

République Algérienne Démocratique et  
Populaire

Ministère de l'Enseignement Supérieur et de la Recherche  
Scientifique

UNIVERSITE MENTOURI-CONSTANTINE  
FACULTE DES SCIENCES EXACTES  
DEPARTEMENT DE MATHÉMATIQUES

N° série :  
N° d'ordre :

## Mémoire

Présenté pour obtenir le diplôme de Magistère  
En Mathématiques

### THEME

ESTIMATION ET TEST DANS LES MODELES  
DE REGRESSION NON PARAMETRIQUE

Option :

Mathématiques Appliquées

Présenté par :

MERABET Farida

Devant le jury :

Président : F.L. Rahmani  
Rapporteur : Z. Mohdeb  
Examineur : F. Messaci  
Examineur : N. Namouchi

Professeur à l'université de Constantine  
Professeur à l'université de Constantine  
Maitre de conférences à l'université de Constantine  
Maitre de conférences à l'université de Constantine

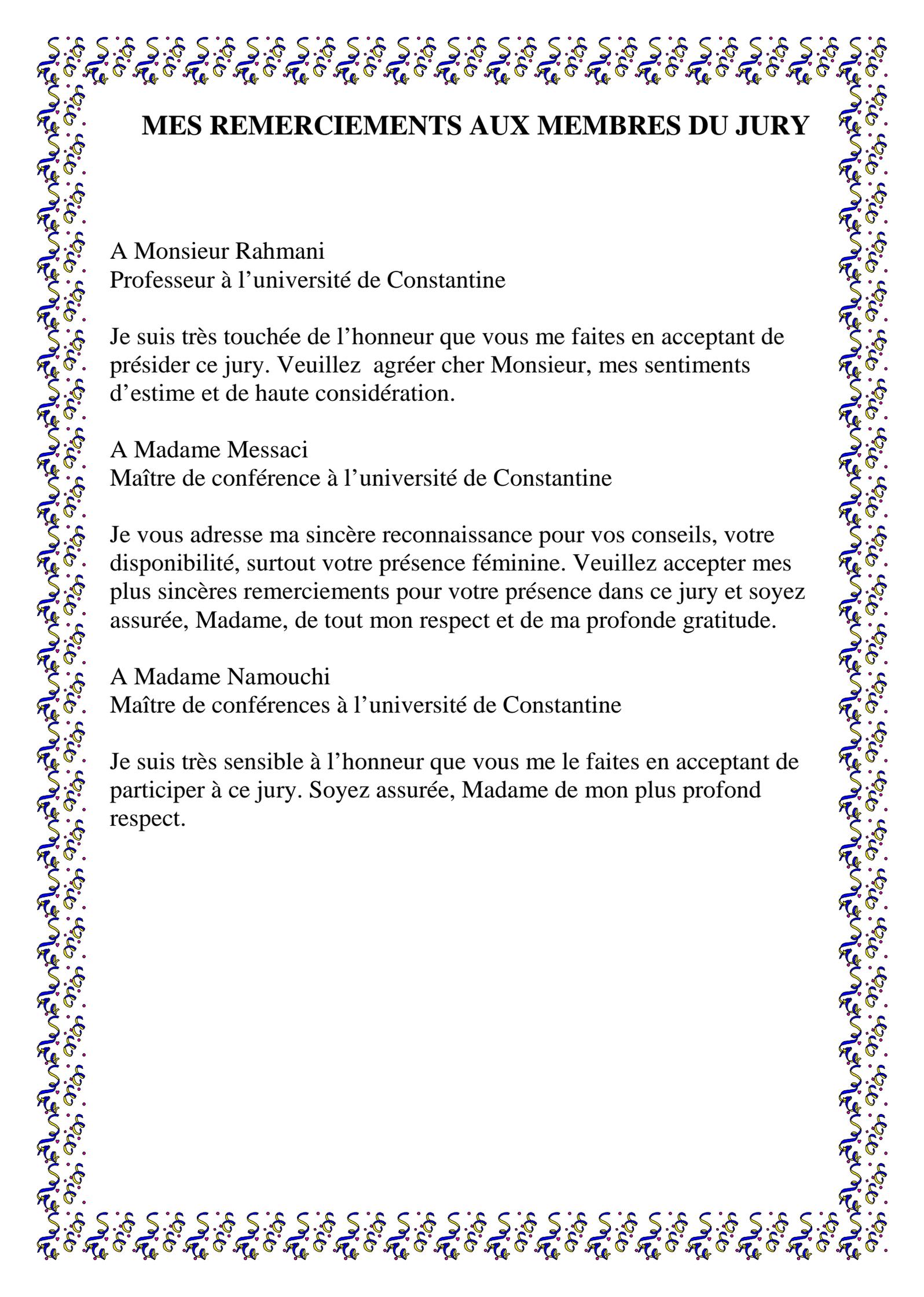
Soutenu le :.../.../...

## REMERCIEMENT

C'est également le lieu pour moi d'exprimer mes remerciements à Monsieur MOHDEB Zaher Professeur à l'université de Constantine, pour avoir accepté diriger ce mémoire et sans qui, la réalisation de ce travail n'aurait pas été possible.

Je le remercie également pour les conseils critiques et le soutien qu'il a su m'apporter dans les moments difficiles de cette recherche.

Veillez acceptez mes plus sincères remerciements pour votre grande disponibilité, votre rigueur scientifique et vos précieux conseils. Soyez assuré, Monsieur, de toute mon estime et de mon profond respect.



## MES REMERCIEMENTS AUX MEMBRES DU JURY

A Monsieur Rahmani  
Professeur à l'université de Constantine

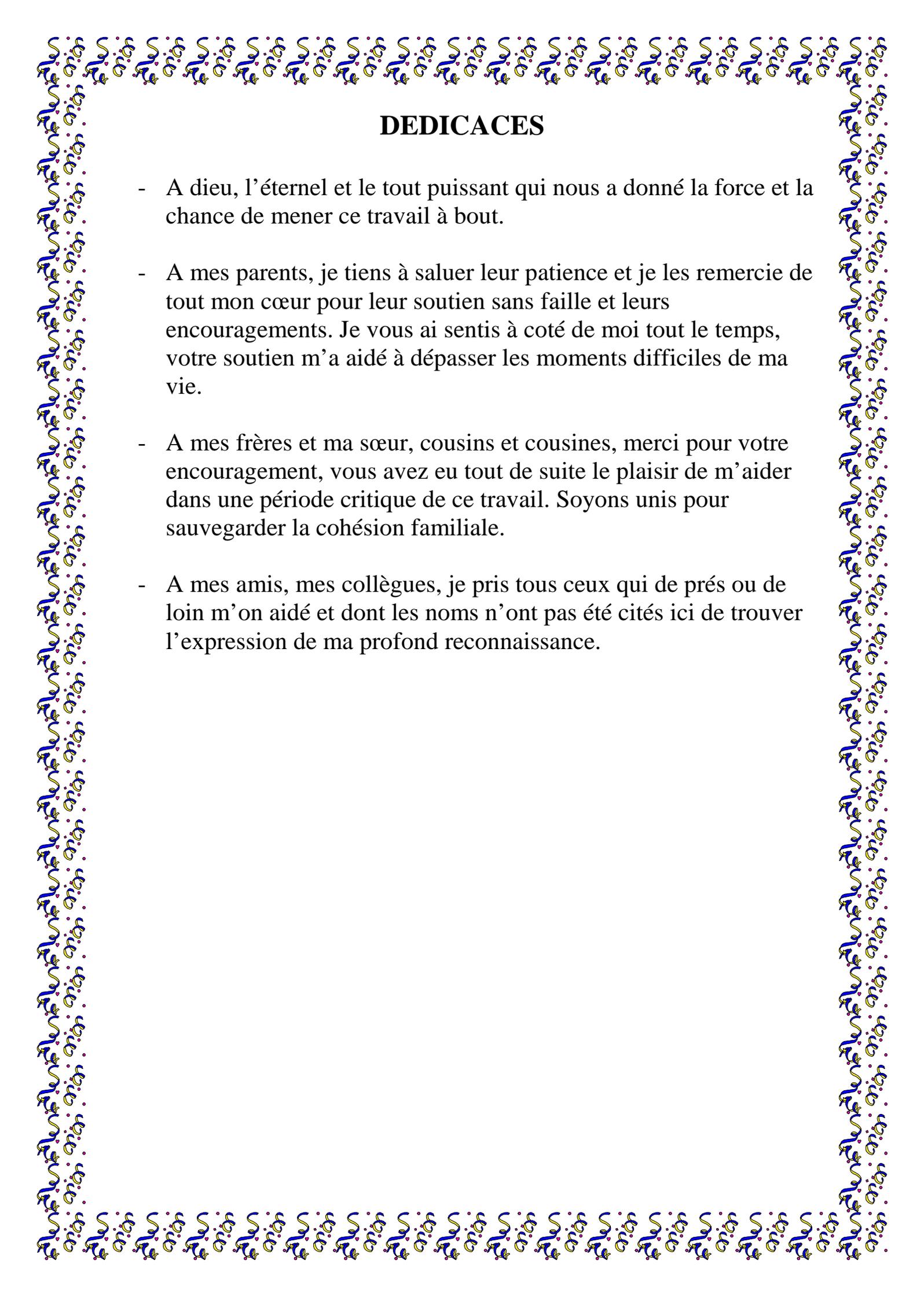
Je suis très touchée de l'honneur que vous me faites en acceptant de présider ce jury. Veuillez agréer cher Monsieur, mes sentiments d'estime et de haute considération.

A Madame Messaci  
Maître de conférence à l'université de Constantine

Je vous adresse ma sincère reconnaissance pour vos conseils, votre disponibilité, surtout votre présence féminine. Veuillez accepter mes plus sincères remerciements pour votre présence dans ce jury et soyez assurée, Madame, de tout mon respect et de ma profonde gratitude.

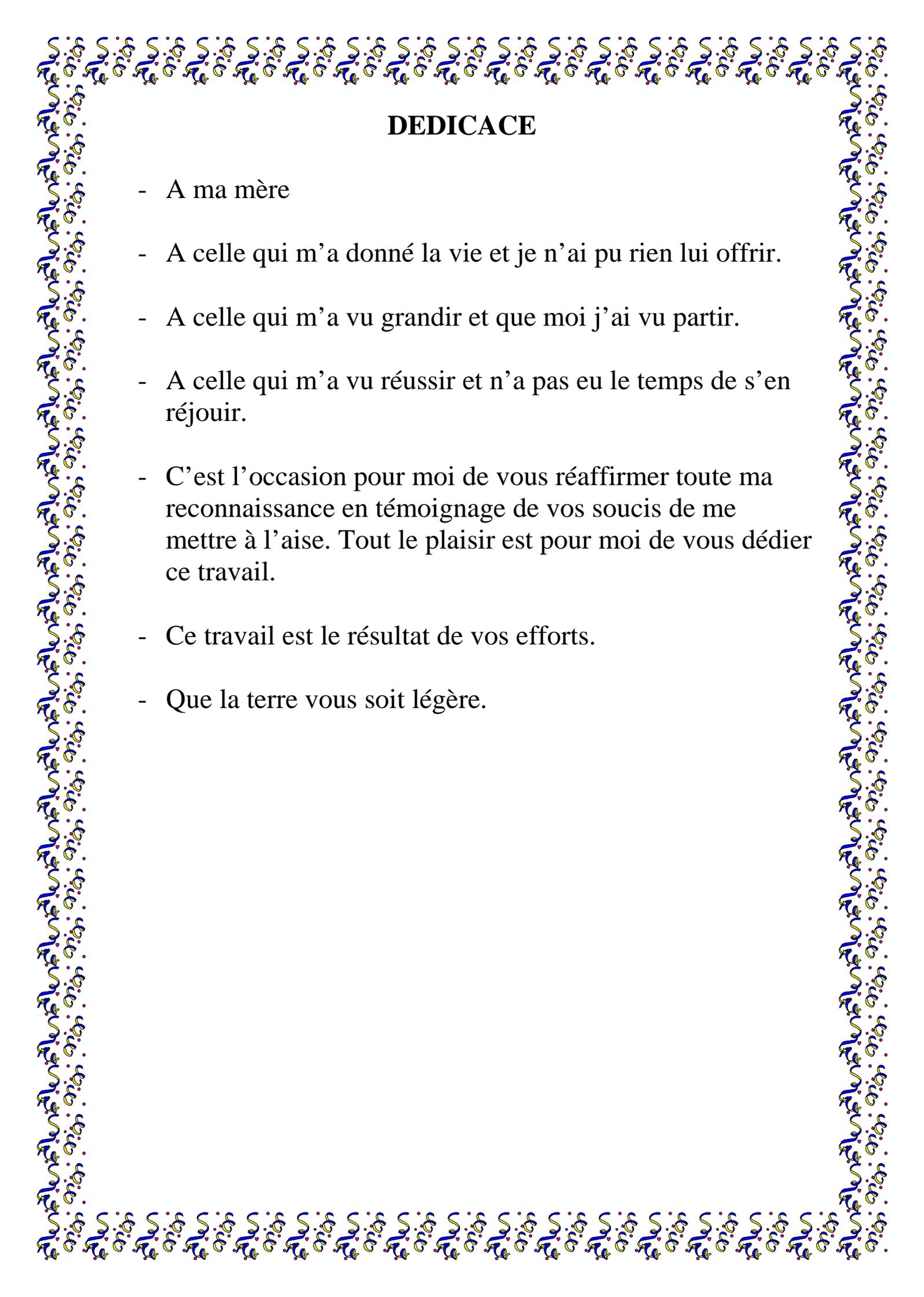
A Madame Namouchi  
Maître de conférences à l'université de Constantine

Je suis très sensible à l'honneur que vous me le faites en acceptant de participer à ce jury. Soyez assurée, Madame de mon plus profond respect.



## DEDICACES

- A dieu, l'éternel et le tout puissant qui nous a donné la force et la chance de mener ce travail à bout.
- A mes parents, je tiens à saluer leur patience et je les remercie de tout mon cœur pour leur soutien sans faille et leurs encouragements. Je vous ai sentis à coté de moi tout le temps, votre soutien m'a aidé à dépasser les moments difficiles de ma vie.
- A mes frères et ma sœur, cousins et cousines, merci pour votre encouragement, vous avez eu tout de suite le plaisir de m'aider dans une période critique de ce travail. Soyons unis pour sauvegarder la cohésion familiale.
- A mes amis, mes collègues, je pris tous ceux qui de près ou de loin m'on aidé et dont les noms n'ont pas été cités ici de trouver l'expression de ma profond reconnaissance.



## DEDICACE

- A ma mère
- A celle qui m'a donné la vie et je n'ai pu rien lui offrir.
- A celle qui m'a vu grandir et que moi j'ai vu partir.
- A celle qui m'a vu réussir et n'a pas eu le temps de s'en réjouir.
- C'est l'occasion pour moi de vous réaffirmer toute ma reconnaissance en témoignage de vos soucis de me mettre à l'aise. Tout le plaisir est pour moi de vous dédier ce travail.
- Ce travail est le résultat de vos efforts.
- Que la terre vous soit légère.

# Table des matières

0.1	<i>Introduction</i> . . . . .	3
<b>1</b>	<b>Le meilleur estimateur</b>	<b>5</b>
1.1	Présentation du modèle . . . . .	5
1.2	Estimation de $P(\lambda)$ et $R(\lambda)$ . . . . .	8
<b>2</b>	<b>Estimation de la fonction de régression par une approximation de la décomposition en série</b>	<b>15</b>
2.1	Introduction . . . . .	15
2.2	Quelques outils de l'espace de fonctions . . . . .	17
2.3	Estimateurs des séries de Fourier généralisés . . . . .	25
2.4	Estimateurs des séries de Fourier . . . . .	33
2.4.1	Forme de l'estimateur . . . . .	34
2.4.2	Etude asymptotique . . . . .	38
2.4.3	Distribution asymptotique . . . . .	44
2.5	Régression polynomiale . . . . .	50
<b>3</b>	<b>Estimation de la variance dans un modèle de régression non para-</b>	

<b>métrique</b>	<b>53</b>
3.1 Introduction . . . . .	53
3.2 Hypothèses et résultats asymptotiques . . . . .	54
<b>4 Test d'un polynôme trigonométrique dans un modèle de régression</b>	<b>63</b>
4.1 Introduction . . . . .	63
4.2 Loi asymptotique de la statistique de test . . . . .	65
4.3 Construction de test . . . . .	69
4.4 Application numérique . . . . .	70

## 0.1 *Introduction*

Les méthodes de régression non paramétrique se sont considérablement développées depuis 1980 , et la régression non paramétrique ( Eubank , 1988 ) .

Notre travail porte sur quelques méthodes d'estimation et test d'hypothèses dans un modèle de régression non paramétrique dans le cas homoscédastique et un échantillonnage fixe . Il s'agit dans un premier temps d'adapter une estimation de la fonction inconnue de régression par une approximation par une série de Fourier dont les coefficients sont estimés par leurs quantités empiriques . Cette estimation est complétée par une étude asymptotique . En général la variance du modèle est inconnue , il faut donc l'estimer , nous proposons un estimateur de la variance basé sur l'estimateur empirique des coefficients de Fourier et nous montrons la consistance de l'estimateur proposé .

