



HAL
open science

**Contribution à l'étude de la régression non
paramétrique et à l'estimation de la moyenne d'un
processus à temps continu**

David Degras

► **To cite this version:**

David Degras. Contribution à l'étude de la régression non paramétrique et à l'estimation de la moyenne d'un processus à temps continu. Mathématiques [math]. Université Pierre et Marie Curie - Paris VI, 2007. Français. NNT: . tel-00201438

HAL Id: tel-00201438

<https://theses.hal.science/tel-00201438>

Submitted on 28 Dec 2007

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : Mathématiques

Option : Statistique

présentée par

David DEGRAS

pour obtenir le grade de

DOCTEUR de l'UNIVERSITE PARIS 6

Sujet de la thèse :

**Contribution à l'étude de la régression non paramétrique
et à l'estimation de la moyenne d'un processus à temps continu**

Soutenue le 7 Décembre 2007 devant le jury composé de :

Directeur de thèse : M. Daniel PIERRE-LOTI-VIAUD Université de Paris 6

Rapporteurs : M. Hervé CARDOT Université de Bourgogne

M. Alois KNEIP Université de Bonn

Président : M. Paul DEHEUVELS Université de Paris 6

Examineurs : Mlle. Delphine BLANKE Université de Paris 6

M. Emmanuel GUERRE Queen Mary, Université de Londres

Remerciements

En tout premier lieu, je remercie du fond du coeur le Professeur Daniel Pierre-Loti-Viaud qui fut mon directeur au long de cette belle aventure. J'ai pu apprécier sa générosité, sa patience, sa disponibilité, ainsi que l'exceptionnelle ouverture d'esprit qui a marqué nos échanges. Qu'il puisse trouver ici le gage de mon respect profond.

Je me tourne ensuite vers les Professeurs Hervé Cardot et Alois Kneip, qui ont accepté d'être les rapporteurs de cette thèse. Par leurs commentaires éclairés, ils placent mes travaux dans une perspective plus vaste et ce faisant, ils m'indiquent la voie de recherches futures. Je les en remercie sincèrement.

Je remercie aussi les Professeurs Paul Deheuvels, Emmanuel Guerre et Mlle. Delphine Blanke qui me font le grand honneur de participer à mon jury de soutenance, ainsi que le Professeur Gérard Biau qui n'a pu se libérer mais a néanmoins eu la gentillesse d'examiner mon manuscrit avec attention. Qu'il me soit permis ici de saluer la diligence et la grande qualité des réponses que le Professeur Deheuvels a bien voulu faire à l'ensemble de mes questions durant ces années.

J'adresse un remerciement particulièrement chaleureux aux secrétaires du LSTA Louise Lamart et Anne Durande : elles sont tout simplement adorables. Les longues journées de travail ont paru moins longues à leurs côtés et le rire n'était jamais loin. Merci également à Pascal Epron qui m'a ouvert les portes de sa bibliothèque.

J'ai eu un vif plaisir à travailler au sein de l'équipe du LSTA, et je tiens à saluer ses membres que j'ai côtoyés au cours de ma thèse, qui m'ont écouté, conseillé, aidé, et avec qui j'ai appris et travaillé au quotidien. En particulier, les Professeurs Denis Bosq, Michel Broniatowki, et Djamel Louani, Messieurs Jérôme Dedecker, Stéphane Gaiffas, Giovanni Peccati, Jean Renaud Pycke, et Philippe Saint-Pierre, Mesdemoiselles Armelle Guilloux et Agathe Guilloux, ainsi que tous les collègues (ex-) doctorants : François-Xavier Lejeune, Amor Keziou, Fateh Chebana, Mohammed Debbarh, Noureddine Berrahou, Véronique Boisson, Leila Bourmouche, Rabah Bennis, Salim Bouzebda, Julien Cornebise, Ravan De Senigon de Roumefort, Lahcen Douge, Kaouthar El Fassi, Issam El Hattab, Olivier Faugeras, Segolen Geffray, Moïse Jérémie, Mamadou Koné, Boris Labrador, Anne Massiani, Nabil Nessigha, Mounir Nisse, Mory Souaré, Gwladys Toulemonde, Tarek Zari, Jean-Baptiste Aubin, Samuela Leoni-Aubin, Pierre Ribereau, Omar Eldakkak, David Blondin, et beaucoup d'autres encore ...

J'éprouve une profonde gratitude pour les chercheurs en neurosciences que j'ai eu la chance de rencontrer au cours de ma première année de thèse. Ces personnes m'ont pris sous leur aile sans être avares de leur temps et grâce à elles, j'ai pu recevoir une solide formation de base à l'analyse statistique des données d'imagerie fonctionnelle et à l'utilisation du logiciel SPM. Que le Professeur Jean Baptiste Poline et le Docteur Serge Stoléru recueillent ici toute ma reconnaissance.

Je me tourne enfin vers ma famille et mes amis, qui me soutiennent sans faille depuis des années. Ils sont trop nombreux pour être tous cités, mais je leur dis ici combien leur gentillesse, leur affection, et leur humour m'ont aidé à avancer. Merci à eux.

Merci à ma collègue et chère amie Roxane, avec qui j'ai connu la joie et la stimulation intellectuelle de la recherche en commun, avec qui j'ai publié mon premier papier, et qui a été à mes côtés à chaque étape de cette thèse.

Merci à Awa pour tout en général, pour sa présence, ses encouragements, et sa sollicitude en particulier.

Merci mille fois à mes parents, Tessa et Marc, qui croient en moi depuis le début, qui m'accompagnent dans chaque jour, et sans qui rien de tout cela ne serait possible.

Et merci à Varenka parce qu'avec elle, la vie est un chemin de bonheur.

À mes parents,

*ceux qui m'ont fait naître dans ce monde,
et ceux qui m'ont aidé à y grandir.*

Résumé

Cette thèse porte sur l'étude de la régression non paramétrique en présence de mesures répétées. D'abord, nous étendons aux estimateurs splines de lissage les vitesses de convergence présentées dans la littérature pour d'autres estimateurs usuels sous différentes hypothèses classiques de dépendance des données. Ensuite, dans le cadre de l'estimation de la moyenne d'un processus aléatoire à temps continu, nous généralisons les résultats existants sur la convergence en moyenne quadratique et nous établissons de nouveaux résultats de normalité asymptotique pour les distributions finies-dimensionnelles. Enfin, dans le cadre d'un échantillon fini et corrélé, nous comparons les performances d'estimateurs construits par moindres carrés ordinaires ou généralisés, nous proposons une méthode efficace de sélection du paramètre de lissage tenant compte de la structure de covariance des données, et à travers des simulations, nous mettons en évidence l'apport du lissage local par rapport au lissage global.

Abstract

The present PhD deals with nonparametric regression using repeated measurements data. On the one hand, the convergence rates of several usual estimators found in the literature under classical dependency assumptions are extended to the smoothing spline estimators. On the other hand, in the context of mean function estimation from continuous-time random processes, the few existing results on mean square convergence are generalized to a large class of linear estimators and new asymptotic normality results are derived for the finite-dimensional distributions of estimators. Finally in the framework of a finite, correlated sample, the ordinary and generalized least squares methods for constructing regression estimators are compared, a new smoothing parameter selection procedure accounting for the covariance structure of the data is presented, and the superiority of local smoothing over global smoothing is shown through simulations.

Publications

1. Degras, D. et Jallet, R. (2005). Convergence de l'estimateur spline cubique de lissage dans un modèle de régression longitudinale avec erreur de type processus. *C. R. Math. Acad. Sci. Paris*, 340 (11), 851–854.
2. Asymptotics for the nonparametric estimation of the mean of a random process. *Soumis à Statistics and Probability Letters*.
3. Nonparametric regression techniques for repeated measurements data. *En préparation*.
4. Approximation discrète des valeurs propres d'un opérateur de covariance sur $C[0, 1]$. *En préparation*.

Exposés et séminaires

1. Sur l'estimation non paramétrique de la fonction moyenne d'un processus aléatoire. *Communication orale aux Journées de la Statistique 2007*.
2. Quelques résultats asymptotiques pour l'estimation non paramétrique de la moyenne d'une fonction aléatoire. *Groupe de travail Processus et Estimation Fonctionnelle du LSTA Paris 6. 2007*.
3. Estimation de la moyenne d'une fonction aléatoire par lissage spline. *Communication écrite aux Journées de la Statistique 2006*.
4. Etude des estimateurs splines dans la régression non paramétrique avec des données corrélées. *Groupe de Travail des Thésards du LSTA Paris 6. 2006*.
5. Modèles linéaires hiérarchiques dans une perspective empirique bayésienne. *Service Hospitalier Frédéric Joliot (CEA-SHFJ). 2003*.

Table des matières

Notations	3
Introduction Générale	5
1 Régression par splines de lissage sur des mesures répétées	21
1.1 Modèle et estimateur	23
1.2 Majoration de l'erreur quadratique moyenne	26
1.2.1 Dispositif expérimental équilibré	26
1.2.2 Dispositif expérimental général	28
1.2.3 Estimation des dérivées de la fonction de régression	29
1.2.4 Résultats numériques pour les splines cubiques	30
1.3 Application : vitesses de convergence	32
1.3.1 Décorrélation asymptotique	33
1.3.2 Processus aléatoire à temps continu	35
1.3.3 Données longitudinales	36
1.4 Démonstrations	37
2 Estimation de la fonction moyenne d'un processus aléatoire	63
2.1 Introduction	63
2.1.1 Historique	63
2.1.2 Modèle et estimateur linéaire	65
2.1.3 Aperçu des résultats	66
2.2 Étude asymptotique des estimateurs linéaires	66
2.2.1 Consistance en moyenne quadratique	67
2.2.2 Normalité asymptotique	69
2.2.3 Application : intervalles de confiance simultanés	70
2.3 Convergence dans quelques cas particuliers	74
2.3.1 Convergence globale	74
2.3.2 Convergence locale	76
2.4 Perspectives de développements	80
2.5 Démonstrations	82

3	Pratique du lissage spline avec des mesures répétées	101
3.1	Estimateur spline de lissage des moindres carrés généralisés	103
3.1.1	Erreur quadratique moyenne de l'estimateur	104
3.1.2	Comparaison des lissages splines MCO et MCG	106
3.2	Sélection du paramètre de lissage	113
3.2.1	Extension des méthodes usuelles	113
3.2.2	La méthode GCV de Rice et Silverman (1991)	116
3.2.3	Une nouvelle méthode pour les splines cubiques	117
3.3	Comparaison des méthodes précédentes	119
4	Estimation adaptative en présence de corrélation	123
4.1	Lissage spline adaptatif	125
4.1.1	Estimateur	125
4.1.2	Noyau autoreproduisant	126
4.1.3	Estimation des paramètres de lissage et de corrélation	127
4.2	P-splines adaptatives	127
4.3	Lissage local vs lissage global : simulations	129
	Annexe A : Rappels	133
	Annexe B : Un résultat d'approximation spectrale	136
	Bibliographie	140

Notations

Ensembles, nombres

$\text{Card}(\Omega)$: cardinal de l'ensemble Ω .

$[x]$: partie entière d'un réel x .

Vecteurs, matrices

Les matrices sont notées en caractères gras majuscules (par exemple \mathbf{A}) et les vecteurs en caractères gras minuscules (par exemple $\mathbf{x} = (x_1, \dots, x_p)'$).

\mathbf{x}', \mathbf{A}' : transposée

$$\|\mathbf{x}\|^2 = \sum_{i=1}^p x_i^2.$$

$$\|\mathbf{A}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{A}\mathbf{x}\|.$$

$\text{tr } \mathbf{A}$: trace de \mathbf{A} .

$\det \mathbf{A}$: déterminant de \mathbf{A} .

$\lambda_1(\mathbf{A}), \lambda_2(\mathbf{A}), \dots$: valeurs propres de \mathbf{A} .

Espaces de fonctions

Soient $m \geq 0$ un entier et $a < b$ deux réels.

$L_2[a, b]$: ensemble des fonctions $g : [a, b] \rightarrow \mathbb{R}$ de carré intégrable.

$C^m[a, b]$: ensemble des fonctions $g : [a, b] \rightarrow \mathbb{R}$ m fois continûment dérivables.

$W_2^m[a, b]$: espace de Sobolev des fonctions $g : [a, b] \rightarrow \mathbb{R}$ dont les dérivées sont absolument continues jusqu'à l'ordre $m - 1$ et dont la dérivée d'ordre m est de carré intégrable. Par convention on pose $W_2^0[a, b] = L_2[a, b]$.

$$|g|_k^2 = \int_a^b (g^{(k)}(t))^2 dt : \text{(semi-) norme d'une fonction } g \in W^m[a, b] \text{ avec } 0 \leq k \leq m.$$

En particulier, $|\cdot|_0$ désigne la norme L_2 classique.

Suites

Soient $(u_p)_{p \geq 1}$ et $(v_p)_{p \geq 1}$ deux suites réelles.

$u_p = o(v_p)$ (prépondérance) : pour tout réel $\epsilon > 0$ on a $u_p \leq \epsilon v_p$ pour p assez grand.

$u_p = \mathcal{O}(v_p)$ (domination) : il existe un réel $C > 0$ tel que $u_p \leq C v_p$ pour p assez grand.

$u_p \sim v_p$ (équivalence asymptotique) : $u_p = \mathcal{O}(v_p)$ et $v_p = \mathcal{O}(u_p)$. En d'autres termes, le rapport $|u_p/v_p|$ reste borné loin de zéro et de l'infini pour p assez grand (si $v_p \neq 0$).

Introduction Générale

Cette introduction présente les thèmes abordés dans la thèse, puis elle donne un aperçu des résultats développés dans chaque chapitre.

Régression non paramétrique en présence d'erreurs corrélées

L'objet de la régression non paramétrique est d'estimer la relation de dépendance qui lie une variable d'intérêt y , dite variable dépendante, à une variable explicative t à partir de couples d'observations $(y_1, t_1), \dots, (y_p, t_p)$ de ces variables sans faire d'hypothèse paramétrique sur la forme de cette dépendance. Un modèle statistique classique pour représenter la dépendance de y par rapport à t est le suivant :

$$y_i = f(t_i) + \varepsilon_i, \quad 1 \leq i \leq p, \quad (1)$$

où f est la fonction de régression (déterministe) que l'on cherche à estimer et les $\varepsilon_1, \dots, \varepsilon_p$ sont des variables aléatoires (v.a.) qui représentent les erreurs de mesure, et plus généralement l'ensemble de la variabilité de y non expliquée par t . Les v.a. $\varepsilon_1, \dots, \varepsilon_p$ sont supposées de moyenne nulle : $\mathbb{E}\varepsilon_i = 0$ pour $1 \leq i \leq p$. L'ensemble des valeurs observées de t constitue le *dispositif expérimental* et lorsque ces valeurs t_1, \dots, t_p sont déterministes, on parle de *régression à effets fixes*.

Pour spécifier la structure de covariance des erreurs aléatoires $\varepsilon_1, \dots, \varepsilon_p$ du modèle (1), une hypothèse courante est celle de décorrélation et d'homoscédasticité des erreurs : autrement dit, $\mathbb{E}(\varepsilon_i \varepsilon_j) = \sigma^2 \delta_{ij}$ pour $1 \leq i, j \leq p$, où σ est un réel strictement positif et δ_{ij} est le symbole de Kronecker. Cette hypothèse peut être renforcée par le fait de supposer que les erreurs $\varepsilon_1, \dots, \varepsilon_p$ sont indépendantes et identiquement distribuées (i.i.d.). Sous ces hypothèses que l'on qualifie de *bruit blanc* (au sens large) et dans le contexte de la régression à effets fixes, les propriétés des estimateurs non paramétriques de f dans le modèle (1) ont été étudiées de manière extensive.

Depuis une vingtaine d'années une littérature importante s'est développée sur l'estimation dans le modèle (1) en s'affranchissant des hypothèses précédentes sur les erreurs qui sont parfois trop restrictives. En effet certains types de données com-

portent intrinsèquement des erreurs de mesure corrélées et/ou hétéroscédastiques, en raison notamment de l'existence de facteurs invariants à travers les mesures ou de l'influence de la variable explicative. C'est le cas notamment des séries chronologiques, des données longitudinales ou clusterisées, des mesures répétées, et des données spatiales ou spatio-temporelles. A ce sujet, Künsch *et al.* (1993) écrivent : "Perhaps most unbelievable to many is the observation that high-quality measurement series from astronomy, physics, chemistry, generally regarded as prototypes of 'i.i.d.' observations, are not independent but long-range correlated."

Influence de la corrélation sur les propriétés asymptotiques des estimateurs

La présence de corrélation dans un ensemble de données implique une mise en commun, une redondance des informations apportées par chaque mesure. Il y a ainsi une perte globale d'information par rapport à la situation d'indépendance. En termes de comportement asymptotique des estimateurs, on peut dire que plus la corrélation entre deux observations lointaines (*i.e.* séparées par de nombreuses autres mesures) est forte, plus la convergence de l'estimateur vers la fonction cible est lente, jusqu'au cas limite où l'estimation n'est plus consistante.

Nous allons illustrer cette idée en reprenant l'exemple du modèle (1) pour lequel il est possible de décrire trois grandes catégories de dépendance des erreurs $\varepsilon_1, \dots, \varepsilon_p$ en fonction de leur impact sur l'estimation lorsque $p \rightarrow \infty$. Avant d'énumérer ces catégories, nous précisons quelques hypothèses et notations. On suppose que la fonction de régression f est définie sur $[0, 1]$ et qu'elle est régulière dans un certain sens. On suppose aussi que les points d'observation t_1, \dots, t_p sont fixés et qu'ils deviennent denses dans $[0, 1]$ lorsque $p \rightarrow \infty$. Pour alléger les notations, nous écrivons y_i, t_i , et ε_i au lieu de y_{ip}, t_{ip} et ε_{ip} , mais nous considérons en fait des données de type tableau triangulaire, c'est-à-dire que les ensembles de points d'observation $\{t_1, \dots, t_p\}$ ne sont pas nécessairement emboîtés les uns dans les autres lorsque p varie, de même pour les mesures $\{y_1, \dots, y_p\}$. Nous présentons maintenant les trois catégories de dépendance des erreurs par force de corrélation croissante :

1. *Décorrélation/Indépendance.* Pour tout entier $p \geq 1$, les v.a. $\varepsilon_1, \dots, \varepsilon_p$ sont décorrélés ou i.i.d. et ont une variance commune $\sigma^2 > 0$. Sous cette hypothèse classique de bruit blanc, si f est dérivable à l'ordre m alors les estimateurs usuels de f sont consistants et leur convergence en moyenne quadratique optimale est de l'ordre de $p^{-2m/(2m+1)}$ (Stone, 1982; Nussbaum, 1985).

2. *Décorrélation asymptotique.* Pour tout entiers $1 \leq i, j \leq p$, les erreurs vérifient $\text{Cov}(\varepsilon_i, \varepsilon_j) = C_p(i, j)$, avec $\sup_{p \in \mathbb{N}^*, 1 \leq i \leq p} C_p(i, i) < \infty$ et $C_p(i, j) \rightarrow 0$ lorsque $|i - j| \rightarrow \infty$ et $p \rightarrow \infty$ (un cas particulier remarquable est celui où $C_p(i, j) = C(i, j)$, *i.e.* la covariance ne dépend que de l'ordre des mesures). Dans cette situation, la covariance de deux mesures en des points fixés $0 \leq s < t \leq 1$ tend vers zéro quand $p \rightarrow \infty$. Les estimateurs de f sont consistants, mais leur convergence est d'autant ralentie par rapport au bruit blanc que la corrélation se propage aux mesures lointaines. Des exemples largement étudiés dans la littérature statistique de dépendances entre les erreurs $\varepsilon_1, \dots, \varepsilon_p$ entrant dans cette catégorie sont la corrélation de courte/longue portée, la mélangeance, et la structure de processus ARMA.
3. *Corrélation persistante.* Pour tout entier $p \geq 1$, les v.a. $\varepsilon_i = \varepsilon(t_i)$, $1 \leq i \leq p$, sont issues d'un processus aléatoire centré $\{\varepsilon(t); t \in [0, 1]\}$ ayant une fonction de covariance non nulle et continue sur $[0, 1] \times [0, 1]$. Dans ce cas, la covariance entre deux mesures y_i et y_j dépend exclusivement des points t_i et t_j où elles sont réalisées. En particulier, elles ne dépendent ni de p , ni de i et j . D'autre part, la covariance entre deux mesures en des points fixés $s, t \in [0, 1]$ ne tend en général pas vers 0 quand $p \rightarrow \infty$. En l'absence de réplication des mesures aux points t_1, \dots, t_p , il n'est donc pas possible d'obtenir un estimateur consistant (voir Hart et Wehrly, 1986, et le Chapitre 2).

Influence de la corrélation sur l'estimation dans le cadre d'un échantillon fini

L'étude asymptotique de la régression non paramétrique dans le cadre de données corrélées fait apparaître un potentiel ralentissement, par rapport au cas du bruit blanc, de la convergence des estimateurs vers la fonction cible. Ce ralentissement est dû au phénomène de corrélation lui-même et non pas à la procédure d'estimation. Cependant, lorsque dans la pratique on dispose d'un ensemble fini de données corrélées, il devient crucial de pouvoir estimer la régression avec un maximum d'efficacité. La question qui vient alors naturellement à l'esprit est de savoir si les méthodes usuelles de construction d'estimateurs, qui sont développées dans le cadre d'erreurs décorréelées, fonctionnent bien dans un contexte de corrélation. En particulier, il s'agit de savoir si pour les estimateurs usuels (estimateurs à noyau, de projection, splines de lissage, polynômes locaux, ondelettes, *etc.*), les méthodes usuelles de choix du paramètre de lissage (largeur de fenêtre, dimension de l'espace de projection, niveau de résolution, *etc.*) de type validation croisée (CV, GCV), maximum de vraisemblance généralisé (GML), estimation sans biais du risque (UBR) ou plug-in, sont efficaces en présence d'erreurs corrélées.

Un corps important de travaux théoriques et d'études de simulations (notamment Hart et Wehrly, 1986; Altman, 1990; Wang, 1998; Opsomer *et al.*, 2001) donne une réponse globalement négative à la question précédente. L'échec des méthodes de sélection de paramètre de lissage basées sur l'hypothèse d'erreurs décorréelées en présence de corrélation tient à ce que dans l'ajustement des données, ces méthodes attribuent les effets aléatoires systématiques induits par la corrélation à la fonction de régression et non pas aux erreurs.

Les constats empiriques précédents, appuyés par la théorie, ont donné lieu à l'adaptation et à l'extension des méthodes existantes de construction d'estimateurs dans le sens d'une prise en compte de la structure de covariance des erreurs (voir entre autres Diggle et Hutchinson, 1989; Altman, 1990; Rice et Silverman, 1991; Kohn *et al.*, 1992; Wang, 1998). Pour estimer cette structure de covariance en général inconnue, il est souvent nécessaire de la modéliser de façon paramétrique ou de faire une hypothèse de stationnarité des erreurs. La construction de l'estimateur se ramène alors à l'estimation simultanée ou séquentielle d'un bon paramètre de lissage d'une part, et des paramètres de la corrélation d'autre part.

La structure de *corrélation persistante* décrite en page 7 appelle des commentaires spécifiques. D'une part, ce type de corrélation étant engendré par un processus aléatoire $\varepsilon = \{\varepsilon(t); t \in [0, 1]\}$, l'estimation de la fonction f du modèle (1) dans ce contexte peut aussi bien être référencée dans la littérature comme le problème de l'estimation de la moyenne d'un processus aléatoire $X = f + \varepsilon$, que comme une régression non paramétrique particulière. La distinction entre ces deux points de vue n'est pas qu'une simple question de vocabulaire : il existe par exemple dans le cadre des processus gaussiens un ensemble de problématiques, de résultats théoriques, et de méthodes d'estimation (*cf.* Grenander, 1981; Beder, 1987) qui n'ont pas d'équivalent en régression. D'autre part, la sévérité de la corrélation des erreurs issues d'un processus aléatoire rend l'estimation ponctuelle de f relativement difficile (certaines trajectoires du processus X peuvent être systématiquement en dessous de f , ou au dessus de f). Dans ces conditions, l'estimation par intervalles ou par bandes de confiance simultanés peut s'avérer particulièrement intéressante. Toutefois, il est difficile de développer efficacement ce type d'estimation dans le cas d'un processus général : il semble nécessaire de spécifier la forme de la distribution du processus. Parmi les rares travaux publiés sur ce sujet, on peut citer l'estimation par bandes de confiance simultanées de l'intensité d'un processus de Poisson non homogène réalisée par Loader (1992).

Estimation adaptative d'une fonction spatialement hétérogène dans un contexte de données corrélées

La régression non paramétrique sur des données corrélées a largement été étudiée. Le champ statistique plus récent de l'estimation adaptative d'une fonction spatialement hétérogène a aussi été exploré par de nombreux chercheurs. Cependant ces deux thèmes ont rarement été examinés *en même temps*. Une des raisons à cela est la difficulté à discerner ce qui, dans la variabilité observée d'un jeu de données, a trait à l'hétérogénéité spatiale de la fonction moyenne de ce qui a trait aux effets aléatoires systématiques dûs notamment à la corrélation et à l'hétéroscédasticité des erreurs. Pour faire de l'inférence statistique, une telle différenciation est nécessaire et n'est rendue possible qu'en spécifiant à la fois un modèle pour le signal et un modèle pour le bruit. Dans ce but, Crainiceanu *et al.* (2007) ont développé un modèle de régression bayésienne dans lequel la fonction de régression et la fonction de variance du processus d'erreur sont représentées à l'aide de splines. Ils proposent un ensemble de lois *a priori* dans le cadre d'un modèle linéaire mixte et ils déterminent les lois postérieures jointes des paramètres par des simulations MCMC. Une autre approche envisageable pour effectuer une estimation doublement adaptative à la corrélation des erreurs et aux variations rapides de la fonction de régression consiste à utiliser une modélisation paramétrique des erreurs. Dans ce cas il reste possible, notamment avec les estimateurs splines, de se placer dans un cadre bayésien empirique permettant de faire de l'inférence (considérer par exemple Ruppert et Carroll, 2000, et les extensions possibles de leur travail à un cadre corrélé).

Chapitre 1

Ce chapitre porte sur l'estimation d'une régression par la méthode des splines de lissage dans un modèle de mesures répétées bruitées. Son résultat principal consiste en un ensemble de majorations de l'erreur quadratique moyenne de l'estimateur qui sont valables pour de nombreuses structures de bruit. Ces majorations généralisent les travaux de Craven et Wahba (1979) et de Ragozin (1983) établis dans le cadre classique du modèle de régression non paramétrique (1) sous l'hypothèse d'erreurs de mesures décorréelées et homoscédastiques. Nous appliquons ces majorations à l'obtention de vitesses de convergence dans le modèle (1) sous diverses hypothèses de dépendance des erreurs (corrélation de courte portée, ARMA, et mélangeance). Les vitesses trouvées étendent les résultats analogues établis pour d'autres estimateurs par Altman (1990), Hall et Hart (1990), Burman (1991), et Wang (1996). Le second résultat important de ce chapitre est le calcul numérique de la norme L_2 d'une spline cubique naturelle.

Dans ce chapitre, nous considérons le modèle de mesures répétées suivant :

$$y_{ij} = f(t_{ij}) + \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq p_i, \quad (2)$$

où les y_{ij} sont des mesures effectuées en des points déterministes t_{ij} , f est une fonction de régression définie sur $[0,1]$, et les ε_{ij} sont des erreurs aléatoires centrées. L'indice i désigne une unité statistique d'une population donnée et l'indice j est un indice de répétition pour différentes valeurs d'une variable explicative $t \in [0, 1]$.

Plutôt que d'exposer le modèle précédent dans toute sa généralité, nous nous limitons dans cette introduction au cas particulier d'observations équilibrées :

$$y_{ij} = f(t_j) + \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq p, \quad (3)$$

pour lequel les erreurs ε_{ij} sont indépendantes à travers l'indice i et les vecteurs aléatoires $(\varepsilon_{i1}, \dots, \varepsilon_{ip})'$, $1 \leq i \leq n$, ont tous la même matrice de covariance \mathbf{V} . De plus, on suppose que les points t_1, \dots, t_p sont distincts et que la variance des v.a. ε_{ij} est uniformément bornée en j et en p .

Les modèles (2) et (3) seront les modèles centraux de cette thèse. Ils sont "riches" en ce sens qu'ils englobent des structures de covariance et des dispositifs expérimentaux très généraux. Ils permettent en outre d'étudier des asymptotiques portant aussi bien sur les nombres p_i , $1 \leq i \leq n$, ou p de points d'observation que sur le nombre n d'unités statistiques.

La fonction f , supposée appartenir à l'espace de Sobolev $W_2^m[0, 1]$ pour un entier $m \geq 1$ donné, est estimée par une spline de lissage \hat{f}_λ d'ordre $2m$ définie comme unique solution du problème variationnel : Trouver $g \in W_2^m[0, 1]$ qui minimise

$$\frac{1}{p} \sum_{j=1}^p (\bar{y}_j - g(t_j))^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt, \quad (4)$$

où $\lambda > 0$ est un paramètre de lissage et $\bar{y}_j = n^{-1} \sum_{i=1}^n y_{ij}$ pour $1 \leq j \leq p$.

Dans l'hypothèse classique où $n = 1$ et $\mathbf{V} = \sigma^2 \mathbf{I}$ dans le modèle (3) (ce cadre est équivalent au modèle (1) sous l'hypothèse de bruit blanc), Craven et Wahba (1979) ont majoré l'erreur quadratique moyenne de l'estimateur \hat{f}_λ aux points d'observation t_1, \dots, t_p . Avec les notations $\mathbf{f} = (f(t_1), \dots, f(t_p))'$ et $\hat{\mathbf{f}}_\lambda = (\hat{f}_\lambda(t_1), \dots, \hat{f}_\lambda(t_p))'$, cette erreur, appelée MASE (pour Mean Average Squared Error), s'écrit $p^{-1} \mathbb{E} \|\mathbf{f} - \hat{\mathbf{f}}_\lambda\|^2$.

Théorème 1. (Craven et Wahba, 1979) *Dans le modèle (1), on suppose que*

- (i) *les points d'observation sont tels que $\int_0^{t_j} w(t) dt = j/p$ pour $1 \leq j \leq p$, où w est une fonction de poids strictement positive et continue sur $[0, 1]$;*
- (ii) *la fonction f est dans l'espace de Sobolev $W_2^m[0, 1]$ pour un $m \geq 1$ donné;*

(iii) les erreurs $\varepsilon_1, \dots, \varepsilon_p$ sont décorrélées et de variance commune $\sigma^2 > 0$.

Alors, en supposant que f n'est pas un polynôme de degré $< m$, l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ est consistant si et seulement si $\lambda = \lambda(p) \rightarrow 0$ et $\lambda p^{2m} \rightarrow \infty$ lorsque $p \rightarrow \infty$. D'autre part, la MASE de cet estimateur vérifie

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 = \mathcal{O}(\lambda) + \mathcal{O}\left(\frac{1}{\lambda^{1/2m} p}\right).$$

Nous généralisons la majoration précédente au cadre du modèle (3) ainsi :

$$\begin{cases} \frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{\lambda}{4} |f|_m^2 + \frac{\text{tr } \mathbf{V}}{np} \\ \frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{\lambda}{4} |f|_m^2 + \frac{M_2 \|\mathbf{V}\|}{np \lambda^{1/2m}} \end{cases}, \quad (5)$$

où M_2 est une constante dépendant du dispositif expérimental et de m qui reste uniformément bornée lorsque $n, p \rightarrow \infty$ sous l'hypothèse (i) du Théorème 1. L'intérêt de la double majoration (5) est que l'inégalité du haut (où figure $\text{tr } \mathbf{V}$) permet d'obtenir la vitesse de convergence optimale dans le cas d'erreurs ε_{ij} de type processus à covariance continue, tandis que celle du bas (où figure $\|\mathbf{V}\|$) fournit la vitesse de convergence optimale dans certains cas de décorrélation asymptotique des erreurs ε_{ij} . Ces vitesses et leurs conditions d'obtention seront précisées plus loin.

Considérons maintenant l'erreur quadratique moyenne intégrée, notée MISE (pour Mean Integrated Squared Error). Dans le cas où $n = 1$ et $\mathbf{V} = \sigma^2 \mathbf{I}$ dans le modèle (3), *i.e.* dans le modèle (1) sous l'hypothèse de bruit blanc, Ragozin (1983) a majoré la MISE de l'estimation des dérivées de la fonction f par les dérivées correspondantes de la spline de lissage \widehat{f}_λ solution de (4).

