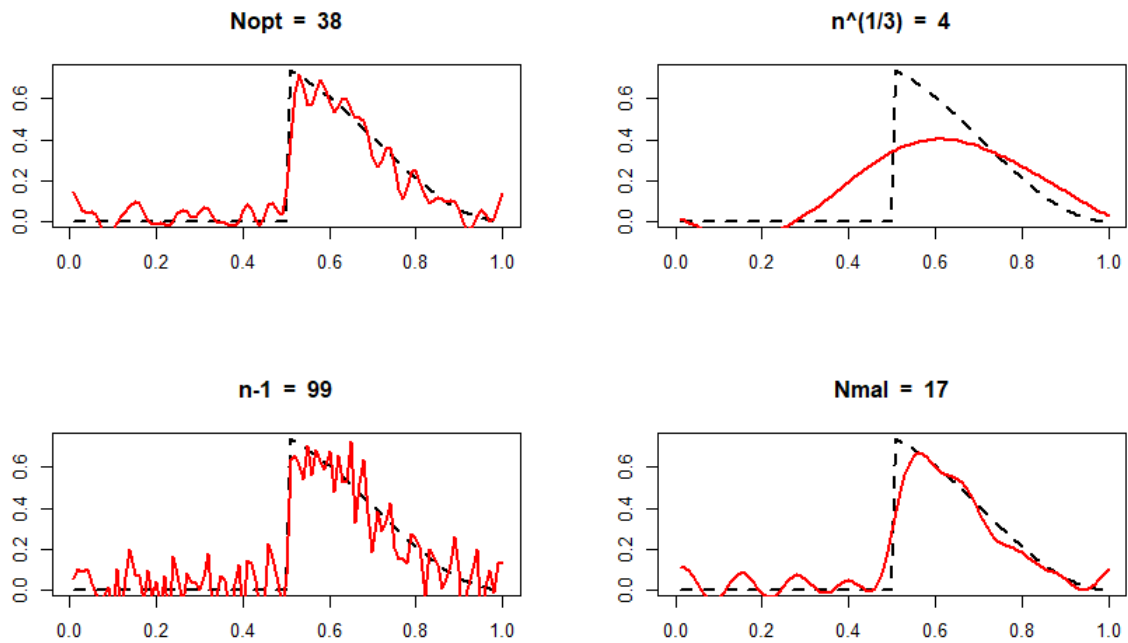
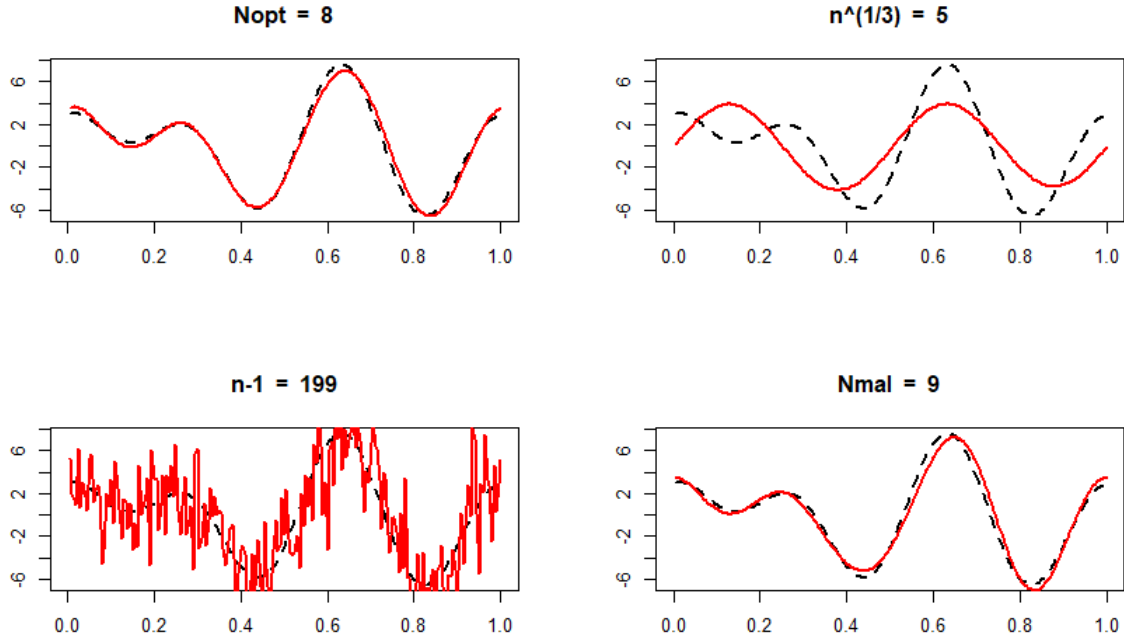


FIGURE 1.2: Représentation de  $f_1$  et de  $\hat{f}_N$  suivant  $N$

FIGURE 1.3: Représentation de  $f_4$  et de  $\hat{f}_N$  suivant  $N$ 

bornes et d'autre part il y a peu de chances que  $f'(0) = f'(1)$ .

Les Figures 1.2 et 1.4 illustrent une fois encore le fait que la partie entière de  $n^{\frac{1}{3}}$  et  $N = n - 1$  ne sont pas adaptés en pratique puisque l'on remarque un écart significatif entre notre fonction  $f$  et son estimateur  $\hat{f}_N$ .

Lorsque  $N = n - 1$ , nous remarquerons aussi que l'estimateur reproduit les données et n'accorde pas d'importance à la régularité de la fonction.

Dans la pratique les fonctions étudiées ne sont pas connues et l'écart-type non plus. On se référera à l'estimation de  $\sigma$  proposé par V. RIVOIRARD dans [RIV09]. On y considère alors le modèle  $Y_i = f(x_i) + \epsilon$  où  $\epsilon \sim \mathcal{N}(0, \sigma^2)$ . Celui-ci n'est pas différent de celui vu en 1 à l'exception que  $\sigma$  soit inconnu, néanmoins nous pouvons estimer cette valeur. Pour cela nous pouvons essayer d'estimer  $\sigma^2$  à l'aide du théorème de Cochran puisque les  $Y_i$  sont gaussiens. Nous noterons, pour tout  $j \in \{1, \dots, n-1\}$ , les valeurs prises par la  $j$ -ième fonction de la base de Fourier sur les  $(x_i)_{i \in \{1, \dots, n\}}$  :

$$\psi_j = (\phi_j(t_1), \dots, \phi_j(t_n))^T$$

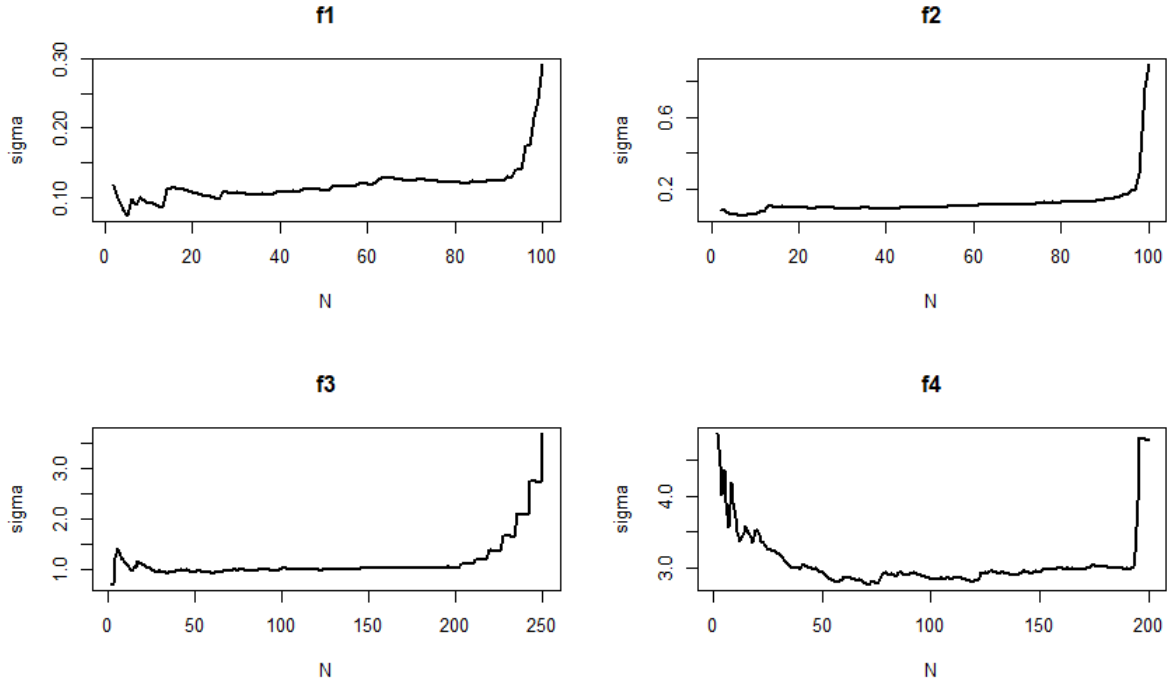
Pour  $q \in \{1, \dots, n-1\}$ , nous introduisons alors l'espace vectoriel  $W_q$  engendré par les  $q$  derniers vecteurs,  $\phi_{n-q}, \dots, \phi_{n-1}$ .

Pour estimer  $\sigma$ , appuyons nous sur l'hypothèse : pour un  $q$  suffisamment grand,

$$F \stackrel{\text{not.}}{=} (f(t_1), \dots, f(t_n))^T \in \text{Vect}(\phi_1, \dots, \phi_{n-q-1})$$

Nous obtenons alors  $F$  qui appartient à l'orthogonal de  $W_q$ , nous pouvons alors appliquer le théorème de Cochran en notant  $\Pi_q$  la matrice de projection sur  $W_q$  pour le produit scalaire euclidien, nous avons :

$$\frac{\|\Pi_q Y_1^n\|_2^2}{\sigma^2} \sim \chi^2(q)$$

FIGURE 1.4: Représentation de  $f_4$  et de  $\hat{f}_N$  suivant  $N$ 

Explicitons les valeurs de cette projection par

$$\|\Pi_q Y_1^n\|_2^2 = \sum_{j=n-q}^{n-1} \left( \sum_{i=1}^n Y_k \frac{\phi_k(x_i)}{\sqrt{n}} \right)^2 = n \sum_{j=n-q}^{n-1} \hat{\theta}_j^2$$

Ceci nous permet de simuler facilement les graphes obtenus dans la figure 1.4. Ainsi considérons un estimateur de  $\sigma$  :

$$\hat{\sigma}_q = \frac{\|\Pi_q Y_1^n\|_2^2}{\sqrt{q}}$$

Cet estimateur dépend du paramètre  $q$  que nous pouvons varier afin d'obtenir la représentation graphique 1.4. Nous pouvons aussi représenter les différentes valeurs prises par  $\hat{\sigma}_q$  sous la forme d'un histogramme qui nous fournit une densité de probabilité du choix de  $\sigma$  comme donné en annexe 10. Il s'agit alors de noter un intervalle large où les valeurs de sigma semblent se stabiliser et tester différentes valeurs de  $N_{mallows}$ , ce que nous illustrerons dans la partie Applications numériques.

Nous remarquerons que dans la représentation de l'évolution de  $\sigma$  en fonction de  $q$  le graphique est instable pour des valeurs faibles de  $q$  et explose lorsque  $q$  est proche de  $n$ . Ceci est dû à l'hypothèse d'orthogonalité faite précédemment pour définir cet estimateur. Notons aussi que nous avons utilisé uniquement des hypothèses sur une base orthogonale, nous pouvons donc fournir la même estimation avec une autre base.