Le document est structuré en quatre chapitre dont le premier est consacré à l'introduction .

Le deuxième chapitre porte sur l'estimation de la fonction de régression basée sur une approximation de la décomposition en série de Fourier et une estimation des coefficients de Fourier de la fonction . Les techniques utilisées sont celles de la régression linéaire avec une étude asymptotique notamment en loi .

Le troisième chapitre est consacré à l'estimation de la variance du modèle . Cette estimation est basée sur le développement en série de Fourier de la fonction de régression et l'approximation étudiée dans le deuxième chapitre . La consistance de cet estimateur est étudiée également dans ce chapitre .

Enfin le quatrième chapitre porte sur la construction de test de l'hypothèse que la fonction de régression est un polynôme trigonométrique de degré au plus  $s$  donné . La statistique de test utilisée est basée sur l'estimation du critère de rejet de l'hypothèse nulle . Le niveau et la puissance du test sont basés sur la loi asymptotique de la statistique de test . Ce chapitre est illustré par une application numérique sur des données du nombre d'accidents survenus aux U.S.A entre 1951 et 1980 .

# Chapitre 1

## Le meilleur estimateur

### 1.1 Présentation du modèle

Nous pouvons établir plusieurs familles d'estimateurs de régression , chaque famille est indexée par un paramètre .

La sélection d'un estimateur est de donner une famille qui exige le choix de la valeur pour le paramètre d'indice .

Supposons que le vecteur des observations soit  $\underline{y} = (y_1, \dots, y_n)'$  et obtenu à partir du modèle suivant

$$\underline{y} = \underline{\mu} + \underline{\varepsilon} \tag{1.1}$$

où  $\underline{\mu} = (\mu_1, \dots, \mu_n)'$  est un vecteur de moyenne inconnue et  $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$

est un vecteur des erreurs aléatoires de moyenne nulle , (non corrélées) , et de variance commune  $\sigma^2$ .

Nous allons considérer une classe d'estimateurs pour  $\underline{\mu}$  telle que  $C(\Lambda) = \{ \underline{\mu}_\lambda : \lambda \in \Lambda \}$  ,  $\Lambda$  est un ensemble d'indices , le paramètre en indice  $\lambda$  peut être

un scalaire ou un vecteur . Nous pouvons aussi supposer que les éléments de  $C(\Lambda)$  sont des estimateurs linéaires , pour chaque  $\lambda$  on lui associe une matrice  $H(\lambda)_{(n \times n)}$  telle que

$$\underline{\mu}_\lambda = H(\lambda)\underline{y}$$

où

$$H(\lambda) = X_\lambda(X'_\lambda X_\lambda)^{-1}X'_\lambda \quad (1.2)$$

et

$$X_\lambda = (x_j(t_i))_{\substack{i=1,\dots,n \\ j=1,\dots,\lambda}} .$$

La matrice  $H(\lambda)$  est symétrique : ( $H' = H$ ) et définie positive . Le problème qui se pose est le choix du bon estimateur de  $\underline{\mu}$  parmi les éléments de  $\{ \underline{\mu}_\lambda : \lambda \in \Lambda \}$

Définissons la fonction de perte par

$$L(\lambda) = \frac{1}{n} \sum_{j=1}^n (\mu_j - \mu_{\lambda j})^2$$

où  $\mu_{\lambda j}$  est le  $j^{\text{ème}}$  élément du  $\underline{\mu}_\lambda$ ,  $nL(\lambda)$  est le carré de la distance euclidienne entre  $\underline{\mu}$  et  $\underline{\mu}_\lambda$ , et qui représente une mesure naturelle entre  $\underline{\mu}$  et  $\underline{\mu}_\lambda$ . Le risque associé à la fonction de perte est

$$R(\lambda) = E [L(\lambda)] \quad .$$

$L(\lambda)$  et  $R(\lambda)$  mesurent l'appréciation de la performance d'un estimateur.

Une autre mesure qui est reliée à  $R(\lambda)$  est le risque de la prédiction . Supposons que nous disposons de  $n$  nouvelles observations  $\underline{y}^* = \underline{\mu} + \underline{\varepsilon}^*$ ,  $\underline{\mu}$  est le même vecteur donné dans la relation (1.1) et  $\underline{\varepsilon}^*$  un vecteur des erreurs aléatoires de moyenne nulle , non corrélées et de variance  $\sigma^2$ .

Pour évaluer la performance du  $\underline{\mu}_\lambda$  comme un prédicteur de la future observation , utilisons

$$P(\lambda) = E \left[ \frac{1}{n} \sum_{j=1}^n (y_j^* - \mu_{\lambda j})^2 \right] \quad . \quad (1.3)$$

Le risque de prédiction dans (1.3) est relié à  $R(\lambda)$  par

$$P(\lambda) = \sigma^2 + R(\lambda) \quad . \quad (1.4)$$

Ainsi un estimateur qui minimise le risque  $R(\lambda)$  peut aussi minimiser  $P(\lambda)$  et inversement.

Nous ne pouvons pas calculer  $L(\lambda)$  ,  $R(\lambda)$  sans connaître  $\underline{\mu}$  .

La fonction de perte , le risque et le risque de prédiction donnent trois mesures possibles de performance pour un estimateur de  $\underline{\mu}$  .

## 1.2 Estimation de $P(\lambda)$ et $R(\lambda)$

Soit  $C(\Lambda) = \{\underline{\mu}_\lambda : \lambda \in \Lambda\}$  une classe d'estimateurs linéaires pour  $\underline{\mu}_\lambda = H(\lambda)\underline{y}$  et considérons une estimation du risque de prédiction définie dans la relation (1.4) , l'estimateur simple de  $E (y_j^* - \mu_{\lambda_j})^2$  est le carré des résiduelles  $y_j - \mu_{\lambda_j}$  . Nous pouvons estimer  $P(\lambda)$  en utilisant l'erreur moyenne quadratique .

$$MSE(\lambda) = \frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda_j})^2 .$$

(*MSE* : mean squared error)

La question posée est : comment exécuter *MSE* comme un estimateur de  $P(\lambda)$  ?

La réponse peut être obtenue par le calcul de la valeur attendue de  $MSE(\lambda)$  et la comparer avec  $P(\lambda)$  , on a

$$MSE(\lambda) = \frac{1}{n} \left[ (\underline{y} - \underline{\mu}_\lambda)' (\underline{y} - \underline{\mu}_\lambda) \right] = \frac{1}{n} \left[ \underline{y}' (I - H(\lambda))^2 \underline{y} \right] ,$$

$$\underline{\mu}_\lambda = X_\lambda \underline{\beta}_\lambda = X_\lambda (X_\lambda' X_\lambda)^{-1} X_\lambda' \underline{y}$$

où

$$X_\lambda = (x_j(t_i))_{\substack{i=1,\dots,n \\ j=1,\dots,\lambda}}$$

et  $\underline{\beta}_\lambda = (X_\lambda' X_\lambda)^{-1} X_\lambda' \underline{y}$  est l'estimateur de  $\underline{\beta} = (\beta_{11}, \beta_{12}, \dots, \beta_{1\lambda})$  obtenu par la méthode des moindres carrés .

$$\begin{aligned}
E(MSE(\lambda)) &= \frac{1}{n} E[\underline{y}'(I - H(\lambda))^2 \underline{y}] \\
&= \frac{1}{n} [\underline{\mu}'(I - H(\lambda))^2 \underline{\mu}] + \frac{1}{n} \sigma^2 tr[(I - H(\lambda))^2] \\
&= \frac{1}{n} [\underline{\mu}'(I - H(\lambda))^2 \underline{\mu}] + \sigma^2 + \frac{1}{n} \sigma^2 tr[H(\lambda)^2] \\
&\quad - 2 \frac{1}{n} \sigma^2 tr[H(\lambda)] .
\end{aligned} \tag{1.5}$$

Par contre

$$\begin{aligned}
P(\lambda) &= \sigma^2 + R(\lambda) = \sigma^2 + \frac{1}{n} \sum_{j=1}^n E(\mu_j - \mu_{\lambda_j})^2 \\
&= \sigma^2 + \frac{1}{n} E(\underline{\mu} - \underline{\mu}_\lambda)'(\underline{\mu} - \underline{\mu}_\lambda) \\
&= \sigma^2 + \frac{1}{n} [\underline{\mu}'(I - H(\lambda))^2 \underline{\mu}] + \frac{1}{n} \sigma^2 tr[H(\lambda)^2] .
\end{aligned} \tag{1.6}$$

Sachant que  $\underline{\mu}_\lambda = H(\lambda)\underline{\mu} + H(\lambda)\underline{\varepsilon}$

En comparant les relations (1.5) et (1.6) nous constatons que  $MSE(\lambda)$  est un estimateur biaisé de  $P(\lambda)$  et son biais est égal à  $-2n^{-1}\sigma^2 tr[H(\lambda)]$ .

Quand la variance  $\sigma^2$  est connue, l'estimateur sans biais de  $P(\lambda)$  donné est

$$\hat{P}(\lambda) = MSE(\lambda) + 2n^{-1}\sigma^2 tr[H(\lambda)] ,$$

et celui de  $R(\lambda)$  est

$$\hat{R}(\lambda) = \hat{P}(\lambda) - \sigma^2 .$$

### Remarque 1

Si nous avons des observations multiples telles que

$$y_{kj} = \mu_j + \varepsilon_{kj} \quad k = 1, 2 \quad j = 1, \dots, n$$

et que  $X_{\lambda(j)}$  représente la matrice des éléments :

$$x_{kr} : k = 1, \dots, j-1, j+1, \dots, n \quad r \in \lambda$$

( $\lambda$  sous ensemble d'indices), et que nous considérons

$$\underline{y}_{(j)} = (y_1, \dots, y_{j-1}, y_{j+1}, \dots, y_n)'$$

alors

$$\underline{\beta}_{\lambda(j)} = \left( X'_{\lambda(j)} X_{\lambda(j)} \right)^{-1} X'_{\lambda(j)} \underline{y}_{(j)}$$

est l'estimateur des moindres carrés des coefficients pour les variables  $X_r$ ,  $r \in \lambda$

calculé sans utilisation de  $(\underline{x}'_j, y_j)$ . Une estimation de  $\mu_j$  est donnée par

$$\mu_{\lambda(j)} = \underline{x}'_{\lambda(j)} \underline{\beta}_{\lambda(j)}$$

avec  $\underline{x}_{\lambda(j)}$  est le vecteur des valeurs pour les variables  $X$  avec les indices dans  $\lambda$  correspondant à la  $j^{\text{ème}}$  réponse. L'estimateur  $\mu_{\lambda(j)}$  est directement relié à  $\mu_{\lambda_j}$  dans le sens que les deux sont les estimateurs des moindres carrés et que l'un d'eux de  $\mu_j$ , entraînant les prédicteurs indexés par  $\lambda$ . La seule différence nécessite  $(\underline{x}'_j, y_j)$ .

On peut montrer que

$$\mu_{\lambda(j)} = \frac{\mu_{\lambda_j} - h_{jj}(\lambda)(y_j - \mu_{\lambda_j})}{(1 - h_{jj}(\lambda))} \quad (1.7)$$

où  $h_{jj}(\lambda)$  est le  $j^{\text{ème}}$  élément de la diagonale de  $H(\lambda)$ .

Revenons au cas général et supposons que  $\mu_{\lambda(j)}$  représente l'analogie de  $\mu_{\lambda_j}$  calculé sans la  $j^{\text{ème}}$  observation; l'estimateur donné par la relation (1.7) est un exemple

vérifiant cette situation . L'estimateur  $\mu_{\lambda(j)}$  est construit à partir du sous- échantillon de taille  $(n - 1)$  pris de l'échantillon initial (données relevées à l'origine) . Si la  $j^{\text{ème}}$  observation est traitée comme une observation supplémentaire , notre précédente discussion suggère l'utilisation de  $(y_j - \mu_{\lambda(j)})^2$  pour estimer  $E (y_j^* - \mu_{\lambda_j})^2$  et estimant  $P(\lambda)$  par

$$CV(\lambda) = \frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda(j)})^2. \quad (1.8)$$

$CV(\lambda)$  est appelé

l'erreur moyenne quadratique de validation croisée (cross-validation mean squared error) .

Une autre méthode utilisée pour sélectionner  $\lambda$  est la validation croisée généralisée ( $GCV$ ) . En supposant que  $tr [H(\lambda)] \leq n$  , le critère  $GCV$  est défini par

$$GCV(\lambda) = \frac{MSE(\lambda)}{\left( \frac{1}{n} tr [I - H(\lambda)] \right)^2} ,$$

c'est à dire

$$GCV(\lambda) = \frac{\frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda_j})^2}{\left( \frac{1}{n} tr [I - H(\lambda)] \right)^2} . \quad (1.9)$$

Le critère  $GCV$  est proposé pour la première fois par Craven et Wahba (1979) pour l'usage dans le contexte d'un problème particulier de la régression non paramétrique.

Wahba (1977,1984) et Golub , Heath et Wahba (1979) avaient montré que  $GCV$  peut être utilisé pour résoudre un certain nombre de problèmes impliquant la sélection d'un paramètre pour minimiser  $R(\lambda)$  ou  $P(\lambda)$  .

Dans la plupart des cas intéressants ,  $GCV$  est proche de  $CV$  .

On peut illustrer ceci par le problème de sélection de variable , en utilisant les relations (1.7) et (1.8) on a

$$\begin{aligned} CV(\lambda) &= \frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda(j)})^2 \\ &= \frac{1}{n} \sum_{j=1}^n \left[ \frac{y_j - \mu_{\lambda_j}}{1 - h_{jj}(\lambda)} \right]^2 , \end{aligned}$$

et que

$$\begin{aligned} GCV(\lambda) &= \frac{\frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda_j})^2}{\left( \frac{1}{n} \text{tr} [I - H(\lambda)] \right)^2} \\ &= \frac{1}{n} \sum_{j=1}^n (y_j - \mu_{\lambda(j)})^2 \left\{ \frac{1 - h_{jj}(\lambda)}{\frac{1}{n} \text{tr} [I - H(\lambda)]} \right\}^2 . \end{aligned}$$

On voit clairement , dans ce cas , que  $GCV$  est une version pondérée de  $CV$  .

La motivation du critère (1.9) provient du théorème appelé Théorème  $GCV$  établi par Craven et Wahba (1979) , (voir aussi Wahba (1977 c) et Golub , Heath et Wahba (1979) pour d'autres variantes du théorème ) .

### **Théorème 1 (le Théorème $GCV$ )**

Soit  $\tau_j(\lambda) = n^{-1} \text{tr} [H(\lambda)^j]$  ,  $j = 1, 2$  , si  $H(\lambda)$  est définie par (1.2) et supposons que  $\tau_1(\lambda) \leq 1$  , alors

$$\frac{| E [GCV(\lambda)] - P(\lambda) |}{R(\lambda)} \leq g(\lambda) ,$$

où

$$g(\lambda) = \frac{2\tau_1(\lambda) + \frac{\tau_1^2(\lambda)}{\tau_2(\lambda)}}{(1 - \tau_1(\lambda))^2} .$$

### Remarque 2

Si  $g(\lambda)$  est proche de 0 , le résultat du théorème montre que la distance entre  $P(\lambda)$  et  $E(GCV(\lambda))$  est relativement petite par rapport à  $R(\lambda)$  . Ce qui signifie , dans ce cas , que  $GCV(\lambda)$  est presque un estimateur sans biais de  $P(\lambda)$  .

### Preuve

Nous avons

$$\begin{aligned} E[GCV(\lambda)] &= \frac{E[MSE(\lambda)]}{\left(1 - \frac{1}{n} \text{tr}[H(\lambda)]\right)^2} \\ &= \frac{\frac{1}{n} E[\underline{y}'(I - H(\lambda))^2 \underline{y}]}{(1 - \tau_1(\lambda))^2} . \end{aligned}$$

Comme  $E(\underline{y}) = \underline{\mu}$  et que la matrice de variance-covariance du vecteur 1 est

$\Sigma = \sigma^2 I_n$  ( $I_n$  étant la matrice identité ( $n \times n$ )) , on a

$$\begin{aligned} E[\underline{y}'(I - H(\lambda))^2 \underline{y}] &= \underline{\mu}' [I - H(\lambda)]^2 \underline{\mu} + \text{tr} \left( [I - H(\lambda)]^2 \Sigma \right) \\ &= \underline{\mu}' [I - H(\lambda)]^2 \underline{\mu} + \sigma^2 \text{tr} \left( [I - H(\lambda)]^2 \right) . \end{aligned}$$

Donc

$$E[GCV(\lambda)] = \frac{\frac{1}{n} \left\{ \underline{\mu}' [I - H(\lambda)]^2 \underline{\mu} + \sigma^2 \text{tr} \left( [I - H(\lambda)]^2 \right) \right\}}{[1 - \tau_1(\lambda)]^2}$$

et d'après la relation (1.6) on a

$$\begin{aligned}
E[GCV(\lambda)] - P(\lambda) &= \frac{R(\lambda)}{(1 - \tau_1(\lambda))^2} + \sigma^2 \frac{1 - 2\tau_1(\lambda)}{(1 - \tau_1(\lambda))^2} - [R(\lambda) + \sigma^2] \\
&= \frac{\tau_1(\lambda)[2 - \tau_1(\lambda)]}{[1 - \tau_1(\lambda)]^2} R(\lambda) - \sigma^2 \frac{\tau_1^2(\lambda)}{[1 - \tau_1(\lambda)]^2} .
\end{aligned}$$

Ceci implique que

$$\begin{aligned}
\frac{|E[GCV(\lambda)] - P(\lambda)|}{R(\lambda)} &= \left| \frac{\tau_1(\lambda)(2 - \tau_1(\lambda))}{[1 - \tau_1(\lambda)]^2} - \frac{\sigma^2 \tau_1^2(\lambda)}{R(\lambda)[1 - \tau_1(\lambda)]^2} \right| \\
&\leq \frac{\tau_1(\lambda)[2 - \tau_1(\lambda)]}{[1 - \tau_1(\lambda)]^2} + \frac{\tau_1^2(\lambda)/\tau_2(\lambda)}{[1 - \tau_1(\lambda)]^2} \\
&\leq g(\lambda) ,
\end{aligned}$$

où les deux dernières inégalités sont obtenues en utilisant le fait que

$$R(\lambda) \geq \sigma^2 \tau_2(\lambda) \text{ et que } \tau_1(\lambda) \leq 1 \text{ entrainant } 1 \leq 2 - \tau_1(\lambda) \leq 2 .$$

On a maintenant trois critères disponibles pour sélectionner la bonne valeur de  $\lambda$  des données ,  $\hat{P}(\lambda)$  (et  $\hat{R}(\lambda) = \hat{P}(\lambda) - \sigma^2$ ) ,  $CV(\lambda)$  et  $GCV(\lambda)$  .