Théorème 2. (Ragozin, 1983) Dans le modèle (1), on suppose que

- (i) les points d'observations sont équidistants : $t_j = j/p, 1 \leq i \leq p$;
- (ii) la fonction f est dans l'espace de Sobolev $W_2^k[0, 1]$, avec $0 < k \leq m$;
- (iii) les erreurs ε_i sont centrées, décorrélées et de variance commune σ^2 .

Soit \widehat{f}_λ l'estimateur spline de lissage d'ordre $2m$ de f . Si le paramètre de lissage $\lambda = \lambda(p)$ est choisi de sorte que $(p\lambda^{1/2m})^{-1} \leq C$ pour une constante C et pour tout $p \geq m$, alors la MISE de la j -ème dérivée ($j < k$) de \widehat{f}_λ vérifie

$$\mathbb{E}|f - \widehat{f}_\lambda|_j^2 := \mathbb{E} \int_0^1 (f^{(j)}(t) - \widehat{f}_\lambda^{(j)}(t))^2 dt \leq P(\lambda + p^{-2m})^{(k-j)/m} |f|_k^2 + \frac{Q\sigma^2}{p\lambda^{(2j+1)/2m}}$$

pour des constantes P, Q ne dépendant que de m, k , et C . En particulier si $\lambda(p) \sim p^{-2m/(2k+1)}$ lorsque $p \rightarrow \infty$, alors

$$\mathbb{E}|f - \widehat{f}_\lambda|_j^2 = \mathcal{O}\left(p^{-2(k-j)/(2k+1)}\right).$$

En fait, on peut montrer que le théorème précédent reste valable dans un dispositif expérimental quasi-uniforme, *i.e.* tel que $\sup_{n \geq 1, p \geq m} (h_{max}/h_{min}) < \infty$, avec $h_{max} = \max_{0 \leq j \leq p} (t_{(j+1)} - t_{(j)})$ et $h_{min} = \min_{1 \leq j \leq p-1} (t_{(j+1)} - t_{(j)})$, où on a noté $t_{(0)} = 0$, $t_{(p+1)} = 1$, et $t_{(1)} < \dots < t_{(p)}$ les points t_1, \dots, t_p ordonnés.

Comme pour le Théorème 1, nous généralisons le Théorème 2 au cadre du modèle (3). Plus précisément, on suppose que $f \in W^k[0, 1]$ pour un entier $k > 0$. Étant donné un entier $m \geq k$, on considère l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p)$. On suppose que $h_{max}^{2m} = \mathcal{O}(\lambda)$ lorsque $n, p \rightarrow \infty$. Alors la MISE de la j -ème dérivée ($j < k$) de \widehat{f}_λ est telle que

$$\begin{cases} \mathbb{E}|f - \widehat{f}_\lambda|_j^2 \leq P(\lambda + h_{max}^{2m})^{(k-j)/m} |f|_k^2 + \frac{Q \operatorname{tr} \mathbf{V}}{np\lambda^{j/m}} \\ \mathbb{E}|f - \widehat{f}_\lambda|_j^2 \leq P(\lambda + h_{max}^{2m})^{(k-j)/m} |f|_k^2 + \frac{Q' \|\mathbf{V}\|}{np\lambda^{(2j+1)/2m}} \end{cases} \quad (6)$$

pour des constantes P, Q, Q' indépendantes de n et p .

Les inégalités (5) et (6) permettent d'obtenir des vitesses de convergence dans le modèle classique (1) pour certaines formes de dépendance des erreurs.

Supposons que $f \in W_2^m[0, 1]$ et considérons l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ de f , avec $\lambda = \lambda(p)$.

Dans le cas où les erreurs $\varepsilon_1, \dots, \varepsilon_p$ ont une corrélation de courte portée, c'est-à-dire qu'il existe un réel $\gamma > 1$ pour lequel $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) \asymp |i - j|^{-\gamma}$ uniformément en i et j ($i \neq j$) lorsque $p \rightarrow \infty$ et que $\sup_{p \in \mathbb{N}^*, 1 \leq j \leq p} \operatorname{Var} \varepsilon_j < \infty$, nous montrons que

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 = \mathcal{O}(\lambda) + \mathcal{O}\left(\frac{1}{p\lambda^{1/2m}}\right) \quad (7)$$

lorsque $p \rightarrow +\infty$. Nous obtenons la même majoration dans le cas où les erreurs $\varepsilon_1, \dots, \varepsilon_p$ sont issues d'un processus ARMA stationnaire et inversible. Des majorations analogues sont valables pour la MISE de \widehat{f}_λ et de ses dérivées d'ordre $j < k$, à condition que le dispositif expérimental soit quasi-uniforme et que le paramètre de lissage ne soit pas trop petit ($\lambda p^{2m} \rightarrow \infty$ lorsque $p \rightarrow \infty$). On constate que la vitesse de convergence optimale en $p^{-2m/(2m+1)}$ de l'estimateur \widehat{f}_λ ne change pas par rapport au cas d'un bruit blanc. Ces résultats étendent les vitesses établies pour les estimateurs à noyau par Altman (1990) et par Hall et Hart (1990) sous des hypothèses un peu plus générales de dépendance des erreurs.

Nous étudions aussi le cas où les erreurs $\varepsilon_1, \dots, \varepsilon_p$ du modèle (1) sont mélangeantes au sens de Burman (1991). Dans cette situation, il existe une suite croissante de réels $(R_p)_{p \geq 1}$ et une fonction φ , resp. α , décroissante telles que

- (i) $R_p \leq p$ et $R_p \rightarrow \infty$ quand $p \rightarrow \infty$;
- (ii) $\int_0^\infty \varphi^{1/2}(t)dt < \infty$, resp. $\int_0^\infty \alpha^{1/2}(t)dt < \infty$;
- (iii) Pour tous entiers $1 \leq j < k \leq p$, on a $|P(B|A) - P(B)| \leq \varphi(R_p(x_k - t_j))$, resp. $|P(A \cap B) - P(A)P(B)| \leq \alpha(R_p(t_k - t_j))$ où A et B sont des événements engendrés respectivement par $\varepsilon_1, \dots, \varepsilon_j$ et $\varepsilon_k, \dots, \varepsilon_p$;
- (iv) $\sup_{p \in \mathbb{N}^*, 1 \leq i \leq p} \mathbb{E}\varepsilon_i^2 < \infty$, resp. $\sup_{p \in \mathbb{N}^*, 1 \leq i \leq p} \mathbb{E}\varepsilon_i^4 < \infty$.

Nous montrons que sous ces conditions sur les erreurs, et si le paramètre de lissage de \hat{f}_λ vérifie $\lambda R_p^{2m} \rightarrow \infty$ lorsque $p \rightarrow \infty$, alors la MASE et la MISE de \hat{f}_λ sont en $\mathcal{O}(\lambda) + \mathcal{O}(R_p^{-1}\lambda^{-1/2m})$ et la vitesse optimale de convergence en $\mathcal{O}(R_p^{-2m/(2m+1)})$ est atteinte pour un choix de λ du même ordre. Ce résultat étend ainsi au cadre des splines de lissage la vitesse établie par Burman (1991) pour les splines de régression.

Nous établissons également deux résultats de convergence dans le cadre d'erreurs issues d'un processus aléatoire à covariance continue, et de données longitudinales issues d'un grand nombre d'individus chacun observé à un petit nombre de reprises.

Les majorations complémentaires de (5) et (6) permettent ainsi de retrouver les vitesses de convergence de la littérature et de les étendre aux splines de lissage dans le cas où les erreurs $\varepsilon_1, \dots, \varepsilon_p$ du modèle (1) ont une corrélation à décroissance rapide (corrélation de courte portée, ARMA) ou modérée (mélange), et aussi dans le cas "opposé" du spectre où la corrélation est très sévère (processus aléatoire avec des mesures répétées). Cependant, ces majorations ne sont pas efficaces pour des erreurs qui ont une structure de corrélation relativement forte tout en restant asymptotiquement décorréliées, comme dans le cas de la corrélation de longue portée. Dans ce cas, l'emploi de techniques plus fines est nécessaire pour montrer que les splines de lissage atteignent la vitesse minimax établie entre autres par Wang (1996).

Un autre résultat important du Chapitre 1 est le calcul numérique de la norme L_2 d'une spline cubique naturelle g ayant ses noeuds aux points t_1, \dots, t_p . En utilisant la représentation de telles fonctions par valeur et par dérivée seconde (voir par exemple Green et Silverman, 1994), nous montrons par un calcul fastidieux mais routinier qu'il existe une matrice Φ de taille $p \times p$, définie positive, ne dépendant que des noeuds t_1, \dots, t_p , telle que

$$|g|_0^2 = \mathbf{g}'\Phi\mathbf{g}, \quad (8)$$

avec les notations $|g|_0^2 = \int_0^1 (g(t))^2 dt$ et $\mathbf{g} = (g(t_1), \dots, g(t_p))'$. Deux applications sont tirées de ce résultat : la première consiste en la majoration de la MISE d'un estimateur spline cubique de lissage dans un dispositif équidistant, avec un calcul fin des constantes numériques multiplicatives. La seconde, plus intéressante, consiste à utiliser l'équation précédente pour estimer avec précision la variance intégrée d'un estimateur spline et intégrer cette estimation dans une procédure de sélection du paramètre de lissage. Cette application sera présentée au Chapitre 3.

Une partie du Chapitre 1 a fait l'objet d'une publication sous la référence :

Degras, D. et Jallet, R. (2005). Convergence de l'estimateur spline cubique de lissage dans un modèle de régression longitudinale avec erreur de type processus. Comptes Rendus de l'Académie des Sciences de Paris, 340 (11), 851–854.

Chapitre 2

Ce chapitre est consacré à l'estimation non paramétrique de la moyenne f d'un processus aléatoire $X = \{X(t); t \in [0, 1]\}$ de variance finie sur $[0, 1]$. Des résultats de consistance, de convergence, et de normalité asymptotique sont établis pour différents estimateurs de f construits à partir d'un échantillon i.i.d. (X_1, \dots, X_n) de X observé en p points déterministes dans l'asymptotique $n, p \rightarrow \infty$. Dans un premier temps, les résultats sont établis pour une classe générale d'estimateurs linéaires comprenant tous les estimateurs usuels. Le résultat de normalité asymptotique est alors appliqué à la construction d'intervalles de confiance simultanés. Dans un second temps, plusieurs résultats de convergence sont formulés pour les estimateurs à noyau et les splines de lissage. Certains de ces résultats permettent de déterminer le paramètre de lissage optimal de chaque estimateur.

Dans ce chapitre, nous indiquons explicitement la dépendance des objets mathématiques considérés par rapport à p et/ou n dans un but de compréhension. Nous considérons ainsi un dispositif expérimental noté $\{t_{1p} < \dots < t_{pp}\}$. La fonction de covariance du processus X est notée R . Cette fonction est inconnue et on la suppose continue sur $[0, 1] \times [0, 1]$.

Nous considérons d'abord un estimateur linéaire de f de forme générale

$$\widehat{f}_{np}(t) = \sum_{i=1}^p W_{nip}(t) \bar{X}_n(t_{ip}) \quad (9)$$

où $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, et W_{n1p}, \dots, W_{npp} sont des fonctions de poids qui dépendent seulement de n, p , et du dispositif expérimental $\{t_{1p}, \dots, t_{pp}\}$.

A l'aide du théorème de Mercer, nous établissons sous de faibles conditions de régularité sur les poids W_{njp} , $1 \leq j \leq p$, et sur la covariance R que pour tout $n \geq 1$, la variance de $\widehat{f}_{np}(t)$ vérifie

$$\text{Var } \widehat{f}_{np}(t) = \frac{1 + o(1)}{n} R(t, t) \quad (10)$$

lorsque $p \rightarrow \infty$, où le $o(1)$ est uniforme par rapport à n . Ce résultat est valable pour l'ensemble des estimateurs usuels, pourvu que le dispositif expérimental soit

suffisamment régulier et que le choix du paramètre de lissage soit raisonnable, et pour la plupart des processus à temps continu usuels. Un corollaire immédiat de l'équivalence asymptotique (10) est que l'estimateur \widehat{f}_{np} est consistant en moyenne quadratique si et seulement si $p \rightarrow \infty$ et $n \rightarrow \infty$, et que dans ce cas sa vitesse de convergence est au mieux de l'ordre de $1/n$. Ces faits sont connus intuitivement par les statisticiens (Hart et Wehrly, 1986; Diggle et Hutchinson, 1989; Wang, 1998), mais ils n'avaient jamais été démontrés dans un cadre aussi général qu'ici.

En utilisant la relation (10), le Théorème Central Limite, et une condition asymptotique sur le biais de l'estimateur, nous pouvons établir la normalité asymptotique de \widehat{f}_{np} en un nombre fini de points η_1, \dots, η_q fixés dans $[0, 1]$. Avec les notations $\widehat{\mathbf{f}}_{np} = (\widehat{f}_{np}(\eta_1), \dots, \widehat{f}_{np}(\eta_q))'$, $\mathbf{f} = (f(\eta_1), \dots, f(\eta_q))'$, et $\boldsymbol{\Sigma} = (R(\eta_i, \eta_j))_{1 \leq i, j \leq q}$, nous avons alors, lorsque $n, p \rightarrow \infty$, la convergence en distribution

$$n^{1/2} \left(\widehat{\mathbf{f}}_{n,p} - \mathbf{f} \right) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}). \quad (11)$$

Une application de cette normalité asymptotique à la construction d'intervalles de confiance simultanés (ICS) aux points η_1, \dots, η_q est ensuite envisagée. En estimant la matrice $\boldsymbol{\Sigma}$ par la matrice de covariance empirique et en utilisant le théorème de Slutsky, nous construisons un ellipsoïde de confiance pour \mathbf{f} au niveau, disons $1 - \lambda$, à l'aide de la loi $\chi^2(q)$, puis la méthode classique de projection de Scheffé conduit aux ICS suivants :

$$\left[\widehat{f}_{np}(\eta_k) \pm \sqrt{\left(\frac{1}{n} \sum_{i=1}^n (X_i(\eta_k) - \bar{X}_n(\eta_k))^2 \right) \frac{z_{1-\lambda}}{n}} \right], 1 \leq k \leq q, \quad (12)$$

pour les $f(\eta_k)$, $1 \leq k \leq q$, où $z_{1-\lambda}$ est le quantile de niveau $1 - \lambda$ de la loi $\chi^2(q)$.

Cette méthode non paramétrique de construction d'ICS est comparée à une méthode paramétrique au cours de la simulation d'un processus de Poisson non homogène et il apparaît que les deux méthodes ont des performances similaires.

Nous observons que les résultats précédents sur l'estimateur général \widehat{f}_{np} ont été exposés par commodité dans le cadre d'un processus défini sur $[0, 1]$, mais qu'ils peuvent être généralisés aux processus définis sur un espace métrique compact (notamment les processus spatiaux ou spatio-temporels) à l'instar du théorème de Mercer et du TCL utilisés dans leur démonstration.

Dans la deuxième partie du Chapitre 2, nous nous intéressons spécifiquement à la convergence des estimateurs splines de lissage et des estimateurs à noyau. Les résultats obtenus portent sur différents modes de convergence et étendent des résultats antérieurs de la littérature (Ragozin, 1983; Hart et Wehrly, 1986;

Biritxinaga, 1990; Nychka, 1995). Nous présentons ici deux de ces résultats qui concernent l'erreur quadratique moyenne ponctuelle (MSE) des estimateurs étudiés.

Commençons par l'estimateur de Gasser-Müller, qui s'écrit sous la forme

$$\widehat{f}_{np}^{GM}(t) := \sum_{j=1}^p \left(\int_{s_{(j-1)p}}^{s_{jp}} K_h(t-s) ds \right) \overline{X}_n(t_{jp}), \quad (13)$$

où $K_h(t) = (1/h)K(t/h)$, avec K un noyau à support compact et $h = h(n, p) > 0$ une largeur de fenêtre. Les points $(s_{jp})_{0 \leq j \leq p}$ sont tels que $s_{0p} = 0, s_{pp} = 1$, et $s_{(j-1)p} \leq t_{jp} \leq s_{jp}$ pour tout $1 \leq j \leq p$. Pour simplifier l'exposé, nous supposons que K est une densité symétrique de support $[-1, 1]$ et que K vérifie une condition de Lipschitz. Nous supposons aussi que la fonction moyenne f est dans $C^2[0, 1]$ et que les (tableaux triangulaires de) points $(t_{ip})_{1 \leq i \leq p}$ et $(s_{ip})_{1 \leq i \leq p}$ ont une répartition quasi-uniforme. Pour la covariance R , nous envisageons les hypothèses suivantes :

- (i) R vérifie une condition de Lipschitz $|R(u, v) - R(u', v')| \leq C(|u - u'| + |v - v'|)$;
- (ii) R admet des dérivées semi-directionnelles au point (t, t) , au sens que $D_{(u,v)}^+ R(t, t) = \lim_{h \rightarrow 0^+} \frac{R(t+uh, t+vh) - R(t, t)}{h}$ existe et est finie pour tout $(u, v) \in \mathbb{R}^2$.

Posons $C_{KR}(t) = \int_{-1}^1 \int_{-1}^1 K(u)K(v)D_{(u,v)}^+ R(t, t)dudv$ et $\sigma_K^2 = \int_{-1}^1 u^2 K(u)du$.

Nous montrons que lorsque $n, p \rightarrow \infty$, la MSE de l'estimateur vérifie

$$\mathbb{E} \left(f(t) - \widehat{f}_{np}^{GM}(t) \right)^2 = \frac{R(t, t) + h C_{KR}(t)}{n} + \frac{h^4 (f''(t))^2}{4} + o(h^4) + o\left(\frac{1}{n}\right), \quad (14)$$

uniformément par rapport à t sur tout compact de $[0, 1]$.

Il est également possible de faire un développement plus fin sous des hypothèses renforcées sur les dérivées de R . L'ensemble de ces résultats généralise alors le travail de Hart et Wehrly (1986) effectué dans l'hypothèse d'une covariance stationnaire. On note que des développements analogues peuvent être établis par les mêmes calculs pour l'estimateur de Priestley et Chao (1972).

Nous examinons maintenant le cas de l'estimateur spline de lissage.

Soit un entier $m \geq 1$. Supposons que

- (i) f admet une dérivée $f^{(2m)}$ hölderienne sur $[0, 1]$;
- (ii) R admet une dérivée croisée $R^{(2m, 2m)}$ hölderienne d'indice β sur $[0, 1] \times [0, 1]$;
- (iii) la fonction de répartition F_p du dispositif expérimental converge uniformément sur $[0, 1]$ vers une fonction $F \in C^2[0, 1]$ de dérivée $F' = \psi > 0$;
- (iv) $\lambda = o(1)$ et $\|F_p - F\|_\infty = o(\lambda^{2+1/2m})$ lorsque $n, p \rightarrow \infty$.

Soit \widehat{f}_λ l'estimateur spline de lissage d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p) > 0$ associé aux données $\overline{X}_n(t_{1p}), \dots, \overline{X}_n(t_{pp})$.

Alors sous une hypothèse technique supplémentaire, la MSE de \widehat{f}_λ vérifie

$$\begin{aligned} \mathbb{E} \left(f(t) - \widehat{f}_\lambda(t) \right)^2 &= \frac{R(t, t) + \lambda(-1)^{m+1} 2R^{(2m,0)}(t, t) + \lambda^2 R^{(2m,2m)}(t, t)}{n} \\ &+ \frac{\lambda^2 (f^{(2m)}(t))^2}{\psi^2(t)} + o(\lambda^2) + o\left(\frac{1}{n}\right). \end{aligned} \quad (15)$$

lorsque $n, p \rightarrow \infty$, uniformément par rapport à t sur tout compact de $[0, 1]$.

La première partie du Chapitre 2 est résumée dans un article soumis au journal Statistics and Probability Letters sous le titre “Asymptotics for the nonparametric estimation of the mean of a random process”.

Chapitre 3

Ce chapitre porte sur la régression par splines de lissage dans un contexte de mesures répétées. L’accent est placé sur la pratique du lissage, et notamment sur le choix de la métrique et du paramètre de lissage utilisés pour la construction de l’estimateur. L’estimateur spline de lissage des moindres carrés généralisés [MCG], qui tient compte de la structure de covariance des erreurs du modèle, est défini. Son erreur quadratique MISE est comparée à celle de l’estimateur des moindres carrés ordinaires [MCO] du Chapitre 1 sur quelques exemples de fonction de régression et de structure de bruit (à notre connaissance, il n’existe pas dans la littérature de comparaison de cette nature). Les principales méthodes de sélection du paramètre de lissage des estimateurs splines en présence de données corrélées sont présentées. Nous construisons une nouvelle méthode de sélection du paramètre de lissage spécifique aux mesures répétées, que nous comparons à la méthode GCV de Rice et Silverman (1991) lors de simulations.

Nous nous plaçons à nouveau dans le cadre du modèle de mesures répétées équilibrées (3) du Chapitre 1. Nous supposons que la régression f est dans l’espace $W_2^m[0, 1]$ pour un entier $m \geq 1$ donné. L’estimateur spline de lissage MCG d’ordre $2m$ et de paramètre $\lambda > 0$ est alors défini comme l’unique solution du problème variationnel : Trouver $g \in W_2^m[0, 1]$ qui minimise

$$\frac{1}{p} (\bar{\mathbf{y}} - \mathbf{g})' \mathbf{V}^{-1} (\bar{\mathbf{y}} - \mathbf{g}) + \lambda \int_0^1 (g^{(m)}(t))^2 dt, \quad (16)$$

avec $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_p)'$. On retrouve bien sûr l’estimateur spline de lissage MCO défini par (4) dans le cas où $\mathbf{V} = \mathbf{I}$.

On obtient facilement des inégalités du type de (5) pour l'estimateur MCG, et il apparaît, conformément à l'intuition, que cet estimateur n'améliore pas la vitesse de convergence de l'estimateur MCO. On peut s'attendre cependant, comme dans le cas du modèle linéaire, à ce que sur un échantillon fini, l'estimation par MCG donne de meilleurs résultats que celle par MCO. Sur les exemples de fonctions f et de matrices de covariance \mathbf{V} traités dans la Section 3.1.2, il apparaît que ce n'est pas le cas (en termes d'erreur quadratique MISE). Une explication est la suivante : de manière globale, les biais au carré des deux estimateurs sont très proches pour tout $\lambda > 0$; par ailleurs, leurs variance sont comparables si le lissage est faible (λ petit), et si le lissage est modéré ou fort (disons $\lambda \geq 10^{-3}$) alors la variance de l'estimateur MCG est significativement plus petite que celle de MCO. Cependant, l'équilibre entre biais au carré et variance qui rend la MISE minimale est atteint pour des valeurs de λ beaucoup plus petites que 10^{-3} (de l'ordre de 10^{-7} ou 10^{-8}). Pour cette raison, les performances optimales des estimateurs MCO et MCG sont très comparables.

La deuxième section du Chapitre 3 présente l'extension au cadre de données corrélées des méthodes usuelles (GML,UBR,GCV) de sélection du paramètre de lissage réalisée par Wang (1998) et Diggle et Hutchinson (1989). Ces méthodes sont développées dans le cadre du modèle (1), c'est-à-dire avec une seule unité statistique observée. Dans ce cadre, il est très difficile d'estimer la structure de covariance \mathbf{V} sans une modélisation paramétrique de cette matrice car sans une telle modélisation, la fonction f et les erreurs $\varepsilon_1, \dots, \varepsilon_p$ sont essentiellement indiscernables. Dans la situation de mesures répétées ($n > 1$ dans le modèle (3)), il devient au contraire possible de choisir un paramètre de lissage en tenant compte de la covariance des données sans avoir à faire d'hypothèse paramétrique sur la forme de \mathbf{V} . Dans ce cadre, nous présentons la méthode GCV "leave one curve out" de Rice et Silverman (1991) et nous contruisons une méthode originale de sélection de λ basé sur le calcul exact de la variance intégrée d'une spline cubique de lissage fait au Chapitre 1. Cette méthode que nous dénotons SPSC consiste à minimiser le score

$$S^*(\lambda) := |\widehat{f}_0 - \widehat{f}_\lambda|_0^2 + \frac{1}{n} \operatorname{tr}(\mathbf{S}_\lambda \Phi \widehat{\mathbf{V}}) + \frac{1}{n} \operatorname{tr}(\Phi \mathbf{S}_\lambda \widehat{\mathbf{V}}), \quad (17)$$

où \widehat{f}_0 est la spline cubique naturelle qui interpole les données $(\bar{y}_1, t_1), \dots, (\bar{y}_p, t_p)$, \mathbf{S}_λ est la matrice $p \times p$ associée à l'estimateur spline cubique \widehat{f}_λ des MCO, Φ est la matrice présentée en (8), et $\widehat{\mathbf{V}}$ désigne la matrice de covariance empirique. Le score $S^*(\lambda)$ estime approximativement la MISE de \widehat{f}_λ à une constante additive près, et on observe que ces deux quantités prennent leur valeur minimale pour des valeurs de λ très voisines.

Des simulations ont été menées pour comparer les performances des méthodes SPSC et GCV de Rice et Silverman (1991). Plusieurs fonctions f ont été choisies, de même que plusieurs structures de bruit, et on a fait varier les dimensions n et p des ensembles de données simulés. Il ressort de ces simulations que les courbes des scores qui définissent les deux méthodes (en tant que fonctions de λ) ont une forme très semblable, qui s'apparente aussi à celle de la MISE. En termes d'erreur L_2 d'estimation, les deux méthodes GCV et SPSC ont des performances similaires. Elles sont particulièrement efficaces, en comparaison du choix de λ qui minimise l'erreur L_2 , dans le cas où les erreurs sont issues d'un processus aléatoire à covariance continue (en l'occurrence un processus de Wiener).

Chapitre 4

Dans le Chapitre 4, nous nous intéressons à un thème récent de la statistique qui est l'estimation adaptative dans la régression non paramétrique sur des données corrélées. Étant donné la nouveauté et la difficulté d'un tel sujet, nous nous limitons à proposer une extension simple de deux méthodes d'estimation qui sont le lissage spline adaptatif (Pintore *et al.*, 2006) et la méthode des P-splines adaptatives (Ruppert et Carroll, 2000). L'extension réalisée consiste à passer dans le modèle (1) d'erreurs de type bruit blanc à des erreurs ayant une structure de covariance paramétrique. L'estimation de la régression se ramène alors à la double estimation de la fonction de lissage adaptative et des paramètres de corrélation des erreurs. Nous étudions ensuite ces deux estimateurs adaptatifs lors de simulations et nous comparons leurs performances à celles d'estimateurs splines utilisant un paramètre de lissage global (*cf.* les méthodes GML, GCV, UBR pour données corrélées du Chapitre 3) en termes d'erreur moyenne d'estimation de la fonction de régression aux points d'observation et des paramètres de corrélation du bruit. Les résultats encourageants obtenus nous permettent d'envisager des recherches ultérieures portant aussi bien sur les applications (implémentation des estimateurs sous forme de package logiciel, applications sur données réelles ...) que sur les propriétés théoriques des estimateurs adaptatifs.

Chapitre 1

Régression par splines de lissage sur des mesures répétées

Ce chapitre porte sur l'estimation d'une régression par la méthode des splines de lissage dans un modèle de mesures répétées bruitées. Son résultat principal consiste en un ensemble de majorations de l'erreur quadratique moyenne de l'estimateur spline de lissage qui sont valables pour de nombreuses structures de bruit. Ces majorations sont appliquées à l'obtention de vitesses de convergence qui étendent les résultats analogues de la littérature établis pour d'autres estimateurs. Un autre résultat important de ce chapitre est le calcul numérique de la norme L_2 d'une spline cubique naturelle, qui permet en particulier de déterminer la variance intégrée d'une spline cubique de lissage (voir aussi Degras et Jallet, 2005). Ce calcul est employé ici pour la majoration numérique de l'erreur d'estimation, et il sera utilisé au Chapitre 3 dans une méthode de sélection du paramètre de lissage.

En Section 1.1, nous présentons le modèle de régression non paramétrique pour mesures répétées (1.1), qui est typiquement utilisé pour le traitement de données longitudinales et de courbes dose-réponse. Quelques applications du modèle (1.1) sont fournies par Besse *et al.* (1997) en pluviométrie, par Müller (1988) dans l'étude de la croissance humaine, et par Hart et Wehrly (1986) en biostatistique. Ce modèle, dans sa généralité, peut s'accommoder de données manquantes ou de mesures faites à des temps différents, mais il englobe aussi le modèle classique de régression (1) présenté dans l'Introduction générale. La construction de l'estimateur spline de lissage \hat{f}_λ de la fonction de régression f pour un $\lambda > 0$ donné fait l'objet de la Proposition 1.1. On notera que cette construction ne tient pas compte de la covariance des erreurs de mesure, par opposition au Chapitre 3.

En Section 1.2, l'erreur quadratique moyenne de \hat{f}_λ est majorée dans les Théorèmes 1.2 à 1.5. Les résultats sont présentés séparément suivant le disposi-

tif expérimental considéré (modèle équilibré (1.4) ou déséquilibré (1.1)) et la mesure d'erreur utilisée (discrétisée, MASE, ou intégrée, MISE). Ils étendent les majorations asymptotiques de Craven et Wahba (1979) et de Ragozin (1983) dans le cas d'un bruit blanc, en améliorant les constantes de la majoration du biais. Une extension à la majoration de l'erreur dans l'estimation des dérivées de f est fournie dans le Théorème 1.6.

Dans la Section 1.2.4 on s'intéresse à l'estimateur spline cubique de lissage, qui est largement utilisé dans la pratique statistique. À l'aide d'une représentation particulière des splines cubiques naturelles, la variance intégrée de l'estimateur est calculée sous forme explicite (Théorème 1.7 et Corollaire 1.1) et des constantes numériques sont fournies pour la majoration de la MISE dans le cas d'une répartition équidistante des points d'observation (Théorème 1.8). Une partie de ces résultats a fait l'objet d'une publication (Degras et Jallet, 2005). D'autre part, une application du calcul de la variance intégrée de l'estimateur à la sélection de paramètre de lissage sera présentée au Chapitre 3.

La Section 1.3 applique les Théorèmes 1.2 à 1.5 à l'obtention de vitesses de convergence pour l'estimateur \hat{f}_λ sous plusieurs hypothèses de bruit et de dispositif expérimental : corrélation de courte portée, processus ARMA, condition de mélange, processus à temps continu, et données longitudinales.

Dans les deux premiers cas, en supposant que f appartient à $W_2^m[0, 1]$ et qu'une seule unité statistique est observée ($n=1$), on constate que la vitesse de convergence de l'estimateur \hat{f}_λ d'ordre $2m$ ne change pas par rapport à la situation d'un bruit blanc : elle est de l'ordre de $p^{-2m/(2m+1)}$ quand le nombre p de points d'observation tend vers l'infini. Ce fait est en accord avec les résultats de Altman (1990) et de Hall et Hart (1990) pour les estimateurs à noyau.

Nous indiquons ensuite les limites de notre travail dans le cas d'une corrélation de longue portée d'indice $\gamma \in]0, 1[$, et nous proposons une méthode alternative pour majorer la variance de l'estimateur. On retrouve alors la vitesse en $p^{-2m\gamma/(2m+\gamma)}$ établie par Hall et Hart (1990) dans un dispositif expérimental équidistant pour l'estimateur à noyau d'une fonction de régression à m dérivées bornées. Pour comparaison, on pourra aussi consulter les travaux de Wang (1996), Yang (1997), et Donoho et Johnstone (1998).

Dans le cas d'un bruit mélangeant, en se basant sur le résultat de Burman (1991) pour les splines de régression, on montre qu'il y a un ralentissement de la convergence quantifié par la vitesse de décroissance des coefficients de mélange. Cette observation a été confirmée indépendamment par Sarda et Vieu (1986) pour l'estimateur à noyau.

L'étude de la régression en présence d'un bruit de type processus aléatoire à temps continu fera l'objet du Chapitre 2, où on trouvera une bibliographie détaillée. Dans ce chapitre, nous donnons simplement un résultat préliminaire, à savoir que la convergence de l'estimateur spline \hat{f}_λ d'ordre $2m$ est en $\mathcal{O}(\lambda) + \mathcal{O}(n^{-1})$ pour la

MASE, avec le terme supplémentaire $\mathcal{O}(p^{-2m})$ pour la MISE, où n est le nombre d'unités statistiques du modèle (1.4).