## Chapitre 2

# Estimateur de Nadaraya - Watson

### 2.1 Exemple de tableau

Dans cette partie intéressons nous à la régression non paramétrique à effets aléatoires. Supposons qu'il existe un couple  $(X, Y)$  de variables aléatoires réelles et dont nous disposons d'un  $n$ -échantillon  $(X_1, Y_1), \dots, (X_n, Y_n)$  qui vérifie le modèle de régression suivant :

$$Y_i = f(X_i) + \epsilon_i, \text{ où } \epsilon_i \sim N(0, \sigma^2) \quad \text{pour } i = 1, \dots, n$$

Sous l'hypothèse que  $\mathbb{E}[Y]$  soit finie, définissons l'estimateur de Nadaraya-Watson. Celui-ci s'inspire directement des estimateurs à noyaux de densité en introduisant une fonction  $K : \mathbb{R} \rightarrow \mathbb{R}$  intégrable, telle que  $\int K(u)du = 1$ , dit *noyau* et d'une *fenêtre*  $h > 0$ . Nous appelons celui-ci estimateur de Nadaraya-Watson (noté  $f_n^{NW}$ ) de la fonction de régression  $f$  défini sur  $[0, 1]$  :

$$f_n^{NW}(x) = \frac{\sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad \mathbb{1}_{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \neq 0}$$

Nous remarquerons d'une part que si le noyau  $K$  est positif alors l'indicatrice vaut toujours 1 et d'autre part qu'il est possible de représenter l'estimateur de Nadaraya-Watson comme une somme pondérée des  $Y_i$  :

$$f_n^{NW}(x) = \sum_{i=1}^n Y_i W_{ni}^{NW}(x)$$

avec des poids définis par :

$$W_{ni}^{NW}(x) = \frac{K\left(\frac{X_i - x}{h}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right)} \quad \mathbb{1}_{\sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \neq 0}$$

Remarquons que les poids  $W_{ni}(x)$  satisfont la relation

$$\sum_{i=1}^n W_{ni}(x) = 1 \quad \forall x \in \mathbb{R}. \quad (2.1)$$

Pour comprendre la formulation de cet estimateur, le livre [TSY00] nous fournit une bonne explication ; il faut d'abord se restreindre à l'étude d'une densité. Supposons que la loi du couple  $(X, Y)$  soit défini par une densité  $p(x, y)$  (par rapport à la mesure de Lebesgue étant dans  $\mathbb{R}^2$ ) et notons

$p(x) = \int p(x, y) dy$  la densité marginale de  $X$ . La fonction de régression de  $Y$  par rapport à  $X$  est alors définie par :

$$f(x) = \mathbf{E}[Y|X = x] = \int y \frac{p(x, y)}{p(x)} dy \quad (2.2)$$

L'égalité (2.2) est directe par définition de la densité conditionnelle. Il s'agit alors d'estimer ces deux densités par la méthode des noyaux, dans le cas de la dimension 1 et 2, définie par les notations précédentes. Nous avons ainsi

$$\begin{aligned} \hat{p}_n(x) &= \frac{1}{nh} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \\ \hat{p}_n(x, y) &= \frac{1}{nh^2} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) K\left(\frac{Y_i - y}{h}\right) \end{aligned}$$

En remplaçant les estimateurs des densités dans l'égalité (2.2) nous obtenons alors une forme plus généralisée de l'estimateur  $f_n^{NW}(x)$  de Nadaraya-Watson.

$$f_n^{NW}(x) = \frac{\int y \hat{p}_n(x, y) dy}{\hat{p}_n(x)} = \frac{1}{nh^2 \hat{p}_n(x)} \sum_{i=1}^n K\left(\frac{X_i - x}{h}\right) \int y K\left(\frac{Y_i - y}{h}\right) dy \quad (2.3)$$

Pour la suite introduisons la notion d'ordre d'un noyau :

**Définition :** Soit  $l \geq 1$  un entier. Nous dirons que  $K : \mathbf{R} \rightarrow \mathbf{R}$  est un **noyau d'ordre  $l$**  si les fonctions  $u \mapsto u^j K(u)$ ,  $j = 0, 1, \dots, l$ , sont intégrables et vérifient

$$\int K(u) du = 1, \quad \int u^j K(u) du = 0, \quad j = 1, \dots, l$$

L'ordre du noyau peut nous informer sur la précision de notre approximation, il est possible par exemple si  $K$  est d'ordre au moins 1 d'obtenir plusieurs résultats utiles si nous souhaitons étudier le biais de l'estimateur de densité :

$$\begin{aligned} \mathbb{E}[\hat{p}_n(x)] &= \frac{1}{h} \mathbb{E}\left[K\left(\frac{X_1 - x}{h}\right)\right] \\ &= \frac{1}{h} \int K\left(\frac{u - x}{h}\right) p(u) du \\ &= \int K(t) p(x + th) dt \\ &\approx p(x) + hp'(x) \int tK(t) dt + \frac{h^2}{2} p''(x) \int t^2 K(t) dt \end{aligned} \quad (2.4)$$

$$\approx p(x) + \frac{h^2}{2} p''(x) \int t^2 K(t) dt \quad (2.5)$$

L'égalité 2.4 est le résultat d'un développement de Taylor à l'ordre 2 de la densité  $p$ . Comme  $K$  est d'ordre au moins 1, la première intégrale est nulle. Ce résultat nous sera utile pour la section abordant le risque quadratique de l'estimateur.

Calculons l'intégrale de l'équation (2.3) dans le cas où  $K$  est un noyau d'ordre au moins 1 :

$$\begin{aligned} \frac{1}{h} \int y K\left(\frac{Y_i - y}{h}\right) dy &= \int \frac{y - Y_i}{h} K\left(\frac{Y_i - y}{h}\right) dy + \int \frac{Y_i}{h} K\left(\frac{Y_i - y}{h}\right) dy \\ &= -h \int u K(u) du + Y_i \int K(u) du \\ &= Y_i \quad \text{car } K \text{ d'ordre } 1 \end{aligned}$$



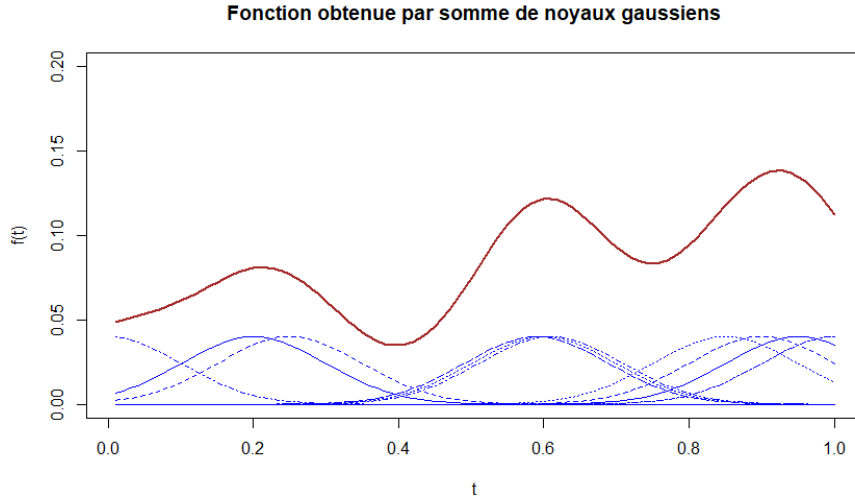


FIGURE 2.1: Exemple d'une fonction obtenue en cumulant les noyaux gaussiens en 10 points de l'intervalle  $[0,1]$  avec  $Y_i = 1 \quad \forall i \in \{1, \dots, 10\}$  pour une fenêtre de  $h = 0.1$

Déduisons de (2.3) une définition plus particulière de l'estimateur de Nadaraya-Watson :

$$f_n^{NW}(x) = \frac{1}{nh\hat{p}_n(x)} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right) \quad (2.6)$$

Posons  $\hat{r}_n(x) = \frac{1}{nh} \sum_{i=1}^n Y_i K\left(\frac{X_i - x}{h}\right)$ , nous avons alors

$$f_n^{NW}(x) = \frac{\hat{r}_n(x)}{\hat{p}_n(x)}$$

De plus si la loi de  $X$  est connue, nous pouvons remplacer le terme  $\hat{p}_n(x)$  par  $p(x)$ . Il est aussi possible d'écrire :

$$\hat{r}_n(x) = \sum_{i=1}^n Y_i W_{ni}^{NW}(x)$$

où

$$W_{ni}^{NW}(x) = \frac{1}{nh} K\left(\frac{X_i - x}{h}\right) \quad (2.7)$$

Dans le cas où les  $X_i$  sont déterministes, l'expression est semblable mais en considérant  $p(x) = 1, \forall x \in [0, 1]$ , c'est cette expression que nous utiliserons dans la suite lors du traitement de nos exemples numériques.

## 2.2 Exemples

Le choix du noyau influence l'allure de la courbe modélisée. Par exemple les noyaux rectangulaires se rapprochent des histogrammes tandis que les paraboliques sont plus lisses. Afin de comprendre ce que représentent les poids et leurs interactions avec l'estimateurs, nous pouvons visualiser chaque poids en un point par une "bosse" dans le cas d'un noyau gaussien.

La figure 2.1 représente les poids avec noyau gaussien apportés à chaque points qui, une fois accumulés, donnent l'allure d'une courbe. Il existe plusieurs noyaux possibles comme ceux abordés. Voici quelques exemples de noyaux classiques :

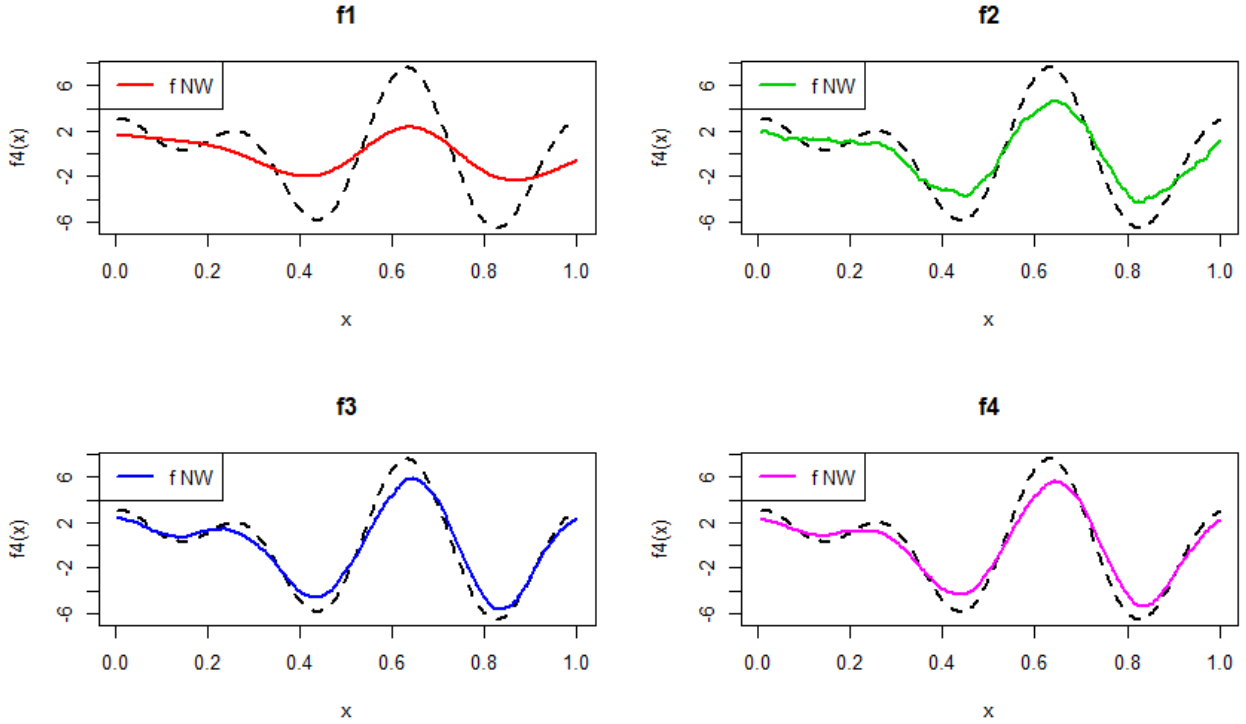


FIGURE 2.2:  $f_4$  est représentée par la fonction en trait plein tandis que les fonctions en pointillées représentent respectivement les estimateurs de Nadaraya-Watson avec un noyau gaussien (haut à gauche), rectangulaire (haut à droite), triangulaire (bas à gauche) et parabolique (bas à droite) pour un certain  $h$  fixé.