La stratégie est de choisir un (ou plusieurs) de ces critères et de les utiliser pour obtenir un estimateur  $\hat{\lambda}$  en minimisant  $P(\lambda)$  (et  $R(\lambda)$ ) .

# Chapitre 2

## Estimation de la fonction de régression par une approximation de la décomposition en série

### 2.1 Introduction

Supposons que nous observons  $n$  valeurs d'une variable réponse  $Y$  ou  $n$  valeurs d'une variable déterministe  $t$ .

Les  $n$  couples de valeurs  $(t_1, y_1), \dots, (t_n, y_n)$  satisfaisant le modèle

$$y_j = \mu(t_j) + \varepsilon_j \quad , \quad j = 1, \dots, n \quad , \quad (2.1)$$

où  $\underline{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)'$  est un vecteur des erreurs aléatoires de moyenne nulle, non corrélées et de variance  $\sigma^2$ ,  $\mu$  est une fonction inconnue de la régression que nous

voulons estimer .

Dans cette partie , nous nous intéressons aux estimateurs de régression non paramétrique sous forme d'estimateurs de séries orthogonales .

Pour motiver notre étude , supposons que le modèle de régression est un modèle paramétrique linéaire, c'est-à dire qu'il existe des fonctions  $x_1, \dots, x_\lambda$  et des coefficients  $\beta_1, \dots, \beta_\lambda$  tels que

$$\mu(t) = \sum_{j=1}^{\lambda} \beta_j x_j(t) . \quad (2.2)$$

L'estimation de  $\mu$  est donc équivalente à l'estimation des  $\beta_j$  ,  $j = 1, \dots, \lambda$  .

L'utilisation d'un modèle linéaire pour  $\mu$  suppose que la forme fondamentale de  $\mu$  est connue , seulement ce n'est pas toujours le cas .

Lorsque la fonction  $\mu$  n'est pas bien connue , on préfère utiliser la classe de fonctions à la laquelle appartient la fonction  $\mu$  . Par exemple , on peut considérer la classe de fonctions lisses , dans le sens où  $\mu$  est continue et différentiable ou bien d'autres hypothèses de même nature .

Lorsque  $\mu$  est supposée appartenir à une classe générale de dimension infinie, le modèle (2.1) est appelé “ modèle de régression non paramétrique ” .

Un tel modèle répond beaucoup plus à la réalité que le modèle linéaire .

Cependant , l'amélioration de notre modèle n'est pas sans pri . Il est naturel , de se demander s'il n'y a pas de parallèle avec le modèle linéaire pour nos ensembles d'hypothèses plus générales .

Par exemple , si  $\mu$  est supposée continue , est ce que nous pouvons écrire la fonc-

tion de régression comme une combinaison linéaire infinie d'une famille de fonctions connues . Il est évident que c'est avantageux , si c'est le cas . Ainsi , estimer la fonction  $\mu$  , revient à estimer les coefficients de régression , en adaptant les techniques classiques d'estimation dans les modèles linéaires pour estimer les fonctions de régressions plus générales .

Il existe dans la littérature plusieurs classes de fonctions dont les éléments admettent une écriture sous forme de modèle linéaire infinie .

Si  $(\beta_j)_{j \geq 1}$  est une suite de carré sommable , la fonction de régression  $\mu$  peut se mettre sous la forme ,

$$\mu(t) = \sum_{j=1}^{\infty} \beta_j x_j(t) \quad , \quad (2.3)$$

où  $(x_j)$  est une famille de fonctions .

Une des approche que nous pouvons utiliser est d'approximer  $\mu$  par

$$\sum_{j=1}^{\lambda} \beta_j x_j(t)$$

pour certain entier  $\lambda$  et ensuite estimer les coefficients  $\beta_j$  pour obtenir une estimation de  $\mu$ , appelée “ estimateur série ” .

La sélection d'une valeur appropriée de  $\lambda$  peut être traitée , comme une catégorie de problèmes traités dans le chapitre précédent .

## 2.2 Quelques outils de l'espace de fonctions

Dans cette section , nous allons aborder plusieurs classes de fonctions jouent un rôle important dans notre étude .

Un des plus importants espaces de fonctions est  $L_2[a, b]$ , l'ensemble de toutes les fonctions de carré intégrable sur l'intervalle  $[a, b]$ . Soient deux fonctions  $\mu_1$  et  $\mu_2$  appartenant à  $L_2[a, b]$ , on définit leur norme par

$$\| \mu_1(t) \| = \left\{ \int_a^b \mu_1(t)^2 dt \right\}^{1/2}, \quad (2.4)$$

et leur produit scalaire par

$$\langle \mu_1, \mu_2 \rangle = \int_a^b \mu_1(t) \mu_2(t) dt. \quad (2.5)$$

Remarquons que la relation (2.5) est bien définie à cause de l'inégalité de Cauchy-Swartz

$$| \langle \mu_1, \mu_2 \rangle |^2 \leq \| \mu_1 \|^2 \| \mu_2 \|^2.$$

L'espace  $L_2[a, b]$  représente une riche collection de fonctions. Cependant, il ne convient pas tout à fait à notre situation. Deux fonctions  $\mu_1$  et  $\mu_2$  sont considérées comme identiques dans  $L_2[a, b]$  si  $\| \mu_1 - \mu_2 \| = 0$ . Donc  $\mu_1$  et  $\mu_2$  sont identiques si elles diffèrent sur des ensembles de mesure de Lebesgue nulle. Ce qui signifie que les éléments de  $L_2[a, b]$  sont des classes d'équivalence de fonctions qui sont égales presque partout par rapport à la mesure de Lebesgue. Ainsi, l'évaluation d'un élément de  $L_2[a, b]$  aux points dans  $[a, b]$ , n'est pas en général une opération bien définie. Quelques restrictions sont alors exigées sur les types de fonctions de régression pour assurer que le modèle (2.1) soit valide. Puisqu'il est raisonnable de supposer que  $\mu$  dans le modèle (2.1) soit une fonction lisse, une démarche consiste à considérer les ensembles de fonctions qui sont lisses dans le sens où elles satisfont les conditions de continuité et de différentiabilité.

Il y a plusieurs sous-ensembles de  $L_2 [a, b]$  possédant le type de propriétés exigées dans un espace de fonctions . L'un d'eux est le suivant

$$C^m [a, b] = \{ \mu / \mu^{(j)} \text{ est continue , } j = 0, 1, \dots, m \} \quad . \quad (2.6)$$

$C^m [a, b]$  est l'ensemble de toutes les fonctions sur  $[a, b]$  possédant  $m$  dérivées continues .

Lorsque  $m = 0$  ,  $C^0 [a, b] = C [a, b]$  qui est l'ensemble des fonctions continues sur  $[a, b]$  .

Il est clair que

$$L_2 [a, b] \supset C [a, b] \supset C^1 [a, b] \supset \dots$$

Dans notre travail , il est nécessaire que  $\mu$  possède  $m$  dérivées continues . Il suffit que  $\mu^{(m)}$  soit de carré intégrable . Ainsi , nous définissons l'espace de Sobolev d'ordre  $m$

$$W_2^m [a, b] = \left\{ \begin{array}{l} \mu / \mu^{(j)} \text{ soit absolument continue , } j = 0, 1, \dots, m - 1 \\ \text{et } \mu^{(m)} \in L_2 [a, b] \end{array} \right\} \quad . \quad (2.7)$$

Notons que  $W_2^m [a, b] \supset C^m [a, b]$  et que  $W_2^0 [a, b] = L_2 [a, b]$  .

Si nous considérons les fonctions de  $W_2^m [a, b]$  , ( $m \geq 1$ )

ou  $C^m [a, b]$  , ( $m \geq 0$ ) , l'ambiguïté correspondant à l'évaluation des points disparaît . Puisque si  $\mu_1$  et  $\mu_2$  appartenant à  $C [a, b]$  et  $\|\mu_1 - \mu_2\| = 0$  ,

signifie que  $\mu_1 \equiv \mu_2$  .

$W_2^m [a, b]$  et  $C^m [a, b]$  peuvent être munis de leurs propres normes et étudiés indépendamment de  $L_2 [a, b]$  .

Autrement dit , le fait qu'ils soient des sous-ensemble de  $L_2 [a, b]$  signifie que les propriétés des fonctions de  $L_2 [a, b]$  que nous allons développer sont applicables également aux éléments de  $W_2^m [a, b]$  et  $C^m [a, b]$  .

Il est facile de vérifier que  $L_2 [a, b]$  est un espace vectoriel .

Pour les espaces vectoriels de dimension finie comme l'espace Euclidien  $\mathbb{R}^p$  , il est toujours possible de représenter , tout élément dans un espace vectoriel comme une combinaison linéaire d'une base .

On peut se poser la question s'il existe une propriété analogue qui sera satisfaite pour un espace vectoriel de dimension infinie telle que  $L_2 [a, b]$  . En fait , ceci est possible .

Pour vérifier ce résultat , il est nécessaire de rappeler quelques définitions .

### **Définition 1**

*Soit  $\mu_1, \mu_2$  deux fonctions de  $L_2 [a, b]$  , on dit que  $\mu_1, \mu_2$  sont orthogonales si  $\langle \mu_1, \mu_2 \rangle = 0$  , que l'on note par  $\mu_1 \perp \mu_2$  .*

### **Définition 2**

*Une suite de fonctions  $\{x_j\}$  est dite orthonormale si pour tout  $i \neq j$  ,  $\langle x_i , x_j \rangle = 0$  et  $\| x_j \| = 1$  pour tout  $j$  .*

### **Définition 3**

*Une suite de fonctions  $\{x_j\}$  est appelée un système complet orthonormal (CONS) si  $\mu \perp x_j$  pour tout  $j$  entraîne que  $\mu \equiv 0$  .*

### **Remarques 1**

a) En rappelant les définitions de l'orthogonalité et de l'orthonormalité

dans  $\mathbb{R}^p$  , on voit clairement à travers les définitions 1 et 2 qu'il existe un parallèle

de ces concepts pour  $L_2 [a, b]$  .

b) La définition 3 montre que pour un système complet orthonormal  $\{x_j\}$  , la seule fonction orthogonale à toutes les fonctions  $x_j$  est la fonction nulle .

### Exemple 1 : Espace de fonctions exponentielles complexes

Considérons le cas où  $L_2 [a, b]$  contient les fonctions à valeurs complexes , aussi bien réelles . Donc toute fonction  $\mu$  de  $L_2 [a, b]$  peut se mettre sous la forme

$$\mu = \mu_{\text{Re}} + i\mu_{\text{Im}} ,$$

$\mu_{\text{Re}}$  étant la partie réelle et  $\mu_{\text{Im}}$  la partie imaginaire et  $i^2 = -1$  . Le conjugué du complexe  $\mu$  est alors défini par  $\mu^* = \mu_{\text{Re}} - i\mu_{\text{Im}}$  .

Lorsque les fonctions complexes sont comprises dans l'espace , le produit scalaire et la norme pour  $L_2 [a, b]$  sont définies par

$$\langle \mu_1, \mu_2 \rangle = \int_a^b \mu_1(t) \mu_2^*(t) dt \quad (2.8)$$

et

$$\| \mu_1 \| = \left\{ \int_a^b | \mu_1(t) |^2 dt \right\}^{1/2} , \quad (2.9)$$

où  $| \mu_1(t) |^2 = \mu_1(t) \mu_1^*(t)$  .

Une des famille de système complet orthonormale est donnée par les exponentielles complexes définies par

$$x_j(t) = \frac{1}{\sqrt{b-a}} e^{2\pi i j t / (b-a)} , \quad j \in \mathbb{Z} . \quad (2.10)$$

### Exemple 2 : Les polynôme de légendre

Un autre système complet orthonormal pour  $L_2 [a, b]$  peut être construit à partir des fonctions polynomiales

$$p_j(t) = t^j, \quad j \in \mathbb{N} .$$

On peut procéder comme suit en prenant

$$x_0(t) = \frac{p_0(t)}{\|p_0\|} \equiv \frac{1}{\sqrt{b-a}} \quad (2.11)$$

et les autres fonctions par la formule de récurrence

$$x_j(t) = \frac{p_j(t) - \sum_{k=0}^{j-1} \langle p_j, x_k \rangle x_k(t)}{\|p_j - \sum_{k=0}^{j-1} \langle p_j, x_k \rangle x_k\|} . \quad (2.12)$$

Les fonctions définies par les relations (2.11) – (2.12) sont appelées les polynômes de légendre .

La méthode utilisée pour construire les polynômes de legendre est appelée

“ méthode d’orthonormalisation de Gram-Schmidt ” . Cette méthode ne s’applique pas seulement aux fonctions polynomiales mais peut s’appliquer également à n’importe quelle suite de fonctions pour construire un système complet orthonormal .

Les deux exemples précédents illustrent bien l’existence d’un système complet orthonormal pour  $L_2 [a, b]$  .

La proposition suivante montre que tout système complet orthonormal nous fournit une base pour  $L_2 [a, b]$  . On peut trouver sa démonstration dans tout livre traitant de l’analyse fonctionnelle , ( par exemple , Luenberger (1969)) .

### Proposition 1

Soit  $\{x_j\}_{j=1}^{\infty}$  un système complet orthonormal pour  $L_2[a, b]$  et pour  $\mu \in L_2[a, b]$ ,  
on définit

$$\beta_j = \langle \mu, x_j \rangle, \quad j = 1, 2, \dots \quad (2.13)$$

Alors  $\sum_{j=1}^{\lambda} \beta_j x_j$  est la meilleure approximation de  $\mu$  dans le sens que

$$\left\| \mu - \sum_{j=1}^{\lambda} \beta_j x_j \right\| \leq \left\| \mu - \sum_{j=1}^{\lambda} \alpha_j x_j \right\|$$

pour tout  $\underline{\alpha} = (\alpha_1, \dots, \alpha_{\lambda})' \in \mathbb{R}^{\lambda}$ .

De plus,

$$\left\| \mu - \sum_{j=1}^{\lambda} \beta_j x_j \right\|^2 \longrightarrow 0 \text{ quand } \lambda \longrightarrow \infty.$$

### Commentaires

La proposition précédente montre que la suite des fonctions  $\sum_{j=1}^{\lambda} \beta_j x_j$  converge vers  $\mu$  en norme dans  $L_2[a, b]$ .

On écrit ceci sous la forme

$$\mu(t) \sim \sum_{j=1}^{\infty} \beta_j x_j(t). \quad (2.14)$$

Pour montrer que  $\mu$  et  $\sum_{j=1}^{\infty} \beta_j x_j$  sont des fonctions identiques, lorsqu'elles sont prises comme éléments de  $L_2[a, b]$ . Cependant, cela ne signifie pas qu'on a la convergence ponctuelle, bien que ça reste vrai dans la plupart des cas que nous

considérons .

D'après la proposition précédente , on peut représenter  $\mu$  , au sens de  $L_2 [a, b]$  , comme une combinaison linéaire infinie d'une base de fonctions  $\{x_j\}$  .

Les coefficients donnés par la formule (2.13) utilisés dans le développement de la relation (2.14) pour  $\mu$  sont appelés “ Coefficients de Fourier généralisés ” , tandis que le développement lui-même est appelé “ série de Fourier généralisée ” pour  $\mu$  .

Les coefficients de Fourier généralisés satisfont

$$\sum_{j=1}^{\infty} |\beta_j|^2 = \|\mu\|^2,$$

qui est connue comme la relation de Parseval .