Le dernier exemple traité dans la Section 1.3 est celui de données longitudinales, qui proviennent typiquement d'un grand nombre d'individus chacun observé un petit nombre de fois. On étudie donc l'asymptotique $n \rightarrow \infty$ avec $\sup_{1 \leq i \leq n} p_i = \mathcal{O}(1)$, p_i désignant le nombre de mesures faites sur l'individu i . On trouve, suivant le dispositif expérimental (ensemble fini ou infini de points d'observation), une convergence optimale de l'estimateur \hat{f}_λ d'ordre $2m$ en $\mathcal{O}(n^{-1})$ ou en $\mathcal{O}(n^{-2m/(2m+1)})$.

Les démonstrations des résultats de ce chapitre sont établies en Section 1.4. Nous en donnons ici un bref aperçu. Pour chacun des Théorèmes 1.2–1.6 et 1.8, on décompose l'erreur quadratique moyenne de l'estimateur en somme du biais au carré et de la variance. La majoration du biais, qui ne dépend pas de la partie aléatoire du modèle, peut être effectuée en reprenant les majorations de Craven et Wahba (1979) et Ragozin (1983) avec des modifications mineures. Le contrôle de la variance se fait par une adaptation des résultats valables dans le cadre d'erreurs décorréelées de variance commune σ^2 . La stratégie adoptée consiste à séparer, dans le terme de variance, l'influence du lissage de celle de la covariance des erreurs. Les techniques utilisées dans les démonstrations proviennent essentiellement du calcul matriciel (Lemme 1.3, théorème min-max de Courant-Fischer, théorème des disques de Gersgorin, *etc.*).

1.1 Modèle et estimateur

Soit le modèle

$$y_{ij} = f(t_{ij}) + \varepsilon_{ij}, \quad 1 \leq i \leq n, \quad 1 \leq j \leq p_i, \quad (1.1)$$

dans lequel les variables aléatoires y_{ij} sont observées en des points déterministes t_{ij} , f est une fonction de régression définie sur $[0,1]$, et les ε_{ij} sont des erreurs aléatoires centrées. L'indice i correspond à une unité statistique (par exemple un individu, une usine, ou un actif financier) et l'indice j est un indice de répétition des mesures à différentes valeurs d'une variable explicative $t \in [0, 1]$.

On suppose que la fonction f appartient à un espace de Sobolev à préciser selon le contexte. On suppose aussi que pour tout $1 \leq i \leq n$, les points $t_{ij}, 1 \leq j \leq p_i$, ont une structure de tableau triangulaire (rigoureusement, ces points devraient être écrits sous la forme t_{ijp_i} , ce que nous ne faisons pas par souci de lisibilité) et que ces p_i points sont distincts. Concernant la structure de covariance du modèle, nous faisons l'hypothèse suivante.

Hypothèse 1.1. *Les mesures y_{ij} sont indépendantes à travers les unités statistiques $1 \leq i \leq n$, et la covariance entre deux mesures en des points fixés t et t' ne dépend pas de l'unité statistique sur laquelle elles sont faites, i.e. si $t_{ij} = t_{i'j'}$ et $t_{ik} = t_{i'k'}$, alors $\text{Cov}(y_{ij}, y_{ik}) = \text{Cov}(y_{i'j'}, y_{i'k'})$.*

L'Hypothèse 1.1 assure que les mesures faites sur chaque unité statistique ont intrinsèquement la même structure de covariance. Ainsi, sous cette hypothèse et pour un dispositif expérimental équilibré ($p_1 = \dots = p_n := p$ et $t_{1j} = \dots = t_{nj}$ pour tout $1 \leq j \leq p$), les vecteurs aléatoires $(y_{i1}, \dots, y_{ip_1})'$, $1 \leq i \leq n$, ont tous la même espérance et la même matrice de covariance. Notons aussi que le modèle (1.1) contient le modèle (1) de l'Introduction générale (cas $n = 1$) et que dans cette perspective plus classique, l'Hypothèse 1.1 nous permet de modéliser librement toutes les structures d'erreur.

Exemple 1.1. *Si les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p_i$, s'écrivent sous la forme $\varepsilon_i(t_{ij})$, où les $\varepsilon_1, \dots, \varepsilon_n$ sont des copies indépendantes d'un processus aléatoire du second ordre $\varepsilon = \{\varepsilon(t), t \in [0, 1]\}$ centré, alors l'Hypothèse 1.1 est vérifiée.*

Notations

Afin de construire et d'étudier l'estimateur spline de lissage du modèle (1.1), quelques notations sont nécessaires. On réécrit d'abord le dispositif expérimental $\{t_{ij}, 1 \leq i \leq n, 1 \leq j \leq p_i\}$ sous forme d'un ensemble $\{t_1 < \dots < t_p\}$ de points distincts et ordonnés de $[0, 1]$. On pose ensuite $t_0 = 0$, $t_{p+1} = 1$, $h_j = t_{j+1} - t_j$ pour $0 \leq j \leq p$, et on note $h_{max} = \max_{0 \leq j \leq p} h_j$ l'écart maximal et $h_{min} = \min_{1 \leq j \leq p-1} h_j$ l'écart minimal du dispositif expérimental. Soit l'ensemble $R_j = \{(i, k) : t_{ik} = t_j\}$ et soit $n_j = \text{Card}(R_j)$ le nombre d'observations au point t_j pour $1 \leq j \leq p$. Notons $N = \sum_{i=1}^n p_i = \sum_{j=1}^p n_j$ le nombre total d'observations et soit $\mathbf{W} = \text{diag}(n_j p / N)$ la matrice $p \times p$ proportionnelle aux fréquences des mesures aux points t_1, \dots, t_p ($\mathbf{W} = \mathbf{I}$ pour un dispositif équilibré).

Soit $\bar{y}_j = n_j^{-1} \sum_{(i,k) \in R_j} y_{ik}$ la moyenne des mesures au point t_j pour $1 \leq j \leq p$, et soit $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_p)'$. Pour $1 \leq j, k \leq p$, posons $v_{jk} = \text{Cov}(\varepsilon_{il}, \varepsilon_{i'l'})$ s'il existe un individu i faisant à la fois l'objet de mesures au point $t_{il} = t_j$ et au point $t_{i'l'} = t_k$ (dans ce cas, v_{jk} ne dépend pas de i, l, l' , d'après l'Hypothèse 1.1) et $v_{jk} = 0$ sinon. Définissons alors la matrice $\mathbf{V} = (v_{jk})_{1 \leq j, k \leq p}$.

Pour tout réel $\lambda > 0$, la matrice $\mathbf{W} + \lambda p \mathbf{\Gamma}$ est définie positive car la matrice \mathbf{W} est diagonale à coefficients strictement positifs, et la matrice d'énergie $\mathbf{\Gamma}$ du Théorème ?? est semi-définie positive. Nous pouvons ainsi définir la matrice de lissage $\mathbf{S}_\lambda = (\mathbf{W} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{W}$, qui n'est pas symétrique sauf dans le cas d'un dispositif équilibré.

Rappelons enfin que pour toute fonction g définie sur $[0, 1]$, on note \mathbf{g} le vecteur $(g(t_1), \dots, g(t_p))'$ de ses valeurs aux points d'observation du modèle.

Estimateur spline de lissage

Dans le modèle (1.1), étant donné un entier $m \geq 0$ et un réel $\lambda > 0$, le problème variationnel qui définit l'estimateur spline de lissage se formule ainsi (cf. la relation (4) de l'Introduction Générale) : Trouver $g \in W_2^m[0, 1]$ qui minimise la quantité

$$\sum_{j=1}^p \frac{n_j}{N} (\bar{y}_j - g(t_j))^2 + \lambda \int_0^1 (g^{(m)}(t))^2 dt. \quad (1.2)$$

Remarque 1.1. *L'estimation de f par lissage spline à partir des n courbes du modèle (1.1) observées aux points t_{ij} se ramène au lissage d'une seule courbe moyenne, avec une pondération liée à la fréquence des mesures en chaque point t_j . Dans le cas où le dispositif expérimental est équilibré (en particulier si $n = 1$), le problème (1.2) mène à la construction classique des splines de lissage sans pondération, comme dans Schoenberg (1964b) ou dans Craven et Wahba (1979).*

Afin d'explicitier la solution du problème de lissage (1.2), nous évoquons le problème analogue d'interpolation contrainte.

Théorème 1.1. (Schoenberg, 1964a) *Soient $p \geq m > 0$ deux entiers, soit $\mathbf{f} = (f_1, \dots, f_p)'$ un vecteur de \mathbb{R}^p et soient t_1, \dots, t_p des points distincts de $[0, 1]$. Le problème d'interpolation contrainte qui consiste à trouver $g \in W_2^m[0, 1]$ telle que*

$$\begin{cases} g(t_i) = f_i, 1 \leq i \leq p, \\ (\forall h \in W_2^m[0, 1]) (h(t_i) = f_i, 1 \leq i \leq p) \Rightarrow (|h|_m^2 \geq |g|_m^2), \end{cases}$$

admet pour unique solution la spline naturelle $S_{0,p}(\mathbf{f})$ d'ordre $2m$ ayant pour noeuds t_1, \dots, t_p et qui interpole les valeurs f_1, \dots, f_p en ces points. Cette fonction est linéaire par rapport à \mathbf{f} et l'application $\mathbf{f} \mapsto |S_{0,p}(\mathbf{f})|_m^2$ est une forme quadratique. On appelle matrice d'énergie et on note $\mathbf{\Gamma}$ la matrice semi-définie positive associée à cette forme quadratique. Si \mathbf{f} représente les valeurs $f(t_1), \dots, f(t_p)$ d'une fonction $f \in W_2^m[0, 1]$, alors

$$\mathbf{f}'\mathbf{\Gamma}\mathbf{f} \leq |f|_m^2 \quad (1.3)$$

avec égalité si et seulement si f et $S_{0,p}(\mathbf{f})$ sont égales.

À l'aide du Théorème 1.1 et de la matrice $\mathbf{\Gamma}$ également définie dans Nussbaum (1985) et Utreras (1988), nous pouvons maintenant construire de l'estimateur spline de lissage. En effet, en appliquant ce théorème au vecteur de données $\bar{\mathbf{y}}$, nous voyons que si la solution de (1.2) existe et si $m \leq p$, alors cette solution est unique et c'est une spline naturelle d'ordre $2m$ ayant pour noeuds les points t_1, \dots, t_p . De plus, nous pouvons à présent formuler le problème (1.2) en termes matriciels : Trouver $\mathbf{g} \in \mathbb{R}^p$ qui minimise

$$\frac{1}{p} (\mathbf{g} - \bar{\mathbf{y}})' \mathbf{W} (\mathbf{g} - \bar{\mathbf{y}}) + \lambda \mathbf{g}'\mathbf{\Gamma}\mathbf{g},$$

où le vecteur \mathbf{g} contient les valeurs d'une spline naturelle g d'ordre $2m$ en ses noeuds t_1, \dots, t_p . En égalant la dérivée de la fonction $\mathbf{g} \mapsto p^{-1}(\mathbf{g} - \bar{\mathbf{y}})' \mathbf{W}(\mathbf{g} - \bar{\mathbf{y}}) + \lambda \mathbf{g}' \mathbf{\Gamma} \mathbf{g}$ à zéro, on obtient l'équation $2(\mathbf{W} + \lambda p \mathbf{\Gamma}) \mathbf{g} - 2 \mathbf{W} \bar{\mathbf{y}} = 0$ qui admet une unique solution $\mathbf{g} = \mathbf{S}_\lambda \bar{\mathbf{y}}$ correspondant à un minimum. Nous pouvons alors conclure.

Proposition 1.1. *Soient m un entier et $\lambda > 0$ un réel donnés. Si $m \leq p$, alors le problème variationnel (1.2) admet une unique solution \hat{f}_λ qui est une spline naturelle d'ordre $2m$ ayant ses noeuds aux points d'observations t_1, \dots, t_p . Cette fonction est déterminée de manière unique par la relation $\hat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \bar{\mathbf{y}} = (\mathbf{W} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{W} \bar{\mathbf{y}}$.*

1.2 Majoration de l'erreur quadratique moyenne

Dans cette section, l'erreur quadratique moyenne de l'estimateur spline de lissage est majorée dans différents contextes (dispositif expérimental équilibré, (1.4), ou déséquilibré, (1.1); mesure d'erreur discrétisée, MASE, ou intégrée, MISE; estimation de la fonction de régression ou de ses dérivées) de manière à couvrir un large éventail de situations. Chaque théorème présenté comporte deux inégalités correspondant à deux majorations complémentaires de la variance, chacune ayant son utilité propre suivant le type de bruit du modèle considéré (voir à ce sujet les applications de la Section 1.3). On note par ailleurs que les majorations sont établies pour un ensemble *fini* de données. Ceci permet ensuite d'envisager plusieurs asymptotiques portant sur les nombres p_1, \dots, p_n de points d'observations et/ou sur le nombre n de courbes observées.

1.2.1 Dispositif expérimental équilibré

Nous étudions ici le modèle de régressions simultanées (1.1) dans la situation où les observations sont équilibrées. Ce modèle s'écrit alors

$$y_{ij} = f(t_j) + \varepsilon_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq p. \quad (1.4)$$

Dans ce cas, l'Hypothèse 1.1 permet une plus grande liberté pour modéliser la covariance des erreurs ε_{ij} que dans le cas déséquilibré. En effet, celle-ci peut alors dépendre des points d'observations, de l'ordre des mesures et de la taille du dispositif expérimental. Autrement dit, la quantité $\text{Cov}(\varepsilon_{ij}, \varepsilon_{ik})$ peut dépendre de t_{ij} et t_{ik} , comme dans le cas déséquilibré, mais aussi de j, k et de p .

Exemple 1.2 (série temporelle). *On suppose dans le modèle (1.4) que pour tout $1 \leq j \leq p$, $(\varepsilon_{1j}, \dots, \varepsilon_{nj})$ est un échantillon i.i.d. de ε_j , où $(\varepsilon_j)_{j \in \mathbb{N}^*}$ est un processus aléatoire centré du second ordre. Alors l'Hypothèse 1.1 est vérifiée.*

Pour établir les majorations des erreurs quadratiques moyennes de l'estimateur spline de lissage \widehat{f}_λ , nous avons besoin d'un résultat sur sa matrice de lissage, déduit du Théorème 5.3 de Utreras (1988). Notons que le même résultat figure dans Craven et Wahba (1979) et Ragozin (1983) sans démonstration complète.

Proposition 1.2. *Dans le modèle équilibré (1.4), pour un entier $m > 0$ donné, il existe des constantes $M_k = M_k(m, h_{\max}/h_{\min})$ vérifiant $M_k \geq m$ pour $k = 1, 2$, telles que pour tout réel $0 < \lambda \leq 1$, la matrice de lissage \mathbf{S}_λ associée à l'estimateur spline \widehat{f}_λ d'ordre $2m$ vérifie*

$$\mathrm{tr} \mathbf{S}_\lambda^k \leq \frac{M_k}{\lambda^{1/2m}}.$$

Nous présentons maintenant deux résultats sur la MASE et la MISE de \widehat{f}_λ , en rappelant auparavant les notations $\mathbf{f} = (f(t_1), \dots, f(t_p))'$, $\widehat{\mathbf{f}}_\lambda = (\widehat{f}_\lambda(t_1), \dots, \widehat{f}_\lambda(t_p))'$, et $\mathbf{V} = (\mathrm{Cov}(\varepsilon_{1j}, \varepsilon_{1k}))_{1 \leq j, k \leq p}$ de la Section 1.1, ainsi que la notation $|g|_k^2 = \int_0^1 (g^{(k)}(t))^2 dt$ valable pour toute fonction $g \in W_2^k[0, 1]$.

Théorème 1.2. *Dans le modèle équilibré (1.4), on suppose que $f \in W_2^m[0, 1]$ pour un entier $m \geq 1$ et que les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$, vérifient l'Hypothèse 1.1. Alors pour tout réel $0 < \lambda \leq 1$, la MASE de l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ est telle que*

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{\lambda}{4} |f|_m^2 + \frac{\mathrm{tr} \mathbf{V}}{np} \quad (1.5a)$$

et

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{\lambda}{4} |f|_m^2 + \frac{M_2 \|\mathbf{V}\|}{np\lambda^{1/2m}}, \quad (1.5b)$$

où M_2 est la constante de la Proposition 1.2.

Remarque 1.2. *La constante de majoration du biais au carré présentée ici améliore d'un facteur $1/4$, resp. $1/2$, celle de Craven et Wahba (1979), resp. Ragozin (1983).*

Théorème 1.3. *Dans le modèle équilibré (1.4), on suppose que $f \in W_2^m[0, 1]$ pour un entier $m \geq 1$ et que les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$, vérifient l'Hypothèse 1.1. Alors il existe des constantes $C = C(m, h_{\max}/h_{\min})$, $D = D(m) \geq 1$ telles que pour tout réel $0 < \lambda \leq 1$, la MISE de l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ vérifie les relations*

$$\mathbb{E} |f - \widehat{f}_\lambda|_0^2 \leq (Cph_{\max}\lambda + Dh_{\max}^{2m}) |f|_m^2 + \left(Ch_{\max} + \frac{Dh_{\max}^{2m}}{p\lambda} \right) \frac{\mathrm{tr} \mathbf{V}}{n} \quad (1.6a)$$

et

$$\begin{aligned} \mathbb{E} |f - \widehat{f}_\lambda|_0^2 &\leq (Cph_{\max}\lambda + Dh_{\max}^{2m}) |f|_m^2 \\ &\quad + \left(CM_2h_{\max} + \frac{DM_1h_{\max}^{2m}}{p\lambda} \right) \frac{\|\mathbf{V}\|}{n\lambda^{1/2m}}, \end{aligned} \quad (1.6b)$$

où M_1, M_2 sont les constantes de la Proposition 1.2, et C est un polynôme d'ordre $2m$ évalué en (h_{\max}/h_{\min}) .

Remarque 1.3. Si les erreurs de mesure $\varepsilon_{ij}, 1 \leq i \leq n, 1 \leq j \leq p$, ont une variance uniformément bornée, alors la variance de l'estimateur \widehat{f}_λ est majorée, aux constantes multiplicatives près, par $1/n$ dans les formules (1.5a), (1.6a) et par $\|\mathbf{V}\|/(np\lambda^{1/2m})$ dans (1.5b) et (1.6b). Ainsi les premières formules, relativement rudimentaires, ne sont utiles que dans une asymptotique $n \rightarrow \infty$, tandis que les secondes permettent une majoration plus précise tenant compte du paramètre de lissage, de la structure de covariance des erreurs et du dispositif expérimental. Cependant, nous verrons la complémentarité de ces formules en Section 1.3.

1.2.2 Dispositif expérimental général

Ayant exposé les résultats de majoration de l'erreur quadratique moyenne dans le cadre simple de données équilibrées, nous revenons maintenant au modèle général (1.1). Pour alléger les notations, nous écrirons dans la suite $\min n_j$ et $\max n_j$ à la place de $\min_{1 \leq j \leq p} n_j$ et $\max_{1 \leq j \leq p} n_j$. La proposition et les deux théorèmes suivants sont les analogues de ceux de la sous-section précédente.

Proposition 1.3. Dans le modèle général (1.1), il existe des constantes $\widetilde{M}_1 = \widetilde{M}_1(m, h_{\max}/h_{\min})$ et $\widetilde{M}_2 = \widetilde{M}_2(m, h_{\max}/h_{\min}, \max n_j / \min n_j)$ strictement positives telles que pour tout réel $0 < \lambda \leq 1$, la matrice de lissage de l'estimateur spline \widehat{f}_λ d'ordre $2m$ vérifie les relations

$$\operatorname{tr} \mathbf{W}\mathbf{S}_\lambda \leq \widetilde{M}_1 \frac{(\max n_j)^2 p}{(\min n_j) N \lambda^{1/2m}} \quad \text{et} \quad \operatorname{tr} \mathbf{S}'_\lambda \mathbf{S}_\lambda \leq \widetilde{M}_2 \frac{(\max n_j)^2 p}{(\min n_j) N \lambda^{1/2m}}.$$

Théorème 1.4. Dans le modèle (1.1), on suppose que $f \in W_2^m[0, 1]$ pour un entier $m \geq 1$ et que les erreurs $\varepsilon_{ij}, 1 \leq i \leq n, 1 \leq j \leq p$, vérifient l'Hypothèse 1.1. Alors pour tout réel $0 < \lambda \leq 1$, la MASE de l'estimateur \widehat{f}_λ d'ordre $2m$ vérifie les relations

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{N\lambda}{(\min n_j)p} |f|_m^2 + \frac{\operatorname{tr} \mathbf{V}}{(\min n_j)p} \quad (1.7a)$$

et

$$\frac{1}{p} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 \leq \frac{N\lambda}{(\min n_j)p} |f|_m^2 + \widetilde{M}_2 \frac{(\max n_j)^2 \|\mathbf{V}_{\text{abs}}\|}{(\min n_j)^2 N \lambda^{1/2m}}, \quad (1.7b)$$

où \widetilde{M}_2 est la constante de la Proposition 1.3 et \mathbf{V}_{abs} est la matrice $(|v_{jk}|)_{1 \leq j, k \leq p}$.

Théorème 1.5. *Dans le modèle (1.1), on suppose que $f \in W_2^m[0, 1]$ pour un entier $m \geq 1$ et que les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p_i$, vérifient l'Hypothèse 1.1. Alors pour tout réel $0 < \lambda \leq 1$, la MISE de l'estimateur \widehat{f}_λ d'ordre $2m$ est telle que*

$$\mathbb{E}|f - \widehat{f}_\lambda|_0^2 \leq \left(\frac{CNh_{\max}\lambda}{\min n_j} + Dh_{\max}^{2m} \right) |f|_m^2 + \left(Ch_{\max} + \frac{Dh_{\max}^{2m}}{p\lambda} \right) \frac{\text{tr } \mathbf{V}}{\min n_j} \quad (1.8a)$$

et

$$\begin{aligned} \mathbb{E}|f - \widehat{f}_\lambda|_0^2 &\leq \left(\frac{CNh_{\max}\lambda}{\min n_j} + Dh_{\max}^{2m} \right) |f|_m^2 \\ &\quad + \left(C\widetilde{M}_2h_{\max} + \frac{D\widetilde{M}_1h_{\max}^{2m}}{p\lambda} \right) \frac{(\max n_j)^2 p \|\mathbf{V}_{\text{abs}}\|}{(\min n_j)^2 N\lambda^{1/2m}}, \end{aligned} \quad (1.8b)$$

où C, D sont les constantes du Théorème 1.3, $\widetilde{M}_1, \widetilde{M}_2$ celles de la Proposition 1.3, et \mathbf{V}_{abs} est la matrice $(|v_{jk}|)_{1 \leq j, k \leq p}$.

Remarque 1.4. *Dans les Théorèmes 1.4 et 1.5, la matrice \mathbf{V}_{abs} apparaît à la place de la matrice \mathbf{V} dans l'étude de la variance de \widehat{f}_λ . Ceci est dû à la difficulté technique à séparer l'influence du dispositif expérimental de celle de la structure de covariance des ε_{ij} . Cependant, pour de nombreux types de bruit (processus de Wiener, pont brownien, etc.), la covariance de deux erreurs est positive et donc $\mathbf{V}_{\text{abs}} = \mathbf{V}$.*

1.2.3 Estimation des dérivées de la fonction de régression

Les résultats des Sections 1.2.1 et 1.2.2 peuvent être étendus dans deux directions. D'une part, on peut envisager l'estimation des dérivées d'ordre $j < k$ d'une fonction de régression $f \in W^k[0, 1]$ par les dérivées correspondantes d'une spline de lissage \widehat{f}_λ . D'autre part, on peut changer l'ordre $2m$ de la spline de lissage utilisée en prenant $m \geq k > 0$, ce qui généralise le choix $m = k$ des sections précédentes. Nous donnons ainsi une généralisation du Théorème 1.3 (MISE avec données équilibrées) basée sur le Théorème 2 de Ragozin (1983).

Théorème 1.6. *Dans le modèle (1.4), on suppose que $f \in W^k[0, 1]$ pour un entier $k > 0$ et que l'Hypothèse 1.1 sur les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$, est valide. Étant donné un entier $m \geq k$, on considère l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p)$. On suppose que $h_{\max}^{2m} = \mathcal{O}(\lambda)$ lorsque $n, p \rightarrow \infty$. Alors la MISE de la j -ème dérivée ($j < k$) de \widehat{f}_λ est telle que*

$$\mathbb{E}|f - \widehat{f}_\lambda|_j^2 \leq P(\lambda + h_{\max}^{2m})^{(k-j)/m} |f|_k^2 + \frac{Q \text{tr } \mathbf{V}}{np\lambda^{j/m}} \quad (1.9a)$$

et

$$\mathbb{E}|f - \widehat{f}_\lambda|_j^2 \leq P(\lambda + h_{max}^{2m})^{(k-j)/m} |f|_k^2 + \frac{Q' \|\mathbf{V}\|}{np\lambda^{(2j+1)/2m}} \quad (1.9b)$$

pour des constantes P, Q, Q' indépendantes de n et p .

Remarque 1.5. La vitesse de convergence de la MISE de $\widehat{f}_\lambda^{(j)}$ ne dépend pas de l'ordre $2m$ de \widehat{f}_λ : dans (1.9b) par exemple, sous la condition $h_{max} = \mathcal{O}(h_{min})$ lorsque $p \rightarrow \infty$, le choix optimal du paramètre $\lambda \sim (\|\mathbf{V}\|/(np))^{2m/(2k+1)}$ mène à la vitesse $\mathcal{O}((\|\mathbf{V}\|/(np))^{(k-j)/(2k+1)})$ lorsque $n, p \rightarrow \infty$.

1.2.4 Résultats numériques pour les splines cubiques

Les fonctions splines cubiques naturelles admettent une représentation dite par valeur/dérivée seconde qui permet d'exprimer la valeur d'une spline cubique naturelle en tout point t de $[0,1]$ à l'aide des seules valeurs de cette fonction et de sa dérivée seconde en ses noeuds (voir le rappel en Annexe A ou Green et Silverman (1994) pour une étude plus exhaustive). De ce fait, la variance intégrée sur $[0,1]$ de l'estimateur peut s'écrire simplement en fonction de la matrice de lissage \mathbf{S}_λ , d'une matrice Φ liée au dispositif expérimental, et de la matrice \mathbf{V} de covariance.

Théorème 1.7. Soient $0 \leq t_1 < \dots < t_p \leq 1$ des points fixés. Alors il existe une matrice définie positive Φ de taille $p \times p$, ne dépendant que de t_1, \dots, t_p , telle que pour toute spline cubique naturelle g ayant ses noeuds aux points t_1, \dots, t_p ,

$$|g|_0^2 = \mathbf{g}' \Phi \mathbf{g},$$

avec la notation $\mathbf{g} = (g(t_1), \dots, g(t_p))'$.

Corollaire 1.1. Dans le modèle équilibré (1.4), sous l'Hypothèse 1.1 sur la covariance des erreurs ε_{ij} , $1 \leq i \leq n, 1 \leq j \leq p$, la variance intégrée de l'estimateur spline cubique de lissage \widehat{f}_λ s'écrit, pour tout $\lambda \geq 0$,

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 = n^{-1} \text{tr}(\mathbf{S}_\lambda \Phi \mathbf{S}_\lambda \mathbf{V}).$$

Nous nous plaçons maintenant dans un dispositif expérimental équidistant qui permet d'utiliser les résultats de Nussbaum (1985) sur l'encadrement des valeurs propres de la matrice d'énergie Γ . À l'aide de ces résultats, nous établissons un lemme d'intérêt indépendant sur la matrice de lissage de l'estimateur spline d'ordre $2m$ valable dès que $p \geq m + 5$.

Lemme 1.1. Dans le modèle équilibré (1.4), on suppose que l'Hypothèse 1.1 est vérifiée et que $t_j = (j-1)/(p-1)$, $1 \leq j \leq p$. On pose $C_1 = (1 + \sum_{l>0} l^{-2m})^{-1} \pi^{2m}$.

Alors pour tout $\lambda > 0$, la matrice de lissage de l'estimateur spline \widehat{f}_λ d'ordre $2m$ vérifie

$$\text{tr } \mathbf{S}_\lambda^2 \leq m + 2 + \frac{1}{(C_1\lambda)^{1/2m}} \int_0^\infty (1 + t^{2m})^{-2} dt.$$

Remarque 1.6. Le Lemme 3.4 de Nussbaum (1985) montre que les valeurs propres $0 = \lambda_1 = \dots = \lambda_m < \lambda_{m+1} < \dots < \lambda_p$ de la matrice d'énergie $\mathbf{\Gamma}$ sont voisines de $p^{-1}(\pi j)^{2m}$ pour $m + 1 \leq j \leq p$. Dans ce sens, la constante C_1 qui minore uniformément par rapport à p les valeurs propres non nulles de $p\mathbf{\Gamma}$ est optimale.

Revenant à l'estimateur spline cubique de lissage, on donne une majoration numérique de la norme de la matrice $\mathbf{\Phi}$ basée sur les propriétés élémentaires de la norme matricielle et sur le théorème des disques de Gersgögin.

Lemme 1.2. Dans le modèle équilibré (1.4), en supposant que $t_j = (j - 1)/(p - 1)$ pour tout $1 \leq j \leq p$, on a $\|\mathbf{\Phi}\| \leq 4,21660(p - 1)^{-1}$. Dans le cas d'une répartition non équidistante, $\|\mathbf{\Phi}\|$ est majorée par le produit de l'écart maximal h_{\max} et d'un polynôme de degré 4 évalué en h_{\max}/h_{\min} .

A l'aide du corollaire et des lemmes précédents, qui servent à majorer la variance intégrée de l'estimateur, et avec la majoration du biais du Théorème 1.3, nous obtenons le résultat suivant.

Théorème 1.8. Dans le modèle équilibré (1.4), on suppose que $p \geq 7$, que $f \in W_2^2[0, 1]$, que l'Hypothèse 1.1 est vérifiée, et que $t_j = (j - 1)/(p - 1)$, $1 \leq j \leq p$. Alors pour tout $\lambda > 0$, la MISE de l'estimateur spline cubique de lissage \widehat{f}_λ vérifie

$$\mathbb{E}|f - \widehat{f}_\lambda|_0^2 \leq \left\{ \frac{17p\lambda}{p-1} + \frac{64}{(p-1)^4} \right\} |f''|_0^2 + \frac{4,21660}{n(p-1)} \text{tr } \mathbf{V} \quad (1.10a)$$

et

$$\mathbb{E}|f - \widehat{f}_\lambda|_0^2 \leq \left\{ \frac{17p\lambda}{p-1} + \frac{64}{(p-1)^4} \right\} |f''|_0^2 + \frac{4,21660}{n(p-1)} \left(4 + \frac{0,31853}{\lambda^{1/4}} \right) \|\mathbf{V}\| \quad (1.10b)$$

Le Théorème 1.8 fournit une majoration de la variance de l'estimateur spline cubique de lissage nettement meilleure que celle du Théorème 1.3 dans le cas $m = 2$. En effet, dans la relation (1.6a) la variance est majorée, au facteur $\text{tr } \mathbf{V}/(n(p - 1))$ près, par $17 + 64/\lambda(p - 1)^4$, contre seulement 4,21660 dans (1.10a). On trouve la même amélioration en passant de (1.6b) à (1.10b).

Cependant, ces majorations de la MISE ne sont efficaces que dans la mesure où l'on sait estimer approximativement les quantités $\|\mathbf{V}\|$ ou $\text{tr } \mathbf{V}$ d'une part, et $|f|_2^2$ d'autre part. L'estimation de la matrice de covariance peut se faire de façon paramétrique, cf. Wang (1998), ou non paramétrique, pour n grand, à l'aide de la matrice de covariance empirique. Par ailleurs on peut estimer la semi-norme $|f|_2^2$ par $|\widehat{f}_\lambda|_2^2$, mais dans ce cas le Théorème 1.6 ne permet pas d'évaluer l'erreur d'estimation.

1.3 Application : vitesses de convergence

Nous appliquons ici les résultats de la section précédente à l'obtention de vitesses de convergence pour l'estimateur spline de lissage des modèles (1.1) et (1.4) sous plusieurs hypothèses de bruit et d'asymptotique.