- $K(u) = \frac{1}{2} \mathbb{1}_{\{|u| \leq 1\}}$  (noyau rectangulaire)
- $K(u) = (1 - |u|) \mathbb{1}_{\{|u| \leq 1\}}$  (noyau triangulaire)
- $K(u) = \frac{3}{4} (1 - u^2) \mathbb{1}_{\{|u| \leq 1\}}$  (noyau parabolique)
- $K(u) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2})$  (noyau gaussien)

Il est facile de vérifier que chaque noyau présenté est positif et d'intégrale 1.

Pour une fenêtre fixée  $h = 0.1$ , nous visualisons avec la figure 2.2, leur représentation graphique au travers d'un jeu de données simulées par la fonction  $f_4(t) = 4 \sin(4\pi t) + 3 \cos(6\pi t) - 2 \sin(6\pi t)$  pour  $\sigma = 3$  et  $n = 200$  définie dans l'introduction.

Néanmoins le choix de la fenêtre  $h$  influence grandement l'allure finale de la courbe et celle-ci n'est pas connue par avance donc doit être estimée. L'ouvrage de [TSY00] fournit un bon exemple de ce problème dont nous observerons l'effet plus tard sur le risque quadratique. En effet si nous choisissons  $K(u) = \frac{1}{2} \mathbb{1}_{\{|u| \leq 1\}}$ , alors en chaque point  $X_i$  nous obtenons :

$$\left\{ \left| \frac{X_i - x}{h} \right| \leq 1 \right\} = \{X_i \in [x - h, x + h]\}$$

En considérant alors que notre fonction est définie sur  $[0, 1]$ , à  $n$  fixé, les deux cas extrêmes pour le choix de la fenêtre sont :

- si  $h \rightarrow \infty$ , notre indicatrice vaut 1 et on a  $\forall x \in [0, 1], W_{ni}^{NW}(x) = 1/n$  nous obtenons que notre estimateur vaut  $f_n^{NW}(x) = (1/n) \sum_{i=1}^n Y_i$ , c'est-à-dire qu'il est constant et de valeur égale à la moyenne des observations.
- si  $h \rightarrow 0$ , en particulier si  $h < 1/2n$ , alors l'estimateur reproduit les observations  $Y_i$  aux points  $X_i$  et s'annule ailleurs. La variance est alors élevée aux points où l'estimateur est nul.

Il s'agira donc de déterminer une fenêtre optimale qui joue entre minimiser le biais ou la variance, c'est donc une problématique similaire au choix d'un  $N_{opt}$  optimal.

Pour résumer, lors d'une étude de données, il faudra

- Choisir un noyau selon les hypothèses sur la fonction à approcher.
- Approcher la fenêtre optimale qui nous permet de minimiser l'erreur quadratique .
- Expliciter l'estimateur et étudier sa pertinence.

## 2.3 Lien avec estimateur par projection sur la base de Fourier

Rappelons que dans le cadre de la partie précédente nous avons défini la statistique

$$\hat{f}_{n,N}(x) = \sum_{j=1}^N \hat{\theta}_j \phi_j(x)$$

où  $(\phi_j)_{1 \leq j \leq N}$  est la base orthonormée de Fourier dans  $\mathbb{L}_2[0, 1]$  et  $\hat{\theta}_j$  sont les estimateurs des coefficients  $\theta_j$  définis par

$$\hat{\theta}_j = \frac{1}{n} \sum_{i=1}^n Y_i \phi_j(X_i)$$

où nous considérons que la répartition des  $X_i$  est uniforme (i.e  $X_i = i/n$ ). Remarquons que les deux estimateurs  $f_n^{NW}$  et  $\hat{f}_{n,N}$  définis sont tous deux linéaires. De plus il est possible de faire une approximation entre ces deux estimateurs, en effet :

$$\begin{aligned} f_{n,N}(x) &= \sum_{j=1}^N \hat{\theta}_j \phi_j(x) \\ &= \sum_{j=1}^N \left( \frac{1}{n} \sum_{i=1}^n Y_i \phi_j(X_i) \right) \phi_j(x) \\ &= \sum_{i=1}^n Y_i \left( \frac{1}{n} \sum_{j=1}^N \phi_j(X_i) \phi_j(x) \right) \end{aligned}$$

En regroupant les termes de cette façon, nous obtenons que :

$$f_{n,N}(x) = \sum_{i=1}^n Y_i W_{ni}(x) \quad \text{où} \quad W_{ni}(x) = \frac{1}{n} \sum_{j=1}^N \phi_j(X_i) \phi_j(x)$$

Si  $N = 1$  nous avons  $W_{ni}(x) = 1/n$ , nous considérons donc  $N > 1$ . Supposons d'abord le cas où  $N = 2m + 1$  :

$$\begin{aligned}
nW_{ni}(x) &= \sum_{j=1}^N \phi_j(X_i) \phi_j(x) \\
&= 1 + \sum_{j=1}^m \phi_{2j}(X_i) \phi_{2j}(x) + \sum_{j=1}^m \phi_{2j+1}(X_i) \phi_{2j+1}(x) \\
&= 1 + 2 \sum_{j=1}^m \cos(2\pi j X_i) \cos(2\pi j x) + 2 \sum_{j=1}^m \sin(2\pi j X_i) \sin(2\pi j x) \\
&= 1 + 2 \sum_{j=1}^m \cos(2\pi j (X_i - x))
\end{aligned}$$

Nous observons que dans le cas où  $x = X_i = i/n$  nous avons  $W_{ni}(x) = (2m+1)/n = N/n$ . Posons  $a = \pi(X_i - x)$ . Nous avons le résultat général suivant, si  $a \neq 0$  :

$$\begin{aligned}
2 \sum_{j=1}^m \cos(2aj) &= \sum_{j=0}^m (e^{i2a})^j + \sum_{j=0}^m (e^{-i2a})^j - 2 \\
&= \frac{1 - e^{i2a(m+1)}}{1 - e^{i2a}} + \frac{1 - e^{-i2a(m+1)}}{1 - e^{-i2a}} - 2 \\
&= \frac{1 - \cos(2a(m+1)) + \cos(2am) - \cos(2a)}{(1 - \cos(2a))} - 2 \\
&= \frac{\cos(2am) - \cos(2a(m+1))}{1 - \cos(2a)} - 1
\end{aligned} \tag{2.8}$$

L'égalité (2.8) s'obtient en mettant l'expression au même dénominateur et par les formules d'Euler. De là, nous pouvons remarquer que  $\cos(2am) - \cos(2a(m+1)) = 2 \sin(a) \sin(a(2m+1)) = 2 \sin(a) \sin(aN)$  et  $1 - \cos(2a) = 1 - (2 \cos^2(a) - 1) = 2 \sin^2(a)$ . Nous obtenons donc :

$$\frac{\cos(2am) - \cos(2a(m+1))}{1 - \cos(2a)} = 2 \frac{\sin(a) \sin(aN)}{1 - \cos(2a)} = \frac{\sin(aN)}{\sin(a)}$$

En reprenant nos poids  $W_{ni}(x)$ , il en résulte que :

$$W_{ni}(x) = \begin{cases} \frac{1}{n} \frac{\sin(\pi(X_i-x)N)}{\sin(\pi(X_i-x))} & \text{si } x \neq X_i \\ N/n & \text{sinon} \end{cases}$$

Ce résultat n'est pas suffisant car nous n'avons pas explicité une fonction  $K(u)$  pour un  $u$  particulier qui nous permettrait d'extraire un noyau mais nous pouvons observer certains résultats.

Tout d'abord nous observons la continuité en tout point (ce qui n'est pas surprenant étant donné que nous sommes partis d'une combinaison linéaire de fonctions continues), de plus  $X_i - x \in [-1, 1]$  et notre fonction  $W_{ni}(x)$  est 1-périodique ce qui nous empêche d'affirmer qu'il s'agit d'un noyau puisque la condition d'intégrale égale à 1 est non vérifiée. Nous pouvons observer la forme générale des poids (2.3) qui nous permet de constater que ces poids ne sont pas positifs et s'annulent en plusieurs points sur  $[0, 1]$ . Nous pouvons aussi se demander si le pic de la fonction en 0.5 illustré pose un problème dans le cas où  $X_i = 0$  ou 1 par périodicité de la fonction. Néanmoins une des hypothèses de la première partie était la relation  $f(0) = f(1)$  et elle correspond bien à ce cas particulier.

Il est donc clair qu'il n'existe pas de noyau  $K$  tel qu'il y ait un lien entre les deux estimateurs. Néanmoins il est possible d'approcher notre fonction par un noyau, en effet il est possible de noter que

$$W_{ni}(x) = \begin{cases} \frac{\sin(\pi(X_i-x)N)}{\pi(X_i-x)N} \cdot \frac{\pi(X_i-x)N}{\sin(\pi(X_i-x))} \cdot \frac{N}{n} & \text{si } x \neq X_i \\ N/n & \text{sinon} \end{cases}$$

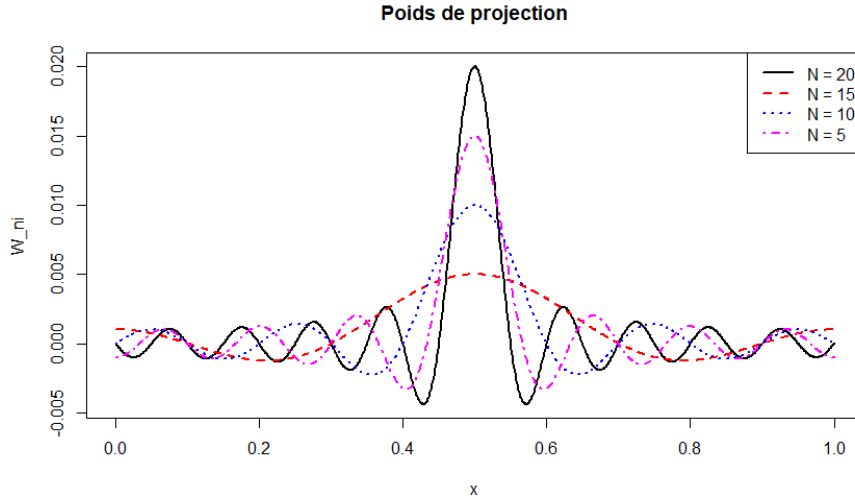


FIGURE 2.3: Représentation du noyau de l'estimateur par projection  $W_{1000,500}(x)$  dans la base de Fourier avec  $n = 1000$ ,  $i = 500$  et  $N = 20, 15, 10$  et  $5$ , il s'agit ici du motif 1-périodique de la fonction. Le parallèle avec la fenêtre  $h$  est clair, le choix d'un  $N$  trop élevé implique une fenêtre petite et se ramène à l'ajustement de l'estimateur de Nadaraya-Watson.