De cette relation , on en déduit que les coefficients de Fourier généralisés des fonctions de  $L_2 [a, b]$  peuvent être considérés comme des éléments d'un ensemble

de toutes les suites de carré sommables . La réciproque est vraie , si  $(\beta_j)_{j \geq 1}$  est une suite de carré sommable , alors  $\mu = \sum_{j=1}^{\infty} \beta_j x_j$  est dans  $L_2 [a, b]$  .

**Exemple : (Suite de l'exemple 1)**

D'après la Proposition 1 , si  $\mu$  est une fonction dans  $L_2 [a, b]$  alors

$$\mu(t) \sim \frac{1}{(b-a)} \sum_{j=-\infty}^{\infty} \beta_j e^{2\pi i j t / (b-a)} , \quad (2.15a)$$

où

$$\beta_j = \int_a^b \mu(t) e^{-2\pi i j t / (b-a)} dt . \quad (2.15b)$$

C'est le développement classique d'une série de Fourier pour  $\mu$  qui a donné lieu aux travaux plus généraux développés dans cette section .

Les fonctions qui nous intéressent sont à valeurs réelles . Dans ce cas ( $\mu$  réelle) , les formules (2.15) peuvent se réécrire sous la forme

$$\begin{aligned}\beta_j^* &= \int_a^b \mu^*(t) e^{2\pi i j t / (b-a)} dt \\ &= \int_a^b \mu(t) e^{2\pi i j t / (b-a)} dt = \beta_{-j} \quad .\end{aligned}$$

Par ailleurs ,

$$\begin{aligned}\mu(t) &\sim \frac{1}{(b-a)} \left[ \beta_0 + \sum_{j=1}^{\infty} (\beta_j e^{2\pi i j t / (b-a)} + \beta_{-j} e^{-2\pi i j t / (b-a)}) \right] \\ &= \frac{1}{(b-a)} \left[ \beta_0 + \sum_{j=1}^{\infty} (\beta_j e^{2\pi i j t / (b-a)} + \beta_j^* e^{-2\pi i j t / (b-a)}) \right] \quad .\end{aligned}$$

L'expression de  $\mu$  peut également s'écrire comme fonction de sinus et cosinus comme suit

$$\mu(t) \sim \frac{1}{(b-a)} \left[ \beta_0 + \sum_{j=1}^{\infty} \left( c_j \cos \frac{2\pi j t}{b-a} + s_j \sin \frac{2\pi j t}{b-a} \right) \right] \quad , \quad (2.16a)$$

où

$$c_j = 2 \int_a^b \mu(t) \cos \frac{2\pi j t}{b-a} dt \quad (2.16b)$$

et

$$s_j = 2 \int_a^b \mu(t) \sin \frac{2\pi j t}{b-a} dt \quad . \quad (2.16c)$$

Notons que  $t$  est un scalaire , mais l'étude reste valable lorsque  $t$  est un vecteur .

## 2.3 Estimateurs des séries de Fourier généralisés

On a vu dans la section précédente que , étant donné un système complet orthonormal  $\{x_j\}_{j \geq 1}$  , toute fonction  $f$  dans  $L_2 [a, b]$  possède un développement en série de

Fourier généralisé  $\sum_{j=1}^{\infty} \beta_j x_j$

où

$$\beta_j = \int_a^b f(t) e^{-2\pi i j t / (b-a)} dt .$$

Supposons que  $\mu$  est dans  $L_2 [a, b]$  et que  $\sum_{j=1}^{\lambda} \beta_j x_j$  converge uniformément en  $t$  vers  $\mu(t)$  quand  $\lambda \rightarrow \infty$  .

Par exemple , on a le développement classique en série de Fourier uniformément en  $t$  pourvu que  $\mu$  soit dans  $C^1 [a, b]$  ,  $\mu(a) = \mu(b)$  et  $\mu'(a) = \mu'(b)$  .

Sous ces conditions sur  $\mu$  , le modèle (2.1) s'écrit

$$y_j = \sum_{k=1}^{\infty} \beta_k x_k(t_j) + \varepsilon_j , \quad j = 1, \dots, n . \quad (2.17)$$

Ainsi , les observations obeissent à un modèle linéaire avec une infinité de coefficients de régression inconnus .

Puisque  $\mu$  est dans  $L_2 [0, 1]$  , on a  $\sum_{j=1}^{\infty} \beta_j^2 < \infty$  et que les  $\beta_j$  doivent décroître vers zéro , lorsque  $j \rightarrow \infty$  . Il semble raisonnable donc de supposer qu'il existe un entier  $\lambda$  tel que

$$\mu =: \sum_{k=1}^{\lambda} \beta_k x_k$$

et ainsi

$$y_j = \sum_{k=1}^{\lambda} \beta_k x_k(t_j) + \varepsilon_j , \quad j = 1, \dots, n . \quad (2.18)$$

On aboutit ainsi à un modèle linéaire à qui on peut appliquer les techniques inférentielles classiques des modèles linéaires .

Pour simplifier , on prend un échantillonnage des points  $t_1, \dots, t_n$  , uniformes sur  $[a, b]$  , par exemple

$$t_j = a + \frac{j-1}{n-1}(b-a) \quad , \quad j = 1, \dots, n \quad .$$

Les coefficients de Fourier  $\beta_j$  ,  $j = 1, \dots, \lambda$  inconnus peuvent être estimés en minimisant l'erreur moyenne quadratique

$$S(\alpha_1, \dots, \alpha_k) = \frac{1}{n} \sum_{j=1}^n \left( y_j - \sum_{k=1}^{\lambda} \alpha_k x_k(t_j) \right)^2 ,$$

par rapport à  $\alpha_1, \dots, \alpha_\lambda$  .

Lorsque les  $t_j$  sont équidistants , cette technique peut être justifiée en utilisant le fait que  $(\beta_1, \dots, \beta_\lambda)$  minimise  $\| \mu - \sum_{k=1}^{\lambda} \alpha_k x_k \|^2$  et que l'erreur moyenne quadratique fournit un même estimateur obtenu de cette norme de la différence .

Si les  $t_j$  ne sont pas équidistants et sont par exemple générés par une densité  $P$  par la relation

$$\int_a^{t_j} P(t) dt = \frac{j-1}{n-1} ,$$

alors l'erreur moyenne quadratique estime

$$\int_a^b \left[ \mu(t) - \sum_{k=1}^{\lambda} \alpha_k x_k(t) \right]^2 P(t) dt ,$$

plutôt que la norme dans  $L_2[a, b]$  de la différence .

Considérons la matrice définie par

$$X_\lambda = \{x_j(t_i)\}_{\substack{i=1, \dots, n \\ j=1, \dots, \lambda}} \quad .$$

Si la matrice  $X_\lambda$  est de rang  $\lambda$  , alors l'unique estimateur minimisant l'erreur moyenne quadratique est donné par

$$\underline{\beta}_\lambda = (\beta_{\lambda 1}, \dots, \beta_{\lambda \lambda})' = (X_\lambda^* X_\lambda)^{-1} X_\lambda^* \underline{y} \quad , \quad (2.19)$$

où  $\underline{y} = (y_1, \dots, y_n)'$  est le vecteur des observations et la notation “\*” signifie la matrice transposée des complexes conjuguées .

L’estimateur de la série de Fourier généralisé de  $\mu(t)$  est alors

$$\mu_\lambda(t) = \sum_{j=1}^{\lambda} \beta_{\lambda j} x_j(t) \quad .$$

Pour estimer la variance de l’erreur , on peut ainsi utiliser

$$MSE(\lambda) = \frac{1}{n} \sum_{j=1}^n \left[ y_j - \sum_{k=1}^{\lambda} \beta_{\lambda k} x_k(t_j) \right]^2 \quad ,$$

ou bien

$$\sigma_\lambda^2 = \frac{n}{n - \lambda} MSE(\lambda) = \frac{1}{n - \lambda} \sum_{j=1}^n \left[ y_j - \sum_{k=1}^{\lambda} \beta_{\lambda k} x_k(t_j) \right]^2 \quad .$$

On définit la fonction perte et le risque , lorsqu’on estime  $\mu$  par  $\mu_\lambda$  , respectivement par :

$$L_n(\lambda) = \frac{1}{n} \sum_{j=1}^n [\mu(t_j) - \mu_\lambda(t_j)]^2$$

et

$$\begin{aligned} R_n(\lambda) &= E [L_n(\lambda)] \quad . \\ &= \frac{1}{n} \sum_{j=1}^n \{ \mu(t_j) - E [\mu_\lambda(t_j)] \}^2 + \frac{\lambda}{n} \sigma^2 \quad . \end{aligned}$$

**Etude du risque  $R_n(\lambda)$**

On dira que  $\mu_\lambda$  est consistant si  $R_n(\lambda) \rightarrow 0$  quand  $n \rightarrow \infty$ .

Le premier terme dans l'expression du risque  $R_n(\lambda)$  est la moyenne empirique des carrés des biais. Intuitivement, cette quantité doit décroître quand  $\lambda$  croît, puisque  $\lambda$  augmente, on prend de plus en plus de termes dans l'approximation

$\mu_\lambda(t)$  de  $\mu(t)$ .

Donc, pour que  $R_n(\lambda)$  converge vers zéro, on doit avoir nécessairement  $\lambda \rightarrow \infty$  quand  $n \rightarrow \infty$ .

Le deuxième terme figurant dans le risque  $R_n(\lambda)$  est  $\frac{\lambda}{n}\sigma^2$ , correspondant à la variance de l'estimateur  $\mu_\lambda$ . Ce terme augmente avec  $\lambda$  de sorte que  $R_n(\lambda)$  soit minné par zéro pourvu que  $\frac{\lambda}{n} \rightarrow 0$ . Ainsi, la consistance de  $\mu_\lambda$  exige que  $\lambda$  augmente avec une vitesse plus petite que  $n$ .

Pour établir, la consistance, on est tenté d'abord de caractériser le comportement asymptotique de la quantité

$$\frac{1}{n} \sum_{j=1}^n [\mu(t_j) - E(\mu_\lambda(t_j))]^2.$$

Nous avons également besoin de conditions explicites sur les éléments du système complet orthonormal.

Intuitivement, son comportement doit dépendre des coefficients de Fourier non estimés et de la possibilité d'estimer les  $\beta_j$  par les  $\beta_{\lambda_j}$ .

Shibata (1981) donne les résultats de la consistance des estimateurs des moindres carrés pour le modèle (2.17). Li (1984) a étudié l'existence des estimateurs consistants pour certaines fonctionnelles des  $\beta_j$ , c'est à dire la valeur de  $\mu$  en un point fixé

$t$  . Il donne une condition nécessaire et suffisante pour l'existence d'un estimateur consistant et construit un estimateur consistant . D'autres estimateurs consistants proches de l'estimateur donnés par la relation (2.18) sont étudiés par Rutkowski (1982) et Rafajlowicz (1987) .

Supposons que nous avons décidé d'utiliser un estimateur de la forme (2.18) pour estimer  $\mu$  .

Il est nécessaire , dans ce cas , de sélectionner une valeur de  $\lambda$  . Le choix de  $\lambda$  influe sur le lissage ou non de l'estimateur en question . Souvent , les grandes valeurs de  $\lambda$  dans les estimateurs entraînent une interpolation proche des  $(t_i, y_i)$  que les petites valeurs .

En particulier , puisque , nous avons supposé que la matrice  $X_\lambda$  est de rang  $\lambda$  le nombre de colonnes dans la relation (2.19) , la valeur  $\lambda = n$  correspond au cas de l'interpolation où  $\mu_\lambda(t_j) = y_j$  ,  $j = 1, \dots, n$  .

Le problème du choix optimal de  $\lambda$  est similaire à la variable de sélection discuté dans le chapitre précédent et ainsi utiliser la même démarche qui est développée . Pour toute valeur particulière de  $n$  , une valeur optimale  $\lambda$  est celle qui minimise la fonction perte  $L_n(\lambda)$  . Pour minimiser  $L_n(\lambda)$  , nous pouvons par exemple utiliser le *GCV* . L'estimateur optimal *GCV* de  $\lambda$  optimal est  $\hat{\lambda}_n$  obtenu en minimisant

$$GCV_n(\lambda) = \frac{\left\{ \frac{1}{n} \sum_{j=1}^n [y_j - \mu_\lambda(t_j)]^2 \right\}}{\left( 1 - \frac{\lambda}{n} \right)^2} . \quad (2.20)$$

Sous certaines conditions , la méthode *GCV* de sélection de la valeur  $\lambda$  est asymptotiquement optimal ( $n \rightarrow \infty$ ) comme le montre le théorème suivant donné par Li

(1987) .

**Théorème 1 (Li , 1987)**

*Supposons que*

*i) Les  $\varepsilon_j$  soient indépendantes et identiquement distribuées avec  $E(\varepsilon_1^4) < \infty$  et qu'il existe une constante  $K$  telle que  $\forall c \geq 0$  ,*

$$\sup_x P(x - c \leq \varepsilon_1 \leq x + c) \leq Kc \ ,$$

*ii)  $\inf_{\lambda \leq n} nR_n(\lambda) \longrightarrow \infty$  quand  $n \longrightarrow \infty$  ,*

et

*iii)  $\inf_{\lambda \leq n} L_n(\lambda) \xrightarrow{P} \cdot$  , quand  $n \rightarrow \infty$  .*

*Alors , si  $\hat{\lambda}_n$  est la valeur qui rend minimum  $GCV_n(\lambda)$  donné par la relation (2.20) , on a*

$$L_n(\hat{\lambda}_n) / \inf_{\lambda \leq n} L_n(\lambda) \xrightarrow{P} 1 \ , \text{ quand } n \rightarrow \infty \quad (2.21)$$

**Remarques 2**

1) Le théorème précédent montre que , l'estimateur  $GCV$  ,  $\hat{\lambda}_n$  et la valeur optimale de  $\lambda$  sont semblable pour  $n$  suffisamment grand , dans le sens où ils ont asymptotiquement la même fonction perte .

Par ailleurs , la formule (2.21) montre aussi que  $L_n(\hat{\lambda}_n)$  tend vers la fonction perte minimale quand  $n \rightarrow \infty$  .

2) -La condition *i*) du théorème précédent porte sur les erreurs .

Si les erreurs ont une densité  $f_\varepsilon$  bornée , alors

$$\begin{aligned} p(x - c \leq \varepsilon \leq x + c) &= \int_{x-c}^{x+c} f_\varepsilon(t) dt \\ &\leq \left[ \sup_t f(t) \right] 2c \end{aligned}$$

et ainsi la relation (2.20) est satisfaite montre que la relation (2.21) est moins restrictive .

- La condition *ii*) montre que la vitesse de convergence optimale de  $R_n(\lambda)$  tendant vers zéro doit être plus petite que  $1/n$  .

- La condition *iii*) exige simplement l'existence d'un estimateur consistant dans le sens que la fonction perte tende vers zéro . Pour établir ceci , nous avons besoin de montrer qu'il existe une suite  $(\lambda_n)$  telle que  $I_n(\lambda_n) \rightarrow 0$  .

Une condition plus forte entrainant *iii*) est que  $\inf_{\lambda \leq n} R_n(\lambda) \rightarrow 0$  quand  $n \rightarrow 0$  . Cette dernière condition est plus facile à vérifier , en pratique .

Dans la section suivante nous allons étudier l'estimateur basé sur la série de Fourier dans le cas particulier où le système complet orthonormale est une famille de fonctions complexes . Nous allons voir qu'il est possible d'établir , la consistance pour  $\mu_\lambda$  dans ce cas .

## 2.4 Estimateurs des séries de Fourier

Dans cette section , nous allons étudier les propriétés de l'estimateur de la série de Fourier résultant de l'utilisation des fonctions complexes pour un système complet orthonormale . Ces résultats correspondent aux concepts classiques des séries de Fourier .

C'est pour cette raison que nous appellerons

“ estimateurs des séries de Fourier ” .

Les méthodes de Fourier sont jusqu'à présent très utilisées dans le domaine de l'analyse des séries temporelles .

La motivation est due au fait que les observations dans le temps , exhibent des fluctuations périodiques . Ces phénomènes périodiques se présentent souvent dans certaines situations , malgré que dans certains cas , ils sont moins fréquents . Par exemple , les ventes en détail présentent des périodicités annuelles .

Le comportement période nous fait penser aux fonctions sinus et cosinus . Donc , il est naturel pour l'analyse de séries temporelles de considérer l'utilisation de modèle de régression comportant des fonctions sinus et cosinus pour décrire le comportement de séries temporelles . L'idée est d'utiliser une combinaison linéaire des fonctions 1 et  $\{\cos(2\pi jt) , \sin(2\pi jt)\}$  ,  $j = 1, \dots, \lambda$  , pour un entier  $\lambda$  non négatif pour modéliser une série.

Nous nous intéressons ici , aux modèles de régression où  $t$  n'est pas forcément le temps .

Pour estimer  $\mu$ , on ajuste un modèle linéaire comportant les premières fonctions

sinus et cosinus . Cette décomposition est définie comme

“ estimateur série de Fourier ” de  $\mu$  .

Pour cela, nous supposons allons faire quelques hypothèses en supposant que

$$[a, b] = [0, 1] \quad (2.22a)$$

et que les  $t_j$  sont uniformément équidistant sur  $[0, 1]$  , c'est à dire que

$$t_j = \frac{j-1}{n} , \quad j = 1, \dots, n \quad (2.22b)$$

### 2.4.1 Forme de l'estimateur

Dans cette section , nous allons donner quelques formes pour l'estimateur série de Fourier  $\mu_\lambda$ , sous l'hypothèse (2.22) .