Dans le cas d'erreurs de mesure asymptotiquement décorréelées (*i.e.* telles que pour tout i , $\text{Cov}(\varepsilon_{ij}, \varepsilon_{ik}) \rightarrow 0$ lorsque $|j-k| \rightarrow \infty$, *cf.* Introduction générale), on peut obtenir un estimateur consistant avec $n = 1$ lorsque $p \rightarrow \infty$ dans le modèle (1.4). Ceci n'est pas possible en présence de corrélation persistante (typiquement avec un bruit de type processus aléatoire). Il faut dans cette situation imposer $n \rightarrow \infty$ pour faire converger la variance de l'estimateur (voir le Chapitre 2 et Hart et Wehrly, 1986) et aussi $p \rightarrow \infty$ pour faire converger le biais au carré sur $[0, 1]$ dans le modèle (1.4). Dans le cas de données longitudinales typiques, on dispose d'un grand nombre d'individus faisant chacun l'objet d'un petit nombre de mesures, ce qui mène à considérer l'asymptotique $n \rightarrow \infty$ avec $\sup_{1 \leq i \leq n} p_i = \mathcal{O}(1)$ dans le modèle (1.1).

Les trois situations précédentes présentent des structures d'erreur et des asymptotiques différentes et pour chacune d'entre elles, on doit utiliser la formule adéquate (1.Xa) ou (1.Xb) d'un des Théorèmes 1.1 – 1.5 afin d'obtenir une bonne vitesse de convergence. C'est la raison pour laquelle ces théorèmes comportent deux majorations différentes.

Pour étudier la convergence de l'estimateur, nous imposons au dispositif expérimental la condition de quasi-uniformité, que nous définissons ici en notant explicitement la dépendance des points d'observations t_j par rapport à p .

Définition 1.1. (Quasi-uniformité) Soit $(x_{jp})_{1 \leq j \leq p, p \in \mathbb{N}^*}$ un tableau triangulaire de points de $[0, 1]$ tel que $x_{1p} < \dots < x_{pp}$ pour tout $p \geq 1$. Pour tout $p \geq 2$, posons $h_{max}(p) = \max_{0 \leq j \leq p} \{x_{(j+1)p} - x_{jp}\}$ et $h_{min}(p) = \min_{1 \leq j \leq p-1} \{x_{(j+1)p} - x_{jp}\}$, avec $x_{0p} = 0$ et $x_{(p+1)p} = 1$. On dit alors que la répartition des points x_{jp} est quasi-uniforme s'il existe une constante $r < +\infty$ telle que pour tout $p \geq 2$,

$$\frac{h_{max}(p)}{h_{min}(p)} \leq r.$$

Par abus de langage, nous dirons d'un dispositif expérimental qu'il est quasi-uniforme au lieu de mentionner le tableau triangulaire correspondant. Notons qu'avec les inégalités $h_{max}(p) \geq h_{min}(p)$, $h_{max}(p) \geq 1/(p+1)$ et $h_{min}(p) \leq 1/(p-1)$ valables pour tout dispositif expérimental, la propriété de quasi-uniformité implique que lorsque $p \rightarrow \infty$,

$$h_{max} \sim h_{min} \sim p^{-1}. \quad (1.11)$$

1.3.1 Décorrélation asymptotique

Nous nous plaçons dans le modèle (1.4) avec $n = 1$ et des erreurs de mesure asymptotiquement décorréelées (cf. modèle (1) sous l'hypothèse 2). Nous écrivons

$$y_i = f(x_i) + \varepsilon_i, \quad 1 \leq i \leq p$$

en gardant à l'esprit que les points d'observation $x_i, 1 \leq i \leq p$, ont une structure de tableau triangulaire ($x_i = x_{ip}$) et que la covariance de deux erreurs ε_i et ε_j peut dépendre de p (tableau triangulaire) ou pas (schéma linéaire).

Remarque 1.7. *Le modèle précédent peut être utilisé dans l'analyse non paramétrique de séries temporelles, cf. Hart (1996) ou Opsomer et al. (2001). Plus précisément, étant donné un échantillon (Z_1, \dots, Z_p) d'une série temporelle $(Z_i)_{i \in I}$ indexée par $I = \mathbf{N}$ ou $I = \mathbf{Z}$, on considère l'estimation des valeurs $\mu_i = \mathbb{E}Z_i$, $1 \leq i \leq p$, de la tendance de cette série. En posant $\varepsilon_i = Z_i - \mu_i$ et en faisant l'hypothèse qu'il existe une fonction régulière f telle que $\mu_i = f(x_i)$ pour des points $x_i, 1 \leq i \leq p$, équidistants dans $[0, 1]$, le problème initial se ramène à l'estimation non paramétrique de f .*

Nous supposons que le dispositif expérimental est quasi-uniforme et nous envisageons plusieurs hypothèses sur la nature du bruit.

Hypothèse 1.2. (Corrélation de courte/longue portée) La variance des erreurs est uniformément bornée, i.e. $\sup_{p \geq 1, 1 \leq i \leq p} \mathbb{E}\varepsilon_i^2 < \infty$, et il existe un réel γ (avec $\gamma > 1$ pour la courte portée et $0 < \gamma \leq 1$ pour la longue portée) tel que $\text{Cov}(\varepsilon_i, \varepsilon_j) \sim |i - j|^{-\gamma}$ uniformément en i et j (pour $i \neq j$) quand $p \rightarrow +\infty$.

Hypothèse 1.3. (ARMA) Les erreurs aléatoires $\varepsilon_1, \dots, \varepsilon_p$, sont issues d'un processus ARMA stationnaire inversible $(\varepsilon_i)_{i \in \mathbf{Z}}$. *Le processus est dit inversible si les racines de sa partie AR sont situées en dehors du disque unité, cf. Bosq et Nguyen (1996) pour une définition détaillée.*

Hypothèse 1.4. (Mélangeance au sens de Burman (1991)) Il existe une suite réelle croissante $(R_p)_{p \geq 1}$ et une fonction φ , resp. α , décroissante telles que

i) $R_p \leq p$ et $R_p \rightarrow \infty$ quand $p \rightarrow \infty$.

ii) $\int_0^\infty \varphi^{1/2}(t)dt < \infty$, resp. $\int_0^\infty \alpha^{1/2}(t)dt < \infty$.

iii) Pour tous entiers $1 \leq j < k \leq p$, on a $|P(B|A) - P(B)| \leq \varphi(R_p(t_k - t_j))$, resp. $|P(A \cap B) - P(A)P(B)| \leq \lambda(R_p(t_k - t_j))$ où A et B sont des événements engendrés respectivement par $\varepsilon_1, \dots, \varepsilon_j$ et $\varepsilon_k, \dots, \varepsilon_p$.

iv) $\sup_{p \geq 1, 1 \leq i \leq p} \mathbb{E}\varepsilon_i^2 < \infty$, resp. $\sup_{p \geq 1, 1 \leq i \leq p} \mathbb{E}\varepsilon_i^4 < \infty$.

Remarque 1.8. *La définition précédente de la mélangeance, adaptée au cadre de la régression, permet de faire varier la dépendance entre deux mesures en fonction de la distance entre les points d'observations d'une part, et en fonction de la taille de l'ensemble de données d'autre part. Burman (1991) a montré que lorsque $p \rightarrow \infty$, la covariance entre deux mesures faites en des points $0 \leq s \neq t \leq 1$ fixés diminue comme $\mathcal{O}(\varphi^{1/2}(|s-t|R_p))$ dans le cas φ -mélangeant, et comme $\mathcal{O}(\lambda^{1/2}(|s-t|R_p))$ dans le cas λ -mélangeant.*

Avec une majoration convenable de la norme $\|\mathbf{V}\|$ de la matrice de covariance des erreurs, nous pouvons déduire des relations (1.5b) et (1.6b) le résultat asymptotique suivant.

Proposition 1.4. *On suppose dans le modèle (1.4) que $n = 1$, que la fonction f appartient à $W_2^m[0, 1]$ pour un entier $m \geq 1$, que le dispositif expérimental est quasi-uniforme et que l'Hypothèse 1.1 sur la covariance des erreurs est valide. Pour un paramètre $\lambda = \lambda(p) > 0$, on considère l'estimateur spline de lissage \hat{f}_λ d'ordre $2m$.*

Si les v.a. $\varepsilon_i, 1 \leq i \leq p$, vérifient l'Hypothèse 1.2 pour un $\gamma > 1$ ou l'Hypothèse 1.3, et si $\lambda p^{2m} \rightarrow \infty$ quand $p \rightarrow \infty$, alors les erreurs quadratiques moyennes MASE et MISE de \hat{f}_λ sont en $\mathcal{O}(\lambda) + \mathcal{O}(p^{-1}\lambda^{-1/2m})$ et la vitesse optimale de convergence en $p^{-2m/(2m+1)}$ est atteinte pour un choix de λ du même ordre.

Si les v.a. $\varepsilon_i, 1 \leq i \leq p$, vérifient l'Hypothèse 1.4, et si $\lambda R_p^{2m} \rightarrow \infty$ quand $p \rightarrow \infty$, alors la MASE et la MISE de \hat{f}_λ sont en $\mathcal{O}(\lambda) + \mathcal{O}(R_p^{-1}\lambda^{-1/2m})$ et la vitesse optimale de convergence en $\mathcal{O}(R_p^{-2m/(2m+1)})$ est atteinte pour un choix de λ du même ordre.

Nous faisons ici quelques commentaires sur les vitesses de convergence obtenues et sur les techniques mathématiques utilisées dans la proposition précédente.

Sous les Hypothèses 1.2 (avec $\gamma > 1$) et 1.3, la vitesse de convergence en $\mathcal{O}(p^{-2m/(2m+1)})$ de l'estimateur spline de lissage est inchangée par rapport au cadre d'un bruit blanc. On montre en fait que cette vitesse reste la même sous l'hypothèse plus générale de décorrélation asymptotique, $\sup_{1 \leq i \leq p} \sum_{j=1}^p |\text{Cov}(\varepsilon_i, \varepsilon_j)| = \mathcal{O}(1)$.

Sous l'Hypothèse 1.4 de bruit mélangeant, on observe un potentiel ralentissement de la convergence du fait que $R_p \leq p$, en accord avec Burman (1991). L'estimateur reste cependant consistant car $R_p \rightarrow \infty$ quand $p \rightarrow \infty$. Nous remarquons par ailleurs que le résultat de la Proposition 1.4 reste valide si, à la place de l'Hypothèse 1.4, on suppose que les erreurs $\varepsilon_1, \dots, \varepsilon_p$, sont issues d'une série temporelle $(\varepsilon_i)_{i \in \mathbb{N}}$ ayant des coefficients de mélange $\varphi_i \downarrow 0$, resp. $\lambda_i \downarrow 0$ (cf. par exemple Roussas et Ionides (1987) pour la définition de ces coefficients dans le cadre des séries temporelles), tels que $\sum_{i=1}^{p/2} \varphi_i^{1/2} \sim R_p/p$, resp. $\sum_{i=1}^{p/2} \lambda_i^{1/2} \sim R_p/p$ quand $p \rightarrow \infty$.

Pour des erreurs exhibant une corrélation de longue portée, les techniques mathématiques utilisées dans ce chapitre ne permettent pas d'obtenir la vitesse de convergence minimax $p^{-2m\gamma/(2m+\gamma)}$ établie par Wang (1996). Nous en donnons ici la raison puis nous indiquons, pour l'estimateur spline de lissage, comment obtenir cette vitesse de convergence.

L'étude de la variance de l'estimateur \hat{f}_λ proposée dans ce chapitre consiste à séparer l'influence du lissage de celle des erreurs. Dans cette section, la variance moyenne $p^{-1} \sum_{i=1}^p \text{Var} \hat{f}_\lambda(x_i) = p^{-1} \text{tr} \mathbf{S}_\lambda^2 \mathbf{V}$ de \hat{f}_λ est majorée par $p^{-1} (\text{tr} \mathbf{S}_\lambda^2) \|\mathbf{V}\|$. Cette stratégie, qui s'avère efficace sous les conditions de la Proposition 1.4, ne fonctionne pas dans le cas d'une corrélation de longue portée. En effet, la meilleure majoration que l'on puisse obtenir dans ce cas pour $\|\mathbf{V}\|$ est $\|\mathbf{V}\| = \mathcal{O}(p^{-\gamma})$, cf. démonstration de la Proposition 1.4. Avec la majoration $p^{-1} \sum_{i=1}^p \{f(x_i) - \hat{f}_\lambda(x_i)\}^2 \leq (\lambda/4) |f|_m^2$ du Théorème 1.2 pour le biais au carré et la majoration $\text{tr} \mathbf{S}_\lambda^2 \leq M_2 \lambda^{-1/2m}$ de la Proposition 1.2, on arrive seulement à la vitesse de convergence $p^{-2m\gamma/(2m+1)}$.

Les travaux de Cox (1983) et Nychka (1995) sur le noyau équivalent de l'estimateur spline de lissage d'ordre $2m$ permettent une étude plus fine de la variance. Plus précisément, en utilisant la conjecture de Nychka (1995) sur la double enveloppe exponentielle du noyau équivalent et de la fonction de Green associée (Théorème 2.1 démontré dans le cas $m = 1$ et conjecturé pour $m > 1$), on peut montrer que $\text{Var} \hat{f}_\lambda(x) = \mathcal{O}(\lambda^{-1/2m} p^{-\gamma})$ uniformément en $x \in [0, 1]$ sous les hypothèses suivantes : i) $\lambda^{1/2m} > C \sup_{[0,1]} |F_p - F|$ pour tout $p \geq 1$ et pour une constante $C > 0$ fixée, où F_p est la fonction de répartition empirique et $F \in C^1[0, 1]$ est une fonction ayant une dérivée strictement positive ; ii) le dispositif expérimental est quasi-uniforme. On retrouve ainsi la vitesse $p^{-2m\gamma/(2m+\gamma)}$ de Wang (1996).

1.3.2 Processus aléatoire à temps continu

Dans le modèle équilibré (1.4), nous supposons que les erreurs aléatoires ε_{ij} sont issues d'un processus aléatoire centré $\varepsilon = \{\varepsilon(t), t \in [0, 1]\}$ dont la fonction de covariance est continue (non nulle), et que les points d'observation ont une répartition quasi-uniforme. On a donc $\varepsilon_{ij} = \varepsilon_i(t_j)$ pour $1 \leq i \leq n$ et $1 \leq j \leq p$, les ε_i étant n réalisations indépendantes du processus ε .

Nous montrons ici que l'erreur quadratique MASE ou MISE de l'estimateur spline de lissage converge vers zéro à une vitesse $\mathcal{O}(n^{-1})$. Ce résultat a été établi par Cardot et Diack (1998) dans le modèle (1.4) pour l'estimateur spline hybride, sous des conditions qui n'englobent pas l'estimateur spline de lissage (en effet, leur résultat de convergence est valide si le nombre k de noeuds de la base de B-splines de l'estimateur hybride reste négligeable devant p dans l'asymptotique, ce qui exclut la possibilité d'identifier l'estimateur spline de lissage à l'estimateur spline hybride). Au Chapitre 2, nous montrerons que cette vitesse de convergence est exactement de l'ordre de n^{-1} et qu'elle s'étend par ailleurs à l'ensemble des estimateurs linéaires.

Nous observons d'abord que $\text{tr } \mathbf{V} = \sum_{j=1}^p \text{Var } \varepsilon(t_j) \leq p \sup_{t \in [0,1]} \text{Var}(\varepsilon(t))$, ce qui fournit une convergence de l'estimateur spline de lissage en $\mathcal{O}(n^{-1})$ avec les inéquations (1.5a) et (1.6a). La même vitesse est valable sous l'hypothèse affaiblie d'une fonction de covariance de ε seulement bornée. Par ailleurs, on peut essayer d'obtenir une meilleure vitesse avec les inéquations (1.5b) et (1.6b). Dans ce but, on montre l'équivalence $\|\mathbf{V}\| \sim p$ quand $p \rightarrow \infty$ grâce à un résultat de théorie spectrale (Lemme 1 de l'Annexe C). Ceci donne une majoration de la variance moyenne aux points d'observations en $n^{-1} \lambda^{-1/2m}$ par application de (1.5b). En optimisant ensuite le choix de λ , il vient une vitesse de convergence en $\mathcal{O}(n^{-2m/(2m+1)})$ qui n'améliore pas la vitesse $\mathcal{O}(n^{-1})$ fournie par (1.5a). Nous retenons donc le résultat suivant.

Proposition 1.5. *Dans le modèle équilibré (1.4), supposons que les erreurs ε_{ij} sont issues d'un processus aléatoire centré $\varepsilon = \{\varepsilon(t), t \in [0, 1]\}$ de fonction de covariance continue, et que les points d'observations ont une répartition quasi-uniforme. Soit \hat{f}_λ l'estimateur spline de lissage de f d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p) > 0$. Alors, sous la condition $\lambda = \mathcal{O}(n^{-1})$, et sous la condition supplémentaire $p^{-2m} = \mathcal{O}(\lambda)$ pour la MISE, les erreurs quadratiques moyennes MASE et MISE de \hat{f}_λ sont en $\mathcal{O}(n^{-1})$ quand $n, p \rightarrow \infty$.*

Remarque 1.9. *Il existe une différence importante entre le cadre de la régression non paramétrique classique (erreurs décorréelées, $n = 1$ et asymptotique $p \rightarrow \infty$) et le cadre considéré dans cette section (mesures répétées avec $n, p \rightarrow \infty$) au niveau des conditions nécessaires et suffisantes de consistance [CNS] portant sur le paramètre de lissage λ . En effet, Craven et Wahba (1979) ont montré dans la première situation qu'une double CNS est ($\lambda \rightarrow 0, \lambda p^{2m} \rightarrow \infty$) (voir page 10). Dans la deuxième situation par contre, la seule CNS de consistance est $\lambda \rightarrow 0$ (ceci découle de la proposition précédente pour la MASE, et on verra au Chapitre 2 que la condition $p^{-2m} = \mathcal{O}(\lambda)$ n'est pas nécessaire pour la MISE). Par conséquent, le paramètre λ peut dans ce cadre tendre vers zéro arbitrairement vite.*

1.3.3 Données longitudinales

Nous examinons maintenant le modèle (1.1) en supposant disposer de mesures issues d'un grand nombre d'individus observés chacun à quelques reprises, les temps d'observations pouvant varier d'un individu à l'autre.

Nous reprenons les notations de la Section 1.1 : n désigne le nombre d'individus, p_i est le nombre de mesures pour l'individu i , $p = p(n)$ est le nombre de points d'observations distincts, n_j est le nombre d'observations au point t_j et N est le nombre total de données. La présente situation correspond à l'asymptotique dans laquelle $n \rightarrow \infty$ avec $\max_{1 \leq i \leq n} p_i = \mathcal{O}(1)$ (il en découle que $n \sim N$ dans l'asymptotique).

Deux cas typiques se présentent alors.

1. L'ensemble des temps d'observations est fixé, de taille p finie et chaque temps d'observation fait approximativement l'objet du même nombre de mesures. En termes mathématiques, on a alors $\min_{1 \leq j \leq p} n_j \sim n$ quand $n \rightarrow \infty$. Exemple : mesure d'un taux quotidien pendant un mois lors d'essais cliniques.
2. L'ensemble des temps d'observations s'accroît indéfiniment alors que le nombre d'individus augmente, chaque temps d'observation ne faisant l'objet que d'un nombre fini de mesures, *i.e.* $\max_{1 \leq j \leq p} n_j = \mathcal{O}(1)$ quand $n \rightarrow \infty$. Exemple : les individus viennent d'eux-mêmes faire l'objet d'une mesure à un temps qui n'est pas contrôlé par l'expérimentateur. Les temps d'acquisition des données sont alors considérés fixes (et non pas aléatoires) si leur variabilité n'est pas pertinente dans l'étude.

Nous déduisons des relations (1.7a) du Théorème 1.4 et (1.8b) du Théorème 1.5 la vitesse de convergence suivante pour l'estimateur spline de lissage.

Proposition 1.6. *Dans le modèle (1.1), on considère l'asymptotique $n \rightarrow \infty$ avec $\max_{1 \leq i \leq n} p_i = \mathcal{O}(1)$. On suppose que $f \in W_2^m[0, 1]$, que les erreurs ε_{ij} vérifient l'Hypothèse 1.1 et la condition $\sup_{1 \leq j \leq p_i, 1 \leq i \leq n} \text{Var } \varepsilon_{ij} = \mathcal{O}(1)$ quand $N \rightarrow \infty$, et que le dispositif expérimental est quasi-uniforme. Pour un réel $\lambda = \lambda(p_1, \dots, p_n, n) > 0$ on s'intéresse à l'estimateur spline de lissage \hat{f}_λ d'ordre $2m$.*

Dans le cas 1, la MASE de \hat{f}_λ est en $\mathcal{O}(\lambda) + \mathcal{O}(n^{-1})$ lorsque $N \rightarrow \infty$. La meilleure vitesse de convergence est alors en $\mathcal{O}(n^{-1})$.

Dans le cas 2, sous la condition $p^{-2m} = \mathcal{O}(\lambda)$, la MISE de l'estimateur \hat{f}_λ est en $\mathcal{O}(\lambda) + \mathcal{O}(n^{-1}\lambda^{-1/2m})$ lorsque $N \rightarrow \infty$ et la vitesse de convergence optimale est en $\mathcal{O}(n^{-2m/(2m+1)})$.

1.4 Démonstrations

Démonstration du Théorème 1.2

Pour étudier la mesure d'erreur quadratique moyenne discrétisée (MASE), nous commençons par décomposer cette quantité en somme du biais au carré et de la variance de $\hat{\mathbf{f}}_\lambda$:

$$p^{-1} \mathbb{E} \left\| \mathbf{f} - \hat{\mathbf{f}}_\lambda \right\|^2 = p^{-1} \left\| \mathbf{f} - \mathbb{E} \hat{\mathbf{f}}_\lambda \right\|^2 + p^{-1} \mathbb{E} \left\| \mathbb{E} \hat{\mathbf{f}}_\lambda - \hat{\mathbf{f}}_\lambda \right\|^2.$$

Ensuite, nous savons par la Proposition 1.1 que $\hat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \bar{\mathbf{y}}$. Notons alors $\bar{\boldsymbol{\varepsilon}} = (\bar{\boldsymbol{\varepsilon}}_1, \dots, \bar{\boldsymbol{\varepsilon}}_p)'$ le vecteur d'erreur égal à $\bar{\mathbf{y}} - \mathbf{f}$. Par la commutativité de la trace matricielle et par la symétrie de $\mathbf{S}_\lambda = (\mathbf{I} + \lambda p \boldsymbol{\Gamma})^{-1}$ dans le modèle équilibré (1.4), il

vient

$$\begin{aligned} p^{-1} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 &= p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|^2 + p^{-1} \mathbb{E} \|\mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}\|^2 \\ &= p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|^2 + p^{-1} \mathbb{E} \operatorname{tr}(\bar{\boldsymbol{\varepsilon}}' \mathbf{S}_\lambda^2 \bar{\boldsymbol{\varepsilon}}) \\ &= p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|^2 + p^{-1} \operatorname{tr}(\mathbf{S}_\lambda^2 \mathbb{E} \bar{\boldsymbol{\varepsilon}} \bar{\boldsymbol{\varepsilon}}'). \end{aligned}$$

D'après l'Hypothèse 1.1 sur la structure de covariance des erreurs, nous voyons que pour tous $1 \leq j, k \leq p$,

$$\mathbb{E} \bar{\boldsymbol{\varepsilon}}_j \bar{\boldsymbol{\varepsilon}}_k = n^{-2} \sum_{i=1}^n \mathbb{E} \varepsilon_{ij} \varepsilon_{ik} = n^{-2} \sum_{i=1}^n v_{jk} = n^{-1} v_{jk}.$$

Par conséquent, la MASE de l'estimateur $\widehat{\mathbf{f}}_\lambda$ s'écrit

$$p^{-1} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 = p^{-1} \mathbf{f}' (\mathbf{I} - \mathbf{S}_\lambda)^2 \mathbf{f} + (np)^{-1} \operatorname{tr}(\mathbf{S}_\lambda^2 \mathbf{V}). \quad (1.12)$$

Pour majorer le terme de biais au carré $p^{-1} \mathbf{f}' (\mathbf{I} - \mathbf{S}_\lambda)^2 \mathbf{f}$, nous allons exprimer les valeurs et vecteurs propres de la matrice $\mathbf{I} - \mathbf{S}_\lambda$ en fonction de ceux de la matrice d'énergie $\boldsymbol{\Gamma}$ définie dans le Théorème ?? de l'Annexe A. Cette matrice classique dans les calculs de splines Soient $\lambda_1 \geq \dots \geq \lambda_p \geq 0$ les valeurs propres de $\boldsymbol{\Gamma}$ et soient $\mathbf{u}_1, \dots, \mathbf{u}_p$, des vecteurs orthonormés associés. Alors, ces vecteurs sont aussi des vecteurs propres de \mathbf{S}_λ pour les valeurs propres $(1 + \lambda p \lambda_1)^{-1}, \dots, (1 + \lambda p \lambda_p)^{-1}$, et ce sont encore des vecteurs propres de $\mathbf{I} - \mathbf{S}_\lambda$ pour les valeurs propres $\lambda p \lambda_1 / (1 + \lambda p \lambda_1), \dots, \lambda p \lambda_p / (1 + \lambda p \lambda_p)$. En décomposant le vecteur \mathbf{f} sur cette base, nous avons

$$(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f} = (\mathbf{I} - \mathbf{S}_\lambda) \sum_{j=1}^p (\mathbf{f}' \mathbf{u}_j) \mathbf{u}_j = \sum_{j=1}^p \frac{\lambda p \lambda_j(\boldsymbol{\Gamma})}{1 + \lambda p \lambda_j(\boldsymbol{\Gamma})} (\mathbf{f}' \mathbf{u}_j) \mathbf{u}_j,$$

ce qui implique

$$\begin{aligned} p^{-1} \mathbf{f}' (\mathbf{I} - \mathbf{S}_\lambda)^2 \mathbf{f} &= p^{-1} \sum_{j=1}^p \left(\frac{\lambda p \lambda_j(\boldsymbol{\Gamma})}{1 + \lambda p \lambda_j(\boldsymbol{\Gamma})} \right)^2 (\mathbf{f}' \mathbf{u}_j)^2 \\ &= \lambda \sum_{j=1}^p \frac{\lambda p \lambda_j(\boldsymbol{\Gamma})}{(1 + \lambda p \lambda_j(\boldsymbol{\Gamma}))^2} \lambda_j(\boldsymbol{\Gamma}) (\mathbf{f}' \mathbf{u}_j)^2. \end{aligned}$$

Comme par ailleurs la quantité $t / (1 + t)^2$ est majorée par $1/4$ pour tout $t \geq 0$, on obtient

$$\begin{aligned} p^{-1} \mathbf{f}' (\mathbf{I} - \mathbf{S}_\lambda)^2 \mathbf{f} &\leq (\lambda/4) \sum_{j=1}^p \lambda_j(\boldsymbol{\Gamma}) (\mathbf{f}' \mathbf{u}_j)^2 \\ &= (\lambda/4) \mathbf{f}' \boldsymbol{\Gamma} \mathbf{f} \\ &\leq (\lambda/4) |f|_m^2, \end{aligned} \quad (1.13)$$

le dernier passage se faisant en utilisant l'hypothèse de régularité $f \in W_2^m(0, 1)$ et la Proposition 1.1.

Remarque 1.10. *Des majorations du biais au carré discrétisé de l'estimateur spline de lissage sont obtenues d'une façon différente dans le Lemme 4.1 de Craven et Wahba (1979) et dans la formule (4.7) de Ragozin (1983), avec les constantes respectives 1 et 1/2 au lieu de 1/4 dans (1.13).*

Pour majorer le terme de variance, nous faisons appel à un lemme de calcul matriciel démontré en Annexe B.

Lemme 1.3. *Soient \mathbf{A} et \mathbf{B} deux matrices symétriques de même taille, avec \mathbf{B} semi-définie positive. Soit $\lambda_{\min}(\mathbf{A})$ la plus petite valeur propre de \mathbf{A} . Alors*

$$\lambda_{\min}(\mathbf{A}) \operatorname{tr} \mathbf{B} \leq \operatorname{tr} \mathbf{AB} \leq \|\mathbf{A}\| \operatorname{tr} \mathbf{B}.$$

En appliquant ce lemme aux matrices \mathbf{S}_λ^2 et \mathbf{V} , il vient

$$\operatorname{tr} \mathbf{S}_\lambda^2 \mathbf{V} \leq \|\mathbf{S}_\lambda^2\| \operatorname{tr} \mathbf{V}. \quad (1.14)$$

Notons alors que $\|\mathbf{S}_\lambda^2\| = \|\mathbf{S}_\lambda\|^2 = 1$. En effet, la matrice $\mathbf{\Gamma}$ étant semi-définie positive, on voit que $\|\mathbf{S}_\lambda\| = \|(\mathbf{I} + \lambda p \mathbf{\Gamma})^{-1}\| \leq 1$ et 0 étant une valeur propre de $\mathbf{\Gamma}$ (voir par exemple Utreras (1983)), on en déduit que

$$\|\mathbf{S}_\lambda\| = 1. \quad (1.15)$$

Il suffit alors d'introduire les formules (1.13) à (1.15) dans l'inégalité (1.12) pour obtenir la majoration (1.5a) du théorème.

Concernant l'inégalité (1.5b), il suffit de permuter les matrices \mathbf{S}_λ^2 et \mathbf{V} dans (1.14) avec le Lemme 1.3. Il vient $\operatorname{tr}(\mathbf{S}_\lambda^2 \mathbf{V}) \leq \|\mathbf{V}\| \operatorname{tr}(\mathbf{S}_\lambda^2)$, puis on complète la démonstration à l'aide de la Proposition 1.2. ■

Démonstration du Théorème 1.3

Pour obtenir une majoration de l'erreur quadratique moyenne intégrée (MISE), nous adaptons les résultats de Ragozin (1983) au contexte d'observations corrélées. Le terme de biais de l'estimateur ne dépendant pas de la partie aléatoire du modèle, nous pouvons reprendre en l'état la majoration du biais de Ragozin (1983). Pour adapter la majoration de la variance, nous faisons essentiellement appel au Lemme 1.3 qui permet de séparer l'influence de la matrice de lissage de celle de la matrice de covariance des données.

Commençons par rappeler la majoration du terme de biais de l'estimateur.

Théorème 1.9. (Ragozin, 1983) Soit $m > 0$ et $0 \leq k \leq m$ des entiers et soient $t_1 < \dots < t_p$ des points de $[0, 1]$, avec $p \geq m$.

Il existe des constantes $C = C(m, h_{\max}/h_{\min})$, $D = D(m)$ et $H = H(m, k)$ telles que pour toute fonction $g \in W_2^m[0, 1]$, en posant $L = Cph_{\max}\lambda/2 + Dh_{\max}^{2m}$,

$$|g - g_\lambda|_k^2 \leq H(1 + L^{2m})^{k/m} L^{(m-k)/m} |g|_m^2,$$

où g_λ est la spline naturelle qui minimise $\{p^{-1}\|\mathbf{g} - \mathbf{h}\|^2 + \lambda|h|_m^2\}$ pour $h \in W_2^m[0, 1]$. Dans le cas où $k = 0$, $H(m, 0) = 1$.

En appliquant ce théorème à la fonction de régression f , la spline associée étant alors $\mathbb{E}\hat{f}_\lambda$, nous pouvons majorer le biais au carré de l'estimateur ($k = 0$) ainsi :

$$|f - \mathbb{E}\hat{f}_\lambda|_0^2 \leq (Cph_{\max}\lambda/2 + Dh_{\max}^{2m})|f|_m^2. \quad (1.16)$$

Il reste à majorer le terme de variance. Dans ce but, nous énonçons un théorème qui relie les normes discrètes et continues d'une fonction $g \in W_2^m[0, 1]$.

Théorème 1.10. (Ragozin, 1983) Avec les notations du Théorème 1.9, on a, pour toute fonction $g \in W_2^m[0, 1]$,

$$h_{\max}^{2k}|g|_k^2 \leq Ch_{\max}\|\mathbf{g}\|^2 + Dh_{\max}^{2m}|g|_m^2.$$

Nous avons aussi besoin d'un lemme permettant de caractériser le vecteur de résidus $\bar{\mathbf{y}} - \hat{\mathbf{f}}_\lambda = \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}$. Nous indiquons sa démonstration afin de l'adapter plus tard.

Lemme 1.4. (Ragozin, 1983) Soit $m > 0$ un entier, soient $t_1 < \dots < t_p$ des points de $[0, 1]$, soit \mathbf{g} un vecteur de \mathbb{R}^p et soit $\lambda > 0$ un réel. Soit aussi g_λ l'unique spline d'ordre $2m$ qui minimise $\{p^{-1}\|\mathbf{g} - \mathbf{h}\|^2 + \lambda|h|_m^2\}$ pour $h \in W_2^m[0, 1]$. Alors, pour toute fonction $h \in W_2^m[0, 1]$, on a la relation

$$p^{-1}\langle \mathbf{g} - \mathbf{g}_\lambda, \mathbf{h} \rangle = \lambda \langle g_\lambda, h \rangle_m. \quad (1.17)$$

Démonstration.