Avec  $h = 1/N$  et  $u = (X_i - x)/h$  on a alors

$$W_{ni}(x) = \frac{1}{nh} K_h\left(\frac{X_i - x}{h}\right) \quad \text{avec} \quad K_h(u) = \frac{\sin(\pi u)h}{\sin(\pi uh)}$$

Une approximation de cette fonction serait le noyau  $K(u) = \frac{\sin(\pi u)}{\pi u}$ . Nous pouvons d'ailleurs vérifier qu'elle est bien intégrable sur  $\mathbb{R}$  et calculer sa valeur :

$$\begin{aligned} \int_{\mathbb{R}} K(u) \, du &= \int_{\mathbb{R}} \frac{\sin(\pi u)}{\pi u} \, du \\ &= \int_{\mathbb{R}} \frac{\sin(x)}{\pi x} \, dx \\ &= \frac{2}{\pi} \int_0^{+\infty} \frac{\sin(x)}{x} \, dx \\ &= 1 \end{aligned} \tag{2.9}$$

L'intégrale de l'équation (2.9) est l'intégrale de Dirichlet, l'intégrande étant la fonction *sinus cardinal* sur la demi-droite des réels positifs. Il s'agit d'une intégrale impropre semi-convergente, c'est-à-dire que la fonction n'est pas intégrable au sens généralisé de Riemann. Ce noyau est donc d'ordre 0 puisque l'intégrale de la fonction  $u \mapsto uK(u)$  n'est pas intégrable sur  $\mathbb{R}$  ce qui n'est pas intéressant pour estimer une fonction par Nadaraya-Watson à l'aide de ce noyau.

## 2.4 Risque quadratique de l'estimateur

Soit  $x_0 \in \mathbb{R}$  fixé. Nous définissons comme dans la première partie le *risque quadratique* d'un estimateur de  $f$  noté  $\hat{f}_n$  par :

$$R(\hat{f}_n) = \mathbb{E} \left[ \int \left( \hat{f}_n(x) - f(x) \right)^2 \, dx \right]$$

Afin d'approcher cette quantité dans le cas de l'estimateur de Nadaraya-Watson, nous poserons les hypothèses suivantes :

- les  $(X_i)$  sont i.i.d de densité  $p$  définie dans la partie (2.2).
- $\mathbb{E}[Y|X=x]$  et  $\mathbb{V}[Y|X=x]$  existent et sont finies.
- $f$  et  $p$  sont bornées et de dérivées continues.
- $h \rightarrow 0$  et  $nh \rightarrow \infty$  quand  $n \rightarrow +\infty$ .
- $K$  est d'ordre 2.

Nous avons défini  $\hat{f}_n$  par la relation 2.1. Afin de calculer le risque quadratique, nous pouvons réécrire notre différence

$$\begin{aligned}\hat{f}_n(x) - f(x) &= \frac{\hat{r}_n(x)}{\hat{p}_n(x)} - f(x) \\ &= \left( \frac{\hat{r}_n(x)}{\hat{p}_n(x)} - f(x) \right) \left( \frac{\hat{p}_n(x)}{p(x)} + \left( 1 - \frac{\hat{p}_n(x)}{p(x)} \right) \right) \\ &= \frac{\hat{r}_n(x) - f(x)\hat{p}_n(x)}{p(x)} + \frac{(\hat{f}_n(x) - f(x))(p(x) - \hat{p}_n(x))}{p(x)}\end{aligned}$$

Ce jeu de réécriture est proposé par L. WASSERMAN ([WAS]). Le second terme est moins dominant que le premier à cause de la répétition de  $\hat{f}_n(x) - f(x)$  (en réitérant l'opération). Ainsi nous pouvons approcher notre différence par le premier terme. Ainsi par le théorème de König-Huygens :

$$R(\hat{f}_n) = \int \mathbb{E} \left[ \hat{f}_n(x) - f(x) \right]^2 + \mathbb{V} \left[ \hat{f}_n(x) - f(x) \right] dx$$

Nous nous contenterons de donner une approximation de ce risque plutôt que de le majorer afin d'étudier son comportement en fonction de la fenêtre  $h$ . Pour cela nous approcherons l'espérance et la variance du terme dominant.

$$\begin{aligned}\mathbb{E}[\hat{r}_n(x)] &= \frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] \quad \text{car i.i.d} \\ &= \frac{1}{h} \mathbb{E} \left[ f(X_1) K \left( \frac{X_1 - x}{h} \right) \right] \quad \text{car } \mathbb{E}[\epsilon_1] = 0 \\ &= \frac{1}{h} \int f(u) K \left( \frac{u - x}{h} \right) p(u) du \\ &= \int f(x + th) K(t) p(x + th) dt\end{aligned}$$

Par un développement de Taylor d'ordre 2 de  $f$  et  $p$  on a :

$$\mathbb{E}[\hat{r}_n(x)] \approx \int \left( f(x) + thf'(x) + \frac{t^2 h^2}{2} f''(x) \right) K(t) \left( p(x) + thp'(x) + \frac{t^2 h^2}{2} p''(x) \right) dt$$

En développant l'expressions intégrées puis en négligeant les termes  $O(h^3)$ , nous obtenons une expression simplifiable par hypothèses sur l'ordre de  $K$  tel que  $\int K(t)dt = 1$  et  $\int tK(t)dt = 0$ . En notant  $\mu_2(K) = \int t^2 K(t)dt$ , il en résulte que :

$$\mathbb{E}[\hat{r}_n(x)] \approx p(x)f(x) + \frac{h^2}{2} f(x)p''(x)\mu_2(K) + \frac{h^2}{2} f''(x)p(x)\mu_2(K) + h^2 f'(x)p'(x)\mu_2(K)$$

D'autre part, avec 2.5, nous avons :

$$\mathbb{E}[f(x)\hat{p}_n(x)] \approx p(x)f(x) + \frac{h^2}{2} f(x)p''(x)\mu_2(K)$$

Ainsi

$$\mathbb{E} \left[ \hat{f}_n(x) - f(x) \right] \approx \mathbb{E} \left[ \frac{\hat{r}_n(x) - f(x)\hat{p}_n(x)}{p(x)} \right] \approx \frac{h^2}{2} \left( f''(x) + 2 \frac{f'(x)p'(x)}{p(x)} \right) \mu_2(K)$$

Nous n'avons pas besoin d'approfondir le calcul si l'on suppose que  $K$  est d'ordre 2, néanmoins lorsque celui-ci est plus élevé nous devons approfondir le calcul sur des hypothèses de régularité sur  $f$  et  $p$ . De plus comme nous considérons que  $h \rightarrow 0$  quand  $n \rightarrow \infty$ , il n'est pas nécessaire d'approfondir les calculs à l'ordre 3 si  $\mu_2(K) \neq 0$ .

Effectuons le calcul de la variance par le même procédé :

$$\mathbb{V} \left[ \hat{f}_n(x) - f(x) \right] \approx \frac{\mathbb{V}[\hat{r}_n(x)] + \mathbb{V}[f(x)\hat{p}_n(x)] - 2Cov(\hat{r}_n(x), f(x)\hat{p}_n(x))}{p(x)^2}$$

En rappelant que  $\sigma^2 = \mathbb{V}[\epsilon_i]$ , séparons le calcul de chaque terme. Notons tout d'abord les résultats suivants où l'approximation est dû au développement de Taylor comme réalisé dans le calcul du biais.

$$\begin{aligned} \frac{1}{h} \mathbb{E} \left[ Y_1^2 K \left( \frac{X_1 - x}{h} \right)^2 \right] &= \frac{1}{h} \mathbb{E} \left[ f(X_1)^2 K \left( \frac{X_1 - x}{h} \right)^2 \right] + \frac{1}{h} \mathbb{E} \left[ \epsilon_1^2 K \left( \frac{X_1 - x}{h} \right)^2 \right] \\ &= \frac{1}{h} \int f(u)^2 K \left( \frac{u - x}{h} \right)^2 p(u) du + \frac{\sigma^2}{h} \int f(u)^2 K \left( \frac{u - x}{h} \right)^2 p(u) du \\ &= \int f(x + th)^2 K(t)^2 p(x + th) dt + \sigma^2 \int f(x + th)^2 K(t)^2 p(x + th) dt \\ &= (f(x)^2 + \sigma^2) p(x) \|K\|_2^2 + O(h^2) \end{aligned}$$

De même :

$$\begin{aligned} \frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] &= \frac{1}{h} \mathbb{E} \left[ f(X_1) K \left( \frac{X_1 - x}{h} \right) \right] \\ &= \frac{1}{h} \int f(u) K \left( \frac{u - x}{h} \right) p(u) du \\ &= \int f(x + th) K(t) p(x + th) dt \\ &= f(x) p(x) + O(h^2) \text{ par définition de } K \end{aligned}$$

De la même façon nous obtenons les approximations

$$\frac{1}{h} \mathbb{E} \left[ K \left( \frac{X_1 - x}{h} \right)^2 \right] = p(x) \|K\|_2^2 + O(h^2)$$

et

$$\frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right)^2 \right] = f(x) p(x) \|K\|_2^2 + O(h^2)$$

Remarquons a que nous nous arrêtons qu'à un développement d'ordre 1 dans ces calculs. En effet comme  $h \rightarrow 0$  nous pouvons considérer que  $h < 1/h$  et donc  $o\left(\frac{1}{nh}\right)$  est un ordre plus grand que  $o\left(\frac{h^2}{nh}\right)$ . De ces résultats intermédiaires, nous avons alors :

$$\begin{aligned}
\mathbb{V}[\hat{r}_n(x)] &= \frac{1}{nh^2} \mathbb{V} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] \\
&= \frac{1}{nh} \left( \frac{1}{h} \mathbb{E} \left[ Y_1^2 K \left( \frac{X_1 - x}{h} \right)^2 \right] \right) - \frac{1}{n} \left( \frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] \right)^2 \\
&= \frac{1}{nh} (f(x)^2 + \sigma^2) p(x) \|K\|_2 - \frac{1}{n} p(x)^2 f(x)^2 + o \left( \frac{1}{nh} \right)
\end{aligned} \tag{2.10}$$

Et

$$\begin{aligned}
\mathbb{V}[f(x)\hat{p}_n(x)] &= \frac{f(x)^2}{nh^2} \mathbb{V} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] \\
&= \frac{f(x)^2}{nh} \left( \frac{1}{h} \mathbb{E} \left[ K \left( \frac{X_1 - x}{h} \right)^2 \right] \right) - \frac{f(x)^2}{n} \left( \frac{1}{h} \mathbb{E} \left[ K \left( \frac{X_1 - x}{h} \right) \right] \right)^2 \\
&= \frac{1}{nh} f(x)^2 p(x) \|K\|_2 - \frac{1}{n} p(x)^2 f(x)^2 + o \left( \frac{1}{nh} \right)
\end{aligned} \tag{2.11}$$

Enfin,

$$\begin{aligned}
Cov(r_n(x), \hat{p}_n(x)) &= \mathbb{E}[\hat{r}_n(x)\hat{p}_n(x)] - \mathbb{E}[\hat{r}_n(x)]\mathbb{E}[\hat{p}_n(x)] \\
&= \mathbb{E} \left[ \frac{1}{n^2 h^2} \sum_{i,j=1}^n Y_i K \left( \frac{X_i - x}{h} \right) K \left( \frac{X_j - x}{h} \right) \right] - p(x)^2 f(x) + o \left( \frac{1}{nh} \right) \\
&= \frac{1}{nh} \underbrace{\frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right)^2 \right]}_{\approx f(x)p(x)\|K\|_2} + \frac{n(n-1)}{n^2} \underbrace{\left( \frac{1}{h} \mathbb{E} \left[ Y_1 K \left( \frac{X_1 - x}{h} \right) \right] \right)}_{\approx f(x)p(x)} \underbrace{\left( \frac{1}{h} \mathbb{E} \left[ K \left( \frac{X_1 - x}{h} \right) \right] \right)}_{\approx p(x)} - p(x)
\end{aligned} \tag{2.12}$$