Posons

$$x_j(t) = e^{2\pi i j t} , \quad j \in \mathbb{Z} ,$$

et pour un entier  $\lambda \geq 0$  , considérons la fonction ajustée  $\sum_{j=-\lambda}^{\lambda} \beta_j x_j(t)$  aux données observées par la méthode des moindres carrés .

Les estimateurs des  $\beta_j$  ont pour expression

$$\underline{\beta}_\lambda = (\beta_{\lambda(-\lambda)}, \dots, \beta_{\lambda\lambda})' = (X_\lambda^* X_\lambda)^{-1} X_\lambda^* \underline{y} , \quad (2.23)$$

où

$$X_\lambda = (e^{2\pi i j r / n})_{\substack{r=0,1,\dots,n-1 \\ j=-\lambda,\dots,\lambda}} .$$

On peut obtenir une forme plus simplifiée en utilisant l'identité

$$\frac{1}{n} \sum_{r=0}^{n-1} e^{2\pi ijr/n} e^{-2\pi ikr/n} = \begin{cases} 1 & \text{si } j = k \\ 0 & \text{si } j \neq k . \end{cases} \quad (2.24)$$

Ainsi  $X_\lambda^* X_\lambda = nI$  (où  $I$  est la matrice identité) . Donc les composantes de  $\underline{\beta}_\lambda$  de la relation (2.23) s'écrivent

$$\beta_{\lambda j} = \frac{1}{n} \sum_{k=1}^n y_k e^{-2\pi ijk/n} \quad , \quad j = -\lambda, \dots, \lambda . \quad (2.25)$$

L'estimateur de  $\mu$  a ainsi pour expression

$$\mu_\lambda(t) = \sum_{j=-\lambda}^{\lambda} \beta_{\lambda j} e^{2\pi ijt} . \quad (2.26)$$

Remarquons que sous l'échantillonnage fixé utilisé , les  $\beta_{\lambda j}$  ne dépendent pas de  $\lambda$  . Nous continuerons quand même à utiliser  $\lambda$  en indice .

L'utilisation d'un échantillonnage régulier des  $t_j$  sur  $[0, 1]$  nous fournit l'orthogonalité des fonctions exponentielles dans la relation (2.26) .

Les  $\beta_{\lambda j}$  peuvent être pris comme des estimateurs des coefficients de Fourier dans le développement en série de Fourier de la fonction  $\mu$  .

En remplaçant les expressions des  $\beta_{\lambda j}$  de la relation (2.25) dans celle de  $\mu_\lambda$  dans la relation (2.26) , on obtient

$$\begin{aligned} \mu_\lambda(t) &= \sum_{j=-\lambda}^{\lambda} \beta_{\lambda j} e^{2\pi ijt} \\ &= \frac{1}{n} \sum_{k=1}^n y_k \sum_{j=-\lambda}^{\lambda} e^{2\pi ijk(t-t_k)} . \end{aligned}$$

En utilisant, maintenant l'identité

$$\sum_{j=-\lambda}^{\lambda} e^{2\pi i j x} = \frac{\sin [\pi(2\lambda + 1)x]}{\sin \pi x} \equiv K_{\lambda}(x) \quad , \quad (2.27)$$

on obtient on a l'expression finale de  $\mu_{\lambda}(t)$

$$\mu_{\lambda}(t) = \frac{1}{n} \sum_{k=1}^n y_k K_{\lambda}(t - t_k) \quad . \quad (2.28)$$

La fonction  $K_{\lambda}(x)$  définie dans la relation (2.27) est appelée le

“ noyau de Dirichlet ”. C'est une fonction périodique de période 1 . Si  $u$  est un entier , on définit  $K_{\lambda}(u) = (2\lambda + 1)$  .

Par ailleurs

$$\sum_{r=1}^n K_{\lambda}(t - t_k) = n \quad .$$

Ainsi , d'après l'équation (2.28) , la fonction  $\mu_{\lambda}(t)$  peut être considérée comme étant la moyenne pondérée des  $y_k$  , dont les poids sont obtenus à partir du noyau de Dirichlet .

Un exemple des graphes du noyau de Dirichlet est illustré dans la figure 2.1 pour différentes valeurs de  $\lambda$  .

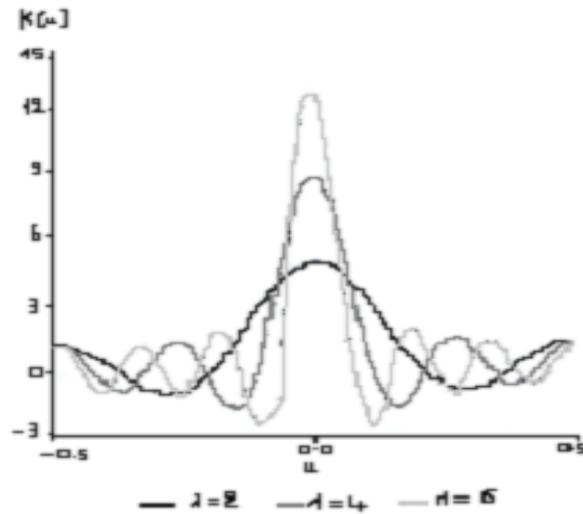


FIG. 2.1 – Exemples des noyaux de Dirichlet

On voit que plus  $\lambda$  est grand, plus la courbe montre un pic en zéro. Ce qui montre que pour les grandes valeurs de  $\lambda$ , les grandes valeurs des poids pour le calcul de  $\mu_\lambda(t)$ , sont au voisinage de  $t$ .

Lorsque  $\lambda$  augmente, on remarque aussi que le noyau oscille de plus en plus à partir du pic.

Le rôle de  $\lambda$  est de déterminer comment dépend l'estimateur du comportement local des observations. La croissance de  $\lambda$  correspond à une augmentation de dépendance locale et qui entraîne une oscillation forte de l'estimateur.

Une alternative est de remplacer le noyau de Dirichlet dans la relation (2.28) par une fonction poids avec un meilleur comportement.

## 2.4.2 Etude asymptotique

Dans cette section , nous allons étudier le comportement asymptotique de la fonction risque quand  $n$  et  $\lambda$  augmentent .

On définit le risque en estimant  $\mu$  avec  $\mu_\lambda$  par

$$R_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} E \left| \mu\left(\frac{k}{n}\right) - \mu_\lambda\left(\frac{k}{n}\right) \right|^2 .$$

Notre objectif est d'étudier le comportement de  $R_n(\lambda)$  comme fonction de  $n$  et de  $\lambda$  , sous certaines restrictions sur  $\mu$  .

Nous commençons par supposer que

$$\mu \in W_{2,per}^2 [0, 1] = \left\{ \mu \left/ \begin{array}{l} \mu \in W_2^2 [0, 1] \text{ et} \\ \mu(0) = \mu(1) , \mu'(0) = \mu'(1) \end{array} \right. \right\} .$$

Donc  $W_{2,per}^2 [0, 1]$  représente une classe de fonctions lisse et périodique . Sous cette hypothèse , la série de Fourier pour  $\mu$  converge vers  $\mu$  , uniformément en  $t$  .

Puisque

$$\begin{aligned} tr X_\lambda (X_\lambda^* X_\lambda)^{-1} X_\lambda^* &= \frac{1}{n} tr X_\lambda^* X_\lambda \\ &= 2\lambda + 1 . \end{aligned}$$

On montre que

$$R_n(\lambda) = \frac{1}{n} \sum_{k=0}^{n-1} \left| \mu\left(\frac{k}{n}\right) - E \left[ \mu_\lambda\left(\frac{k}{n}\right) \right] \right|^2 + \left( \frac{2\lambda + 1}{n} \right) \sigma^2 . \quad (2.29)$$

Puisque la somme dans la relation (2.29) est non négative , il s'ensuit que  $R_n(\lambda)$  est minoré par zéro si  $\frac{\lambda}{n} \longrightarrow 0$  , quand  $n \rightarrow \infty$  . Donc une condition nécessaire pour

que le risque tende vers zéro est que le nombre de termes dans “ l’estimateur série de Fourier ” ait un ordre plus petit que la taille de l’échantillon  $n$  .

Pour cela , commençons par écrire une expression convenable pour  $\mu(t) - E(\mu_\lambda(t))$

.

Puisque  $\mu$  peut s’écrire sous forme d’une série de Fourier , on a

$$\mu\left(\frac{k}{n}\right) = \sum_{j=-\infty}^{\infty} \beta_j e^{2\pi i j k/n} , \quad k = 0, 1, \dots, n-1 , \quad (2.30)$$

où

$$\beta_j = \int_0^1 \mu(t) e^{-2\pi i j t} dt .$$

Il est clair que  $E[\mu_\lambda(t)]$  dépend de  $E(\beta_{\lambda_j})$  ,  $j = -\lambda, \dots, \lambda$  .

En prenant l’espérance de  $\beta_{\lambda_j}$  de la relation (2.25) et en utilisant la relation (2.30)

, on a

$$\begin{aligned} E(\beta_{\lambda_j}) &= \frac{1}{n} \sum_{k=0}^{n-1} \mu\left(\frac{k}{n}\right) e^{-2\pi i j k/n} \\ &= \sum_{r=-\infty}^{\infty} \beta_r \left\{ \frac{1}{n} \sum_{k=0}^{n-1} e^{-2\pi i k(j-r)/n} \right\} \\ &= \beta_j + \sum_{r \neq 0} \beta_{j+nr} . \end{aligned} \quad (2.31)$$

En utilisant les relations (2.30) et (2.31) , on a

$$\mu(t) - E[\mu_\lambda(t)] = \sum_{|j| > \lambda} \beta_j e^{2\pi i j t} - \sum_{|J| \leq \lambda} \left( \sum_{r \neq 0} \beta_{j+nr} \right) e^{2\pi i j t} . \quad (2.32)$$

Par ailleurs , la relation (2.31) nous donne l'expression pour le biais en estimant  $\beta_j$  par  $\beta_{\lambda_j}$  , en effet

$$\text{Biais}\beta_{\lambda_j} = E\beta_{\lambda_j} - \beta_j = \sum_{r \neq 0} \beta_{j+nr} .$$

Maintenant , en remplaçant la formule (2.32) dans la relation (2.29) et en utilisant la relation (2.24) , on a

$$R_n(\lambda) = \sum_{|j| > \lambda} |\beta_j|^2 + \left( \frac{2\lambda + 1}{n} \right) \sigma^2 + \sum_{|j| > \lambda} \beta_j \sum_{r \neq 0} \theta_{j+nr}^* , \quad (2.33)$$

où

$$\theta_j = \begin{cases} \beta_j & \text{si } |j| > \lambda \\ \beta_j - E(\beta_{\lambda_j}) & \text{si } |j| \leq \lambda , \end{cases}$$

avec la condition que  $2\lambda + 1 < n$  .

Cette expression montre comment  $R_n(\lambda)$  dépend des coefficients de Fourier

de  $\mu$  . En particulier , en utilisant la relation de Parseval , il est clair que  $\sum_{|j| > \lambda} |\beta_j|^2$  est la norme dans  $L_2 [0, 1]$  de l'erreur de l'approximation de  $\mu$  par la somme partielle

$$\sum_{j=-\lambda}^{\lambda} \beta_j e^{2\pi i j t} \text{ de sa série de Fourier .}$$

Pour  $n$  suffisamment grand , le dernier terme de la relation (2.33) est en général négligable relativement aux deux autres . Donc , asymptotiquement , une sélection optimale de  $\lambda$  correspond à un balancement de l'erreur stochastique ou du terme  $\left( \frac{2\lambda + 1}{n} \right) \sigma^2$  et de l'erreur déterministe de l'approximation de  $\mu$  par la somme partielle de sa série de Fourier .

Pour étudier le comportement asymptotique de  $R_n(\lambda)$  , il devient nécessaire de faire hypothèses sur les coefficients de Fourier de  $\mu$  . D'après la relation de

Parseval ,  $|\beta_j|^2 \longrightarrow 0$  quand  $j \rightarrow \infty$  . Cependant , plus d'hypothèses précises concernant la vitesse de convergence des  $|\beta_j|$  peuvent être faites pour des fonctions lisses .

Par exemple , si  $\mu \in W_{2,per}^2[0,1]$  , sa seconde dérivée est dans  $L_2[0,1]$  et par l'intégration par parties montre que

$$\begin{aligned} \beta_j &= \int_0^1 \mu(t) e^{-2\pi i j t} dt \\ &= -\frac{1}{(2\pi j)^2} \int_0^1 \mu''(t) e^{-2\pi i j t} dt \\ &= -\frac{\beta_j''}{(2\pi j)^2} . \end{aligned}$$

Puisque

$$|\mu''|^2 = \sum_{j=-\infty}^{\infty} |\beta_j''|^2 = \sum_{j=-\infty}^{\infty} (2\pi j)^4 |\beta_j|^2 < \infty ,$$

il s'ensuit que

$$|\beta_j|^2 = o\left(\frac{1}{j^4}\right) .$$

Donc , le carré du module des coefficients de Fourier pour toute fonction dans  $W_{2,per}^2[0,1]$  , doit tendre vers zéro plus vite que  $j^{-4}$ .

Une des possibilités de modéliser les coefficients de Fourier pour  $\mu$  est de supposer que

$$|\beta_j|^2 \sim C |j|^{-5-\delta} \tag{2.34}$$

où  $C$  et  $\delta$  sont des constantes positives .

Il est clair que si les  $\beta_j$  vérifient cette condition , il s'ensuit que  $|\mu''| < \infty$  . En plus avec les résultats des calculs concernant la continuité et la différentiabilité de la limite d'une suite de fonctions uniformément convergente , il est clair que la condition (2.34) suffit pour assurer que  $\mu \in W_{2,per}^2 [0, 1]$  .

Sous la condition (2.34) , l'amplitude du biais de  $\beta_{\lambda j}$  devient comme une constante multiple de

$$\sum_{r \neq 0} |j + nr|^{-(5+\delta)/2} .$$

Pour étudier cette somme , il suffit d'estimer la taille de  $\sum_{r=1}^{\infty} (nr \pm j)^{-(5+\delta)/2}$  . Ceci peut être fait , en utilisant une approximation de l'intégrale pour obtenir

$$\begin{aligned} \sum_{r=1}^{\infty} |nr \pm j|^{-(5+\delta)/2} &\leq \int_1^{\infty} (nx \pm j)^{-(5+\delta)/2} dx \\ &= \frac{2}{(3+\delta)} (n \pm j)^{-(3+\delta)/2} \\ &= o\left(\frac{1}{n^{(5+\delta)/2}}\right) , \end{aligned} \tag{2.35}$$

uniformément en  $|j| \leq \frac{n-1}{2}$  .

Un calcul simple utilisant la relation (2.35) montre que le dernier terme dans la relation (2.33) est de l'ordre  $o(n^{-(5+\delta)/2})$  .

Une approximation par une intégrale peut être utilisée pour montrer que lorsque  $\lambda \rightarrow \infty$  on a avec la relation (2.34) pour montrer que lorsque  $\lambda \rightarrow \infty$  ,

$$\sum_{|j| > \lambda} |\beta_j|^2 \sim 2C_1 \lambda^{-4-\delta} , \tag{2.36}$$

où  $C_1 = \frac{C}{4+\delta}$  .

En combinant les résultats précédentes avec la relation (2.33) , on a l'expression asymptotique de  $R_n(\lambda)$  .

Si  $\lambda \rightarrow \infty$  ,  $n \rightarrow \infty$  telle que  $\frac{\lambda}{n} \rightarrow 0$  , alors

$$R_n(\lambda) \sim 2C_1\lambda^{-4-\delta} + \frac{2\lambda}{n}\sigma^2 . \quad (2.37)$$

Pour déterminer la valeur asymptotique optimale de  $\lambda$  , on minimise  $R_n(\lambda)$  dans la relation (2.37) , en annulant sa dérivée pour obtenir

$$\lambda_{opt} \propto n^{1/(5+\delta)} . \quad (2.37a)$$

En remplaçant cette quantité dans la relation (2.37) , on a

$$R_n(\lambda_{opt}) = O(n^{-(4+\delta)/(5+\delta)}) . \quad (2.37b)$$

Voici quelque remarques concernant les deux relations précédentes :

- (i)  $\mu_\lambda$  est un estimateur consistant de  $\mu$  pourvu que  $n$  ,  $\lambda \rightarrow \infty$  et  $\lambda/n \rightarrow 0$  .
- (ii) La vitesse optimale du risque est plus petite que  $1/n$  .
- (iii) D'après la relation (2.37a) , le nombre de termes contenus dans l'estimateur " série de Fourier " augmente en restant inférieure à la taille l'échantillon .

Par exemple , pour avoir une idée de l'ordre de grandeur :

$$100^{1/5} = 2.5 , 1000^{1/5} = 4 , 10000^{1/5} = 6.3 \text{ et } 100000^{1/5} = 10 .$$

Ceci ,bien sûr , ne veut pas dire que si  $n = 1000$  , on utilise approximativement 4 termes dans notre estimateur séries de Fourier , puisque cette situation néglige la constante de proportionnalité dans la relation (2.37a) et la valeur de  $\delta$  dans la relation (2.34) .