Considérons la norme définie sur l'espace de Hilbert $\mathbb{R}^p \oplus W_2^m[0, 1]$ par $\|(\mathbf{x}, h)\|^2 = p^{-1}\|\mathbf{x}\|^2 + \lambda|h|_0^2$ (somme de la norme euclidienne et de la norme L_2). Alors la spline g_λ est l'unique fonction h qui minimise $\|(\mathbf{g}, 0) - (\mathbf{h}, h^{(m)})\|^2$ dans $W_2^m[0, 1]$. Le sous-espace $H = \{(\mathbf{h}, h^{(m)}), h \in W_2^m[0, 1]\}$ étant un convexe fermé de $\mathbb{R}^p \oplus W_2^m[0, 1]$ pour la norme considérée, on en déduit que $(\mathbf{g}_\lambda, g_\lambda^{(m)})$ est le projeté orthogonal de $(\mathbf{g}, 0)$ sur H . Par un argument standard, $(\mathbf{g}_\lambda, g_\lambda^{(m)}) - (\mathbf{g}, 0)$ est orthogonal à H , ce qui fournit (1.17). \square

Appliquons le lemme précédent au vecteur d'erreur $\mathbf{g} = \bar{\boldsymbol{\varepsilon}} = \bar{\mathbf{y}} - \mathbf{f}$. Dans ce cas, la spline g_λ correspondante n'est autre que l'estimateur centré $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$, et on a $\mathbf{g}_\lambda = \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}$. En prenant $h = \widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$, il vient $p^{-1} \langle \bar{\boldsymbol{\varepsilon}} - \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}, \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}} \rangle = \lambda |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2$. Forts de ces résultats, nous pouvons majorer le terme de variance :

$$\begin{aligned} \mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 &\leq Ch_{max} \mathbb{E}\|\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2 + Dh_{max}^{2m} \mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2 \\ &= Ch_{max} \mathbb{E}\bar{\boldsymbol{\varepsilon}}' \mathbf{S}_\lambda^2 \bar{\boldsymbol{\varepsilon}} + Dh_{max}^{2m} (p\lambda)^{-1} \mathbb{E}\bar{\boldsymbol{\varepsilon}}' (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}} \\ &= Ch_{max} n^{-1} \text{tr} \mathbf{S}_\lambda^2 \mathbf{V} + Dh_{max}^{2m} (np\lambda)^{-1} \text{tr} (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{S}_\lambda \mathbf{V}. \end{aligned}$$

Mais comme les matrices \mathbf{S}_λ et \mathbf{V} sont symétriques, la matrice $\mathbf{S}_\lambda \mathbf{V} \mathbf{S}_\lambda$ est semi-définie positive, et donc $\text{tr} \mathbf{S}_\lambda^2 \mathbf{V} = \text{tr} \mathbf{S}_\lambda \mathbf{V} \mathbf{S}_\lambda \geq 0$. Par conséquent,

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 \leq Ch_{max} n^{-1} \text{tr} \mathbf{S}_\lambda^2 \mathbf{V} + Dh_{max}^{2m} (np\lambda)^{-1} \text{tr} \mathbf{S}_\lambda \mathbf{V}.$$

Ensuite, pour établir la majoration de la variance dans (1.6a), on utilise successivement le Lemme 1.3 et l'équation (1.15) :

$$\begin{aligned} \mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 &\leq n^{-1} (Ch_{max} \|\mathbf{S}_\lambda^2\| \text{tr} \mathbf{V} + Dh_{max}^{2m} (p\lambda)^{-1} \|\mathbf{S}_\lambda\| \text{tr} \mathbf{V}) \\ &\leq n^{-1} (Ch_{max} + Dh_{max}^{2m} (p\lambda)^{-1}) \text{tr} \mathbf{V}. \end{aligned} \quad (1.18)$$

Pour établir la la majoration de la variance dans (1.6b), on utilise le Lemme 1.3 et la Proposition 1.2 pour écrire

$$\begin{aligned} \mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 &\leq n^{-1} (Ch_{max} \|\mathbf{V}\| \text{tr} \mathbf{S}_\lambda^2 + Dh_{max}^{2m} (p\lambda)^{-1} \|\mathbf{V}\| \text{tr} \mathbf{S}_\lambda) \\ &\leq n^{-1} \lambda^{-1/2m} (Ch_{max} M_2 + Dh_{max}^{2m} (p\lambda)^{-1} M_1) \|\mathbf{V}\|. \end{aligned} \quad (1.19)$$

La réunion des inégalités (1.16), (1.18) et (1.19) établit alors le Théorème 1.3. ■

Démonstration de la Proposition 1.3

Les deux majorations des quantités $\text{tr} \mathbf{W} \mathbf{S}_\lambda$ et $\text{tr} \mathbf{S}'_\lambda \mathbf{S}_\lambda$ sont obtenues par une légère modification de la technique de calcul du cas équilibré. Pour toute matrice \mathbf{M} de taille $p \times p$ admettant des valeurs propres réelles, notons $\lambda_1(\mathbf{M}) \geq \dots \geq \lambda_p(\mathbf{M})$ ses valeurs propres. Avec le Lemme 1.3, on a

$$\begin{aligned} \text{tr} \mathbf{W} \mathbf{S}_\lambda &= \text{tr} \mathbf{W} (\mathbf{W} + \lambda p \boldsymbol{\Gamma})^{-1} \mathbf{W} \\ &= \text{tr} \mathbf{W}^2 (\mathbf{W} + \lambda p \boldsymbol{\Gamma})^{-1} \\ &\leq \|\mathbf{W}\|^2 \sum_{j=1}^p \lambda_j^{-1} (\mathbf{W} + \lambda p \boldsymbol{\Gamma}) \\ &= \left(\frac{(\max n_j) p}{N} \right)^2 \sum_{j=1}^p \lambda_j^{-1} (\mathbf{W} + \lambda p \boldsymbol{\Gamma}). \end{aligned} \quad (1.20)$$

Le Corollaire 1 du théorème min-max de Courant-Fischer (*cf.* Annexe B page 134), qui porte sur l'encadrement des valeurs propres d'une somme de matrices symétriques, permet d'affirmer que

$$\lambda_j(\mathbf{W} + \lambda p \mathbf{\Gamma}) \geq \lambda_p(\mathbf{W}) + \lambda_j(\lambda p \mathbf{\Gamma}) = \frac{(\min n_j)p}{N} + \lambda p \lambda_j(\mathbf{\Gamma}). \quad (1.21)$$

Ensuite, d'après Utreras (1988), on sait que

$$\lambda_{p-m+1}(\mathbf{\Gamma}) = \dots = \lambda_p(\mathbf{\Gamma}) = 0, \quad (1.22a)$$

et

$$C_1 \frac{j^{2m}}{p} \leq \lambda_{p+1-j}(\mathbf{\Gamma}) \leq C_2 \frac{j^{2m}}{p} \quad \text{pour tout } m+1 \leq j \leq p, \quad (1.22b)$$

où $C_k = C_k(m, h_{max}/h_{min})$, $k = 1, 2$, sont des constantes strictement positives. Par comparaison de l'intégrale convergente d'une fonction décroissante et de la série associée, il vient

$$\begin{aligned} \frac{(\min n_j)p}{N} \sum_{j=m+1}^p \lambda_j^{-1}(\mathbf{W} + \lambda p \mathbf{\Gamma}) &\leq \sum_{j=m+1}^p \left[1 + C_1 \lambda \frac{N}{(\min n_j)p} j^{2m} \right]^{-1} \\ &\leq \int_0^\infty \left(1 + \lambda C_1 \frac{N}{(\min n_j)p} t^{2m} \right)^{-1} dt \end{aligned}$$

puis, avec un changement de variable,

$$\sum_{j=m+1}^p \lambda_j^{-1}(\mathbf{W} + \lambda p \mathbf{\Gamma}) \leq \left(\frac{N}{(\min n_j)p} \right)^{1-1/2m} (C_1 \lambda)^{-1/2m} \int_0^\infty \frac{du}{1+u^{2m}}. \quad (1.23)$$

On pose $C = C_1^{-1/2m} \int_0^\infty (1+u^{2m})^{-1} du$. En utilisant les relations (1.21), (1.22a) et (1.23) dans (1.20), on trouve alors

$$\begin{aligned} \text{tr } \mathbf{W} \mathbf{S}_\lambda &\leq \left(\frac{(\max n_j)p}{N} \right)^2 \left(\frac{mN}{(\min n_j)p} + C \left(\frac{N}{(\min n_j)p} \right)^{1-1/2m} \lambda^{-1/2m} \right) \\ &= \frac{(\max n_j)^2 p}{(\min n_j)N} \left(m + C \left(\frac{(\min n_j)p}{N\lambda} \right)^{1/2m} \right). \end{aligned} \quad (1.24)$$

En remarquant que $(\min n_j)p \leq N$, on peut ensuite trouver une constante \widetilde{M}_1 vérifiant la Proposition 1.3 pour une valeur "raisonnable" de λ , par exemple $\lambda \leq 1$. L'obtention d'une constante \widetilde{M}_2 telle que $\text{tr } \mathbf{S}'_\lambda \mathbf{S}_\lambda \leq \widetilde{M}_2 \frac{(\max n_j)^2 p}{(\min n_j)N\lambda^{1/2m}}$ se fait de manière analogue. ■

Démonstration du Théorème 1.4

Rappelons d'abord que pour un dispositif expérimental déséquilibré, la matrice de lissage s'écrit $\mathbf{S}_\lambda = (\mathbf{W} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{W}$. Par la décomposition en biais au carré plus variance (cf. (1.12)), la MASE s'écrit

$$p^{-1} \mathbb{E} \|\mathbf{f} - \widehat{\mathbf{f}}_\lambda\|^2 = p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|^2 + p^{-1} \text{tr} \mathbf{S}_\lambda (\text{Var} \bar{\boldsymbol{\varepsilon}}) \mathbf{S}'_\lambda. \quad (1.25)$$

Pour majorer le terme de biais de l'estimateur, il suffit d'adapter le Lemme 4.1 de Craven et Wahba (1979). Cette adaptation sera faite au Chapitre 3 par le Lemme 3.1, dans le contexte de du lissage spline par moindres carrés généralisés. Nous l'établissons ici pour l'estimateur spline de lissage classique dans le cadre du dispositif déséquilibré.

Nous avons vu en section 1.1 que l'estimateur spline \widehat{f}_λ minimise le critère $p^{-1} (\bar{\mathbf{y}} - \mathbf{g})' \mathbf{W} (\bar{\mathbf{y}} - \mathbf{g}) + \lambda |g|_m^2$, où g varie dans $W_2^m[0, 1]$. De même, l'espérance $\mathbb{E} \widehat{f}_\lambda$ minimise ce critère si on remplace le vecteur $\bar{\mathbf{y}}$ par \mathbf{f} . Par conséquent, en appliquant cette propriété de minimisation à $g = f$, il vient

$$\frac{1}{p} (\mathbf{f} - \mathbb{E} \widehat{\mathbf{f}}_\lambda)' \mathbf{W} (\mathbf{f} - \mathbb{E} \widehat{\mathbf{f}}_\lambda) + \lambda |\mathbb{E} \widehat{f}_\lambda|_m^2 \leq \frac{1}{p} (\mathbf{f} - \mathbf{f})' \mathbf{W} (\mathbf{f} - \mathbf{f}) + \lambda |f|_m^2 = \lambda |f|_m^2.$$

En observant que $\mathbf{f} - \mathbb{E} \widehat{\mathbf{f}}_\lambda = (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}$ et que $|\mathbb{E} \widehat{f}_\lambda|_m^2 \geq 0$, on en déduit :

$$\frac{1}{p} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|_{\mathbf{W}}^2 \leq \lambda |f|_m^2.$$

Ensuite, en appliquant l'inégalité $\|\mathbf{x}\|_{\mathbf{W}}^2 = \mathbf{x}' \mathbf{W} \mathbf{x} \geq \|x\|^2 / \|\mathbf{W}^{-1}\|$ au vecteur $\mathbf{x} = (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}$ de \mathbb{R}^p , nous obtenons

$$\begin{aligned} p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|^2 &\leq p^{-1} \|(\mathbf{I} - \mathbf{S}_\lambda) \mathbf{f}\|_{\mathbf{W}}^2 \|\mathbf{W}^{-1}\| \\ &\leq \lambda \|\mathbf{W}^{-1}\| |f|_m^2 \\ &= \frac{N\lambda}{(\min_{1 \leq j \leq p} n_j)p} |f|_m^2. \end{aligned} \quad (1.26)$$

La majoration de la variance se fait en écrivant d'abord, à l'aide du Lemme 1.3,

$$\text{tr} \mathbf{S}_\lambda (\text{Var} \bar{\boldsymbol{\varepsilon}}) \mathbf{S}'_\lambda \leq \begin{cases} (\text{tr} \mathbf{S}'_\lambda \mathbf{S}_\lambda) \|\text{Var} \bar{\boldsymbol{\varepsilon}}\| \\ (\text{tr} (\text{Var} \bar{\boldsymbol{\varepsilon}})) \|\mathbf{S}'_\lambda \mathbf{S}_\lambda\| \end{cases}. \quad (1.27)$$

On majore alors $\text{tr} \mathbf{S}'_\lambda \mathbf{S}_\lambda$ avec la Proposition 1.3. D'autre part on montre sans difficulté que la matrice de lissage vérifie encore la relation (1.15) dans un dispositif déséquilibré : il suffit d'écrire \mathbf{S}_λ sous la forme $(\mathbf{I} + \lambda p \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$ et de remarquer que les vecteurs propres associés à la valeur propre 0 de $\mathbf{\Gamma}$ sont encore vecteurs propres

de \mathbf{S}_λ pour la valeur propre 1. On a donc $\|\mathbf{S}'_\lambda \mathbf{S}_\lambda\| = \|\mathbf{S}_\lambda\| = 1$. Ensuite, la matrice $\text{Var } \bar{\boldsymbol{\varepsilon}}$ s'explique facilement à l'aide des coefficients de la matrice $\mathbf{V} = (v_{kl})$. Il vient

$$[\text{Var } \bar{\boldsymbol{\varepsilon}}]_{kl} = \frac{\text{Card}(\mathbf{R}_k \cap \mathbf{R}_l)}{n_k n_l} v_{kl},$$

où \mathbf{R}_k est l'ensemble $\{i : \exists j, x_{i,j} = x_k\}$ et $n_k = \text{Card}(\mathbf{R}_k)$ pour tout $1 \leq k \leq p$. On peut maintenant écrire

$$\text{tr}(\text{Var } \bar{\boldsymbol{\varepsilon}}) = \sum_{j=1}^p \frac{1}{n_j} v_{jj} \leq \frac{\text{tr } \mathbf{V}}{\min_{1 \leq j \leq p} n_j}. \quad (1.28)$$

Par ailleurs, un simple calcul permet de séparer dans la norme $\|\text{Var } \bar{\boldsymbol{\varepsilon}}\|$ l'influence de la structure de covariance et celle du dispositif expérimental :

$$\begin{aligned} \|\text{Var } \bar{\boldsymbol{\varepsilon}}\| &= \sup_{\|x\|=1} \left| \sum_{1 \leq k, l \leq p} [\text{Var}(\bar{\boldsymbol{\varepsilon}})]_{kl} x_k x_l \right| \\ &\leq \sup_{\|x\|=1} \sum_{1 \leq k, l \leq p} \frac{\text{Card}(\mathbf{R}_k \cap \mathbf{R}_l)}{n_k n_l} |v_{kl}| \cdot |x_k| \cdot |x_l|. \end{aligned}$$

Or pour tout $1 \leq k, l \leq p$, on a

$$\frac{\text{Card}(\mathbf{R}_k \cap \mathbf{R}_l)}{n_k n_l} \leq \frac{\min(n_k, n_l)}{n_k n_l} = \frac{1}{\max(n_k, n_l)} \leq \frac{1}{\min_{1 \leq j \leq p} n_j}.$$

Par conséquent, en notant $\mathbf{V}_{abs} = (|v_{jk}|)_{1 \leq j, k \leq p}$, nous avons finalement

$$\begin{aligned} \|\text{Var } \bar{\boldsymbol{\varepsilon}}\| &\leq \frac{1}{\min_{1 \leq j \leq p} n_j} \sup_{\|x\|=1} \sum_{1 \leq k, l \leq p} |v_{kl}| \cdot |x_k| \cdot |x_l| \\ &= \frac{1}{\min_{1 \leq j \leq p} n_j} \sup_{\|x\|=1} \sum_{1 \leq k, l \leq p} |v_{kl}| x_k x_l \\ &\leq \frac{\|\mathbf{V}_{abs}\|}{\min_{1 \leq j \leq p} n_j}. \quad (1.29) \end{aligned}$$

L'application des relations (1.15), (1.28), (1.29) et de la Proposition 1.3 dans la formule (1.27) établit alors les majorations de la variance annoncées, ce qui démontre le Théorème 1.4. ■

Démonstration du Théorème 1.5

Dans le cas d'un dispositif déséquilibré, la majoration du biais de l'estimateur spline \widehat{f}_λ s'effectue en adaptant le Théorème 1.9. Partant du Théorème 1.10 appliqué à la fonction $f - \widehat{f}_\lambda$, on écrit

$$|f - \mathbb{E}\widehat{f}_\lambda|_0^2 \leq Ch_{max} \|\mathbf{f} - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2 + Dh_{max}^{2m} |f - \mathbb{E}\widehat{f}_\lambda|_m^2.$$

Pour majorer le terme de biais discret, on reprend la relation (1.26). Nous adaptons ensuite le Lemme 1.4 au cas d'un dispositif expérimenté. Il suffit de remplacer dans la démonstration la norme $\|(\mathbf{x}, g)\|^2 = p^{-1}\|\mathbf{x}\|^2 + \lambda|g|_0^2$ définie sur $\mathbb{R}^p \oplus W_2^m[0, 1]$ par $\|(\mathbf{x}, g)\|_{\mathbf{W}}^2 = p^{-1}\mathbf{x}'\mathbf{W}\mathbf{x} + \lambda|g|_0^2$. La relation (1.17) est alors généralisée par

$$(\bar{\mathbf{y}} - \widehat{\mathbf{f}}_\lambda)' \mathbf{W} \mathbf{g} = p\lambda \langle \widehat{f}_\lambda, g \rangle_m \quad (1.30)$$

pour toute fonction $g \in W_2^m[0, 1]$. On prend ensuite l'espérance de cette formule et on l'utilise avec $g = f - \mathbb{E}\widehat{f}_\lambda$. Il vient

$$(\mathbf{f} - \mathbb{E}\widehat{\mathbf{f}}_\lambda)' \mathbf{W} (\mathbf{f} - \mathbb{E}\widehat{\mathbf{f}}_\lambda) = p\lambda \langle \widehat{f}_\lambda, f - \mathbb{E}\widehat{f}_\lambda \rangle_m. \quad (1.31)$$

Puisque $|f|_m^2 = |f - \mathbb{E}\widehat{f}_\lambda|_m^2 + |\mathbb{E}\widehat{f}_\lambda|_m^2 + 2\langle \widehat{f}_\lambda, f - \mathbb{E}\widehat{f}_\lambda \rangle_m$, on peut faire, au regard de (1.31), la majoration $|f - \mathbb{E}\widehat{f}_\lambda|_m^2 \leq |f|_m^2$. Rassemblant les résultats précédents, on obtient

$$|f - \mathbb{E}\widehat{f}_\lambda|_0^2 \leq \left(\frac{CNh_{max}\lambda}{\min_{1 \leq j \leq p} n_j} + Dh_{max}^{2m} \right) |f|_m^2. \quad (1.32)$$

Pour majorer la variance, nous appliquons le Théorème 1.10 à l'estimateur centré $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$, puis nous utilisons (1.30) en remplaçant $\bar{\mathbf{y}}$ par $\bar{\boldsymbol{\varepsilon}}$, \widehat{f}_λ par $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$ et en prenant $g = \widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$. Il s'en suit que

$$\begin{aligned} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 &\leq Ch_{max} \|\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2 + Dh_{max}^{2m} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2 \\ &= Ch_{max} \|\mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}\|^2 + Dh_{max}^{2m} (p\lambda)^{-1} ((\mathbf{I} - \mathbf{S}_\lambda)' \bar{\boldsymbol{\varepsilon}})' \mathbf{W} \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}. \end{aligned}$$

On prend ensuite l'espérance de cette inégalité. On a

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 \leq Ch_{max} \text{tr} \mathbf{S}_\lambda \text{Var}(\bar{\boldsymbol{\varepsilon}} \mathbf{S}'_\lambda + Dh_{max}^{2m} (p\lambda)^{-1} \text{tr}(\mathbf{I} - \mathbf{S}_\lambda)' \mathbf{W} \mathbf{S}_\lambda \text{Var} \bar{\boldsymbol{\varepsilon}}). \quad (1.33)$$

Le premier terme du majorant correspond au terme de variance de la MASE (Théorème 1.4). Le second terme est majoré par $Dh_{max}^{2m} (p\lambda)^{-1} \text{tr} \mathbf{W} \mathbf{S}_\lambda \text{Var} \bar{\boldsymbol{\varepsilon}}$ (en effet la matrice $\mathbf{S}'_\lambda \mathbf{W} \mathbf{S}_\lambda \text{Var} \bar{\boldsymbol{\varepsilon}}$ est le produit de deux matrices définies positives, donc sa trace est positive). On peut alors écrire : $\text{tr} \mathbf{W} \mathbf{S}_\lambda \text{Var} \bar{\boldsymbol{\varepsilon}} \leq \text{tr} \mathbf{W} \mathbf{S}_\lambda \|\text{Var} \bar{\boldsymbol{\varepsilon}}\|$ ou $\text{tr} \mathbf{W} \mathbf{S}_\lambda \text{Var} \bar{\boldsymbol{\varepsilon}} \leq \|\mathbf{W} \mathbf{S}_\lambda\| \text{tr} \text{Var} \bar{\boldsymbol{\varepsilon}}$ avec le Lemme 1.3. Ensuite, l'étude des quantités $\text{tr} \text{Var} \bar{\boldsymbol{\varepsilon}}$ et $\|\text{Var} \bar{\boldsymbol{\varepsilon}}\|$ a été faite dans la preuve du Théorème 1.4. Restent à majorer $\|\mathbf{W} \mathbf{S}_\lambda\|$ et $\text{tr} \mathbf{W} \mathbf{S}_\lambda$. D'après l'équation (1.15), la première norme est majorée par $\|\mathbf{W}\| = p \max_{1 \leq j \leq p} n_j / N$, et la seconde quantité peut être bornée à l'aide de la Proposition 1.3. Le théorème est alors démontré. ■

Démonstration du Théorème 1.6

Dans un dispositif expérimental équilibré, l'estimateur spline de lissage \widehat{f}_λ défini par la Proposition 1.1 est l'unique fonction qui minimise la quantité $\left\{ p^{-1} \sum_{j=1}^p (\bar{y}_j - g(t_j))^2 + \lambda |g|_m^2 \right\}$ pour g variant dans $W_2^m[0, 1]$, cf. problème (1.2). Il correspond ainsi à l'estimateur spline "classique", construit sans pondération sur les observations, par opposition au cadre déséquilibré dans lequel interviennent les poids n_j/N à la place de p^{-1} . Par conséquent, le biais de la j -ème dérivée de \widehat{f}_λ , qui ne dépend pas des erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$ peut être majoré par l'application directe du résultat suivant.

Théorème 1.11. (Ragozin, 1983) *Soit $\{x_{jp}; 1 \leq j \leq p; p \geq 1\}$ un tableau triangulaire de points de $[0, 1]$ ayant une répartition quasi-uniforme, i.e. tel que $h_{\max}/h_{\min} \leq r$ pour tout $p \geq 1$ pour une constante $r \geq 1$ donnée. Pour toute fonction $g \in W_2^k[0, 1]$ et pour tous entiers $0 \leq j < k \leq m \leq p$, soit g_λ la spline de lissage d'ordre $2m$ qui minimise $\left\{ p^{-1} \sum_{j=1}^p (g(t_j) - h(t_j))^2 + \lambda |h|_m^2 \right\}$ pour $h \in W_2^m[0, 1]$. Alors g_λ satisfait à la relation*

$$|g - g_\lambda|_j^2 = \mathcal{O} \left((\lambda + h_{\max}^{2m})^{(k-j)/m} |g|_k^2 \right), \quad (1.34)$$

où le \mathcal{O} ne dépend que de m, k, j et r .

Pour majorer la variance intégrée de $\widehat{f}_\lambda^{(j)}$, nous utilisons le résultat suivant.

Lemme 1.5. (Ragozin, 1983) *Pour tous entiers k, m , tels que $0 \leq k \leq m$, il existe une constante $\gamma = \gamma(m, k)$ telle que pour toute fonction $g \in W_2^k[0, 1]$ et pour tout $x \in [0, 1]$,*

$$x^{2k} |g|_k^2 \leq \gamma (|g|_0^2 + x^{2m} |g|_m^2).$$

En appliquant ce lemme à $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$ au point $x = \lambda^{1/2m}$, nous obtenons ainsi

$$|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_j^2 \leq \gamma \left(\lambda^{-j/m} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 + \lambda^{(m-j)/m} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2 \right), \quad (1.35)$$

puis nous prenons l'espérance de cette inégalité. Pour finir, on majore $|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2$ et $|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2$ comme dans la fin du Théorème 1.3 (formules (1.18), (1.19) et Lemme 1.4). Il vient

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_j^2 \leq \frac{\gamma}{n} \left(\frac{Ch_{\max} + Dh_{\max}^{2m}(p\lambda)^{-1}}{\lambda^{j/m}} + \frac{\lambda^{(m-j)/m}}{p\lambda} \right) \text{tr } \mathbf{V},$$

c'est-à-dire, compte tenu de la majoration $h_{\max} \leq h_{\min} \sup_{p \geq 1} (h_{\max}/h_{\min}) \leq p^{-1} \sup_{p \geq 1} (h_{\max}/h_{\min}) < \infty$ (quasi-uniformité du dispositif expérimental) et de la condition initiale $\sup_{n \geq 1, p \geq m} (h_{\max}^{2m}/\lambda) < \infty$,

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_j^2 \leq \frac{Q \operatorname{tr} \mathbf{V}}{np\lambda^{j/m}}. \quad (1.36)$$

On a posé $Q = \gamma(C \sup_p(h_{\max}/h_{\min}) + D \sup_{n,p}(h_{\max}^{2m}/\lambda) + 1)$, C et D étant les constantes du Théorème 1.3. On montre de même que

$$\mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_j^2 \leq \frac{Q' \|\mathbf{V}\|}{np\lambda^{2j+1/2m}}, \quad (1.37)$$

avec $Q' = CM_2 \sup_p(h_{\max}/h_{\min}) + DM_1 \sup_{n,p}(h_{\max}^{2m}/\lambda) + M_1$, les constantes M_1 et M_2 étant celles de la Proposition 1.2. ■

Démonstration du Théorème 1.7

L'outil essentiel de cette démonstration est la représentation par valeur/dérivée seconde d'une spline cubique naturelle, que nous rappelons ici.

Soit $p \geq 4$ un entier et soient $t_1 < \dots < t_p$ des réels. Posons $h_j = t_{j+1} - t_j$, $1 \leq j \leq p-1$, et $g_j = g(t_j)$, $\gamma_j = g''(t_j)$, $1 \leq j \leq p$, pour toute fonction g deux fois dérivable.

Proposition 1.7. (Green et Silverman, 1994) *Soit g une spline cubique naturelle ayant ses noeuds aux points $t_1 < \dots < t_p$. Alors pour tout $1 \leq j \leq p-1$ et pour tout $t_j \leq t \leq t_{j+1}$, nous avons*

$$g(t) = \frac{(t-t_j)g_{j+1} + (t_{j+1}-t)g_j}{h_j} - \frac{1}{6}(t-t_j)(t_{j+1}-t) \left[\left(1 + \frac{t-t_j}{h_j}\right) \gamma_{j+1} + \left(1 + \frac{t_{j+1}-t}{h_j}\right) \gamma_j \right].$$

Pour tout $t \leq t_1$, nous avons

$$g(t) = g_1 - (t_1 - t)g'(t_1) = g_1 - (t_1 - t) \left(\frac{g_2 - g_1}{h_1} - \frac{h_1 \gamma_2}{6} \right).$$

Enfin, pour tout $t \geq t_p$, nous avons

$$g(t) = g_p + (t - t_p)g'(t_p) = g_p + (t - t_p) \left(\frac{g_p - g_{p-1}}{h_{p-1}} + \frac{h_{p-1} \gamma_{p-1}}{6} \right).$$

Pour toute spline cubique naturelle g ayant pour noeuds $t_1 < \dots < t_p$, il existe une caractérisation utile des vecteurs $\mathbf{g} = (g(t_1), \dots, g(t_p))'$ et $\boldsymbol{\gamma} = (\gamma_2, \dots, \gamma_{p-1})'$ (on rappelle que dans ce cas $\gamma_1 = \gamma_p = 0$ car par définition, g est affine sur $]-\infty, t_1]$ et sur $[t_p, +\infty[$). Afin d'énoncer cette caractérisation, nous introduisons deux matrices bandes (*i.e.* dont les coefficients non nuls sont regroupés sur un petit nombre de diagonales autour de la principale) indexées de façon non standard pour simplifier les écritures.