En simplifiant les membres nous obtenons donc :

$$Cov(\hat{r}_n(x), \hat{p}_n(x)) = \frac{1}{nh} f(x)p(x)\|K\|_2 - \frac{1}{n} p(x)^2 f(x) + o \left( \frac{1}{nh} \right)$$

En injectant les approximations des égalités 2.10, 2.11 et 2.4 dans 2.4, nous obtenons finalement :

$$\mathbb{V}[\hat{f}_n(x) - f(x)] = \frac{\sigma^2}{nhp(x)} \|K\|_2 + o \left( \frac{1}{nh} \right)$$

Rassemblons l'étude du biais et de la variance afin d'approcher le risque quadratique défini en 2.4 par :

$$R(\hat{f}_n) \approx \frac{h^4}{4} \mu_2^2(K) \int f''(x) + 2f'(x)p'(x)p^{-1}(x) dx + \frac{\sigma^2}{nh} \|K\|_2 \int p^{-1}(x) dx$$

Nous en déduisons que la fenêtre  $h$  optimale qui minimise le risque quadratique est :

$$h_{opt} = \left[ \frac{\sigma^2 \|K\|_2^2 \int p^{-1}(x) dx}{\int \{2f'(x)p'(x)p^{-1}(x) + f''(x)\}^2 dx \mu_2^2(K)} \right]^{1/5} n^{-1/5}$$

Dans la pratique il est possible d'approcher cette valeur optimale par diverses méthodes :



- Par la méthode du "plug-in" qui consiste à remplacer les inconnues telles que les fonctions  $f'$  ou  $p'$  par des estimateurs associés.
- Par validation croisée en considérant

$$h_{CV} = \underset{h}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^N (Y_i - \widehat{f^{-i}}(X_i))^2$$

Le package KernSmooth [KER20] nous permet d'approcher une fenêtre  $h$  par la méthode du "plug-in" pour le noyau gaussien. Pour illustrer le comportement de notre estimateur par des données, nous utiliserons les fonctions présentées dans l'introduction. Connaissant la fonction d'origine  $f$ , nous pouvons étudier l'évolution du risque quadratique selon les valeurs de  $h$ . La figure 2.4 exprime ce risque, nous pouvons remarquer que la fenêtre optimale est relativement bien estimée par la fonction dpill fournie par le package. Nous noterons aussi que les graphes ont une forme convexe qui illustre le dilemme illustré dans la section exemple où le choix d'un  $h$  trop petit reproduit les données tandis qu'une valeur trop élevée se rapproche d'une moyenne des données.

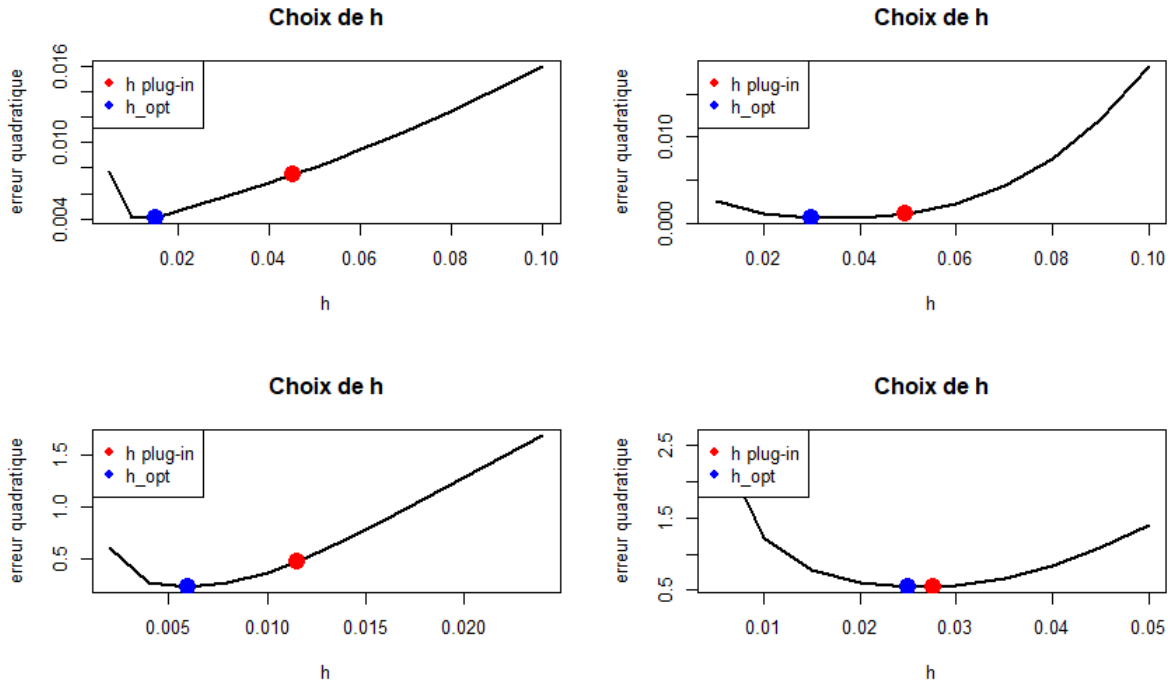


FIGURE 2.4: Évolution du risque quadratique  $R(\hat{f}_n)$  pour les fonctions  $f_1$  (haut à gauche),  $f_2$  (haut à droite),  $f_3$  (bas à gauche) et  $f_4$  (bas à droite). En bleu la fenêtre qui minimise ce risque et en rouge celui fournit par la fonction dpill. On notera que l'échelle est différente selon les fonctions afin d'avoir un aperçu plus détaillé du risque.

La figure 2.5 nous permet de visualiser l'estimateur de Nadaraya-Watson appliqué à nos fonctions de l'introduction, elle illustre aussi la régularité de l'estimateur. Celui obtenu pour un  $h$  optimal qui minimise le risque quadratique semble moins lisse que pour le  $h$  obtenu par "plug-in". En effet le  $h$  que l'on estime par la formule 2.4 suppose une régularité sur la fonction  $f$  et  $p$ . Dans la pratique nous avons considéré l'estimateur 2.7 avec  $p(x) = 1$ , puisqu'ici nous pouvons supposer que la répartition de la variable  $X$  uniforme.

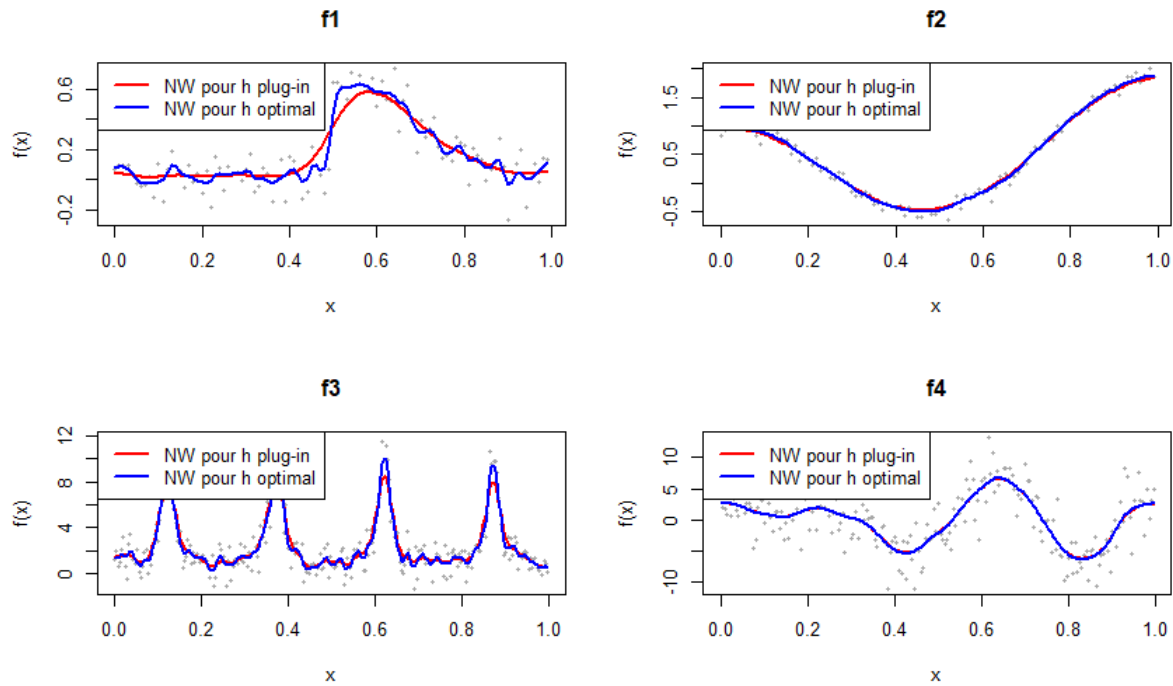


FIGURE 2.5: Estimateur de Nadaraya-Watson pour les fonctions  $f_1$  (haut à gauche),  $f_2$  (haut à droite),  $f_3$  (bas à gauche) et  $f_4$  (bas à droite). En bleu l'estimateur calculé à partir du  $h$  obtenu précédemment et en rouge à partir de celui fournit par la fonction `dpill`.

Nous remarquerons pour la  $f_2$  que l'estimateur n'a pas de difficulté pour les fonctions à caractère non périodique contrairement à l'estimateur par projection. Nous illustrerons d'avantage la question du choix de l'estimateur dans la partie suivante où nous pourrons traiter un même problème sous deux angles différents.

## Chapitre 3

# Applications numériques

### 3.1 Étude des moyennes mensuelles de température

Tout d'abord nous allons nous intéresser aux moyennes mensuelles de température de Nottingham entre 1920 et 1939. Comme l'illustre la figure 3.1, le nuage de points semble être anarchique s'il est pris hors contexte. En effet la température varie de façon périodique en fonction des mois, supposer à tort l'indépendance entre ces deux variables (mois et température) serait une erreur. Afin de donner  $N$  optimal nous devons tout d'abord estimer la variance des données. Nous remarquons que le graphe sépare en trois phases selon  $N$ , nous allons donc considérer les valeurs de  $\sigma$  aux alentours de  $\sigma = 2$  et  $\sigma = 8$ . Le tableau suivant nous révèle trois valeurs de  $N$  différentes : 237, 82, 42. La première

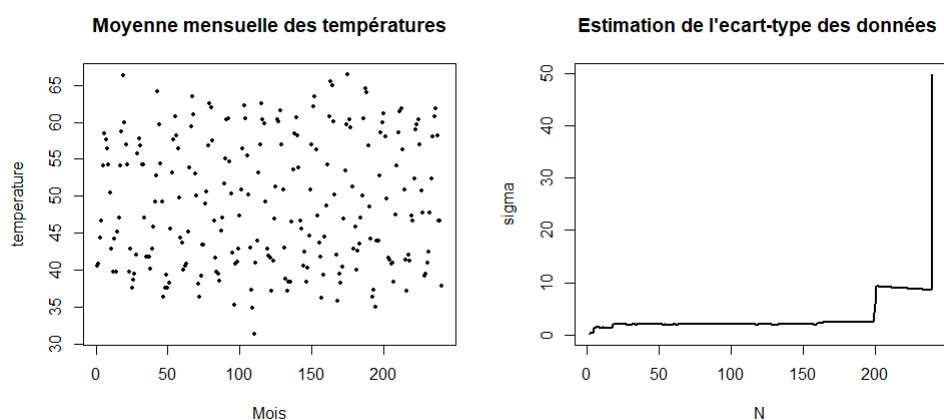


FIGURE 3.1: Représentation des données par défaut sur Rstudio (nottem) qui représente la moyenne des températures de Nottingham en fonction des mois. À gauche la répartition de l'échantillon et à droite l'estimation de la variance

étant le cas où l'estimateur reproduit les données et n'est donc pas intéressant car égale à la taille de l'échantillon. Nous retiendrons donc 82 et 42.