### 2.4.3 Distribution asymptotique

L'équation  $\mu_\lambda(t) = \frac{1}{n} \sum_{k=1}^n y_k K_\lambda(t-t_k)$  donnée par la relation (2.28) est une moyenne pondérée de variables aléatoires non corrélées . On peut donc , espérer que le théorème central limite peut s'appliquer aux estimateurs des séries de Fourier .

Dans cette section , on va étudier le comportement probabiliste de  $\mu_\lambda$  quand

$n, \lambda \longrightarrow \infty$  . On va supposer que les erreurs ( $\varepsilon$ ) sont indépendantes et identiquement distribuées (i.i.d.) avec  $E(|\varepsilon|^{2+\alpha}) < +\infty, \forall \alpha > 0$  .

L'outil qu'on va utiliser pour étudier la distribution asymptotique de  $\mu_\lambda(t)$  est le théorème central limite de Liapunov .

#### **Théorème 2 (de Liapunov )**

*Soit  $\{V_{jn} : 1 \leq j \leq n; n = 1, 2, \dots\}$  un tableau triangulaire de variables aléatoires indépendantes pour tout  $j = 1, \dots, n$  et pour tout  $n = 1, 2, \dots$  .*

*Posons*

$$\sigma_n = \left\{ \sum_{j=1}^n E [V_{jn} - E (V_{jn})]^2 \right\}^{1/2}$$

*et pour tout  $\theta > 2$  ,*

$$C_n(\theta) = \sum_{j=1}^n E |V_{jn} - E (V_{jn})|^\theta .$$

*Alors , si  $C_n(\theta)/\sigma_n^\theta \longrightarrow 0$  quand  $n \longrightarrow \infty$  ,*

*la distribution asymptotique de  $\sum_{j=1}^n [V_{jn} - E (V_{jn})] / \sigma_n$  est une loi normale standard*

#### **Remarque 3**

Remarquons que l'utilisation du théorème de Liapunov n'exige pas que les variables dans la somme  $\sum_{j=1}^n [V_{jn} - E(V_{jn})]$  soient identiquement distribuées .

C'est précisément notre cas . En utilisant la relation (2.28) , on a

$$\begin{aligned} \mu_\lambda(t) - E[\mu_\lambda(t)] &= \frac{1}{n} \sum_{j=1}^n [y_j - E(y_j)] K_\lambda(t - t_j) \\ &= \frac{1}{n} \sum_{j=1}^n \varepsilon_j K_\lambda(t - t_j) \quad , \end{aligned}$$

où  $K_\lambda(\cdot)$  est donnée par la relation (2.27) .

Ainsi ,  $\mu_\lambda(t) - E[\mu_\lambda(t)]$  s'écrit sous la forme  $\sum_{j=1}^n [V_{jn} - E(V_{jn})]$  avec des variables  $V_{jn}$  indépendantes mais ne sont pas identiquement distribuées définies par

$$V_{jn} = \frac{1}{n} \varepsilon_j K_\lambda(t - t_j)$$

On a

$$E(V_{jn}) = 0 \text{ et } E(V_{jn}^2) = \frac{\sigma^2}{n^2} |K_\lambda(t - t_j)|^2,$$

avec  $\sigma^2 = \text{Var}(\varepsilon_j)$  .

Donc

$$\begin{aligned} \sigma_n^2 &= \sum_{j=1}^n E(V_{jn}^2) = \frac{\sigma^2}{n^2} \sum_{j=1}^n |K_\lambda(t - t_j)|^2 \\ &= \frac{\sigma^2}{n^2} \sum_{j=1}^n \sum_{r=-\lambda}^{\lambda} \sum_{k=-\lambda}^{\lambda} e^{2\pi i r(t-t_j)} e^{-2\pi i k(t-t_j)} \\ &= \frac{\sigma^2}{n} (2\lambda + 1) \quad , \end{aligned}$$

avec la dernière égalité obtenue , en utilisant la relation (2.24) .

De même , en utilisant la relation (2.24) et le fait que  $|K_\lambda(u)| \leq (2\lambda + 1)$  , on a

$$\begin{aligned} \sum_{j=1}^n E |V_{jn}|^{2+\alpha} &= \frac{1}{n^{2+\alpha}} E |\varepsilon_1|^{2+\alpha} \sum_{j=1}^n |K_\lambda(t - t_j)|^{2+\alpha} \\ &< O\left(\left(\frac{\lambda}{n}\right)^{1+\alpha}\right) . \end{aligned}$$

Une application du théorème de Liapovov avec  $\theta = 2 + \alpha$  montre que  $\sum_{j=1}^n \frac{V_{jn}}{\sigma_n}$  est asymptotiquement normale standard , pourvu que

$$\sum_{j=1}^n \frac{E |V_{jn}|^{2+\alpha}}{[\sigma^2(2\lambda + 1)/n]^{1+\alpha/2}} = O((\lambda/n)^{\alpha/2})$$

converge vers zéro quand  $n \rightarrow \infty$  .

Ce résultat est résumé par le théorème suivant .

### **Théorème 3**

*Soit  $\lambda, n \rightarrow \infty$  tels que  $(\lambda/n) \rightarrow 0$  . Alors pour tout  $t$  fixé.*

$$\sqrt{n} \frac{\mu_\lambda(t) - E[\mu_\lambda(t)]}{\sigma\sqrt{2\lambda + 1}}$$

*converge en loi vers une variable aléatoire normale standard .*

### **Remarque 4**

Notons que sous certaines conditions , on a

$$\sqrt{n} \frac{\mu_\lambda(t) - \mu(t)}{\sigma\sqrt{2\lambda + 1}}$$

tend en loi vers une variable aléatoire normale standard , quand  $n \rightarrow \infty$

d'après le corollaire suivant .

Il suffit , en effet de décomposer

$$\sqrt{n} \frac{\mu_\lambda(t) - \mu(t)}{\sigma\sqrt{2\lambda+1}} = \sqrt{n} \frac{\mu_\lambda(t) - E[\mu_\lambda(t)]}{\sigma\sqrt{2\lambda+1}} + \sqrt{n} \frac{E[\mu_\lambda(t)] - \mu(t)}{\sigma\sqrt{2\lambda+1}} .$$

Le premier terme du second membre tend en loi vers une variable aléatoire normale standard d'après le théorème précédent et si le second terme converge vers zéro , alors on a le résultat .

D'après l'étude faite dans la section précédente , on a

$$\begin{aligned} |\mu(t) - E[\mu_\lambda(t)]| &= \left| \sum_{|j|>\lambda} \beta_j e^{2\pi i j t} - \sum_{|j|\leq\lambda} \left( \sum_{r\neq 0} \beta_{j+rn} \right) e^{2\pi i j t} \right| \\ &\leq \sum_{|j|>\lambda} |\beta_j| + \sum_{|j|\leq\lambda} \left| \sum_{r\neq 0} \beta_{j+rn} \right| \\ &= O\left(\lambda^{-(3+\delta)/2}\right) + O\left(\lambda n^{-(5+\delta)/2}\right) , \end{aligned}$$

pour vu que la relation (2.34) soit vérifiée .

Si  $\frac{\lambda}{n} \rightarrow 0$  , le premier terme domine le second , donc

$$\sqrt{n} \frac{|\mu(t) - E[\mu_\lambda(t)]|}{\sigma\sqrt{2\lambda+1}} \leq \sqrt{n} O\left(\lambda^{-(4+\delta)/2}\right)$$

uniformément en  $t$  . Ce qui donne le Corollaire suivant(2.34)

### Corollaire 1

*Supposons que  $\mu \in W_{2,per}^2[0,1]$  et que ses coefficients de Fourier satisfont la condition (2.34) . Alors , pour tout  $t$  fixé ,  $\sqrt{n} \frac{\mu_\lambda(t) - \mu(t)}{\sigma\sqrt{2\lambda+1}}$  converge en loi vers une variable aléatoire normale standard , quand  $n \rightarrow \infty$  et  $\lambda \rightarrow \infty$  tel que  $n\lambda^{-(4+\delta)} \rightarrow 0$*

**Remarque 5**

Ce Corollaire nous permet de construire un intervalle de confiance pour  $\mu(t)$ .

Comme c'est un résultat asymptotique, on a pour une grande taille

d'échantillon, un intervalle de confiance pour  $\mu(t)$ , de coefficient de confiance

$p = 1 - \alpha$ , donné par

$$\left[ \mu_\lambda(t) - Z_{\alpha/2} \sigma \sqrt{\frac{2\lambda + 1}{n}}, \mu_\lambda(t) + Z_{\alpha/2} \sigma \sqrt{\frac{2\lambda + 1}{n}} \right],$$

où  $Z_{\alpha/2}$  est telle que  $P(Z > Z_{\alpha/2}) = \frac{\alpha}{2}$ , avec  $Z \sim N(0, 1)$ . Cela suppose, bien sûr que  $\sigma$  est connu.

Si  $\sigma$  est inconnu, il faut l'estimer. Un des estimateurs les plus simples est

construit, en répétant les observations en chaque  $t_j$ . Par exemple, nous avons

les observations

$$y_{rj} = \mu(t_j) + \varepsilon_{rj}, \quad r = 1, \dots, n_j, \quad j = 1, \dots, n.$$

Il est facile de vérifier que

$$S^2 = \frac{1}{N - n} \sum_{j=1}^n \sum_{r=1}^{n_j} (y_{rj} - \bar{y}_j)^2,$$

avec  $\bar{y}_j = \frac{1}{n_j} \sum_{r=1}^{n_j} y_{rj}$  et  $N = \sum_{j=1}^n n_j$ , est un estimateur sans biais de  $\sigma^2$ .

Lorsqu'il n'y a pas de répétitions d'observations en chaque  $t_j$ , des estimateurs consistants peuvent être construits, en utilisant la différence des observations.

Par exemple, lorsque  $n$  est pair et  $\mu$  est suffisamment régulière

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n/2} (y_{2j} - y_{2j-1})^2$$

est un estimateur consistant de  $\sigma^2$ .

Un autre estimateur de  $\sigma^2$  est construit en utilisant la similitude entre les séries de Fourier et la régression linéaire, il est donné par

$$\sigma_\lambda^2 = \frac{1}{n - (2\lambda + 1)} \sum_{j=1}^n [y_j - \mu_\lambda(t_j)]^2.$$

Calcul de la moyenne et la variance de  $\hat{\sigma}_\lambda^2$ .

On montre que

$$\begin{aligned} E(\sigma_\lambda^2) &= \frac{1}{1 - \frac{2\lambda + 1}{n}} \left[ R_n(\lambda) + \sigma^2 - 2\sigma^2 \frac{(2\lambda + 1)}{n} \right] \\ &\sim R_n(\lambda) + \sigma^2 - 2\sigma^2 \frac{(2\lambda + 1)}{n} \\ &\sim \sigma^2, \end{aligned}$$

pourvu que  $\lambda, n \rightarrow +\infty$  tels que  $\frac{\lambda}{n} \rightarrow 0$ .

Ainsi,  $\sigma_\lambda^2$  est asymptotiquement sans biais de  $\sigma^2$ .

Si, par exemple, les  $\varepsilon$  sont iid  $N(0, \sigma^2)$ ,  $\sigma_\lambda^2$  est un estimateur consistant de  $\sigma^2$ .

En effet, sous la condition de normalité, on a

$$\begin{aligned} Var(\sigma_\lambda^2) &= \frac{1}{[n - (2\lambda + 1)]^2} \{2\sigma^4 [n - 3(2\lambda + 1)] + \sigma^2 n R_n(\lambda)\} \\ &= O\left(\frac{1}{n}\right). \end{aligned}$$

Pourvu que  $\lambda$  tende vers l'infini avec une vitesse de telle sorte que  $R_n(\lambda) \rightarrow 0$  quand  $n \rightarrow \infty$ .

## 2.5 Régression polynomiale

Dans cette section, nous allons étudier la régression polynomiale, du point de vue des approximations de séries.

Posons  $p_j(t) = t^j$ , pour tout  $j = 0, 1, 2, \dots$  et considérons les polynômes de légendre définis par

$$x_0(t) = \frac{p_0(t)}{\|p_0(t)\|} = \frac{1}{\sqrt{(b-a)}}$$

et

$$x_j(t) = \frac{p_j(t) - \sum_{k=0}^{j-1} \langle p_j, x_k \rangle x_k(t)}{\left\| p_j(t) - \sum_{k=0}^{j-1} \langle p_j, x_k \rangle x_k(t) \right\|}.$$

Les fonctions  $x_j(t)$ ,  $j \in \mathbb{N}$  sont appelés les polynômes de légendre et forment un système complet orthonormal dans  $L_2[a, b]$ . Si la série de Fourier généralisée pour  $\mu$  converge uniformément, on peut alors écrire le modèle (2.1) comme un modèle de régression polynomiale infinie. La convergence d'une telle série dépend de la convergence d'une série de Fourier classique pour une fonction provenant de  $\mu$ . Ceci a pour conséquence que les estimateurs de régression polynomiale sont aussi proches des estimateurs de séries de Fourier classiques. Pour estimer  $\mu$ , on peut alors approximer sa série de Fourier de légendre par sa somme partielle  $\sum_{j=0}^{\lambda-1} \beta_j x_j$  et estimer les  $\beta_j$  par la méthode des moindres carrés et sont appelés

“ estimateur de régression polynomiale ”.

Une des motivations de l'utilisation des estimateurs de régression polynomiale découle du théorème de Taylor suivant .

### **Théorème de Taylor**

Soit  $\mu(t) \in W_2^m [a, b]$  , alors il existe des coefficients  $\theta_0, \dots, \theta_{m-1}$  tels que

$$\mu(t) = \sum_{j=0}^{m-1} \theta_j t^j + \int_a^b \frac{(t-u)_+^{m-1}}{(m-1)!} \mu^{(m)}(u) du \ ,$$

où

$$(x)_+^r = \begin{cases} x^r & \text{si } x \geq 0 \\ 0 & \text{si } x < 0 \end{cases} .$$

### **Preuve :**

Il suffit d'écrire

$$\mu(t) = \int_a^b (t-u)_+^0 \mu'(u) du - \mu(a)$$

et ensuite intégrer par parties .

### **Remarque 6**

Le théorème de Taylor nous montre que pour un entier non négatif  $\lambda$  , le reste du développement de la fonction  $\mu(t)$  .

$$R_\lambda(t) = \frac{1}{(\lambda-1)!} \int_a^b (t-u)_+^{\lambda-1} \mu^{(\lambda)}(u) du$$

est uniformément à t petit et nous pouvons ainsi approximer le modèle par

$$y_k = \sum_{j=0}^{\lambda-1} \theta_j t_k^j + \varepsilon_k \ , \quad k = 1, \dots, n \ .$$

# Chapitre 3

## Estimation de la variance dans un modèle de régression non paramétrique

### 3.1 Introduction

On considère le modèle de régression non paramétrique donné par

$$y_j = \mu(t_j) + \varepsilon_j \quad , \quad j = 1, \dots, n \quad ,$$

où les  $(t_j)$  est un échantillonnage fixé de l'intervalle  $[0, 1]$  ,  $\mu$  étant une fonction réelle inconnue définie sur  $[0, 1]$  et les  $(\varepsilon_j)$  est une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne nulle et de variance

inconnue  $\sigma^2$  .

Notre travail consiste à estimer la variance  $\sigma^2$  basé sur les coefficients de Fourier de la fonction de régression  $\mu$ . Plus précisément, soit

$$c_k = \int_0^1 e^{-2\pi ikt} \mu(t) dt, \quad k \in \mathbb{Z}, \quad (3.1)$$

les coefficients de Fourier de  $\mu$ .

L'estimateur que nous utilisons est basé sur les estimateurs empiriques  $\hat{c}_k$  de  $c_k$  définis par

$$\hat{c}_k = \frac{1}{n} \sum_{j=1}^n Y_j e^{-2\pi i k j/n}, \quad k \in \mathbb{Z}. \quad (3.2)$$

Plus précisément, soit

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n |Y_j - \hat{\mu}(j/n)|^2, \quad (3.3)$$

où  $\hat{\mu}$  est défini comme suit.