Soit $\mathbf{Q} = (q_{ij})$ la matrice de dimensions $p \times (p-2)$, indexée par $1 \leq i \leq p$ et $2 \leq j \leq p-1$, définie par

$$\begin{cases} q_{jj} = -\frac{1}{h_{j-1}} - \frac{1}{h_j}, & q_{j,j-1} = \frac{1}{h_{j-1}}, & q_{j+1,j} = \frac{1}{h_j} & \text{pour } 2 \leq j \leq p-1, \\ q_{ij} = 0 & \text{si } |i-j| \geq 2. \end{cases} \quad (1.38)$$

Soit aussi $\mathbf{R} = (r_{ij})$ la matrice $(p-2) \times (p-2)$ symétrique indexée par $2 \leq i, j \leq p-1$, telle que

$$\begin{cases} r_{jj} = \frac{1}{3}(h_{j-1} + h_j) & \text{pour } 2 \leq j \leq p-1, \\ r_{j+1,j} = r_{j,j+1} = \frac{1}{6} h_j & \text{pour } 2 \leq j \leq p-2, \\ r_{ij} = 0 & \text{si } |i-j| \geq 2. \end{cases} \quad (1.39)$$

La matrice \mathbf{R} étant à diagonale strictement dominante, *i.e.* $|r_{ii}| > \sum_{j \neq i} |r_{ij}|$ pour tout $2 \leq i \leq p-1$, elle est définie positive et inversible (*cf.* par exemple Horn et Johnson (1990)). On peut donc définir la matrice $p \times p$ symétrique $\mathbf{K} = \mathbf{Q}\mathbf{R}^{-1}\mathbf{Q}'$. Il vient alors :

Proposition 1.8. (Green et Silverman, 1994) *Soit g une spline cubique naturelle ayant ses noeuds aux points t_1, \dots, t_p , avec $-\infty \leq a \leq t_1 < \dots < t_p \leq b \leq \infty$. Alors*

$$\mathbf{Q}'\mathbf{g} = \mathbf{R}\boldsymbol{\gamma}$$

et

$$\int_a^b \{g''(t)\}^2 dt = \boldsymbol{\gamma}'\mathbf{R}\boldsymbol{\gamma} = \mathbf{g}'\mathbf{K}\mathbf{g}.$$

Remarque 1.11. *La matrice \mathbf{K} n'est autre que la matrice d'énergie $\boldsymbol{\Gamma}$ dans le cas $m = 2$, *cf.* Théorème 1.1.*

Nous pouvons maintenant aborder le calcul de l'intégrale $|g|_0^2 = \int_0^1 \{g(t)\}^2 dt$. Commençons par faire la décomposition

$$\begin{aligned} |g|_0^2 &= \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \{g(t)\}^2 dt + \int_0^{t_1} \{g(t)\}^2 dt + \int_{t_p}^1 \{g(t)\}^2 dt \\ &:= I_1 + I_2 + I_3. \end{aligned} \quad (1.40)$$

Il s'agit alors de calculer les quantités I_1 , I_2 et I_3 en utilisant les deux propositions précédentes et le résultat classique

Lemme 1.6. *Soient $k, l > 0$ deux entiers et soient a, b deux réels ($a \leq b$). Alors*

$$\int_a^b (b-t)^{k-1} (t-a)^{l-1} dt = (b-a)^{k+l-1} \frac{(k-1)! (l-1)!}{(k+l-1)!}.$$

Nous écrivons d'abord, avec la Proposition 1.7,

$$I_1 = \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \left\{ \frac{(t-t_j)g_{j+1} + (t_{j+1}-t)g_j}{h_j} - \frac{1}{6}(t-t_j)(t_{j+1}-t) \left[\left(1 + \frac{t-t_j}{h_j}\right) \gamma_{j+1} + \left(1 + \frac{t_{j+1}-t}{h_j}\right) \gamma_j \right] \right\}^2 dt,$$

c'est-à-dire, en développant le carré dans l'intégrale,

$$\begin{aligned} I_1 &= \sum_{j=1}^{p-1} \frac{1}{h_j^2} \int_{t_j}^{t_{j+1}} \{ (t-t_j)^2 g_{j+1}^2 + (t_{j+1}-t)^2 g_j^2 \} dt \\ &+ \sum_{j=1}^{p-1} \frac{1}{h_j^2} \int_{t_j}^{t_{j+1}} \{ 2(t-t_j)(t_{j+1}-t)g_{j+1}g_j \} dt \\ &+ \frac{1}{36} \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \left\{ 2(t-t_j)^2(t_{j+1}-t)^2 \left(1 + \frac{t-t_j}{h_j}\right) \left(1 + \frac{t_{j+1}-t}{h_j}\right) \gamma_{j+1}\gamma_j \right\} dt \\ &+ \frac{1}{36} \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)^2(t_{j+1}-t)^2 \left(1 + \frac{t-t_j}{h_j}\right)^2 \gamma_{j+1}^2 \right\} dt \\ &+ \frac{1}{36} \sum_{j=1}^{p-1} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)^2(t_{j+1}-t)^2 \left(1 + \frac{t_{j+1}-t}{h_j}\right)^2 \gamma_j^2 \right\} dt \\ &- \frac{1}{3} \sum_{j=1}^{p-1} \frac{1}{h_j} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)^2(t_{j+1}-t) \left(1 + \frac{t-t_j}{h_j}\right) g_{j+1}\gamma_{j+1} \right\} dt \\ &- \frac{1}{3} \sum_{j=1}^{p-1} \frac{1}{h_j} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)^2(t_{j+1}-t) \left(1 + \frac{t_{j+1}-t}{h_j}\right) g_{j+1}\gamma_j \right\} dt \\ &- \frac{1}{3} \sum_{j=1}^{p-1} \frac{1}{h_j} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)(t_{j+1}-t)^2 \left(1 + \frac{t-t_j}{h_j}\right) g_j\gamma_{j+1} \right\} dt \\ &- \frac{1}{3} \sum_{j=1}^{p-1} \frac{1}{h_j} \int_{t_j}^{t_{j+1}} \left\{ (t-t_j)(t_{j+1}-t)^2 \left(1 + \frac{t_{j+1}-t}{h_j}\right) g_j\gamma_j \right\} dt. \end{aligned}$$

Utilisons à présent le Lemme 1.6 sur les intervalles $[t_j, t_{j+1}]$ pour calculer les coefficients suivants :

$$\begin{aligned} \int_{t_j}^{t_{j+1}} (t - t_j)^2 dt &= \int_{t_j}^{t_{j+1}} (t_{j+1} - t)^2 dt = \frac{h_j^3}{3}, \\ \int_{t_j}^{t_{j+1}} (t - t_j)(t_{j+1} - t) dt &= \frac{h_j^3}{6}, \\ \int_{t_j}^{t_{j+1}} (t - t_j)^2 (t_{j+1} - t)^2 \left(1 + \frac{t - t_j}{h_j}\right)^2 dt &= \frac{8}{105} h_j^5, \\ \int_{t_j}^{t_{j+1}} (t - t_j)^2 (t_{j+1} - t)^2 \left(1 + \frac{t_{j+1} - t}{h_j}\right)^2 dt &= \frac{8}{105} h_j^5, \end{aligned}$$

$$\begin{aligned} \int_{t_j}^{t_{j+1}} (t - t_j)^2 (t_{j+1} - t)^2 \left(1 + \frac{t_{j+1} - t}{h_j}\right) \left(1 + \frac{t - t_j}{h_j}\right) dt &= \frac{31}{420} h_j^5, \\ \int_{t_j}^{t_{j+1}} (t - t_j)^2 (t_{j+1} - t) \left(1 + \frac{t_{j+1} - t}{h_j}\right) dt &= \frac{7}{60} h_j^4, \\ \int_{t_j}^{t_{j+1}} (t - t_j)(t_{j+1} - t)^2 \left(1 + \frac{t - t_j}{h_j}\right) dt &= \frac{7}{60} h_j^4, \\ \int_{t_j}^{t_{j+1}} (t - t_j)^2 (t_{j+1} - t) \left(1 + \frac{t - t_j}{h_j}\right) dt &= \frac{2}{15} h_j^4, \\ \int_{t_j}^{t_{j+1}} (t - t_j)(t_{j+1} - t)^2 \left(1 + \frac{t_{j+1} - t}{h_j}\right) dt &= \frac{2}{15} h_j^4. \end{aligned}$$

Avec ce qui précède, on obtient

$$\begin{aligned} I_1 &= \frac{1}{3} \sum_{j=1}^{p-1} h_j (g_{j+1}^2 + g_j^2 + g_{j+1}g_j) \\ &\quad + \frac{1}{36} \sum_{j=1}^{p-1} h_j^5 \left(\frac{8}{105} (\gamma_{j+1}^2 + \gamma_j^2) + \frac{31}{210} \gamma_{j+1}\gamma_j \right) \\ &\quad - \frac{1}{3} \sum_{j=1}^{p-1} h_j^3 \left(\frac{7}{60} (g_{j+1}\gamma_j + g_j\gamma_{j+1}) + \frac{2}{15} (g_{j+1}\gamma_{j+1} + g_j\gamma_j) \right). \quad (1.41) \end{aligned}$$

Intéressons nous à présent au calcul de I_2 . En posant $h_0 = t_1 - 0 = t_1$ et en appliquant la Proposition 1.7 et le Lemme 1.6, on a les égalités

$$\begin{aligned}
I_2 &= \int_0^{t_1} \left\{ g_1 - (t_1 - x) \left(\frac{g_2 - g_1}{h_1} - \frac{h_1}{6} \gamma_2 \right) \right\}^2 dx \\
&= h_0 g_1^2 - 2g_1 \left(\frac{g_2 - g_1}{h_1} - \frac{h_1 \gamma_2}{6} \right) \int_0^{t_1} (t_1 - x) dx \\
&\quad + \left(\frac{g_2 - g_1}{h_1} - \frac{h_1 \gamma_2}{6} \right)^2 \int_0^{t_1} (t_1 - x)^2 dx \\
&= h_0 g_1^2 - h_0^2 g_1 \left(\frac{g_2 - g_1}{h_1} - \frac{h_1 \gamma_2}{6} \right) + \frac{h_0^3}{3} \left(\frac{g_2 - g_1}{h_1} - \frac{h_1 \gamma_2}{6} \right)^2 \\
&= \left(h_0 + \frac{h_0^2}{h_1} + \frac{h_0^3}{3h_1^2} \right) g_1^2 + \frac{h_0^3}{3h_1^2} g_2^2 - \left(\frac{h_0^2}{h_1} + \frac{2h_0^3}{3h_1^2} \right) g_1 g_2 \\
&\quad + \frac{h_0^3 h_1^2}{108} \gamma_2^2 + \left(\frac{h_0^2 h_1}{6} + \frac{h_0^3}{9} \right) g_1 \gamma_2 - \frac{h_0^3}{9} g_2 \gamma_2. \tag{1.42}
\end{aligned}$$

Le calcul de I_3 est analogue à celui de I_2 . Dans la formule précédente il suffit de remplacer h_0 par $h_p = 1 - t_p$ et h_1 par h_{p-1} , g_1 par g_{p-1} et g_2 par g_{p-2} , et γ_2 par γ_{p-1} . On obtient, de manière symétrique,

$$\begin{aligned}
I_3 &= \int_0^{t_1} \left\{ g_p + (x - t_p - x) \left(\frac{g_p - g_{p-1}}{h_p} - \frac{h_p \gamma_{p-1}}{6} \right) \right\}^2 dx \\
&= \left(h_p + \frac{h_p^2}{h_{p-1}} + \frac{h_p^3}{3h_{p-1}^2} \right) g_p^2 + \frac{h_p^3}{3h_{p-1}^2} g_{p-1}^2 - \left(\frac{h_p^2}{h_{p-1}} + \frac{2h_p^3}{3h_{p-1}^2} \right) g_p g_{p-1} \\
&\quad + \frac{h_p^3 h_{p-1}^2}{108} \gamma_{p-1}^2 + \left(\frac{h_p^2 h_{p-1}}{6} + \frac{h_p^3}{9} \right) g_p \gamma_{p-1} - \frac{h_p^3}{9} g_{p-1} \gamma_{p-1}. \tag{1.43}
\end{aligned}$$

A présent, nous allons écrire la formule (1.40) sous forme matricielle.

Soit $\mathbf{A} = (a_{ij})$ la matrice symétrique tribande de taille $p \times p$ définie par

$$\left\{ \begin{array}{l} a_{11} = h_0 + \frac{h_1}{3} + \frac{h_0^2}{h_1} + \frac{h_0^3}{3h_1^2}, \quad a_{22} = \frac{h_1}{3} + \frac{h_2}{3} + \frac{h_0^3}{3h_1^2}, \\ a_{12} = a_{21} = \frac{1}{2} \left(\frac{h_1}{3} - \frac{h_0^2}{h_1} - \frac{2h_0^3}{3h_1^2} \right), \\ a_{jj} = \frac{1}{3} (h_{j-1} + h_j) \text{ pour } 3 \leq j \leq p-2, \\ a_{j,j+1} = a_{j+1,j} = \frac{h_j}{6} \text{ pour } 2 \leq j \leq p-2, \\ a_{p-1,p} = a_{p,p-1} = \frac{1}{2} \left(\frac{h_{p-1}}{3} - \frac{h_p^2}{h_{p-1}} - \frac{2h_p^3}{3h_{p-1}^2} \right), \\ a_{p-1,p-1} = \frac{h_{p-1}}{3} + \frac{h_{p-2}}{3} + \frac{h_p^3}{3h_{p-1}^2}, \quad a_{pp} = h_p + \frac{h_{p-1}}{3} + \frac{h_p^2}{h_{p-1}} + \frac{h_p^3}{3h_{p-1}^2}, \\ a_{ij} = 0 \text{ si } |i - j| \geq 2. \end{array} \right. \quad (1.44)$$

Soit $\mathbf{B} = (b_{ij})$ la matrice symétrique tribande de dimensions $(p-2) \times (p-2)$ indexée par $2 \leq i, j \leq p-1$, et définie par

$$\left\{ \begin{array}{l} b_{22} = \frac{1}{108} \left(\frac{8(h_1^5 + h_2^5)}{35} + h_0^3 h_1^2 \right), \\ b_{jj} = \frac{8(h_{j-1}^5 + h_j^5)}{108 \cdot 35} \text{ pour } 3 \leq j \leq p-2, \\ b_{j,j+1} = b_{j+1,j} = \frac{1}{2} \left(\frac{31h_j^5}{108 \cdot 70} \right) \text{ pour } 2 \leq j \leq p-2, \\ b_{p-1,p-1} = \frac{1}{108} \left(\frac{8(h_{p-1}^5 + h_{p-2}^5)}{35} + h_p^3 h_{p-1}^2 \right), \\ b_{ij} = 0 \text{ si } |i - j| \geq 2. \end{array} \right. \quad (1.45)$$

Soit enfin $\mathbf{C} = (c_{ij})$ la matrice tribande de dimensions $p \times (p-2)$ indexée par $1 \leq i \leq p$ et $2 \leq j \leq p-1$, et définie par

$$\left\{ \begin{array}{l} c_{12} = \frac{1}{180} (7h_1^3 - 20h_0^3 - 30h_0^2 h_1), \quad c_{22} = \frac{1}{180} (20h_0^3 + 8h_1^3 + 8h_2^3), \\ c_{j,j+1} = c_{j+1,j} = \frac{7h_j^3}{180} \text{ pour } 2 \leq j \leq p-2, \\ c_{jj} = \frac{8(h_{j-1}^3 + h_j^3)}{180} \text{ pour } 3 \leq j \leq p-2, \\ c_{p-1,p-1} = \frac{1}{180} (20h_p^3 + 8h_{p-1}^3 + 8h_{p-2}^3), \\ c_{p,p-1} = \frac{1}{180} (7h_{p-1}^3 - 20h_p^3 - 30h_p^2 h_{p-1}), \\ c_{ij} = 0 \text{ si } |i - j| \geq 2. \end{array} \right. \quad (1.46)$$

En rassemblant les équations (1.41)–(1.43) et en les exprimant sous forme matricielle à l'aide des matrices \mathbf{A} , \mathbf{B} et \mathbf{C} précédentes, la formule (1.40) devient

$$|g|_0^2 = \mathbf{g}'\mathbf{A}\mathbf{g} + \boldsymbol{\gamma}'\mathbf{B}\boldsymbol{\gamma} - \mathbf{g}'\mathbf{C}\boldsymbol{\gamma}.$$

D'après la Proposition 1.8, on a $\boldsymbol{\gamma} = \mathbf{R}^{-1}\mathbf{Q}'\mathbf{g}$. Avec la symétrie de \mathbf{R} , on en déduit que

$$\begin{aligned} |g|_0^2 &= \mathbf{g}'\mathbf{A}\mathbf{g} + \mathbf{g}'\mathbf{Q}\mathbf{R}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{Q}'\mathbf{g} - \mathbf{g}'\mathbf{C}\mathbf{R}^{-1}\mathbf{Q}'\mathbf{g} \\ &= \mathbf{g}'\left(\mathbf{A} + \mathbf{Q}\mathbf{R}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{Q}' - \mathbf{C}\mathbf{R}^{-1}\mathbf{Q}'\right)\mathbf{g} \\ &= \mathbf{g}'\left(\mathbf{A} + \mathbf{Q}\mathbf{R}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{Q}' - \frac{\mathbf{C}\mathbf{R}^{-1}\mathbf{Q}' + \mathbf{Q}\mathbf{R}^{-1}\mathbf{C}'}{2}\right)\mathbf{g}. \end{aligned} \quad (1.47)$$

Il reste à poser, pour obtenir la relation $|g|_0^2 = \mathbf{g}'\boldsymbol{\Phi}\mathbf{g}$ voulue,

$$\boldsymbol{\Phi} := \mathbf{A} + \mathbf{Q}\mathbf{R}^{-1}\mathbf{B}\mathbf{R}^{-1}\mathbf{Q}' - \frac{\mathbf{C}\mathbf{R}^{-1}\mathbf{Q}' + \mathbf{Q}\mathbf{R}^{-1}\mathbf{C}'}{2}. \quad (1.48)$$

Il est clair que $\boldsymbol{\Phi}$ est symétrique, par construction et par symétrie de \mathbf{A} , \mathbf{B} , et de \mathbf{R} . On montre que cette matrice est définie positive en considérant, pour tout vecteur $\mathbf{g} \in \mathbb{R}^p$, l'unique spline d'interpolation g associée à \mathbf{g} et aux noeuds t_1, \dots, t_p . On a $\mathbf{g}'\boldsymbol{\Phi}\mathbf{g} = |g|_0^2 \geq 0$, et si $\mathbf{g}'\boldsymbol{\Phi}\mathbf{g} = 0$, alors la fonction g est nulle, ce qui implique que $\mathbf{g} = 0$. La démonstration du théorème est alors achevée. ■

Démonstration du Corollaire 1.1

Ce corollaire est une application directe du Théorème 1.7 à la fonction spline cubique naturelle $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$. On rappelle que dans le modèle équilibré (1.4), la matrice de poids des observations \mathbf{W} est égale à l'identité et que la matrice de lissage $\mathbf{S}_\lambda = (\mathbf{I} + \lambda p \boldsymbol{\Gamma})^{-1}$ est symétrique. Il vient alors

$$\begin{aligned} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 &= \left(\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\right)' \boldsymbol{\Phi} \left(\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\right) \\ &= (\mathbf{S}_\lambda(\bar{\mathbf{y}} - \mathbf{f}))' \boldsymbol{\Phi} (\mathbf{S}_\lambda(\bar{\mathbf{y}} - \mathbf{f})) \\ &= \bar{\boldsymbol{\varepsilon}}' \mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}}. \end{aligned} \quad (1.49)$$

En prenant l'espérance de la formule précédente, on obtient

$$\mathbb{E}|\mathbf{f}_\lambda - \widehat{\mathbf{f}}_\lambda|_0^2 = \mathbb{E}\bar{\boldsymbol{\varepsilon}}' \mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{S}_\lambda \bar{\boldsymbol{\varepsilon}} = \text{tr}\{\mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{S}_\lambda \mathbb{E}(\bar{\boldsymbol{\varepsilon}}\bar{\boldsymbol{\varepsilon}}')\} = n^{-1} \text{tr} \mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{S}_\lambda \mathbf{V},$$

ce qu'il fallait démontrer. ■

Démonstration du Lemme 1.1

Le Lemme 1.1, qui majore la trace de la matrice \mathbf{S}_λ^2 , se ramène à la minoration des valeurs propres $\lambda_1 \geq \dots \geq \lambda_{p-m} > 0$ de la matrice d'énergie $\mathbf{\Gamma}$ (on rappelle que $\lambda_{p-m+1} = \dots = \lambda_p = 0$, cf. (1.22a)). Une fois le minorant obtenu, il est aisé de majorer la trace précédente en reprenant pas à pas la démonstration de la Proposition 1.3. L'apport du présent résultat, inspiré du Lemme 3.4 de Nussbaum (1985), est de donner une valeur explicite à la constante C_1 de la formule (1.22b) qui minore les valeurs propres de $\mathbf{\Gamma}$.

En notant $[x]$ la partie entière d'un réel x , nous partons de l'inégalité

$$\frac{(1-p^{-1})p}{(\pi j)^{2m}} \lambda_{p-j+1} \geq \left(\sum_{l \in \mathbb{Z}} (2[(j-m-1)/2]j^{-1} + 2l(p-1)j^{-1})^{-2m} \right)^{-1}, \quad (1.50)$$

tirée de la preuve du Lemme 3.4 de Nussbaum (1985), et valable pour tout $p \geq m+5$ et $j = m+3, \dots, p-2$.

Nous allons majorer la série du terme de droite en la décomposant sur $\{\mathbb{Z} - \mathbb{N}\} \cup \{0\} \cup \mathbb{N}^*$.

Commençons par fixer $p \geq m+5$ et par écrire, pour tout $j = m+3, \dots, p-2$,

$$\begin{aligned} \sum_{l < 0} (2[(j-m-1)/2]j^{-1} + 2l(p-1)j^{-1})^{-2m} \\ = \sum_{l > 0} (2l(p-1)j^{-1} - 2[(j-m-1)/2]j^{-1})^{-2m}. \end{aligned}$$

Comme $2[(j-m-1)/2] \leq (j-m-1) < 2l(p-1)$ pour tout $l > 0$, il suit que

$$\begin{aligned} \sum_{l < 0} (2[(j-m-1)/2]j^{-1} + 2l(p-1)j^{-1})^{-2m} \\ \leq \sum_{l > 0} \left(\frac{2l(p-1) + m+1}{j} - 1 \right)^{-2m}. \end{aligned}$$

Le majorant précédent est maximal pour $j = p-2$. Il vient alors

$$\sum_{l < 0} (2[(j-m-1)/2]j^{-1} + 2l(p-1)j^{-1})^{-2m} \leq \sum_{l > 0} (2l-1)^{-2m}. \quad (1.51)$$

Par ailleurs, un calcul simple permet de montrer que

$$\begin{aligned} \sum_{l > 0} (2[(j-m-1)/2]j^{-1} + 2l(p-1)j^{-1})^{-2m} &\leq \sum_{l > 0} (2l(p-1)j^{-1})^{-2m} \\ &\leq \sum_{l > 0} (2l)^{-2m}. \end{aligned} \quad (1.52)$$

En observant que le terme $2\lfloor(j-m-1)/2\rfloor j^{-1}$ correspondant à l'indice $l=0$ dans la série de (1.50) est inférieur à 1, et en regroupant cette majoration avec les inégalités (1.51) et (1.52), on obtient

$$\sum_{l \in \mathbb{Z}} \left(2\lfloor(j-m-1)/2\rfloor j^{-1} + 2l(p-1)j^{-1} \right)^{-2m} \leq 1 + \sum_{l>0} l^{-2m}.$$

Injectant ensuite cette relation dans (1.50), nous en déduisons que

$$j^{-2m} p \lambda_{p-j+1} \geq \left(1 + \sum_{l>0} l^{-2m} \right)^{-1} \frac{\pi^{2m}}{1-p^{-1}} \geq C_1, \quad (1.53)$$

pour tout $j = m+3, \dots, p-2$, avec $C_1 = C_1(m) = \left(1 + \sum_{l>0} l^{-2m} \right)^{-1} \pi^{2m}$.

Maintenant, d'après les relations (1.22a) et (1.22b), la matrice $\mathbf{\Gamma}$ possède seulement quatre autres valeurs propres non nulles ($\lambda_1 \geq \lambda_2 \geq \lambda_{p-m-1} \geq \lambda_{p-m} > 0$). On peut ainsi faire la majoration

$$\text{tr } \mathbf{S}_\lambda^2 = \sum_{j=1}^p (1 + \lambda p \lambda_j)^{-2} \leq \sum_{j=m+1}^{p-2} (1 + \lambda C_1 j^{2m})^{-2} + (m+2) \quad (1.54)$$

et finir le calcul comme dans la Proposition 1.3. ■

Démonstration du Lemme 1.2

Afin de majorer la norme $\|\Phi\|$ Dans le cas où les points d'observations du modèle (1.4) sont équidistants ($t_j = (j-1)/(p-1)$ pour tout $1 \leq j \leq p$), les matrices \mathbf{A} , \mathbf{B} , \mathbf{C} , \mathbf{Q} , et \mathbf{R} définies par (1.44)–(1.46) et (1.38)–(1.39), ont une forme simple. On a dans ce cas $h_0 = h_p = 0$ et $h_j = 1/(p-1)$ pour $1 \leq j \leq p-1$. Ainsi, il vient

$$\mathbf{A} = \frac{(p-1)^{-1}}{6} \begin{pmatrix} 6 & 1 & 0 & \cdots & 0 \\ 1 & 4 & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & 4 & 1 \\ 0 & \cdots & 0 & 1 & 6 \end{pmatrix}, \quad \mathbf{B} = \frac{(p-1)^{-5}}{15120} \begin{pmatrix} 64 & 31 & 0 & \cdots & 0 \\ 31 & 64 & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & 64 & 31 \\ 0 & \cdots & 0 & 31 & 64 \end{pmatrix},$$

$$\mathbf{C} = \frac{(p-1)^{-3}}{180} \begin{pmatrix} 7 & 0 & \cdots & 0 \\ 16 & \ddots & \ddots & 0 \\ 7 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 7 \\ \vdots & \ddots & \ddots & 16 \\ 0 & \cdots & 0 & 7 \end{pmatrix}, \mathbf{Q} = (p-1) \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -2 & \ddots & \ddots & 0 \\ 1 & \ddots & \ddots & 0 \\ 0 & \ddots & \ddots & 1 \\ \vdots & \ddots & \ddots & -2 \\ 0 & \cdots & 0 & 1 \end{pmatrix},$$

$$\text{et enfin } \mathbf{R} = \frac{(p-1)^{-1}}{6} \begin{pmatrix} 4 & 1 & 0 & \cdots & 0 \\ 1 & \ddots & \ddots & 0 & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \ddots & 1 \\ 0 & \cdots & 0 & 1 & 4 \end{pmatrix}.$$

On rappelle que la matrice \mathbf{A} est $p \times p$, que \mathbf{B} et \mathbf{R} sont $(p-2) \times (p-2)$, et que \mathbf{C} et \mathbf{Q} sont $p \times (p-2)$.

Nous rappelons ici quelques propriétés élémentaires des normes matricielles, *cf.* par exemple Ciarlet (1982).

Proposition 1.9. *Soient deux matrices \mathbf{U} et \mathbf{W} de dimensions compatibles. Alors*

$$\|\mathbf{UW}\| \leq \|\mathbf{U}\| \|\mathbf{W}\|, \quad \|\mathbf{U}\| = \|\mathbf{U}'\mathbf{U}\|^{1/2}, \quad \text{et} \quad \|\mathbf{U}\| = \|\mathbf{U}'\|.$$

Nous pouvons alors écrire

$$\begin{aligned} \|\Phi\| &= \left\| \mathbf{A} + \mathbf{QR}^{-1}\mathbf{BR}^{-1}\mathbf{Q}' - \frac{1}{2}(\mathbf{QR}^{-1}\mathbf{C}' + \mathbf{CR}^{-1}\mathbf{Q}') \right\| \\ &\leq \|\mathbf{A}\| + \|\mathbf{QR}^{-1}\mathbf{BR}^{-1}\mathbf{Q}'\| + \frac{1}{2}(\|\mathbf{QR}^{-1}\mathbf{C}'\| + \|\mathbf{CR}^{-1}\mathbf{Q}'\|) \\ &\leq \|\mathbf{A}\| + \|\mathbf{Q}\|^2 \|\mathbf{R}^{-1}\|^2 \|\mathbf{B}\| + \|\mathbf{Q}\| \|\mathbf{R}^{-1}\| \|\mathbf{C}\| \\ &= \|\mathbf{A}\| + \|\mathbf{Q}'\mathbf{Q}\| \|\mathbf{R}^{-1}\|^2 \|\mathbf{B}\| + \|\mathbf{Q}'\mathbf{Q}\|^{1/2} \|\mathbf{R}^{-1}\| \|\mathbf{C}'\mathbf{C}\|^{1/2}. \end{aligned} \quad (1.55)$$

Il reste à majorer les normes des matrices carrées \mathbf{A} , \mathbf{B} , $\mathbf{C}'\mathbf{C}$, $\mathbf{Q}'\mathbf{Q}$ et \mathbf{R}^{-1} . Cette opération s'effectue sans difficulté à l'aide du théorème des disques de Gersgorin énoncé ci-après.

Théorème 1.12. (Gersğorin) Soit $\mathbf{M} = (m_{ij})_{1 \leq i, j \leq n}$ une matrice carrée complexe. Alors chaque valeur propre de \mathbf{M} est située dans au moins un des disques

$$\mathcal{D}_i = \left\{ z \in \mathbb{C} : |z - m_{ii}| < \sum_{\substack{j \neq i \\ j=1}}^n |m_{ij}| \right\}, \quad i = 1, \dots, n.$$

Il vient d'abord

$$\begin{aligned} \|\mathbf{A}\| &\leq \frac{1}{6(p-1)} \max(6 + |1|, 4 + |1| + |1|) = \frac{7}{6(p-1)}, \\ \|\mathbf{B}\| &\leq \frac{1}{15120(p-1)^5} \max(64 + |31|, 64 + |31| + |31|) = \frac{7}{840(p-1)^5}. \end{aligned} \quad (1.56)$$

Puis on calcule les matrices $\mathbf{C}'\mathbf{C}$ et $\mathbf{Q}'\mathbf{Q}$ à partir des matrices tribandes \mathbf{C} et \mathbf{Q} . On obtient des matrices bandes d'ordre 5, de taille $(p-2) \times (p-2)$, définies respectivement par

$$[\mathbf{C}'\mathbf{C}]_{ij} = \frac{1}{180^2(p-1)^6} \begin{cases} 7^2 + 16^2 + 7^2 = 210 & \text{si } i = j, \\ 16 \cdot 7 + 7 \cdot 16 = 224 & \text{si } |i - j| = 1, \\ 7 \cdot 16 = 112 & \text{si } |i - j| = 2, \\ 0 & \text{si } |i - j| > 2. \end{cases}$$

$$\text{et } [\mathbf{Q}'\mathbf{Q}]_{ij} = (p-1)^2 \begin{cases} 1^2 + (-2)^2 + 1^2 = 6 & \text{si } i = j, \\ (-2) \cdot 1 + 1 \cdot (-2) = -4 & \text{si } |i - j| = 1, \\ 1 \cdot 1 = 1 & \text{si } |i - j| = 2, \\ 0 & \text{si } |i - j| > 2. \end{cases}$$

L'application du Théorème de Gersğorin fournit alors les majorations

$$\begin{aligned} \|\mathbf{C}'\mathbf{C}\| &\leq \frac{1}{180^2(p-1)^6} (210 + 2 \cdot |224| + 2 \cdot |112|) = \frac{770}{180^2(p-1)^6}, \\ \|\mathbf{Q}'\mathbf{Q}\| &\leq (p-1)^2 (6 + 2 \cdot |-4| + 2 \cdot |1|) = 16(p-1)^2. \end{aligned} \quad (1.57)$$

Enfin, pour majorer $\|\mathbf{R}^{-1}\|$, on procède indirectement en minorant la plus petite valeur propre de \mathbf{R} , notée $\lambda_{\min}(\mathbf{R})$. L'inverse du minorant trouvé est alors un majorant de $\|\mathbf{R}^{-1}\|$. Avec le théorème de Gersğorin, nous pouvons écrire $\lambda_{\min}(\mathbf{R}) \geq \{\min(4 - |1|, 4 - |1| - |1|)\} / (6(p-1)) = 1/(3(p-1))$ et par conséquent,

$$\|\mathbf{R}^{-1}\| \leq 3(p-1). \quad (1.58)$$

En insérant les formules (1.56)–(1.58) dans (1.55), on parvient à la majoration

$$\|\Phi\| \leq \frac{1}{p-1} \left(\frac{7}{6} + \frac{7 \cdot 16 \cdot 3^2}{840} + \frac{4 \cdot 3 \cdot \sqrt{770}}{180} \right) \leq \frac{4,21660}{p-1},$$

ce qui termine la démonstration du lemme. ■

Démonstration du Théorème 1.8

Le Théorème 1.8 résulte du Théorème 1.3 pour la majoration du biais au carré (cas $m = 2$ où $f \in W^2[0, 1]$ est estimée par une spline cubique) et des Corollaires 1.1 et 1.2 ainsi que du Lemme 1.1 pour l'étude de la variance.

La majoration du terme de biais se ramène au calcul effectif des constantes C et D du Théorème 1.3. Dans ce but, nous utilisons la relation

$$|g|_k^2 \leq 4(k+1) \binom{2k}{k} \frac{h_{max}}{h_{min}^{2k}} \sum_{j=1}^p \{g(t_j)\}^2 + 4(k+1)^2 h_{max}^2 |g|_{k+1}^2 \quad (1.59)$$

tirée de la preuve du Théorème 1.10 de Ragozin (1983) et valable pour tout entier $k \geq 0$, toute fonction $g \in W_2^{k+1}[0, 1]$ et tous points $0 \leq t_1 < \dots < t_p \leq 1$. On applique alors deux fois cette formule ($k = 0, 1$) à la fonction $g = f - \mathbb{E}f_\lambda$, en se rappelant que $h_{min} = h_{max} = 1/(p-1)$ par l'hypothèse d'équidistance des points $t_j = (j-1)/(p-1)$, $1 \leq j \leq p$. Il vient

$$|f - \mathbb{E}f_\lambda|_0^2 \leq \frac{68}{p-1} \|\mathbf{f} - \mathbb{E}\mathbf{f}_\lambda\|^2 + \frac{64}{(p-1)^4} |f - \mathbb{E}f_\lambda|_2^2.$$

Il reste à utiliser la majoration $|f - \mathbb{E}\hat{f}_\lambda|_m^2 \leq |f|_m^2$ faite dans la démonstration du Théorème 1.5 (étant valable pour un dispositif déséquilibré et pour $m > 0$ quelconque, elle l'est *a fortiori* dans le cadre équilibré pour $m = 2$) et la majoration (1.13) du biais discret. Nous obtenons ainsi l'inégalité

$$|f - \mathbb{E}f_\lambda|_0^2 \leq \left(\frac{17p\lambda}{p-1} + \frac{64}{(p-1)^4} \right) |f|_2^2. \quad (1.60)$$

Pour majorer la variance de l'estimateur, on utilise d'abord le Corollaire 1.1 et le Lemme 1.3. Il en résulte que

$$\mathbb{E}|f_\lambda - \mathbb{E}\hat{f}_\lambda|_0^2 = n^{-1} \text{tr}(\mathbf{S}_\lambda \Phi \mathbf{S}_\lambda \mathbf{V}) \leq n^{-1} \min \left\{ \text{tr} \mathbf{S}_\lambda^2 \|\Phi\| \|\mathbf{V}\|, \|\mathbf{S}_\lambda\|^2 \|\Phi\| \text{tr} \mathbf{V} \right\}.$$

On fait alors appel au Corollaire 1.2, au Lemme 1.1 et à l'équation (1.15) pour écrire

$$\mathbb{E}|\hat{f}_\lambda - \mathbb{E}\hat{f}_\lambda|_0^2 \leq \frac{4, 21660}{n(p-1)} \left(4 + \frac{1}{(C_1\lambda)^{1/4}} \int_0^\infty \frac{dt}{(1+t^{2m})^2} \right) \|\mathbf{V}\| \quad (1.61)$$

et

$$\mathbb{E}|\hat{f}_\lambda - \mathbb{E}\hat{f}_\lambda|_0^2 \leq \frac{4, 21660}{n(p-1)} \text{tr} \mathbf{V}. \quad (1.62)$$

Restent à calculer la constante $C_1 = (1 + \sum_{l>0} l^{-4})^{-1} \pi^4$ et l'intégrale précédente. La somme $\sum_{l>0} l^{-4}$ est égale à $\pi^4/90$ (fonction zeta de Riemann). De plus, un calcul de primitive donne $\int_0^\infty (1+t^{2m})^{-2} dt = 3\sqrt{2}\pi/16$. Numériquement, il vient l'inégalité $\text{tr} \mathbf{S}_\lambda^2 \leq 4 + 0,318531/\lambda^{1/4}$, ce qui termine la démonstration. ■

Démonstration de la Proposition 1.4.