$\sigma$	1	2	3	4	8	9	10
$N_{mallows}$	237	82	42	42	42	42	42

La figure 3.2 représente en haut les fonctions estimées des données. À première vue il semble qu'il n'y ait pas de différences flagrantes selon les valeurs de  $\sigma$  ou  $N$  choisies. Néanmoins le graphe du couple des résidus en fonction des valeurs de la fonction estimée pour  $s = 4$  traduit une légère tendance

parabolique qui montre que le modèle n'est pas représentatif. En effet l'estimateur par projection pour  $s = 2$  représente d'avantage les données puisque la répartition des couples est uniforme sans tendance.

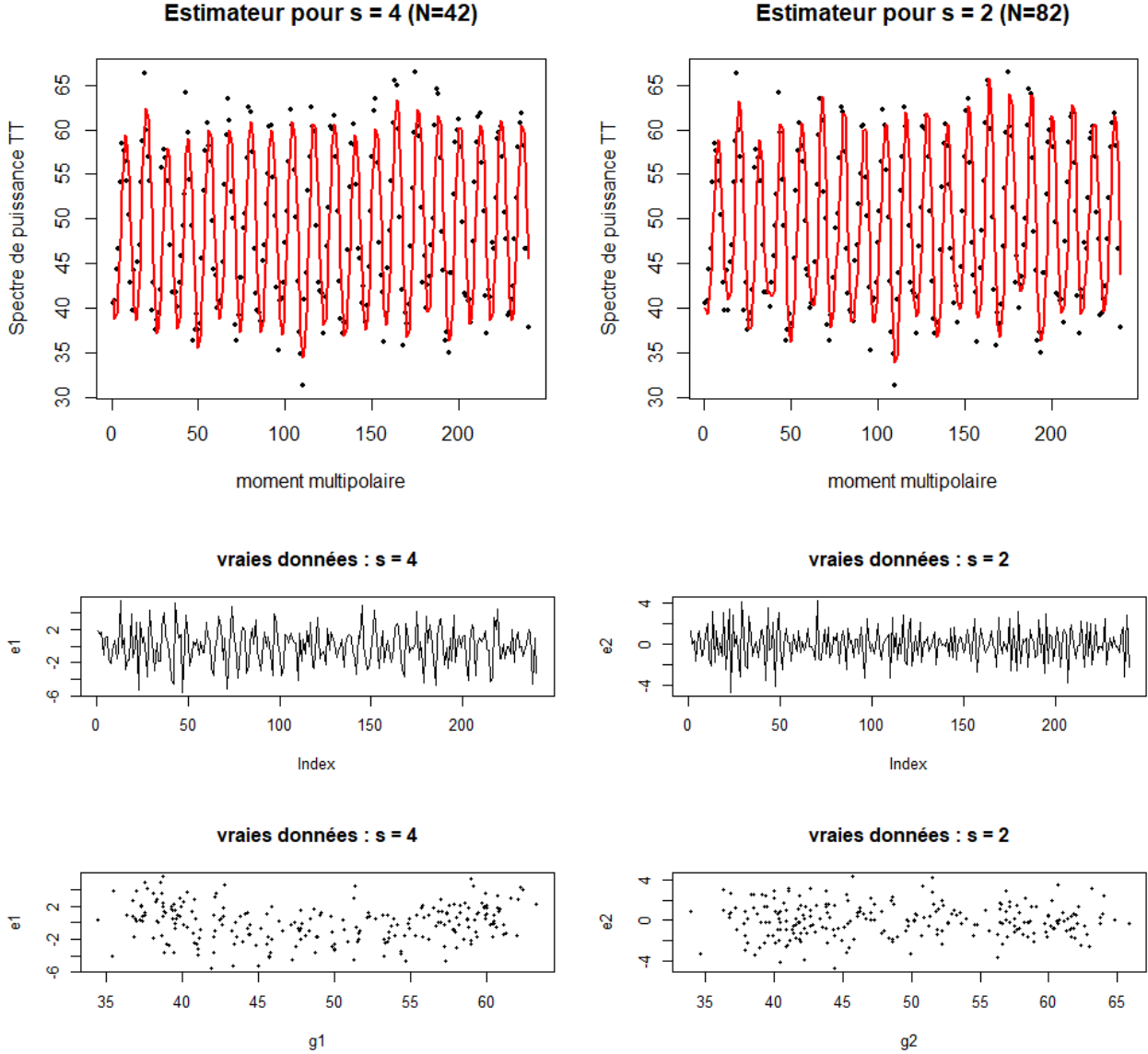


FIGURE 3.2: Etude de l'estimateur de projection qui sépare selon les valeurs de  $\sigma$  (ou  $N$ ) choisies, à gauche pour  $s = 4$  (i.e  $N = 82$ ) et à droite  $s = 2$  (i.e  $N = 42$ ). En haut est représentée l'allure de l'estimateur de projection par rapport aux données, au centre l'erreur entre les données réelles et celles de l'estimateur et en bas les couples  $(\hat{f}(X_i), \hat{e}(X_i))$  pour  $k = 1, \dots, n$

Nous allons maintenant étudier ce jeu de données (centré sur la moyenne) avec l'estimateur de Nadaraya-Watson en utilisant le  $h$  calculé par "plug-in" pour le noyau gaussien. La figure 3.3 (haut) permet de constater que l'estimateur n'est pas fiable dans le cas où il n'existe pas de tendance particulière, en effet le nuage de points est semblable au cas où les variables  $X$  et  $Y$  sont indépendantes, ainsi :

$$f(x) = \mathbb{E}[Y|X = x] = E[Y]$$

Ce qui est cohérent avec le résultat obtenu puisque  $\mathbb{E}[Y] \in [48.86; 49.22]$  à 95%. Néanmoins l'estimateur est plus fidèle aux données lorsque nous modifions manuellement  $h \approx 0.002$ , l'estimation de  $h$  par "plug-in" est effectué sous hypothèses que la fonction soit lisse ce qui justifie d'autant plus ce phénomène. Ainsi l'étude des graphes des résidus revient à étudier le comportement des données autour de leur moyenne, le graphe des erreurs (bas à gauche) montre une périodicité dans celles-ci, ce qui nous confirme l'idée que notre estimateur est insuffisant pour expliquer les données.

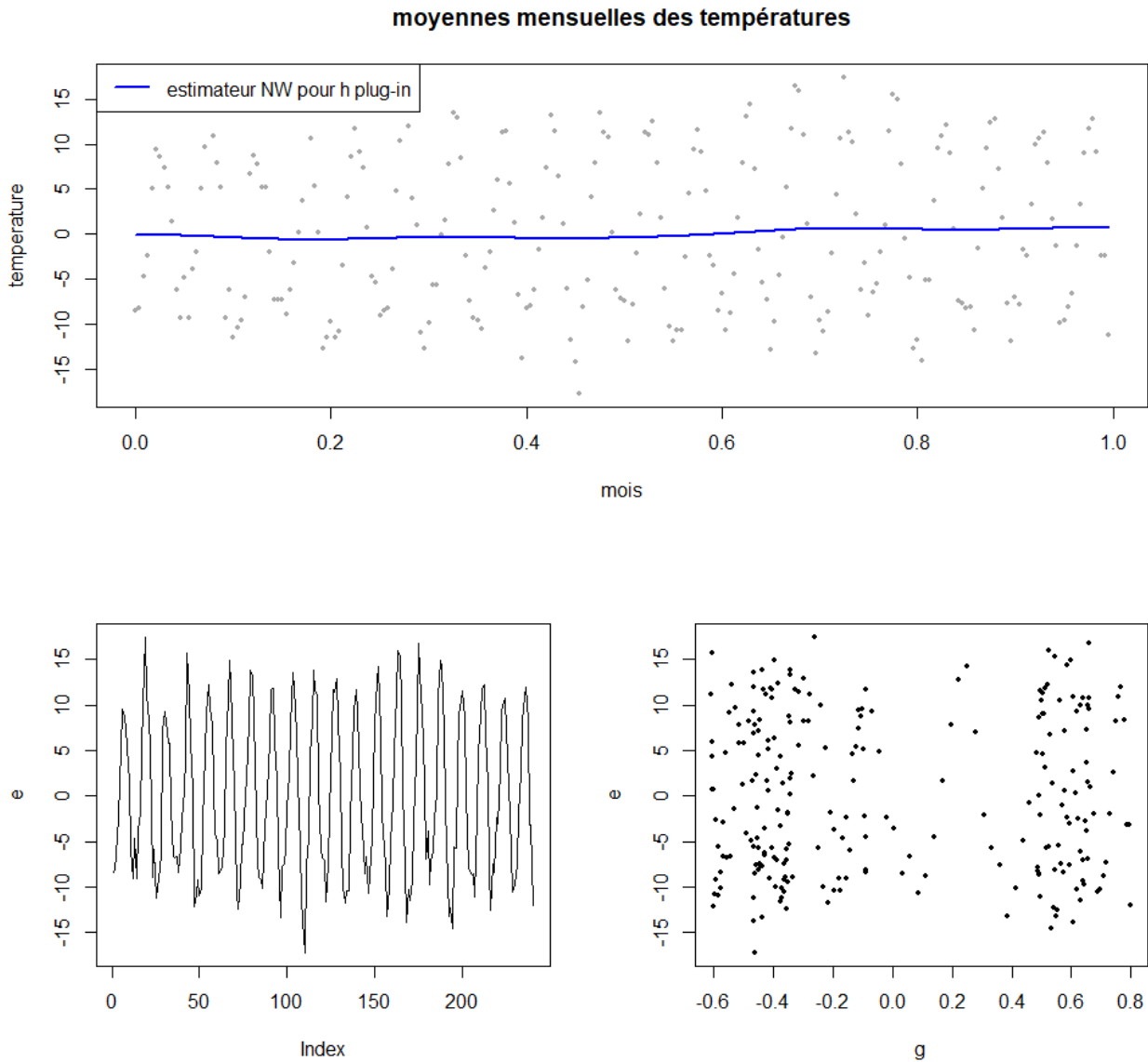


FIGURE 3.3: Aperçu de l'étude du jeu de données nottem pour l'estimateur de Nadaraya-Watson. On représente ici l'estimateur obtenu pour le  $h$  fournit par dpill (haut), l'erreur entre l'estimateur et les données réelles (bas à gauche) et le nuage de points  $(\hat{f}_n(X_i), \hat{e}(X_i))$  (bas à droite).

Pour conclure l'étude de ces données, nous pouvons facilement comprendre que le choix de l'estimateur dépend fortement de l'allure à priori et du contexte dans lesquels les données sont fournies. Conjeturons alors que l'estimateur par projection sur des fonctions périodiques est plus efficace que l'estimateur de Nadaraya-Watson qui a besoin de régularité dans la fonction à estimer.