On considère une suite d'entier  $q = q(n)$  telle que  $\lim_{n \rightarrow \infty} q(n) = +\infty$  et on pose

$$\hat{\mu}(t) = \sum_{|k| \leq q} \hat{c}_k e^{2\pi ikt}. \quad (3.4)$$

## 3.2 Hypothèses et résultats asymptotiques

Les hypothèses que nous utilisons sont les suivantes

(A<sub>1</sub>) -  $\mu$  satisfait la condition de Lipschitz d'ordre  $\delta$ , avec  $\frac{1}{2} < \delta \leq 1$ , c'est-à-dire qu'il existe une constante positive  $M$  telle que

$$|\mu(s) - \mu(t)| \leq M |s - t|^\delta, \quad \forall s, t \in [0, 1].$$

(A<sub>2</sub>) -  $(\varepsilon_j)$ ,  $j = 1, \dots, n$  sont des variables aléatoires i.i.d de moyenne nulle et de variance inconnue  $\sigma^2$ .

$$(A_3) - \lim_{n \rightarrow \infty} \frac{q(n)}{\sqrt{n}} = 0 \text{ et } \sum_{k \in \mathbb{Z}} |c_k| < \infty .$$

$$(A_4) - \lim_{n \rightarrow \infty} \left\{ \sqrt{p(n)} \sum_{|k| > q(n)} |c_k| \right\} = 0 .$$

$$(A_5) - \lim_{n \rightarrow \infty} \left\{ \frac{p(n)q^2(n)}{n} \right\} = 0 .$$

On a alors la Proposition suivante

**Proposition 1**

Supposons que  $E(\varepsilon_1)^4 < \infty$  et que les conditions  $(A_1)$  et  $(A_2)$  sont satisfaites ,  
alors pour l'estimateur  $\hat{\sigma}^2$  défini par la relation (3.3) , on a les résultats suivants

(i) Si la condition  $(A_3)$  est vérifiée , alors

$$\hat{\sigma}^2 \xrightarrow{P} \sigma^2 , \text{ quand } n \rightarrow \infty .$$

(ii) Si les conditions  $(A_4)$  et  $(A_5)$  sont vérifiées , alors

$$\sqrt{p} (\hat{\sigma}^2 - \sigma^2) \xrightarrow{P} 0 , \text{ quand } n \rightarrow \infty .$$

Avant de démontrer la proposition précédente montrons d'abord les lemmes suivants . Pour cela , notons  $\tilde{c}_k$  le  $k^{\text{ème}}$  coefficient de Fourier discret de la fonction de régression  $\mu$  défini par

$$\tilde{c}_k = \frac{1}{n} \sum_{j=1}^n \mu(j/n) e^{-2\pi i j k / n} . \quad (3.4)$$

On pose

$$\hat{c}_k - \tilde{c}_k = \alpha_k - i\beta_k , \quad (3.5)$$

où

$$\alpha_k = \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cos(2\pi j k / n) \quad (3.6)$$

et

$$\beta_k = \frac{1}{n} \sum_{j=1}^n \varepsilon_j \sin(2\pi jk/n) . \quad (3.7)$$

Remarquons que

$$E(\hat{c}_k) = \tilde{c}_k .$$

### Lemme 1

Pour  $\alpha_k$  et  $\beta_k$  définis respectivement par les relations (3.6) et (3.7), on a

$$\begin{aligned} \text{Var}(\alpha_k) = E(\alpha_k^2) &= \begin{cases} \frac{\sigma^2}{2n} & \text{si } 0 < |k| < n/2 \\ \frac{\sigma^2}{n} & \text{si } k = 0 \text{ ou } |k| = n/2 \end{cases} , \\ \text{Var}(\beta_k) = E(\beta_k^2) &= \begin{cases} \frac{\sigma^2}{2n} & \text{si } 0 < |k| < n/2 \\ 0 & \text{si } k = 0 \text{ ou } |k| = n/2 \end{cases} , \\ \text{Cov}(\alpha_k, \beta_{k'}) &= 0 \quad \forall k \in \mathbb{Z}, k' \in \mathbb{Z} \end{aligned}$$

$$\text{Cov}(\alpha_k, \alpha_{k'}) = \text{Cov}(\beta_k, \beta_{k'}) = 0 \text{ si } |k| \neq |k'| .$$

### Démonstration du Lemme 1

Il suffit de remarquer que

$$\sum_{j=1}^n \left[ \cos\left(\frac{2\pi jk}{n}\right) \right]^2 = \begin{cases} \frac{n}{2} & \text{si } 0 < |k| < \frac{n}{2} \\ 0 & \text{si } k = 0 \text{ ou } |k| = \frac{n}{2} \end{cases} ,$$

$$\sum_{j=1}^n \left[ \sin\left(\frac{2\pi jk}{n}\right) \right]^2 = \begin{cases} \frac{n}{2} & \text{si } 0 < |k| < \frac{n}{2} \\ 0 & \text{si } k = 0 \text{ ou } |k| = \frac{n}{2} \end{cases} ,$$

$$\sum_{j=1}^n \cos\left(\frac{2\pi jk}{n}\right) \sin\left(\frac{2\pi jk'}{n}\right) = 0 \quad \forall k \in \mathbb{Z} , \forall k' \in \mathbb{Z} ,$$

$$\sum_{j=1}^n \cos\left(\frac{2\pi jk}{n}\right) \cos\left(\frac{2\pi jk'}{n}\right) = 0 \quad \text{si } |k| \neq |k'| ,$$

$$\sum_{j=1}^n \sin\left(\frac{2\pi jk}{n}\right) \sin\left(\frac{2\pi jk'}{n}\right) = 0 \quad \text{si } |k| \neq |k'| \quad .$$

**Lemme 2**

Si la condition (A<sub>1</sub>) est vérifiée , alors

$$E |\hat{c}_k - c_k|^2 \leq \frac{C}{n} \quad ,$$

où  $C$  est une constante positive .

**Démonstration du Lemme 2**

On a

$$\begin{aligned} |\hat{c}_k - c_k|^2 &= |\hat{c}_k - \tilde{c}_k|^2 + |\tilde{c}_k - c_k|^2 \\ &\quad + \left[ (\hat{c}_k - \tilde{c}_k) \overline{(\tilde{c}_k - c_k)} + \overline{(\hat{c}_k - \tilde{c}_k)} (\tilde{c}_k - c_k) \right] \quad . \end{aligned} \quad (3.8)$$

Par ailleurs , on a

$$\begin{aligned} |\tilde{c}_k - c_k| &= \left| \frac{1}{n} \sum_{j=1}^n \mu(j/n) e^{-2\pi i j k / n} - \int_0^1 \mu(t) e^{-2\pi i k t} dt \right| \\ &= \left| \sum_{j=1}^n \left\{ \frac{1}{n} \mu(j/n) e^{-2\pi i j k / n} - \int_{(j-1)/n}^{j/n} \mu(t) e^{-2\pi i k t} dt \right\} \right| \\ &\leq \sum_{j=1}^n \left\{ \int_{(j-1)/n}^{j/n} |\mu(j/n) e^{-2\pi i j k / n} - \mu(t) e^{-2\pi i k t}| dt \right\} \quad . \end{aligned}$$

Puisque  $\mu$  satisfait la condition de Lipschitz d'ordre  $\delta$  , il est facile de montrer qu'il existe une constante  $M > 0$  telle que

$$\left| \mu\left(\frac{j}{n}\right) e^{-2\pi i j k / n} - \mu(t) e^{-2\pi i k t} \right| \leq M \left( \frac{j}{n} - t \right)^\delta$$

pour

$$\frac{j-1}{n} \leq t \leq \frac{j}{n} \quad , \quad j = 1, \dots, n \quad .$$

Il s'ensuit que

$$\int_{(j-1)/n}^{j/n} \left| \mu\left(\frac{j}{n}\right) e^{-2\pi i j k/n} - \mu(t) e^{-2\pi i k t} \right| dt \leq \frac{M}{(\delta+1)n^{\delta+1}} \quad .$$

Donc

$$|\tilde{c}_k - c_k| \leq \frac{M}{(\delta+1)n^\delta} \quad . \quad (3.9)$$

Par ailleurs d'après la relation (3.5) et le Lemme 1 , on a

$$E |\hat{c}_k - \tilde{c}_k|^2 = E(\alpha_k^2) + E(\beta_k^2) = \frac{\sigma^2}{n}$$

et donc

$$E |\hat{c}_k - \tilde{c}_k| \leq \frac{\sigma}{\sqrt{n}} \quad . \quad (3.10)$$

Ainsi , d'après les relations (3.7) – (3.8) – (3.9) , on a

$$E |\hat{c}_k - c_k|^2 \leq \frac{C}{n} \quad ,$$

où  $C$  est une constante donnée par

$$C = \sigma^2 + \frac{M^2}{(\delta+1)^2} + \frac{2M\sigma}{\delta+1} \quad .$$

### Démonstration de la Proposition 1

Soit  $\hat{c}_k = \frac{1}{n} \sum_{j=1}^n Y_j e^{-2\pi i k j/n}$  , l'estimateur du  $k^{\text{ième}}$  coefficient de Fourier  $c_k$  de la fonction de régression

$$\mu(t) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi i k t}$$

et

$$\hat{\mu}(t) = \sum_{|k| \leq q} \hat{c}_k e^{2\pi i k t}$$

l'estimateur de  $\mu$ .

On a

$$\begin{aligned} \hat{\sigma}^2 &= \frac{1}{n} \sum_{j=1}^n |Y_j - \hat{\mu}(j/n)|^2 \\ &= \frac{1}{n} \sum_{j=1}^n \varepsilon_j^2 + \frac{1}{n} \sum_{j=1}^n \varepsilon_j \left[ \mu(j/n) - \hat{\mu}(j/n) + \overline{\mu(j/n) - \hat{\mu}(j/n)} \right] \\ &\quad + \frac{1}{n} \sum_{j=1}^n |\hat{\mu}(j/n) - \mu(j/n)|^2 \\ &= \overline{\varepsilon^2} + F_n + G_n \end{aligned}$$

où

$$\overline{\varepsilon^2} = \frac{1}{n} \sum_{j=1}^n \varepsilon_j^2, \quad (3.11)$$

$$F_n = \frac{1}{n} \sum_{j=1}^n \varepsilon_j \left[ \mu(j/n) - \hat{\mu}(j/n) + \overline{\mu(j/n) - \hat{\mu}(j/n)} \right], \quad (3.12)$$

et

$$G_n = \frac{1}{n} \sum_{j=1}^n |\hat{\mu}(j/n) - \mu(j/n)|^2. \quad (3.13)$$

On a

$$E(\overline{\varepsilon^2} - \sigma^2) = \frac{1}{n} [E(\varepsilon_1^4) - \sigma^4] \longrightarrow 0 \text{ quand } n \rightarrow \infty.$$

Pour démontrer (i) et (ii), il suffit de montrer que  $F_n \xrightarrow{P} 0$  et  $G_n \xrightarrow{P} 0$  (respectivement  $\sqrt{p}F_n \xrightarrow{P} 0$  et  $\sqrt{p}G_n \xrightarrow{P} 0$ ) quand  $n \rightarrow \infty$ .

On a

$$\begin{aligned} E |F_n| &\leq \frac{2}{n} \sum_{j=1}^n \sqrt{E(\varepsilon_j)^2 E |\hat{\mu}(j/n) - \mu(j/n)|^2} \\ &\leq \frac{2\sigma}{n} \sum_{j=1}^n \sqrt{E |\hat{\mu}(j/n) - \mu(j/n)|^2} \end{aligned} \quad (3.14)$$

et

$$E |G_n| \leq \frac{1}{n} \sum_{j=1}^n E |\hat{\mu}(j/n) - \mu(j/n)|^2 . \quad (3.15)$$

On a

$$\begin{aligned} |\hat{\mu}(j/n) - \mu(j/n)|^2 &= \left| \sum_{|k| \leq q} (\hat{c}_k - c_k) e^{2\pi i k j/n} - \sum_{|k| > q} c_k e^{2\pi i k j/n} \right|^2 \\ &\leq \left( \sum_{|k| \leq q} |\hat{c}_k - c_k| + \sum_{|k| > q} |c_k| \right)^2 \\ &\leq \sum_{|k| \leq q} |\hat{c}_k - c_k|^2 + \sum_{-q \leq k \neq k' \leq q} |\hat{c}_k - c_k| |\hat{c}_{k'} - c_{k'}| \\ &\quad + \left( \sum_{|k| > q} |c_k| \right)^2 + 2 \left( \sum_{|k| > q} |c_k| \right) \left( \sum_{|k| \leq q} |\hat{c}_k - c_k| \right) , \end{aligned}$$

alors

$$\begin{aligned} E |\hat{\mu}(j/n) - \mu(j/n)|^2 &\leq \sum_{|k| \leq q} E |\hat{c}_k - c_k|^2 + \sum_{-q \leq k \neq k' \leq q} \sqrt{E |\hat{c}_k - c_k| E |\hat{c}_{k'} - c_{k'}|} \\ &\quad + \left( \sum_{|k| > q} |c_k| \right)^2 + 2 \left( \sum_{|k| > q} |c_k| \right) \left( \sum_{|k| \leq q} E |\hat{c}_k - c_k| \right) . \end{aligned}$$

D'après le Lemme 2 , il existe une constante  $C$  telle que

$$\begin{aligned} E |\hat{\mu}(j/n) - \mu(j/n)|^2 &\leq (2q+1) \frac{C}{n} + 2q(2q+1) \frac{C}{n} + \left( \sum_{|k| > q} |c_k| \right)^2 \\ &\quad + 2 \left( \sum_{|k| > q} |c_k| \right) (2q+1) \sqrt{\frac{C}{n}} . \end{aligned} \quad (3.16)$$

En utilisant les relations (3.14) – (3.15) – (3.16) et les conditions (A<sub>3</sub>) (respectivement (A<sub>4</sub>) et (A<sub>5</sub>)) , on en déduit que  $\lim_{n \rightarrow \infty} E|F_n| = 0$  et  $\lim_{n \rightarrow \infty} E|G_n| = 0$ , (respectivement  $\lim_{n \rightarrow \infty} \sqrt{p}E|F_n| = 0$  et  $\lim_{n \rightarrow \infty} \sqrt{p}E|G_n| = 0$ ) .

Ce qu'il fallait démontrer .

# Chapitre 4

## Test d'un polynôme

## trigonométrique dans un modèle

## de régression

### 4.1 Introduction

Cette partie est consacrée à la construction de test d'hypothèse d'un polynôme trigonométrique dans un modèle de régression dans le cas homoscédastique .

Plus précisément , on considère le modèle de régression suivant

$$Y_j = \mu(t_j) + \varepsilon_j \quad , \quad j = 1, \dots, n \quad ,$$

où  $\mu$  est une fonction réelle inconnue définie sur l'intervalle  $[0, 1]$  , les  $(t_j)$  est échantionnage fixé sur  $[0, 1]$  et les  $(\varepsilon_j)$  sont des variables aléatoires centrées i.i.d de

variance inconnue  $\sigma^2$  .

Soit

$$c_k = \int_0^1 e^{-2\pi ikt} \mu(t) dt \quad , \quad k \in \mathbb{Z} \quad (4.1)$$

les coefficients de Fourier de  $\mu$  .

Il s'agit de tester l'hypothèse que la fonction  $\mu$  est un polynôme trigonométrique de degré au plus  $s$  , c'est-à-dire le test de l'hypothèse nulle

$$H_0 : c_k = 0 \quad , \quad \forall k \geq s + 1 \quad \text{contre} \quad H_1 : \exists k \geq s + 1 \quad \text{tel que} \quad c_k \neq 0 \quad .$$

Il est clair que  $H_0$  est vraie si et seulement si

$$\sum_{|k| \geq s+1} |c_k|^2 = 0 \quad .$$

Pour construire la statistique de test qui est une estimation de ce critère , on considère une suite de nombre entiers  $p = p(n)$  telle que  $p(n) \rightarrow \infty$  quand  $n \rightarrow \infty$  et une estimation empirique  $\hat{c}_k$  des coefficients de Fourier  $c_k$  , donnée par

$$\hat{c}_k = \frac{1}{n} \sum_{j=1}^n y_j e^{-2\pi i k j / n} \quad , \quad k \in \mathbb{Z} \quad . \quad (4.2)$$

La statistique de test que nous utilisons est  $\hat{T}_{n,p}$  la suivante

$$\hat{T}_{n,p} = \sum_{s+1 \leq |k| \leq p} |\hat{c}_k|^2 \quad , \quad (4.3)$$

et on rejette  $H_0$  si  $\hat{T}_{n,p} > t$  pour une valeur  $t > 0$  .

## 4.2 Loi asymptotique de la statistique de test

Les hypothèse que nous utilisons sont les suivantes

(A<sub>1</sub>)  $\mu$  satisfait la condition de Lipschitz d'ordre  $\delta$  avec  $\frac{1}{2} \leq \delta \leq 1$ .

(A<sub>2</sub>) Les  $(\varepsilon_j)$ ,  $j = 1, \dots, n$  sont des variables aléatoires indépendantes de loi normale centrée et de variance inconnue  $\sigma^2$ .

(A<sub>3</sub>)  $\lim_{n \rightarrow \infty} \{n^{-2\delta+1}p(n)\} = 0$ .

On a la loi asymptotique de la statistique de test donnée par le théorème suivant.

### Théorème 1

Si (A<sub>1</sub>), (A<sub>2</sub>) et (A<sub>3</sub>) sont vérifiées, alors on a

$$\frac{n \sum_{s+1 \leq |k| \leq p} |\hat{c}_k - c_k|^2 - 2(p-s)\sigma^2}{2\sigma^2\sqrt{p-s}} \xrightarrow{\mathcal{L}} N(0, 1).$$

### Remarque 1

Puisque la région critique est définie par  $\hat{T}_{n,p} = \sum_{s+1 \leq |k| \leq p} |\hat{c}_k|^2 > t$ , le théorème précédent donne le niveau et la puissance du test lorsque  $\sigma^2$  est connue.

### Démonstration du théorème 1

Soit

$$\tilde{c}_k = \frac{1}{n} \sum_{j=1}^n \mu(j/n) e^{-2\pi i j k / n} \quad (4.4)$$

le coefficient de Fourier discret de  $\mu$ .