Pour obtenir des vitesses de convergence dans le cas d'erreurs asymptotiquement décorréées, il suffit d'appliquer les relations (1.5b) et (1.6b) des Théorèmes 1.2 et 1.3 puis de majorer $\|\mathbf{V}\|$ à l'aide du théorème de Gersgorin (Théorème 1.12).

Avec la condition $\lambda p^{2m} \rightarrow \infty$ quand $p \rightarrow \infty$ d'une part, et l'hypothèse de quasi-uniformité du dispositif d'autre part, on voit que le biais intégré ou discrétisé de l'estimateur \widehat{f}_λ est comme $\mathcal{O}(\lambda)$. Les mêmes conditions et hypothèses assurent que la variance intégrée ou discrétisée de \widehat{f}_λ est en $\mathcal{O}(\|\mathbf{V}\|/(p\lambda^{1/2m}))$. Il reste à majorer la norme $\|\mathbf{V}\|$.

À l'aide du Théorème 1.12, on écrit d'abord

$$\|\mathbf{V}\| \leq \max_{1 \leq i \leq p} \left(\text{Var } \varepsilon_i + \sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right). \quad (1.63)$$

On fait alors la majoration

$$\begin{aligned} & \max_{1 \leq i \leq p} \left(\text{Var } \varepsilon_i + \sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right) \\ & \leq \max_{1 \leq i \leq p} \text{Var } \varepsilon_i + \max_{1 \leq i \leq p} \left(\sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right). \end{aligned} \quad (1.64)$$

L'hypothèse de variance uniformément bornée ($\sup_{p \geq 1} \max_{1 \leq i \leq p} \text{Var } \varepsilon_i < \infty$) permet de majorer le terme de gauche du majorant par une constante finie indépendante de p . Nous étudions maintenant la somme de droite sous les différentes hypothèses de bruit de la Proposition 1.4.

Sous l'Hypothèse 1.2 avec $\gamma > 1$ (corrélation de courte portée), la quantité $|\text{Cov}(\varepsilon_i, \varepsilon_j)|$ est de l'ordre de $|i - j|^{-\gamma}$ quand $p \rightarrow \infty$, uniformément en i et j (pour $i \neq j$). Par conséquent, nous en déduisons que

$$\max_{1 \leq i \leq p} \left(\sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right) = \mathcal{O} \left(\max_{1 \leq i \leq p} \left(\sum_{1 \leq j \leq p, j \neq i} |i - j|^{-\gamma} \right) \right).$$

Ensuite, par décroissance de la fonction $t \mapsto t^{-\gamma}$ sur \mathbb{R}_+ , la somme précédente

est maximale pour $i = \lfloor p/2 \rfloor$. Il vient

$$\begin{aligned} \max_{1 \leq i \leq p} \left(\sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right) &= \mathcal{O} \left(\sum_{1 \leq j \leq p, j \neq \lfloor p/2 \rfloor} |\lfloor p/2 \rfloor - j|^{-\gamma} \right) \\ &= \mathcal{O} \left(2 \sum_{j=1}^{\lfloor p/2 \rfloor} j^{-\gamma} \right) = \mathcal{O}(1), \end{aligned} \quad (1.65)$$

puisque $\gamma > 1$. En joignant les relations (1.63)–(1.65) et l’hypothèse de variance des erreurs uniformément bornée, il vient donc $\|\mathbf{V}\| = \mathcal{O}(1)$. On obtient ainsi les majorations et vitesses de convergence voulues.

Notons que si la corrélation est de longue portée ($0 < \gamma < 1$), on a alors $\|\mathbf{V}\| = \mathcal{O}(\sum_{j=1}^p j^{-\gamma}) = \mathcal{O}(p^{1-\gamma})$. La variance discrétisée ou intégrée de \hat{f}_λ est alors seulement majorée par $\mathcal{O}(p^{-\gamma} \lambda^{-1/2m})$, ce qui ne permet pas d’obtenir la vitesse minimax $p^{-2m\gamma/(2m+\gamma)}$ annoncée par Hall et Hart (1990) et Wang (1996).

Sous l’Hypothèse 1.3 (erreurs de type processus ARMA stationnaire inversible), nous savons (*cf.* par exemple Brockwell et Davis (1987)) que l’autocorrélation des erreurs vérifie, pour tous entiers $i, j \geq 0$, la relation

$$\text{Cov}(\varepsilon_i, \varepsilon_{i+j}) = \sum_{k=1}^K \alpha_k \frac{j^{m_k-1}}{|r_k|^j}, \quad (1.66)$$

où r_1, \dots, r_K sont les racines de multiplicité respective m_1, \dots, m_K du polynôme associé à la partie AR du processus, et où les $\alpha_1, \dots, \alpha_K$ sont des coefficients constants. En exploitant la stationnarité des erreurs et la relation (1.66), nous obtenons

$$\begin{aligned} \max_{1 \leq i \leq p} \left(\sum_{1 \leq j \leq p, j \neq i} |\text{Cov}(\varepsilon_i, \varepsilon_j)| \right) &\leq 2 \sum_{j=1}^p |\text{Cov}(\varepsilon_1, \varepsilon_j)| \\ &\leq 2 \sum_{j=1}^{p-1} \sum_{k=1}^K |\alpha_k| \frac{j^{m_k-1}}{|r_k|^j} = 2 \sum_{k=1}^K \sum_{j=1}^{p-1} |\alpha_k| \frac{j^{m_k-1}}{|r_k|^j} \\ &\leq 2 \sum_{k=1}^K \sum_{j=1}^{\infty} |\alpha_k| \frac{j^{m_k-1}}{|r_k|^j} < \infty, \end{aligned}$$

car par hypothèse d’inversibilité du processus ARMA, les racines r_1, \dots, r_K ont toutes leur modules $|r_k| > 1$. Nous concluons à nouveau que $\|\mathbf{V}\| = \mathcal{O}(1)$.

Nous passons maintenant à l'Hypothèse 1.4 d'un bruit mélangeant. Seule la φ -mélangeance est considérée ici, la α -mélangeance se traitant de façon analogue. Nous employons d'abord le Théorème 5.1 de Roussas et Ionnides (1987) repris par Burman (1991) : pour tous les entiers i, j, p tels que $1 \leq i \neq j \leq p$, on a

$$|\text{Cov}(\varepsilon_i, \varepsilon_j)| \leq \sqrt{\text{Var } \varepsilon_i} \sqrt{\text{Var } \varepsilon_j} \sqrt{\varphi(R_p |t_j - t_i|)}. \quad (1.67)$$

Avec la décroissance de φ et la quasi-uniformité du dispositif expérimental (celle-ci implique qu'il existe une constante finie $C > 0$ indépendante de p telle que pour tous $1 \leq i, j \leq p$, $|t_j - t_i| \geq C|j - i|/p$), la majoration (1.63) devient

$$\begin{aligned} \|\mathbf{V}\| &\leq \left(\max_{1 \leq i \leq p} \text{Var } \varepsilon_i \right) \left(1 + \max_{1 \leq i \leq p} \left(\sum_{j=1, j \neq i}^p \sqrt{\varphi(CR_p |i - j|/p)} \right) \right) \\ &\leq \left(\max_{1 \leq i \leq p} \text{Var } \varepsilon_i \right) \left(1 + 2 \sum_{j=1}^{\lfloor p/2 \rfloor} \sqrt{\varphi(CR_p j/p)} \right) \\ &\leq \left(\max_{1 \leq i \leq p} \text{Var } \varepsilon_i \right) \left(1 + \int_0^{\lfloor p/2 \rfloor} \sqrt{\varphi(CR_p x/p)} dx \right) \\ &\leq \left(\max_{1 \leq i \leq p} \text{Var } \varepsilon_i \right) \left(1 + \frac{2p}{CR_p} \int_0^{CR_p/2} \sqrt{\varphi(u)} du \right). \end{aligned} \quad (1.68)$$

Dans le calcul précédent, on a utilisé la décroissance de φ pour la première majoration, pour la recherche du maximum de $\sum_{j=1, j \neq i}^p \sqrt{\varphi(CR_p |i - j|/p)}$ par rapport à i , et pour la comparaison de $\sum_{j=1}^{\lfloor p/2 \rfloor} \sqrt{\varphi(CR_p j/p)}$ avec l'intégrale associée. Le dernier passage résulte du changement de variable $u = CR_p x/p$ et de l'inégalité $\lfloor p/2 \rfloor \leq p/2$.

Nous déduisons ainsi, grâce à l'hypothèse de variance uniformément bornée ($\sup_{p \geq 1} \max_{1 \leq i \leq p} \mathbb{E} \varepsilon_i^2 < \infty$) et aux hypothèses $\int_0^\infty \varphi^{1/2}(x) dx < \infty$ et $R_p \leq p$, que

$$\|\mathbf{V}\| = \mathcal{O}(p/R_p) \quad (1.69)$$

lorsque $p \rightarrow \infty$. La démonstration est alors complète. ■

Démonstration de la Proposition 1.6

Dans cette proposition, nous considérons dans le modèle (1.1) des données longitudinales telles que $n \rightarrow \infty$, avec $\max_{1 \leq i \leq n} p_i = \mathcal{O}(1)$. Nous établissons alors des vitesses de convergence dans deux cas typiques.

Dans le cas 1 ($p < \infty$ et $\min n_j \sim n \sim N$), la mesure d'erreur appropriée est la MASE, étant donné que l'ensemble des points d'observation (ou temps de mesure) est fixé indépendamment de n . Pour l'asymptotique considérée, d'après la relation (1.7a) du Théorème 1.4, le biais au carré de l'estimateur spline de lissage est en $\mathcal{O}(N\lambda/(\min n_j p)) = \mathcal{O}(\lambda)$ et sa variance est en $\mathcal{O}(\text{tr } \mathbf{V}/(\min n_j p)) = \mathcal{O}((\min n_j)^{-1}) = \mathcal{O}(n^{-1})$. Ceci permet alors de conclure.

Dans le cas 2 ($\max n_j = \mathcal{O}(1)$), nous voyons que $N = \sum_{j=1}^p n_j \leq p \max n_j = \mathcal{O}(p)$. Comme par ailleurs $p \leq N$, il vient $p \sim n \sim N$ dans l'asymptotique, c'est à dire que le nombre de points distincts d'observation, le nombre d'individus et le nombre total de données sont du même ordre. En combinant le fait que $p = p(n) \rightarrow \infty$ et la quasi-uniformité du dispositif expérimental, nous constatons que l'ensemble des points d'observation devient dense dans $[0,1]$ quand $n \rightarrow \infty$. Par conséquent, nous pouvons nous intéresser à la mesure d'erreur MISE.

La quasi-uniformité du dispositif et la condition $\max n_j = \mathcal{O}(1)$ garantissent que les constantes $\widetilde{M}_1(m, h_{\max}/h_{\min})$ et $\widetilde{M}_2(m, h_{\max}/h_{\min}, \max n_j/\min n_j)$ de la Proposition 1.3 restent bornées loin de zéro et de l'infini quand $n \rightarrow \infty$. La relation (1.8b) du Théorème 1.5 majore alors le biais au carré de l'estimateur \widehat{f}_λ par $\mathcal{O}((Nh_{\max}\lambda)/(\min n_j))$ et sa variance par

$$\frac{1}{p} \sum \text{Var}(\widehat{f}_\lambda(t_j)) = \mathcal{O} \left(\left\{ h_{\max} + h_{\max}^{2m}/(p\lambda) \right\} \frac{(\max n_j)^2 p \|\mathbf{V}_{abs}\|}{(\min n_j)^2 N \lambda^{1/2m}} \right). \quad (1.70)$$

En utilisant ensuite la condition $p^{-2m} = \mathcal{O}(\lambda)$ et la relation (1.11), ces majorants deviennent respectivement $\mathcal{O}(\lambda)$ et $\mathcal{O}(\|\mathbf{V}_{abs}\|/(N\lambda^{1/2m}))$.

Il reste à majorer la norme $\|\mathbf{V}_{abs}\|$. Étant donné des entiers $1 \leq j \leq p$ et $n \geq 1$, la condition $\max n_j = \mathcal{O}(1)$ signifie que le nombre d'individus faisant l'objet d'une mesure au point t_j est majoré par une constante finie $m_1 > 0$ indépendante de j et de p , et la condition $\max_{1 \leq i \leq n} p_i = \mathcal{O}(1)$ établit que pour chacun de ces individus i , le nombre de mesures est majoré par une constante finie $m_2 > 0$ indépendante de i et de n . Nous en déduisons que la matrice $\mathbf{V}_{abs} = (|v_{jk}|)_{1 \leq j, k \leq p}$ (on rappelle que $v_{jk} = \text{Cov}(\varepsilon_{il}, \varepsilon_{il'})$ s'il existe un individu i faisant à la fois l'objet de mesures au point $x_{il} = t_j$ et au point $x_{il'} = t_k$, et $v_{jk} = 0$ sinon, cf. Section 1.1) n'a qu'un nombre fini de coefficients non nuls sur sa j -ème ligne, ce nombre étant majoré par $m_1 m_2$. Chacun des coefficients étant majoré par la quantité $\max_{1 \leq j \leq p_i, 1 \leq i \leq n} \text{Var } \varepsilon_{ij}$, elle-même majorée par une constante finie $C > 0$ indépendante de n et de p_1, \dots, p_n d'après l'hypothèse de variance uniformément bornée, nous pouvons, à l'aide du Théorème de Gersgorin (cf. (1.63)), faire l'évaluation

$$\|\mathbf{V}_{abs}\| \leq m_1 m_2 C \quad (1.71)$$

pour toutes les valeurs de n et de p . Nous déduisons de ce qui précède que la variance intégrée de \widehat{f}_λ est en $\mathcal{O}(N^{-1}\lambda^{-1/2m}) = \mathcal{O}(n^{-1}\lambda^{-1/2m})$, ce qui finit d'établir la proposition. ■

Chapitre 2

Estimation de la fonction moyenne d'un processus aléatoire

Ce chapitre est consacré à l'estimation non paramétrique de la fonction moyenne d'un processus aléatoire $X = \{X(t); t \in [0, 1]\}$ de variance finie sur $[0, 1]$. Des résultats de consistance, de convergence, et de normalité asymptotique sont établis pour différents estimateurs de $f = \mathbb{E}X$ construits à partir d'un échantillon i.i.d. (X_1, \dots, X_n) de X observé en p points déterministes, dans l'asymptotique $n, p \rightarrow \infty$.

2.1 Introduction

2.1.1 Historique

La modélisation d'un phénomène par un processus aléatoire à temps continu et les problèmes associés de prévision et d'estimation interviennent dans de nombreux domaines appliqués, comme l'étude de la croissance humaine (Müller, 1988), l'étude de courbes dose-réponse (Grizzle et Allen, 1969; Verbyla *et al.*, 1999), la chimométrie (Diggle et Hutchinson, 1989), la chimie moléculaire (Ycart *et al.*, 1990), la pluviométrie (Besse *et al.*, 1997), ou le traitement du signal (Kay, 1993). Ce type de modélisation est particulièrement adapté au traitement de mesures répétées sur un échantillon aléatoire d'une population (Hart et Wehrly, 1986). Par ailleurs, le processus à temps continu est l'objet essentiel du champ statistique des données fonctionnelles, qui connaît un essor considérable depuis quelques années (voir, parmi une longue liste, Ramsay et Silverman, 1997; Cardot, 2000; Besse *et al.*, 2005; Masry, 2005; Ferraty et Vieu, 2004, 2006, et les références bibliographiques qu'ils contiennent).

Du point de vue théorique, plusieurs travaux ont porté sur l'estimation de la moyenne f d'un processus aléatoire $X = \{X(t); t \in [0, 1]\}$. Parzen (1961) a été parmi

les premiers à étudier l'estimation par maximum de vraisemblance dans le cas où X est gaussien et où sa fonction de covariance est connue. Il a construit un tel estimateur en supposant que l'espace du paramètre f est de dimension finie, et il a souligné l'impossibilité d'une telle construction dans le cas où cet espace est de dimension infinie, la vraisemblance n'étant alors pas bornée. Grenander (1981), Beder (1987), et Anilkumar (1994) ont remédié à ce problème en introduisant un critère de vraisemblance restreinte (en anglais, "sieve estimator"). Beder (1987) a établi plusieurs résultats de consistance pour son estimateur basé sur un échantillon i.i.d. (X_1, \dots, X_n) de X lorsque $n \rightarrow \infty$. Dans le même cadre gaussien, Berger et Wolpert (1983) ont exposé le phénomène de Stein selon lequel l'estimateur minimax usuel X est inadmissible, et ils ont proposé de meilleurs estimateurs basés sur des développements de type Karhunen-Loève. Peele et Kimeldorf (1977) se sont quant à eux intéressés à plusieurs problèmes d'estimation minimax ou d'estimation linéaire sans biais de variance minimale.

Un autre cas étudié dans la littérature est celui où un échantillon (X_1, \dots, X_n) de X est observé en un nombre fini p de points déterministes et où la fonction de covariance R est inconnue. En supposant que X est stationnaire au second ordre et que sa fonction d'autocorrélation est régulière, Hart et Wehrly (1986) ont donné un développement asymptotique de la variance d'un estimateur à noyau lorsque $n, p \rightarrow \infty$. Ils ont observé que l'estimateur n'est consistant que si $n \rightarrow \infty$, que dans ce cas sa variance ponctuelle est de l'ordre de n^{-1} , et ils ont fourni un choix théorique optimal de la largeur de fenêtre h . Sous l'hypothèse d'une covariance R bornée, Cardot et Diack (1998) et Degras et Jallet (2005) ont établi une convergence analogue en $\mathcal{O}(n^{-1})$ pour la variance intégrée, respectivement, des estimateurs splines hybrides et splines de lissage de f lorsque $n, p \rightarrow \infty$. Sous la même hypothèse pour R et en supposant de plus que f appartient à l'espace de Sobolev $W_2^m[0, 1] = \{g : g^{(0)}, \dots, g^{(m-1)} \text{ abs. continue ; } g^{(m)} \in L_2[0, 1]\}$, Biritxinaga (1987) a démontré la consistance presque sûre dans $W_2^m[0, 1]$ de trois estimateurs splines de f lorsque $n, p \rightarrow \infty$ au sens d'une certaine filtration de \mathbb{N}^2 .

Pour finir, nous signalons deux problèmes statistiques connexes à l'estimation de la moyenne d'un processus aléatoire, à savoir l'analyse en composantes principales fonctionnelles [ACPF] et la régression non paramétrique à effets fixes (*i.e.* avec des points d'observations déterministes) avec erreurs corrélées. Précisons tout d'abord que les techniques d'ACPF (voir par exemple Dauxois *et al.*, 1982; Besse, 1991; Ramsay et Silverman, 1997) portent dans leur majorité plus sur l'estimation des composantes aléatoires d'un processus que sur l'estimation de sa moyenne, qui en général est supposée connue. Une exception notable est l'approche de Rice et Silverman (1991) qui, pour l'estimateur spline de lissage de la moyenne, fournit une méthode simple de sélection du paramètre de lissage ("GCV leave one curve out") tenant compte de la structure de covariance des données sans avoir à estimer celle-ci.

D'autre part, l'estimation de la moyenne d'un processus aléatoire peut être vue comme une régression particulière. Cependant, le cadre des processus aléatoires impose une corrélation très forte entre les observations; il implique en particulier que la corrélation de deux mesures faites en des points fixés $s \neq t \in [0, 1]$ ne tend en général pas vers zéro à mesure que la taille du dispositif expérimental tend vers l'infini. Au contraire, les formes de dépendance étudiées dans la régression à effets fixes avec erreurs corrélées, comme la corrélation de courte ou longue portée et la mélangeance (*cf.* Section 1.3.1), impliquent la décorrélation asymptotique. Il en résulte des différences fondamentales entre ces deux cadres quant à la consistance et à la vitesse de convergence des estimateurs. Parmi la vaste littérature sur la régression à effets fixes avec erreurs corrélées, nous pouvons citer les articles de Hall et Hart (1990), Burman (1991), Wang (1998), et Opsomer *et al.* (2001) pour l'étude de vitesses de convergence, et les articles de Roussas *et al.* (1992), Csörgó et Mielniczuk (1995), Liebscher (1999) et Truong-Van et Bru (2001) concernant la normalité asymptotique d'estimateurs.

2.1.2 Modèle et estimateur linéaire

Soit $X = \{X(t); t \in [0, 1]\}$ un processus aléatoire de fonction moyenne f et de fonction de covariance R toutes deux inconnues. On suppose que ces fonctions sont continues sur leurs domaine de définition, et on s'intéresse à l'estimation non paramétrique de f à partir d'un échantillon i.i.d. (X_1, \dots, X_n) de X observé en des points $t_{1p} < \dots < t_{pp}$ fixés dans $[0, 1]$.

Remarque 2.1. *Le présent modèle est un cas particulier du modèle de régression (1.4) de la Section 1.3.2. Par ailleurs, l'hypothèse de continuité sur R exclut le modèle de régression classique $X(t) = f(t) + \varepsilon(t)$ pour lequel ε est un bruit blanc, la covariance $R(s, t) = \sigma^2 \delta_{st}$ (δ_{st} est le symbole de Kronecker) d'un tel bruit n'étant pas continue.*

Les X_1, \dots, X_n jouant un rôle symétrique et le dispositif expérimental étant équilibré, nous considérons un estimateur linéaire de forme générale

$$\hat{f}_{np}(t) = \sum_{i=1}^p W_{nip}(t) \bar{X}_n(t_{ip}) \quad (2.1)$$

où $\bar{X}_n = n^{-1} \sum_{i=1}^n X_i$, et W_{n1p}, \dots, W_{npp} sont des fonctions de poids qui dépendent seulement de n, p , et du dispositif expérimental $\{t_{1p}, \dots, t_{pp}\}$. Remarquons que l'ensemble des estimateurs non paramétriques usuels (à noyau, de projection, splines, etc.) entre dans le cadre de la formule (2.1).

2.1.3 Aperçu des résultats

Dans la Section 2.2 figurent plusieurs résultats asymptotiques sur l'estimateur général \widehat{f}_{np} défini par (2.1), supposé continu et asymptotiquement sans biais. Le résultat principal (Théorème 2.2), obtenu sous de faibles hypothèses de régularité sur la fonction de covariance R , est que lorsque $p \rightarrow \infty$, la variance de $\widehat{f}_{np}(t)$ est équivalente à $n^{-1}R(t, t)$ pour tout $t \in [0, 1]$, uniformément par rapport à n . Ce résultat permet d'obtenir en toute généralité la condition nécessaire et suffisante de consistance $[n, p \rightarrow \infty]$ établie jusqu'alors dans un seul cas particulier par Hart et Wehrly (1986). Plus généralement, on montre à l'aide du Théorème 2.2 que lorsque $n, p \rightarrow \infty$, l'écart quadratique moyen entre l'estimateur \widehat{f}_{np} et la moyenne empirique \overline{X}_n décroît comme $\mathcal{O}(n^{-1})$. Ce fait, utilisé de concert avec le Théorème Central Limite, permet d'établir la normalité asymptotique de \widehat{f}_{np} lorsque $n, p \rightarrow \infty$ (Théorème 2.3). Une application de cette normalité asymptotique à l'obtention d'intervalles de confiances simultanés est alors proposée puis étudiée par simulation en Section 2.2.3.

Dans la Section 2.3, la convergence spécifique des estimateurs splines et des estimateurs à noyaux est étudiée. Les résultats fournis portent sur différents modes de convergence et étendent des résultats antérieurs de la littérature. Le Théorème 2.6 étend les résultats de Hart et Wehrly (1986) sur l'estimateur à noyau de Gasser-Müller au cadre d'un processus à covariance non stationnaire. Le Théorème 2.5 généralise l'article de Biritxinaga (1990) sur la consistance de l'interpolation spline dans les espaces de Sobolev. Enfin, les Théorèmes 2.4 et 2.7 sont des extensions au cadre processus des articles respectifs de Ragozin (1983) et de Nychka (1995) sur l'erreur quadratique des estimateurs splines dans le cadre classique d'erreurs décorréelées. Un point important est que tous les résultats présentés font apparaître explicitement l'influence du paramètre de lissage propre à chaque estimateur et permettent, le cas échéant, de déterminer un paramètre optimal dans l'asymptotique.

Les démonstrations des résultats de ce chapitre sont fournies en Section 2.5.

2.2 Étude asymptotique des estimateurs linéaires

Au Chapitre 1, nous avons montré dans le cadre de processus aléatoires que la variance d'un certain estimateur de la régression f (en l'occurrence une spline de lissage) était en $\mathcal{O}(n^{-1})$ lorsque $n, p \rightarrow \infty$ (voir la Proposition 1.5). Avant de montrer que la variance de l'estimateur linéaire général décrit par (2.1) est équivalente à $n^{-1} \text{Var} X(t) = n^{-1}R(t, t)$ lorsque $n, p \rightarrow \infty$, nous présentons un exemple pour lequel l'équivalence précédente est une égalité exacte valable pour un ensemble fini de données.

Exemple 2.1. Dans le cadre de la Section 2.1.2, soit $X = f + \sum_{k=0}^d P_k \xi_k$ un processus aléatoire, avec $d \geq 1$ un entier, $P_k, 1 \leq k \leq d$, des polynômes de degré respectif k , et $\xi_k, 1 \leq k \leq d$, des v.a. centrées, décorréélées et de variance 1. Soit \widehat{f}_λ la spline de lissage d'ordre $2m$ (avec $p \geq m > d$ et $\lambda > 0$) associée aux observations $\overline{X}_n(t_{1p}), \dots, \overline{X}_n(t_{pp})$. Du fait que le lissage spline d'ordre $2m$ est linéaire par rapport aux données et qu'il laisse invariant les polynômes de degré $< m$, l'estimateur centré $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$ est la spline de lissage associée au processus centré $\overline{X}_n - f$, et comme ce processus est une combinaison linéaire (aléatoire) de polynômes de degré $< m$, il suit que $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda = \overline{X}_n - f$. Par conséquent nous avons l'égalité $\text{Var } \widehat{f}_\lambda(t) = n^{-1}R(t, t)$, valable pour tous $\lambda > 0, t \in [0, 1]$, et pour tous $n \geq 1, p \geq m$.

2.2.1 Consistance en moyenne quadratique

Nous donnons dans cette section des développements asymptotiques de la variance moyenne aux points d'observation (Théorème 2.1) et de la variance ponctuelle (Théorème 2.2) de l'estimateur \widehat{f}_{np} défini par (2.1), lorsque $p \rightarrow \infty$. Ces développements sont valables de manière uniforme par rapport à $n \geq 1$ (voir la Remarque 2.5), ce qui permet de traiter à la fois le cas n fixé et le cas $n \rightarrow \infty$, et d'obtenir ainsi un résultat de consistance en même temps qu'une vitesse de convergence (Corollaire 2.1).

L'outil mathématique principal servant à établir les résultats de cette section est le théorème de Mercer (rappelé en page 82) qui permet de représenter la fonction de covariance $R(s, t)$, continue par hypothèse, comme la limite uniforme de la série $\sum_{k=1}^{\infty} \lambda_k \varphi_k(s) \varphi_k(t)$, où $\lambda_1, \lambda_2, \dots$ et $\varphi_1, \varphi_2, \dots$ sont les valeurs propres et fonctions propres de l'opérateur de covariance associé à R .

Théorème 2.1. *Supposons que*

(i) *la fonction de répartition F_p du dispositif expérimental converge simplement sur $[0, 1]$ vers une limite F lorsque $p \rightarrow \infty$.*

Supposons aussi que lorsque $p \rightarrow \infty$ les poids de l'estimateur \widehat{f}_{np} vérifient, uniformément par rapport à n ,

(ii) $\forall g \in C[0, 1], \quad p^{-1} \sum_{i=1}^p \left(\sum_{j=1}^p W_{njp}(t_{ip}) g(t_{jp}) - g(t_{ip}) \right)^2 = o(1);$

(iii) $\max_{1 \leq i \leq p} \sum_{i=1}^p |W_{njp}(t_{ip})| = \mathcal{O}(1).$

Alors pour tout $n \geq 1$, la variance moyenne de \widehat{f}_{np} aux points d'observation vérifie

$$\frac{1}{p} \sum_{j=1}^p \text{Var } \widehat{f}_{np}(t_j) = \frac{1 + o(1)}{n} \int_0^1 R(t, t) dF(t) \quad (2.2)$$

lorsque $p \rightarrow \infty$, où le $o(1)$ est uniforme par rapport à n .

Nous proposons maintenant une version ponctuelle du résultat précédent sous des hypothèses de régularité renforcées sur R et sur les poids W_{n1p}, \dots, W_{npp} de \widehat{f}_{np} . Ces hypothèses sont vérifiées par de nombreux processus aléatoires et par l'ensemble des estimateurs usuels dans le cadre d'un dispositif expérimental régulier (voir les Remarques 2.2 et 2.3). Dans la suite, nous ferons appel à la notion de fonction hölderienne rappelée ci-après.

Définition 2.1. (Fonctions hölderiennes) Une fonction $g : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ est dite hölderienne s'il existe un réel $C > 0$ et un indice $\beta \in]0, 1]$ tels que $|g(s) - g(t)| \leq C \|s - t\|^\beta$ pour tous $s, t \in U$, où $\|\cdot\|$ désigne une des normes équivalentes de \mathbb{R}^d .

Nous avons alors le résultat suivant.

Théorème 2.2. Soit $t \in [0, 1]$. Considérons les conditions suivantes sur R :

- (i) R est hölderienne d'indice $\gamma \in]0, 1]$;
- (ii) les fonctions propres $\varphi_1, \varphi_2, \dots$ associées à R vérifient $\sup_{k \geq 1} \|\varphi_k\|_\infty < \infty$, où $\|\cdot\|_\infty$ est la norme du supremum.

Considérons aussi les conditions asymptotiques suivantes sur \widehat{f}_{np} (comprises comme devant être réalisées uniformément par rapport à n lorsque $p \rightarrow \infty$) :

- (iii) $\forall g \in C[0, 1], \sum_{i=1}^p W_{nip}(t)g(t_{ip}) - g(t) = o(1)$;
- (iv) $\sum_{i=1}^p |W_{nip}(t)| = \mathcal{O}(1)$;
- (v) $\sum_{i=1}^p |W_{nip}(t)| \cdot |t - t_{ip}|^\gamma = o(1)$, où γ est la constante de la condition (i).

Supposons que l'un des deux groupes de conditions (i)-(iii)-(iv)-(v) ou (ii)-(iii)-(iv) est vérifié. Alors, pour tout $n \geq 1$, la variance de $\widehat{f}_{np}(t)$ est telle que

$$\text{Var } \widehat{f}_{np}(t) = \frac{1 + o(1)}{n} R(t, t) \quad (2.3)$$

lorsque $p \rightarrow \infty$, où le $o(1)$ est uniforme par rapport à n .