### 3.2 Étude des données prélevées par la WMAP

Nous nous intéressons à un jeu de données récupéré sur le site <https://lambda.gsfc.nasa.gov/>. Il s'agit de données prélevées la première année par la sonde d'anisotropie micro-ondes Wilkinson (WMAP) où nous représentons le spectre de puissance TT en fonction des moments multipolaires. On remarquera que la dispersion des données présentée sur la figure 3.4 n'est pas uniforme. Cela nous permettra de remarquer comment se comportent nos estimations dans ce cas.

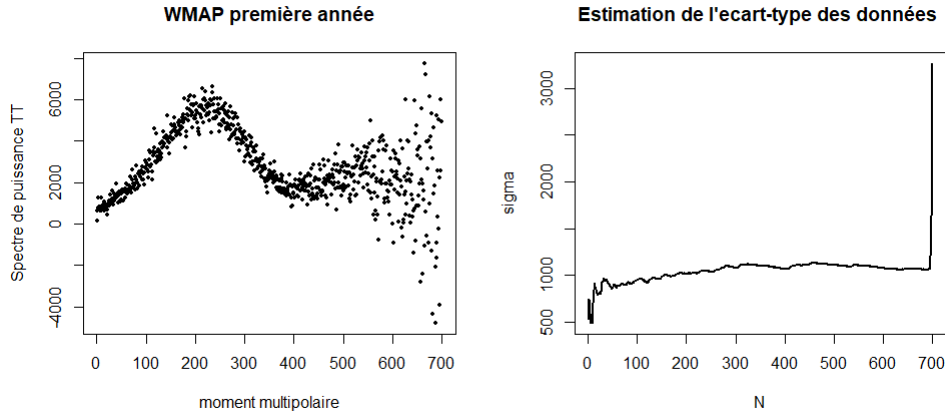


FIGURE 3.4: Données prélevées sur le site de la NASA (<https://lambda.gsfc.nasa.gov/>) qui représente le spectre de puissance TT en fonction des moments multipolaires (gauche) et l'estimation de sa variance en fonction de  $N$  (droite).

De façon analogue à l'étude des températures réalisées dans la première partie, nous allons devoir estimer la valeur de  $\sigma$  supposée finie afin d'estimer un  $N_{opt}$  optimal pour l'estimateur de projection. De cette étude nous remarquons que les valeurs de  $\sigma$  semblent être stabilisées entre 900 et 1200.

Le tableau ci-dessous nous suggère un  $N_{mallows}$  optimal qui s'oriente vers deux valeurs : 7 et 14. Pour chacune d'elles il est possible de représenter la fonction et étudier l'erreur absolue entre les données et la fonction estimée ainsi que la répartition des erreurs.

$\sigma$	900	950	1000	1050	1100	1150	1200
$N_{mallows}$	15	14	14	7	7	7	7

L'étude présentée par la figure 3.5 nous montre un résultat satisfaisant. Nous remarquerons que l'hypothèse  $f(0) = f(1)$  crée une légère perturbation de l'estimateur à l'origine. De plus les graphes des résidus montrent bien que l'hypothèse d'homoscédasticité n'est pas respectée. Néanmoins il n'y a aucune tendance ou périodicité particulière. Enfin les graphes des couples montre que les faibles valeurs de  $\hat{f}$  sont plus réparties que celles les plus élevées. Ceci est également induit par un variance non constante des données.

Le modèle est donc imparfait et des hypothèses plus solide seraient à introduire tel que

$$Y_i = f(X_i) + \epsilon_i, \quad \mathbb{E}[\epsilon_i] = 0 \text{ et } \mathbb{V}[\epsilon_i] = \sigma_i, \quad i = 1, \dots, n$$

que nous n'aborderons pas ici. Néanmoins ce jeu de données nous permet de tester l'estimateur de Nadaraya-Watson. Ici les données semblent avoir une tendance particulière contrairement à celles vues

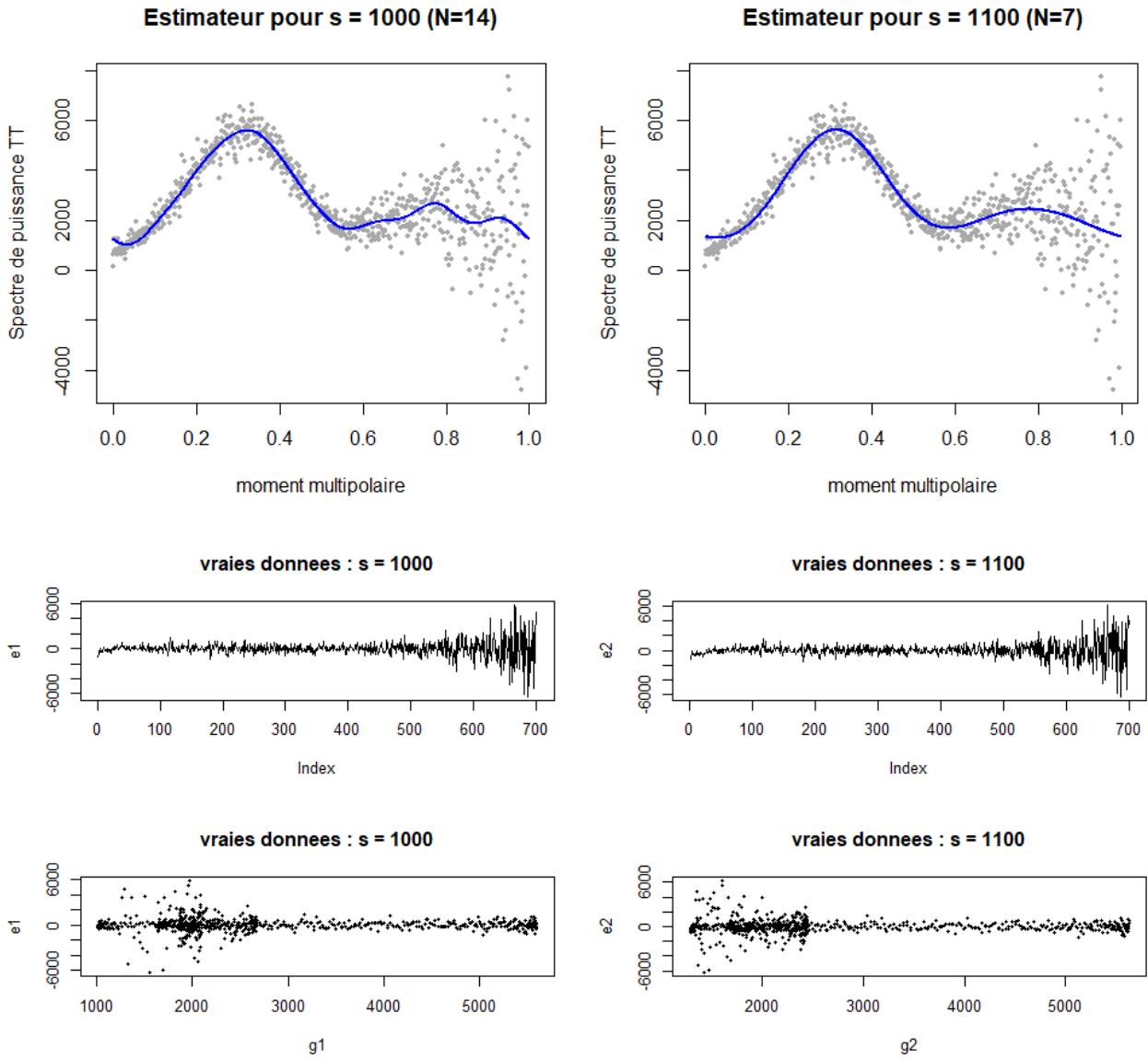


FIGURE 3.5: Etude de l'estimateur de projection qui sépare selon les valeurs de  $\sigma$  (ou  $N$ ) choisies, à gauche pour  $s = 1000$  (i.e  $N = 14$ ) et à droite  $s = 1100$  (i.e  $N = 7$ ). En haut est représenté l'allure de l'estimateur de projection par rapport aux données, au centre l'erreur entre les données réelles et celles de l'estimatrice et en bas les couples  $(\hat{f}(X_i), \hat{e}(i))$  pour  $k = 1, \dots, n$

précédemment. Ainsi nous pouvons comparer visuellement les résultats fournis par les deux estimateurs à l'aide de nos fonctions utilisées dans la partie précédente. La figure 3.6 nous donne une synthèse de notre étude.

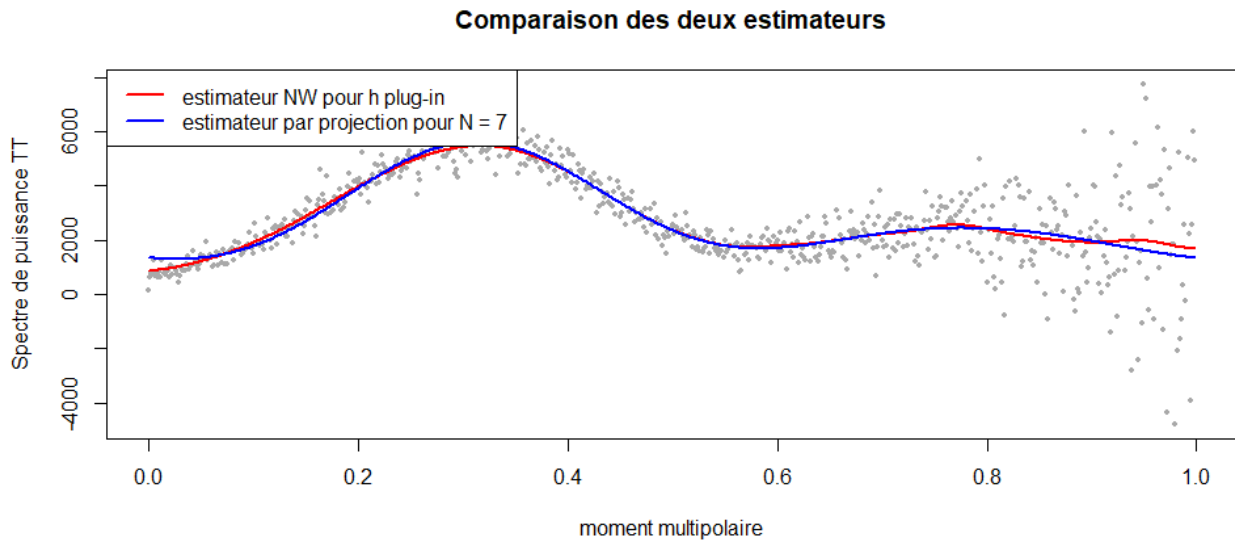


FIGURE 3.6



# Conclusion

Nous avons pu dans ce mémoire découvrir deux méthodes d'estimation pour le modèle de régression mais cette liste n'est pas exhaustive. Nous pourrions choisir différentes bases de projection ou d'autres noyaux comme celui introduit dans la section 2. Il existe d'autres méthodes comme l'estimation par polynômes locaux qui utilisent des hypothèses de régularités de  $f$ .

Nous avons, dans un second temps pu appliquer nos recherches sur des jeux de données réelles. Nous retiendrons des deux estimateurs abordés que :

- Si nous possédons un indice sur l'existence d'une périodicité et que les données semblent vérifier  $f(0) = f(1)$ , alors nous choisirons l'estimateur par projection.
- Si nous n'avons pas d'a priori sur la fonction ou qu'elle ne vérifie pas  $f(0) = f(1)$  alors l'estimateur de Nadaraya-Watson sera pertinent.

Nous pouvons aussi noter qu'il est possible de généraliser ce dernier dans de plus grandes dimensions ce qui permet de traiter des domaines très divers. Néanmoins comme beaucoup d'estimateurs, les grandes dimensions dispersent d'avantage les données et nous pouvons nous retrouver dans le cas de la figure 3.5 où l'estimateur passe à côté d'une information importante.