$p = p(n)$  une suite de nombre entiers telle que  $p(n) \rightarrow \infty$  quand  $n \rightarrow \infty$ . On

pose

$$\Lambda_{n,p} = \sum_{s+1 \leq |k| \leq p} |\hat{c}_k - \tilde{c}_k|^2 \quad . \quad (4.5)$$

Soit  $\alpha_k$  et  $\beta_k$  sont donnés par

$$\alpha_k = \frac{1}{n} \sum_{j=1}^n \varepsilon_j \cos(2\pi jk/n) \quad , \quad (4.6)$$

$$\beta_k = \frac{1}{n} \sum_{j=1}^n \varepsilon_j \sin(2\pi jk/n) \quad . \quad (4.7)$$

On a

$$\sqrt{2n}(\alpha_{s+1}, \beta_{s+1}, \dots, \alpha_p, \beta_p)' \sim N(0, \sigma^2 I_{2(p-s)}) \quad ,$$

où  $I_{2(p-s)}$  est la matrice identique de  $2(p-s) \times 2(p-s)$  .

Alors

$$\frac{n}{\sigma^2} \Lambda_{n,p} = \frac{2n}{\sigma^2} \sum_{k=s+1}^p (\alpha_k^2 + \beta_k^2) \sim \chi^2(2(p-s)) \quad .$$

Soit  $(\xi_i)$  une suite de variables aléatoires indépendantes et identiquement distribuées d'une loi  $\chi^2(1)$  , d'après le théorème central limite , nous avons

$$\frac{S_{n,p} - 2(p-s)}{2\sqrt{p-s}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1) \quad ,$$

où  $S_{n,p} = \sum_{k=1}^{2(p-s)} \xi_k$  .

Puisque  $\frac{n}{\sigma^2} \Lambda_{n,p}$  et  $S_{n,p}$  ont la même loi pour toute valeur  $n$  , alors

$$\frac{\frac{n}{\sigma^2} \Lambda_{n,p} - 2(p-s)}{2\sqrt{p-s}} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} N(0, 1) \quad .$$

Nous avons

$$\frac{n \sum_{s+1 \leq |k| \leq p} |\hat{c}_k - c_k|^2 - 2(p-s)\sigma^2}{2\sigma^2\sqrt{p-s}} = U_n + V_n + W_n \quad ,$$

où

$$U_n = \frac{n \sum_{s+1 \leq |k| \leq p} |\hat{c}_k - \tilde{c}_k|^2 - 2(p-s)\sigma^2}{2\sigma^2\sqrt{p-s}} ,$$

$$V_n = \frac{n}{2\sigma^2\sqrt{p-s}} \sum_{s+1 \leq |k| \leq p} |\hat{c}_k - c_k|^2$$

et

$$W_n = \frac{n}{2\sigma^2\sqrt{p-s}} \sum_{s+1 \leq |k| \leq p} \left[ (\hat{c}_k - \tilde{c}_k) \overline{(\tilde{c}_k - c_k)} + \overline{(\hat{c}_k - \tilde{c}_k)} (\tilde{c}_k - c_k) \right] .$$

Pour montrer le résultat , il suffit de montrer que  $\lim_{n \rightarrow \infty} V_n = 0$  et  $W_n \xrightarrow{P} 0$  quand  $n \rightarrow \infty$  .

Nous avons

$$\begin{aligned} |\tilde{c}_k - c_k| &= \left| \frac{1}{n} \sum_{j=1}^n \mu(j/n) e^{-2\pi i j k / n} - \int_0^1 \mu(t) e^{-2\pi i k t} dt \right| \\ &\leq \sum_{j=1}^n \left\{ \int_{(j-1)/n}^{j/n} |\mu(j/n) e^{-2\pi i j k / n} - \mu(t) e^{-2\pi i k t}| dt \right\} . \end{aligned}$$

Puisque  $\mu$  satisfait la condition de Lipschitz d'ordre  $\delta$  et la fonction exponentielle est continuellement différentiable , il est facile de montrer qu'il existe une constante  $M > 0$  telle que

$$\left| \mu\left(\frac{j}{n}\right) e^{-2\pi i j k / n} - \mu(t) e^{-2\pi i k t} \right| \leq M \left(\frac{j}{n} - t\right)^\delta .$$

Il s'ensuit que

$$\int_{(j-1)/n}^{j/n} \left| \mu\left(\frac{j}{n}\right) e^{-2\pi i j k / n} - \mu(t) e^{-2\pi i k t} \right| dt \leq \frac{M}{(\delta + 1)n^{\delta+1}} .$$

Donc

$$|\tilde{c}_k - c_k| \leq \frac{M}{(\delta + 1)n^\delta} \quad . \quad (4.8)$$

D'après la relation (4.8) nous avons

$$V_n \leq \frac{M^2}{(\delta + 1)^2 \sigma^2 n^{2\delta - 1}} \sqrt{p - s} \quad .$$

Alors , d'après la condition (A<sub>3</sub>) , nous avons  $\lim_{n \rightarrow \infty} V_n = 0$  .

Par ailleurs ,

$$E |W_n| \leq \frac{n}{\sigma^2 \sqrt{p - s}} \sum_{s+1 \leq |k| \leq p} |\tilde{c}_k - c_k| E |\hat{c}_k - \tilde{c}_k| \quad .$$

En notant que

$$\hat{c}_k - \tilde{c}_k = \alpha_k - i\beta_k$$

et que

$$E(\alpha_k^2) = \begin{cases} \frac{\sigma^2}{2n} & \text{si } 0 < |k| < n/2 \\ \frac{\sigma^2}{n} & \text{si } k = 0 \text{ ou } |k| = n/2 , \end{cases}$$

$$E(\beta_k^2) = \begin{cases} \frac{\sigma^2}{2n} & \text{si } 0 < |k| < n/2 \\ 0 & \text{si } k = 0 \text{ ou } |k| = n/2 . \end{cases}$$

On a

$$E |\hat{c}_k - \tilde{c}_k|^2 = E(\alpha_k^2) + E(\beta_k^2) = \frac{\sigma^2}{n}$$

et ainsi

$$E |\hat{c}_k - \tilde{c}_k| \leq \frac{\sigma}{\sqrt{n}} \quad . \quad (4.9)$$

En utilisant les relations (4.8) et (4.9) , on déduit que

$$E |W_n| \leq \frac{2M}{(\delta + 1)\sigma} \frac{\sqrt{p-s}}{n^{\delta-\frac{1}{2}}} .$$

Donc , d'après la condition (A<sub>3</sub>) , on en déduit que  $W_n \xrightarrow{P} 0$  quand  $n \rightarrow \infty$  .

### 4.3 Construction de test

Le résultat du théorème 1 peut être utilisé pour construire le test

de l'hypothèse  $H_0$  , mais en pratique la variance  $\sigma^2$  est inconnue , il faut donc l'estimer . On peut utiliser l'estimateur  $\hat{\sigma}^2$  de la variance  $\sigma^2$  introduit dans le chapitre précédent , alors le corollaire suivant montre le comportement asymptotique du test sous l'hypothèse  $H_0$  .et donné par

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^n |Y_j - \hat{\mu}(j/n)|^2 ,$$

où  $\hat{\mu}$  est défini par

$$\hat{\mu}(t) = \sum_{|k| \leq q} \hat{c}_k e^{2\pi i k t} ,$$

avec  $q = q(n)$  une suite de nombre entiers telle que  $q(n) \rightarrow +\infty$  quand  $n \rightarrow \infty$  .

On a besoin d'une autre hypothèse pour énoncer un corollaire du théorème précédent

$$(A_4) \lim_{n \rightarrow \infty} \left\{ \sqrt{p(n)} \sum_{|k| > q(n)} |c_k| \right\} = 0 ,$$

$$(A_5) \lim_{n \rightarrow \infty} \left\{ \frac{p(n)q^2(n)}{n} \right\} = 0 .$$

**Corollaire**

Si les conditions du théorème 1 et les conditions (A<sub>4</sub>) et (A<sub>5</sub>) sont vérifiées alors , sous l'hypothèse  $H_0 : c_k = 0, \forall k \geq s + 1$  , on a

$$\hat{T}_p = \frac{n \sum_{s+1 \leq |k| \leq p} |\hat{c}_k|^2 - 2(p-s)\hat{\sigma}^2}{2\hat{\sigma}^2 \sqrt{p-s}} \xrightarrow{\mathcal{L}} N(0, 1) .$$

### Démonstration

Conséquence du théorème 1 et la proposition énoncée dans le chapitre 3 .

## 4.4 Application numérique

Nous allons appliquer le test de l'hypothèse  $H_0 : c_k = 0, \forall k \geq s + 1$  aux données du nombre d'accidents mensuels survenus aux U.S.A entre 1951 et 1980 .

Ces données sont relevées mensuellement et sont relevées dans le tableau suivant

t	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
y	4737	5117	5091	3468	4320	3825	3673	3694	3708	3333
t	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
y	3367	3614	3362	3655	3963	4405	4595	5045	5700	5716
t	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
y	5138	5010	5353	6074	5031	5648	5506	4230	4827	3885

Tableau (4.1)

On applique le corollaire 1 et l'estimation de la variance  $\sigma^2$  par  $\hat{\sigma}^2$  pour déterminer la valeur critique de la région de rejet de  $H_0$  .

Nous avons pris les valeurs de  $s = 1, 2, \dots, 10$  et  $p - s = 1, \dots, 5$  . Les valeurs de statistique de test sont données dans le tableau suivant

$\hat{T}_p$	$p - s = 1$	$p - s = 2$	$p - s = 3$	$p - s = 4$	$p - s = 5$
$s = 1$	3.0453	1.4688	2.3696	2.5098	1.8615
$s = 2$	-0.9681	0.7488	1.1398	0.5585	0.7580
$s = 3$	2.0270	2.0805	1.2039	1.3325	1.0267
$s = 4$	0.9153	0.0411	0.3683	0.1344	0.0743
$s = 5$	-0.8572	-0.1961	-0.3732	-0.3744	0.1409
$s = 6$	0.5798	0.1490	0.0626	0.5861	0.4362
$s = 7$	-0.3691	-0.3333	0.3420	0.1978	-0.0822
$s = 8$	-0.1023	0.6798	0.4415	0.0927	0.0612
$s = 9$	1.0637	0.6130	0.1661	0.1196	0.3802
$s = 10$	-0.1968	-0.5488	-0.4761	-0.1068	-0.2881

Tableau (4.2)

On constate que pour  $s \geq 4$ , l'hypothèse  $H_0$  est acceptée; pour  $s = 3$  avec  $p - s = 1$  et  $2$ ,  $H_0$  est rejetée.  $H_0$  est acceptée pour  $p - s = 3, 4, 5$ .

On décide que  $H_0$  est acceptée pour  $s = 4$ .

### Remarque 2

Le tableau (4.2) montre qu'au niveau significatif  $\alpha = 0,05$ , l'hypothèse nulle  $H_0$ :

“ $\mu$  est un polynôme trigonométrique de degré au plus de 4”

$$\hat{\mu}(t) = \sum_{|k| \leq 4} \hat{c}_k e^{2\pi i k t} = \sum_{k=-4}^4 \hat{c}_k e^{2\pi i k t}$$

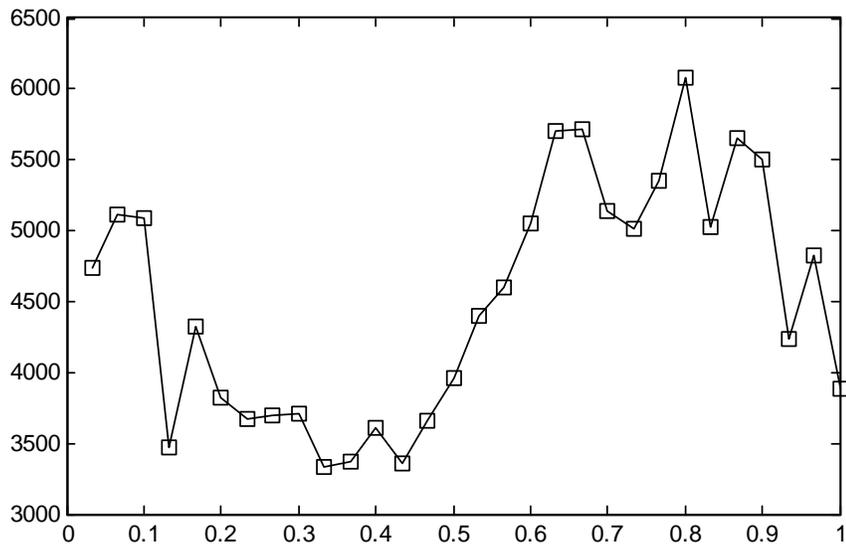


FIG. 4.1 – Le nombre d'accidents dans l'USSA entre 1951 et 1980

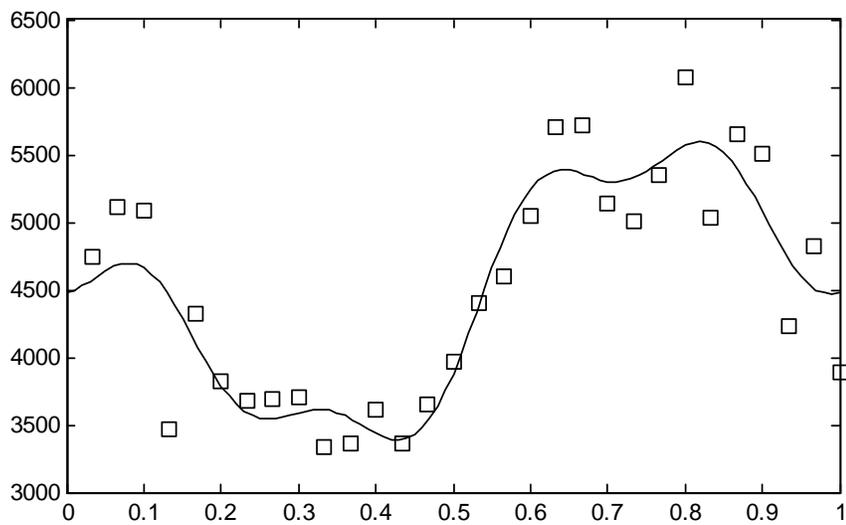


FIG. 4.2 – Ajustement des données des accidents mensuels survenus aux U.S.A par un polynôme trigonométrique de degré au plus 4

## **Conclusion**

L'objet de ce travail consistait à mettre en place une procédure d'estimation et de sélection de modèles pour la fonction de régression dans le cadre statistique de la régression non paramétrique. Cette étude peut servir à résoudre d'une manière concrète un grand nombre de problème physique, économique et biomédical.

**Bibliographie**

- [1] : Eubank,R.L. (1988), Spline Smothing and Nonparametric Regression.  
Marcel Dekker, New-York .
- [2] : Eubank,R.L. and Spiegelman,C.H.(1990), Testing the Goodness-of-Fit of  
a linear Model Via Nonparametric Regression Techniques. J. Amer. Statist.  
Assoc.85, 410,387-392.
- [3] : Eubank,R.L. and Hart, J. D.(1992),Testing Goodness-of-fit in Regression  
Using Nonparametric via Order Selection Criteria. Ann. Statist, 20, 1412-  
1425.
- [4] : Jayasuriya, B.R.(1996), Testing for Polynomial Regression Using  
Nonparametric Regression Techniques J. Amer., 80, 391, 683-691.
- [5] : Mohdeb, Z., Makkadem, A.(2004), On the use of nonparametric  
regression for testing linear hypotheses. Ann, I.S.U.P., 48,3,63-77.

## **Abstract**

The object of our work is based on some methods of estimation and hypotheses test in a nonparametric regression model. More precisely, first, it is about estimating the function of regression by an approximation of the decomposition of Fourier series and to estimate the Fourier coefficients. Then, we construct hypothesis tests which the regression function is a trigonometric polynomial of order at most given value  $s$

**Key-words:** Nonparametric regression, Empirical Fourier coefficient, Nonparametric test.

## الملخص

الهدف من هذا العمل هو دراسة بعض طرق التقدير و اختبار الفرضيات بالنسبة إلى نموذج الانحدار الغير الوسيطي . بالضبط يتعلق الأمر بتقدير دالة الانحدار بطريقة تحويلات سلاسل فوريي و هذه الطريقة تمكن من تقدير معاملات فوريي . بعدها يتم إنشاء اختبار الفرضية أن دالة الانحدار هي دالة مثلثيه درجتها على الأكثر عدد  $s$ .

مفتاح الكلمات : الانحدار الغير وسيطي ، معاملات فوريي ، الاختبار الغير وسيطي .

## Résumé

L'objet de notre travail porte sur quelques méthodes d'estimation et test d'hypothèses dans un modèle de régression non paramétrique. Plus précisément, il s'agit, dans un premier temps d'estimer la fonction de régression par une approximation de la décomposition en série de Fourier et d'en estimer les coefficients de Fourier de la fonction. Ensuite, on construit des tests de l'hypothèse que la fonction de régression est un polynôme trigonométrique de degré au plus  $s$  donné.

**Mot-clé :** Régression non paramétrique, Coefficients empiriques de Fourier, Test non paramétrique.