# Bibliographie

- [RIV09] V. RIVOIRARD et G. STOLTZ, *Statistique mathématique en action*. Edition Vuibert, Novembre 2009.
- [TSY00] A. TSYBAKOV, *Introduction à l'estimation non paramétrique*. Edition Springer, 2000.
- [DIO10] C. DION, *Estimation non-paramétrique de la densité de variables aléatoires cachées*. Université Grenoble Alpes, 2010.
- [LOU10] S. LOUSTAU, *Estimation non-paramétrique et Apprentissage statistique*. <https://www.math.univ-angers.fr/~loustau/larochelle.pdf>, 2010.
- [CHA20] G. CHAGNY, *Estimation non paramétrique d'une fonction de régression avec des bases déformées : le point de vue de la sélection de modèles*. <https://hal.archives-ouvertes.fr/hal-00519556v1/document>, 2020.
- [MAT12] C. MATIAS, *Introduction à la statistique non paramétrique*. [http://cmatias.perso.math.cnrs.fr/Docs/atelier\\_stat\\_np\\_3\\_estim\\_densite.pdf](http://cmatias.perso.math.cnrs.fr/Docs/atelier_stat_np_3_estim_densite.pdf), 2012.
- [KER20] M. WANS, *Package 'KernSmooth'*. <https://cran.r-project.org/web/packages/KernSmooth/KernSmooth.pdf>, 2020.
- [LEJ85] M. LEJEUNE *Estimation non-parametrique par noyaux : regression polynomiale mobiles*. <https://cran.r-project.org/web/packages/KernSmooth/KernSmooth.pdf>, 1985.
- [WAS] L. WASSERMAN *Nonparametric Regression*. <http://www.stat.cmu.edu/~larry/=sml/NonparRegression.pdf>,
- [BON14] M. LEJEUNE *Estimation non parametrique : Quelques (bonnes ?) pratiques dans l'R*. <http://www.thibault.laurent.free.fr/presentation2013-05-14/PratiquesNP2.pdf>. 2014.
- [WIK] *Apprentissage non parametrique en regression*. <https://www.math.univ-toulouse.fr/~besse/Wikistat/pdf/st-m-app-non-param.pdf>.



# Annexes

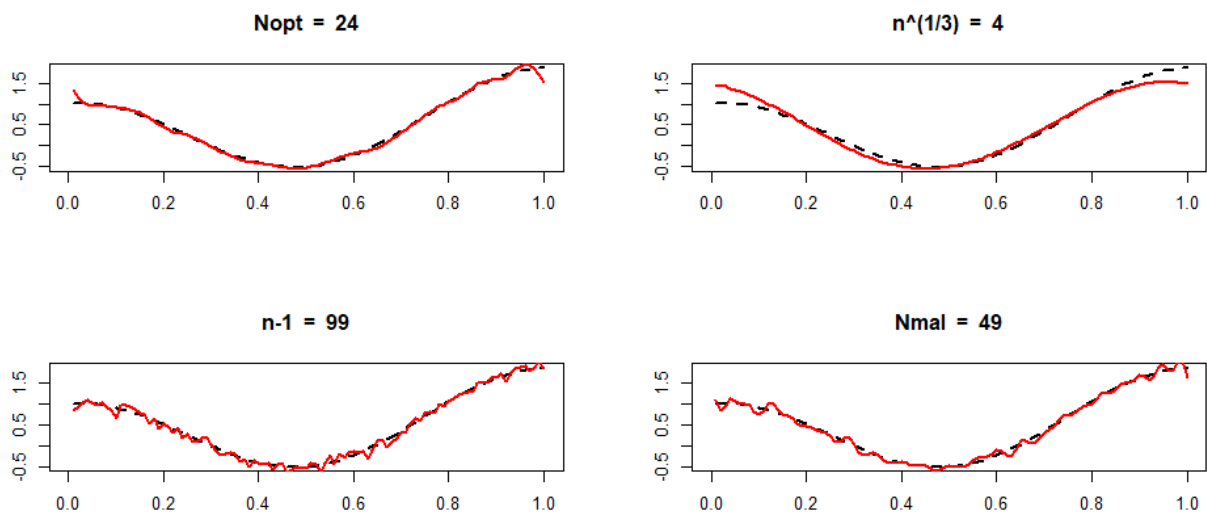


FIGURE 7: Estimation par projection de la fonction  $f_2$  pour les différentes valeurs de  $N$  définis

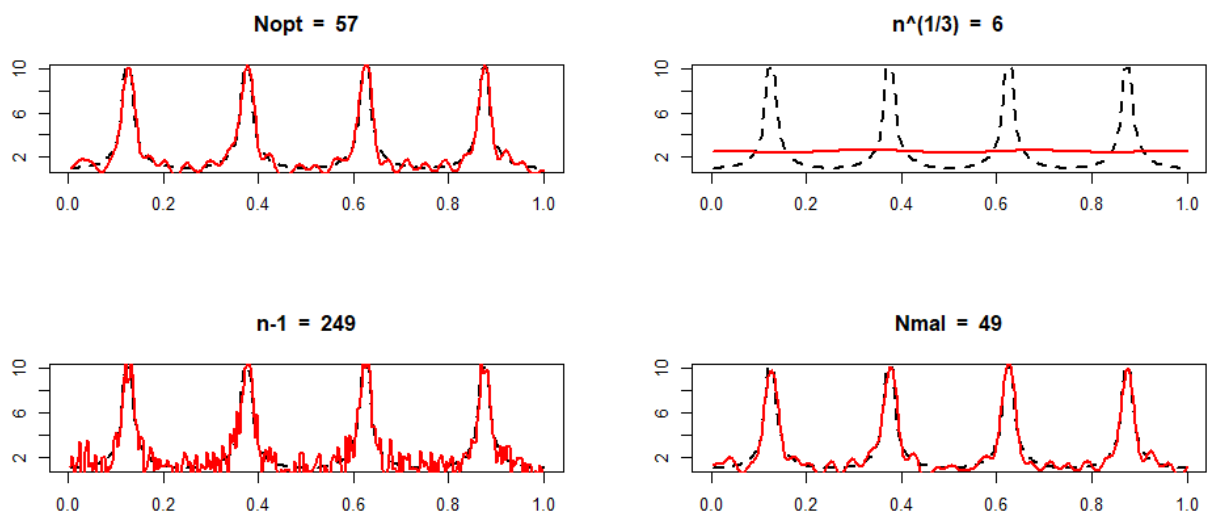


FIGURE 8: Estimation par projection de la fonction  $f_3$  pour les différentes valeurs de  $N$  définis

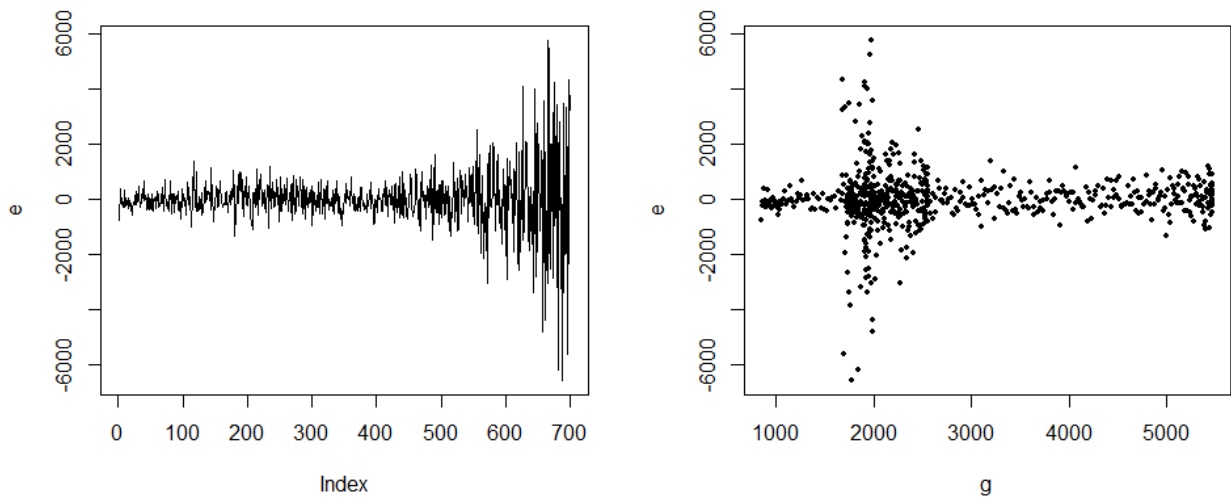


FIGURE 9: Étude des résidus pour l'estimateur de Nadaraya-Watson pour les données de WMAP de première année

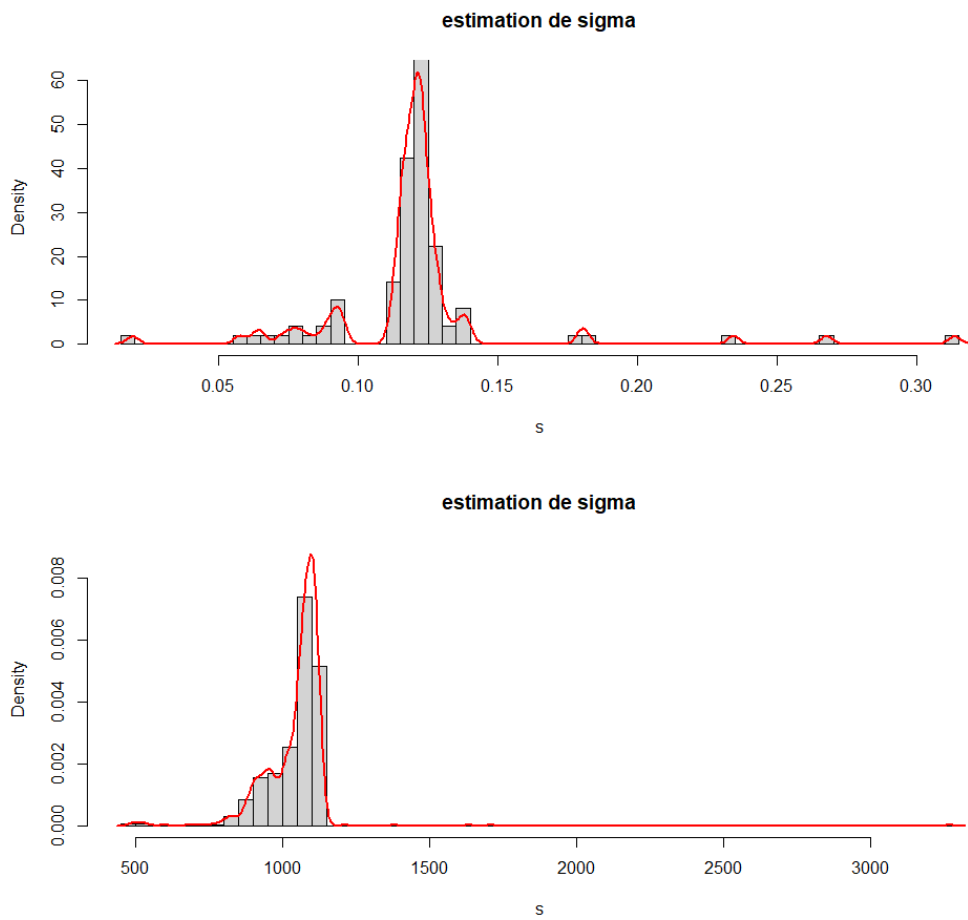


FIGURE 10: Estimation de sigma par une autre représentation pour les données simulées par  $f_1$  (haut) et pour les données de WMAP (bas)