Remarque 2.2. Les conditions (i) ou (ii) précédentes sont vérifiées par de nombreux processus aléatoires comme par exemple le processus de Wiener standard ou fractionnaire, le pont brownien, le processus d'Ornstein-Uhlenbeck ou le processus de Poisson.

Remarque 2.3. Aucune hypothèse n'est faite sur le dispositif expérimental dans le théorème précédent car pour la plupart des estimateurs linéaires usuels, plusieurs types de dispositifs réguliers permettent de satisfaire aux conditions (iii), (iv) et/ou (v). Citons par exemple les dispositifs uniformes (i.e. équidistants), quasi-uniformes (voir page 32, un exemple étant celui de points t_1, \dots, t_p tels que $\int_0^{t_i} \psi(t)dt = i/p$ pour tous $1 \leq i \leq p$, où ψ est une densité continue et strictement positive), ou encore les dispositifs admettant une fonction de répartition limite de classe C^1 . Par ailleurs, la condition (iii) de biais asymptotiquement nul est vérifiée par l'ensemble des estimateurs linéaires usuels pour un "bon" choix du paramètre de lissage.

Remarque 2.4. À l'aide du théorème de convergence dominée de Lebesgue, on peut intégrer (2.3) et obtenir la relation $\int_0^1 \text{Var} \widehat{f}_{np}(t) dt = (1 + o(1)) n^{-1} \int_0^1 R(t, t) dt$ lorsque $p \rightarrow \infty$, uniformément par rapport à n .

Remarque 2.5. L'énoncé du Théorème 2.1, resp. du Théorème 2.2, reste valide si l'on remplace à la fois dans les hypothèses (ii)-(iii) et dans le résultat (2.2), resp. dans les hypothèses (iii)-(iv)-(v) et dans le résultat (2.3), l'expression "uniformément par rapport à n " par "pour un entier $n \geq 1$ donné" ou par "lorsque $n, p \rightarrow \infty$ ". Du reste, les conditions uniformes des Théorèmes 2.1 et 2.2 sur les poids de l'estimateur \widehat{f}_{np} se ramènent en général à une condition d'uniformité par rapport à n du paramètre de lissage de \widehat{f}_{np} lorsque $p \rightarrow \infty$.

Pour obtenir plus de détails sur les propriétés des fonctions de poids et sur le biais asymptotique d'estimateurs linéaires, le lecteur pourra consulter Gasser et Müller (1979), Härdle *et al.* (1998), Györfi *et al.* (2002), et Tsybakov (2004) en rapport avec, respectivement, l'estimation par noyaux, par ondelettes, par projection, et par polynômes locaux.

Nous déduisons à présent du Théorème 2.2 et de la continuité de f une condition nécessaire et suffisante de consistance. Cette condition, qui apparaît sans preuve dans de nombreux articles (voir par exemple Altman, 1990; Hart, 1991; Wang, 1998), a été établie théoriquement par Hart et Wehrly (1986) dans le cas particulier de l'estimateur à noyau de Gasser-Müller (voir la Section 2.3.2 pour plus de détails), et a été observée lors de simulations par Diggle et Hutchinson (1989).

Corollaire 2.1. *Sous les hypothèses du Théorème 2.1 (resp. 2.2), l'estimateur \widehat{f}_{np} est consistant en moyenne quadratique discrétisée (resp. ponctuelle) si et seulement si $p \rightarrow \infty$ et $n \rightarrow \infty$. Sa vitesse de convergence est alors au plus de l'ordre de $1/n$.*

2.2.2 Normalité asymptotique

Nous établissons ici la normalité asymptotique de l'estimateur \widehat{f}_{np} en des points fixés $0 \leq \eta_1 < \dots < \eta_q \leq 1$ lorsque $n, p \rightarrow \infty$, à l'aide du Théorème 2.2 et du Théorème Central Limite [TCL].

Supposons que les conditions d'application du Théorème 2.2 sont remplies pour tout $t \in [0, 1]$, et supposons de plus que le biais au carré de \widehat{f}_{np} vérifie

$$(\mathbb{E} \widehat{f}_{np}(t) - f(t))^2 = o(n^{-1}) \quad (2.4)$$

pour tout $t \in [0, 1]$ lorsque $n, p \rightarrow \infty$. Notons que la relation (2.4) est vérifiée par l'ensemble des estimateurs usuels pour un bon choix du paramètre de lissage, à

condition que la fonction moyenne f soit régulière (typiquement, $f \in C^m[0, 1]$ pour un entier $m \geq 1$) et que n soit petit devant p dans l'asymptotique (voir la Section 2.3.2 pour plus de détails).

Décomposons l'écart quadratique moyen entre $\widehat{f}_{np}(t)$ et $\overline{X}_n(t)$ en

$$\mathbb{E}(\widehat{f}_{np}(t) - \overline{X}_n(t))^2 = (\mathbb{E}\widehat{f}_{np}(t) - f(t))^2 + \text{Var}(\widehat{f}_{np}(t) - \overline{X}_n(t))$$

puis décomposons à nouveau le terme de variance et écrivons, avec le Théorème 2.2,

$$\begin{aligned} \text{Var}(\widehat{f}_{np}(t) - \overline{X}_n(t)) &= \text{Var}\widehat{f}_{np}(t) - 2\text{Cov}(\widehat{f}_{np}(t), \overline{X}_n(t)) + \text{Var}\overline{X}_n(t) \\ &= n^{-1} \left((1 + o(1))R(t, t) - 2 \sum_{i=1}^p W_i(t)R(t_i, t) + R(t, t) \right). \end{aligned}$$

La continuité de R et la condition (iii) du Théorème 2.2 impliquent que la somme $\sum_{i=1}^p W_i(t)R(t_i, t)$ apparaissant dans la relation précédente est égale à $R(t, t) + o(1)$ lorsque $n, p \rightarrow \infty$. On obtient ainsi la relation $\text{Var}(\widehat{f}_{np}(t) - \overline{X}_n(t)) = o(n^{-1})$, et avec la formule (2.4), nous en déduisons que

$$\mathbb{E}(\widehat{f}_{np}(t) - \overline{X}_n(t))^2 = o(n^{-1}) \quad (2.5)$$

pour tout $t \in [0, 1]$.

Soient alors $\widehat{\mathbf{f}}_{np}, \mathbf{f}$, et $\overline{\mathbf{X}}_n$ les vecteurs des valeurs respectives de \widehat{f}_{np}, f , et \overline{X}_n aux points η_1, \dots, η_q , et soit Σ la matrice de covariance $(R(\eta_i, \eta_j))_{1 \leq i, j \leq q}$. En appliquant la forme vectorielle du TCL à $\overline{\mathbf{X}}_n$, on obtient la convergence en distribution $n^{1/2}(\overline{\mathbf{X}}_n - \mathbf{f}) \xrightarrow{d} N(\mathbf{0}, \Sigma)$, où $N(\mathbf{0}, \Sigma)$ désigne la loi normale multivariée centrée de covariance Σ , et avec (2.5) qui implique la convergence $n^{1/2}(\widehat{\mathbf{f}}_{np} - \overline{\mathbf{X}}_n) \xrightarrow{d} \mathbf{0}$, il vient enfin le résultat suivant.

Théorème 2.3. *Supposons que les conditions d'application du Théorème 2.2 sont remplies et que le biais au carré de l'estimateur \widehat{f}_{np} vérifie (2.4) pour tout $t \in [0, 1]$. Alors, lorsque $n, p \rightarrow \infty$, nous avons*

$$n^{1/2} \left(\widehat{\mathbf{f}}_{n,p} - \mathbf{f} \right) \xrightarrow{d} N(\mathbf{0}, \Sigma). \quad (2.6)$$

2.2.3 Application : intervalles de confiance simultanés

Nous fournissons à présent une application du Théorème 2.3 à l'obtention d'intervalles de confiance simultanés approchés (ICS) pour la fonction f .

Soient $X_i(t_{jp}), 1 \leq i \leq n, 1 \leq j \leq p$, des observations issues du modèle de la Section 2.1.2, avec n et p grands. On se donne pour objectif la construction d'ICS

au niveau de confiance $1 - \lambda$ pour les valeurs $f(\eta_k)$, $1 \leq k \leq q$, de f en des points fixés que l'on prend parmi les points d'observation t_{jp} , $1 \leq j \leq p$, afin de simplifier l'exposé. On suppose que les hypothèses du Théorème 2.3 sont rendues valides par le processus X dont sont issues les données et par l'estimateur \hat{f}_{np} considéré. On note $\hat{\Sigma} = n^{-1} \sum_{i=1}^n (\mathbf{X}_i - \bar{\mathbf{X}}_n)(\mathbf{X}_i - \bar{\mathbf{X}}_n)'$ la matrice de covariance empirique, avec $\mathbf{X}_i = (X_i(\eta_1), \dots, X_i(\eta_q))'$ pour $1 \leq i \leq n$. D'après la Loi Forte des Grands Nombres, la matrice $\hat{\Sigma}$ converge presque sûrement vers Σ lorsque $n \rightarrow \infty$. En utilisant le théorème de Slutsky dans (2.6), nous pouvons donc affirmer que la variable aléatoire $n(\hat{\mathbf{f}}_{np} - \mathbf{f})' \hat{\Sigma}^{-1} (\hat{\mathbf{f}}_{np} - \mathbf{f})$ suit approximativement une loi du χ^2 à q degrés de liberté, notée $\chi^2(q)$. Remarquons que si $\hat{\Sigma}$ n'est pas inversible, en particulier si $q \geq n$, il est nécessaire d'ajouter une petite perturbation à cette matrice afin que la variable aléatoire précédente soit bien définie. Nous construisons alors un ellipsoïde de confiance pour \mathbf{f} à partir de la formule

$$P \left(n (\hat{\mathbf{f}}_{np} - \mathbf{f})' \hat{\Sigma}^{-1} (\hat{\mathbf{f}}_{np} - \mathbf{f}) \leq z_{1-\lambda} \right) \approx 1 - \lambda, \quad (2.7)$$

où $z_{1-\lambda}$ est le quantile de niveau $(1 - \lambda)$ de la loi $\chi^2(q)$. En désignant par $\hat{\Sigma}_{kk} = n^{-1} \sum_{i=1}^n (X_i(\eta_k) - \bar{X}(\eta_k))^2$ le k -ème terme de la diagonale de $\hat{\Sigma}$ et en exploitant la méthode classique de Scheffé (1959), nous obtenons ensuite les ICS suivants :

$$P \left(f(\eta_k) \in \left[\hat{f}_{np}(\eta_k) \pm \sqrt{\hat{\Sigma}_{kk} z_{1-\lambda}/n} \right], 1 \leq k \leq q \right) \geq 1 - \lambda. \quad (2.8)$$

Nous comparons à présent cette méthode ("méthode A") à une méthode paramétrique standard ("méthode B") lors de la simulation d'un processus de Poisson non homogène X de fonction moyenne $f(t) = 25(1 - \exp(-4t))$.

On utilise le logiciel R pour simuler n réalisations indépendantes de X aux points $t_{jp} = (2j-1)/2p$; $1 \leq j \leq p$, avec les différentes valeurs $n = 10, 20, 50, 100$ et $p = 200$. On calcule des ICS de niveau de confiance 95% aux points $\eta_k = t_{\lfloor kp/q \rfloor, p}$, $1 \leq k \leq q$, pour différentes valeurs de $q = q(n)$ ($\lfloor \cdot \rfloor$ désigne la partie entière).

La méthode A est implémentée avec un estimateur spline cubique de lissage de paramètre 10^{-7} selon la formule (2.8).

La méthode B exploite le fait que X est à accroissements indépendants et que $X(t) - X(s)$ suit une loi de Poisson de paramètre $f(t) - f(s)$ pour tous $0 \leq s < t \leq 1$. Ainsi, les variables aléatoires $n(\bar{X}_n(\eta_k) - \bar{X}_n(\eta_{k-1}))$, $1 \leq k \leq q$, sont indépendantes entre elles et sont distribuées selon la loi de Poisson de paramètre $n(f(\eta_k) - f(\eta_{k-1}))$, pour $1 \leq k \leq q$ respectivement, avec la notation $\eta_0 = 0$. D'après le TCL, on peut approcher les lois de Poisson précédentes par des lois normales pour n grand, ce qui permet d'écrire en notation matricielle

$$\mathbf{A} \bar{\mathbf{X}}_n \stackrel{d}{\approx} N(\mathbf{A} \mathbf{f}, n^{-1} \mathbf{\Delta}),$$

où $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ désigne la loi normale multivariée de moyenne $\boldsymbol{\mu}$ et de matrice de covariance $\boldsymbol{\Sigma}$, \mathbf{A} est la matrice de taille $q \times q$ et de coefficients $a_{ij} = 1$ si $i = j$, $a_{ij} = -1$ si $i - j = 1$, et $a_{ij} = 0$ sinon, et $\boldsymbol{\Delta}$ est la matrice diagonale de taille $q \times q$ et de coefficients non nuls les $\Delta_{kk} = f(\eta_k) - f(\eta_{k-1})$, $1 \leq k \leq q$. Par une manipulation simple, on obtient que le vecteur $n^{1/2}\boldsymbol{\Delta}^{-1/2}\mathbf{A}(\bar{\mathbf{X}}_n - \mathbf{f})$ suit approximativement la loi $N(0, \mathbf{I})$, où \mathbf{I} désigne la matrice identité. De plus, en utilisant le fait que la matrice diagonale $\widehat{\boldsymbol{\Delta}}$ de coefficients non nuls les $\widehat{\Delta}_{kk} = \bar{X}_n(\eta_k) - \bar{X}_n(\eta_{k-1})$, $1 \leq k \leq q$, est un estimateur convergent de $\boldsymbol{\Delta}$ (pour n grand) et en faisant appel au théorème de Slutsky, nous pouvons écrire que

$$n^{1/2}\widehat{\boldsymbol{\Delta}}^{-1/2}\mathbf{A}(\bar{\mathbf{X}}_n - \mathbf{f}) \stackrel{d}{\approx} N(0, \mathbf{I}).$$

Nous pouvons à nouveau, comme dans (2.7) et (2.8), déterminer un ellipsoïde de confiance pour \mathbf{f} avec la loi $\chi^2(q)$, puis transformer cet ellipsoïde en ICS par la méthode de Scheffé. Nous obtenons alors successivement

$$P\left(n(\bar{\mathbf{X}}_n - \mathbf{f})'\mathbf{A}'\widehat{\boldsymbol{\Delta}}^{-1}\mathbf{A}(\bar{\mathbf{X}}_n - \mathbf{f}) \leq z_{1-\lambda}\right) \approx 1 - \lambda,$$

où $z_{1-\lambda}$ est le quantile de niveau $(1 - \lambda) = 0.95$ de la loi $\chi^2(q)$, puis

$$P\left(f(\eta_k) \in \left[\bar{X}_n(\eta_k) \pm \sqrt{(z_{1-\lambda}/n)(\mathbf{A}^{-1}\widehat{\boldsymbol{\Delta}}(\mathbf{A}^{-1})')_{kk}}\right], 1 \leq k \leq q\right) \geq 95\%.$$

Après un calcul de routine (on voit que \mathbf{A}^{-1} est triangulaire inférieure avec ses coefficients diagonaux et subdiagonaux égaux à 1), nous avons finalement les ICS suivants pour la méthode paramétrique B :

$$P\left(f(\eta_k) \in \left[\bar{X}_n(\eta_k) \pm \sqrt{(z_{1-\lambda}/n)\bar{X}_n(\eta_k)}\right], 1 \leq k \leq q\right) \geq 95\%. \quad (2.9)$$

Remarque 2.6. En comparant les ICS de (2.8) à ceux de (2.9), on observe deux petites différences : d'une part, la valeur $f(\eta_k)$ ($1 \leq k \leq q$) est estimée par $\widehat{f}_{np}(\eta_k)$ dans un cas et par $\bar{X}_n(\eta_k)$ dans l'autre, et d'autre part, la variance du processus X au point η_k , égale au paramètre $f(\eta_k)$ de la loi de Poisson que suit $X(\eta_k)$, est estimée par la variance empirique $n^{-1}\sum_{i=1}^n(X_i(\eta_k) - \bar{X}_n(\eta_k))^2$ dans un cas et par la moyenne empirique $\bar{X}_n(\eta_k)$ dans l'autre. On peut donc s'attendre à des performances comparables pour chaque méthode, en notant toutefois qu'un estimateur de lissage peut donner de meilleurs résultats que le processus \bar{X}_n dans l'estimation des $f(\eta_k)$. D'autre part, l'apport majeur de la méthode B par rapport à la méthode A est de permettre la construction d'ICS en dehors des points d'observation t_{jp} , $1 \leq j \leq p$. Il est vrai que lorsque p est grand, cet apport devient relativement négligeable.

Pour chacune des valeurs de n considérées, la simulation a été répétée 500 fois. Les résultats en termes de couverture simultanée (fréquence des occurrences pour lesquelles chacun des ICS contient la vraie valeur de f) et en termes d'amplitude moyenne des ICS sont donnés dans le Tableau 2.1. Des ICS typiques sont fournis dans la Figure 2.1.

n	q	Couverture (%)		Amplitude	
		Méthode A	Méthode B	Méthode A	Méthode B
10	10	99.8	98.6	11.99	11.63
20	10	100	99.4	8.46	8.38
50	20	100	100	6.87	6.84
100	20	100	100	4.86	4.85
100	50	100	100	7.03	7.00
100	100	100	100	9.49	9.45

TAB. 2.1 – ICS de niveau de confiance $\geq 95\%$. Fréquence de couverture simultanée des valeurs de f aux points équidistants $\eta_k, 1 \leq k \leq q$, et amplitude moyenne des intervalles lors de 500 simulations d'un échantillon i.i.d. (X_1, \dots, X_n) de X aux points équidistants $t_{jp}, 1 \leq j \leq p$ ($p = 200$).

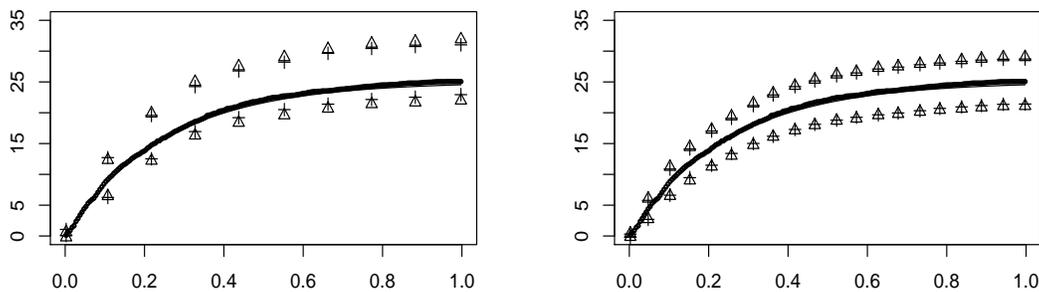


FIG. 2.1 – Intervalles de confiance simultanés. Les points désignent la trajectoire moyenne observée et la ligne pleine est la fonction f . Les ICS de la méthode A sont représentés par des croix et ceux de la méthode B par des triangles. À gauche : $n = 20, p = 200$; à droite : $n = 50, p = 200$.

Globalement, on observe que la correction de la méthode de Scheffé est très conservatrice : la fréquence de couverture simultanée des ICS est nettement supérieure au niveau nominal 95%, au détriment de leur précision. Dans les simulations précédentes, les intervalles de confiance non simultanés obtenus pour chaque méthode, qui ne sont pas présentés ici, ont une fréquence de couverture simultanée qui varie entre 73% et 93% avec une amplitude qui représente seulement 15% à 45%

de l'amplitude des ICS correspondants. Cette fréquence de couverture simultanée élevée est vraisemblablement due à la forte corrélation du processus de Poisson X ($\text{Corr}(X(s), X(t)) = \sqrt{f(s)/f(t)} = \sqrt{(1 - \exp(-4s))/(1 - \exp(-4t))}$ pour $s < t$). D'autre part, les ICS de la méthode B (paramétrique) sont légèrement plus petits que ceux de la méthode A, et la méthode A assure une fréquence de couverture simultanée légèrement meilleure que celle de la méthode B pour les petites valeurs de n .

Dans l'ensemble, les deux méthodes offrent des performances comparables. Ceci atteste de l'efficacité de la méthode A dans la mesure où elle n'a pas à sa disposition l'information paramétrique sur le processus X qu'utilise la méthode B. Nous pouvons donc à l'issue de ces simulations retenir la méthode A, basée sur le Théorème 2.3, comme une méthode valide de construction d'intervalles de confiance simultanés pour la moyenne d'un processus aléatoire.

2.3 Convergence dans quelques cas particuliers

2.3.1 Convergence globale

Nous donnons ici deux résultats de convergence globale valables pour les estimateurs splines de lissage ou d'interpolation de f (voir l'Introduction Générale et le Chapitre 1 pour la définition de ces estimateurs).

Étant donnés des entiers $n, p, m \geq 1$ tels que $p \geq m$ et un paramètre de lissage $\lambda(n, p) \geq 0$, nous notons simplement \hat{f}_λ la spline de lissage ou d'interpolation d'ordre $2m$ associée à λ et aux données $\bar{X}_n(t_{1p}), \dots, \bar{X}_n(t_{pp})$. En utilisant les résultats de Ragozin (1983) sur l'estimation par splines, et la représentation en série d'une dérivée partielle $R^{(q,q)}(s, t) = (\partial^{2q}/\partial^q s \partial^q t)R(s, t)$ de la fonction de covariance R due à Kadota (1967), nous obtenons la convergence en moyenne quadratique suivante.

Théorème 2.4. *Supposons que*

- (i) *le dispositif expérimental est quasi-uniforme ;*
- (ii) *la fonction moyenne f est dans $W_2^{m_1}[0, 1]$ pour un entier $m_1 \geq 1$;*
- (iii) *la fonction de covariance admet une dérivée partielle $R^{(m_2, m_2)}$ continue sur $[0, 1]^2$ pour un entier $m_2 \geq 1$.*

Considérons l'estimateur spline \hat{f}_λ d'ordre $2m$, avec $m \geq \max(m_1, m_2)$, et de paramètre $\lambda = \lambda(n, p) \geq 0$. Alors quand $n, p \rightarrow \infty$, la MISE de \hat{f}_λ vérifie

$$\begin{aligned} \mathbb{E}|f - \hat{f}_\lambda|_0^2 &= \mathcal{O}\left((\lambda + p^{-2m})^{m_1/m}\right) |f|_{m_1}^2 + \frac{1}{n} \int_0^1 R(t, t) dt \\ &+ \mathcal{O}\left(\frac{(\lambda + p^{-2m})^{m_2/2m}}{n}\right) \left(\int_0^1 R^{(m_2, m_2)}(t, t) dt\right)^{1/2}. \end{aligned} \quad (2.10)$$

L'expression asymptotique (2.10) indique une différence fondamentale par rapport au cas d'erreurs décorréelées, où X est la somme de f et d'un bruit blanc. En effet, dans ce dernier cas nous savons par les Théorèmes 2 ou 1.3 que l'estimateur \widehat{f}_λ est consistant si et seulement si $\lambda \rightarrow 0$ et $np\lambda^{1/2m} \rightarrow \infty$ lorsque $n, p \rightarrow \infty$, et que dans ce cas son biais au carré est en $\mathcal{O}((\lambda + p^{-2m})^{m_1/m})$ tandis que sa variance est en $\mathcal{O}((np\lambda^{1/2m})^{-1})$.

Au contraire, dans la formule (2.10) ce sont des puissances positives de λ qui apparaissent à la fois dans la majoration du biais au carré et de la variance de \widehat{f}_λ . On peut donc obtenir une estimation consistante sans lissage. Ceci nous amène à conjecturer, pour l'estimation de la moyenne d'un processus aléatoire quelconque, un équilibre entre biais et variance très différent du cas du bruit blanc. Cette conjecture sera confirmée par les Théorèmes 2.6 et 2.7 qui donnent des équivalents asymptotiques du biais et de la variance des estimateurs à noyau et splines de lissage.

Nous nous contentons pour l'instant d'observer que dans le contexte de processus aléatoires, le fait de lisser plus augmente certainement le biais, mais ne diminue pas nécessairement la variance, comme l'indique l'Exemple 2.1 dans lequel la variance de $\widehat{f}_\lambda(t)$ ne dépend pas de λ . Pour autant, une certaine quantité de lissage apporte de meilleures performances dans l'estimation, autant sur le plan théorique (*cf.* Hart et Wehrly, 1986, et la Section 2.3.2) que sur le plan pratique (voir la supériorité du lissage spline sur l'interpolation dans les simulations de Biritxinaga, 1987).

Nous fournissons à présent un résultat de convergence presque sûre pour l'estimateur spline de lissage ou d'interpolation, lorsque le processus X vérifie une loi du logarithme itéré fonctionnelle dans $C[0, 1]$. Ce résultat généralise le travail de Biritxinaga (1990) sur la consistance de l'interpolation spline dans les espaces de Sobolev. En notant $S_n = X_1 + \dots + X_n - nf$ la somme partielle centrée associée à l'échantillon i.i.d. de X et $\log_2 n = \log_+ \log_+ n$, avec $\log_+ n = \log(\max(e, n))$, nous précisons maintenant la loi du logarithme itéré considérée.

Définition 2.2. *On dit que le processus X à valeurs dans $C[0, 1]$ vérifie la version bornée de la loi du logarithme itéré dans $C[0, 1]$ si la limite (non-aléatoire) $\Lambda(X) = \limsup_{n \rightarrow \infty} \frac{\|S_n\|_\infty}{(2n \log_2 n)^{1/2}}$ est finie.*

Le lecteur intéressé par la loi du logarithme itéré dans les espaces de Banach séparables pourra consulter avec profit le livre de Ledoux et Talagrand (1991) et les articles de Kuelbs (1976) et de Carmona et Kôno (1976). Deux exemples fondamentaux de processus vérifiant la version bornée de la loi du logarithme itéré dans $C[0, 1]$ sont le processus de Wiener et le pont brownien standards. En fait, ces processus vérifient la version compacte de cette loi qui est plus forte que la version bornée (voir par exemple Deheuvels *et al.*, 2007).

Théorème 2.5. *Soient m, k deux entiers tels que $0 \leq k < m$. Supposons que*

- (i) X vérifie la version bornée de la loi du logarithme itéré dans $C[0, 1]$;
- (ii) $f \in W_2^m[0, 1]$;
- (iii) le dispositif expérimental est quasi-uniforme ;

Considérons l'estimateur spline de lissage ou d'interpolation \widehat{f}_λ d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p) \geq 0$. Alors, lorsque $n, p \rightarrow \infty$, l'estimateur $\widehat{f}_\lambda^{(k)}$ converge presque sûrement vers $f^{(k)}$ à la vitesse suivante :

$$|f - \widehat{f}_\lambda|_k^2 = \mathcal{O}\left((\lambda + p^{-2m})^{(m-k)/m}\right) |f|_m^2 + \mathcal{O}\left(\frac{p^{2k} \log_2 n}{\max(1, \lambda p^{2m})n}\right), \quad (2.11a)$$

et à l'ordre m , nous avons presque sûrement

$$|f - \widehat{f}_\lambda|_m^2 = o(1) |f|_m^2 + \mathcal{O}\left(\frac{\log_2 n}{\max(p^{-2m}, \lambda)n}\right). \quad (2.11b)$$

Contrairement au Théorème 2.4 dans lequel la majoration asymptotique de la MISE ne semble pas justifier le lissage (le majorant étant rendu minimal par l'estimateur d'interpolation), il apparaît ici, concernant la convergence presque sûre, que dans certains cas, le majorant asymptotique de l'erreur d'estimation est rendu minimal avec une certaine quantité de lissage. Plus précisément, si l'une des deux conditions asymptotiques suivantes :

- (i) $n^{-1}(\log_2 n)p^{2k} \rightarrow 0$ et $n^{-1}(\log_2 n)p^{2(2m^3 - 2mk^2 + k^3)} \rightarrow \infty$,
- (ii) $n^{-1}(\log_2 n)p^{2k} \not\rightarrow 0$ et $n^{-1}(\log_2 n)p^{2(k-m)} \rightarrow 0$,

est valide lorsque $n, p \rightarrow \infty$, alors on peut montrer que le majorant optimal dans (2.11a) est obtenu pour le choix $\lambda \asymp (n^{-1}(\log_2 n))^{m/(2m-k)} p^{-2(m-k)}$ et qu'il est de l'ordre de $p^{-2(m-k)} + (n^{-1}(\log_2 n)p^{-2(m-k)})^{m/(2m-k)}$.

2.3.2 Convergence locale

Le Théorème 2.2, valable pour un estimateur linéaire général, peut être rendu plus précis en particulierisant l'estimateur, dans le sens où on peut faire apparaître explicitement le paramètre de lissage de l'estimateur dans le développement asymptotique de sa variance. Nous considérons ici les exemples des estimateurs à noyau de Gasser-Müller et des splines de lissage.

A. Estimateur à noyau.

Nous donnons dans cette partie une expression de l'erreur quadratique moyenne ponctuelle (MSE) de l'estimateur à noyau de Gasser-Müller lorsque $n, p \rightarrow \infty$, en notant qu'un développement similaire s'obtient sans difficulté pour l'estimateur de Priestley et Chao (1972). Nous abordons ensuite le problème du choix optimal de la largeur de fenêtre. Les résultats présentés permettent de généraliser le travail de

Hart et Wehrly (1986) portant sur les processus à covariance stationnaire.

Commençons par rappeler la définition de l'estimateur de Gasser-Müller. Soit K un noyau à support compact et soit $h > 0$ une largeur de fenêtre. Posons $K_h(t) = (1/h)K(t/h)$ pour tout $t \in [0, 1]$. Définissons aussi une suite de points $(s_{ip})_{1 \leq i \leq p}$ telle que $s_{0p} = 0, s_{pp} = 1$, et $s_{(i-1)p} \leq t_{ip} \leq s_{ip}$ pour tout $1 \leq i \leq p$. L'estimateur de Gasser-Müller est alors défini pour tout $t \in [0, 1]$ par

$$\widehat{f}_{np}^{GM}(t) := \sum_{i=1}^p \left(\int_{s_{(i-1)p}}^{s_{ip}} K_h(t-s) ds \right) \bar{X}_n(t_{ip}). \quad (2.12)$$

Pour simplifier l'exposé, nous supposons que K est une densité symétrique ayant pour support $[-1, 1]$ et vérifiant une condition de Lipschitz $|K(s) - K(t)| \leq C|s - t|$. Nous supposons aussi que la fonction moyenne f est dans $C^2[0, 1]$ et que les points $(t_{ip})_{1 \leq i \leq p}$ d'une part, et $(s_{ip})_{1 \leq i \leq p}$ d'autre part, ont une répartition quasi-uniforme (voir page 32 pour la définition de cette notion).

Nous rappelons à présent l'expression du biais asymptotique de \widehat{f}_{np}^{GM} (Gasser et Müller, 1979). Sous les hypothèses précédentes nous avons, pour tout $t \in]0, 1[$,

$$f(t) - \mathbb{E}\widehat{f}_{np}^{GM}(t) = \frac{h^2(f''(t))^2}{2} + o(h^2) + \mathcal{O}\left(\frac{1}{p}\right) \quad (2.13)$$

lorsque $n, p \rightarrow \infty$ avec $h = h(n, p) \rightarrow 0$.

On montre aussi, par une simple adaptation du Théorème 1 de Hart et Wehrly (1986), que la variance asymptotique de \widehat{f}_{np}^{GM} s'écrit

$$\text{Var} \widehat{f}_{np}^{GM}(t) = \frac{1}{n} \int_{-1}^1 \int_{-1}^1 K(u)K(v)R(t+uh, t+vh)dudv + \mathcal{O}\left(\frac{1}{np}\right). \quad (2.14)$$

Nous présentons maintenant un résultat original qui développe la double intégrale précédente sous différentes hypothèses de régularité sur R .

Lemme 2.1. *Soit $t \in]0, 1[$. Considérons les hypothèses*

- (i) *R vérifie une condition de Lipschitz $|R(u, v) - R(u', v')| \leq C(|u - u'| + |v - v'|)$ pour tous réels $0 \leq u, u', v, v' \leq 1$;*
- (ii) *R admet des dérivées semi-directionnelles au point (t, t) , au sens que $D_{(u,v)}^+ R(t, t) = \lim_{h \rightarrow 0^+} \frac{R(t+uh, t+vh) - R(t, t)}{h}$ existe et est finie pour tous réels u, v ;*
- (iii) *R est deux fois continûment différentiable au voisinage de (t, t) .*

Posons $C_{K,R,t} = \int_{-1}^1 \int_{-1}^1 K(u)K(v)D_{(u,v)}^+ R(t, t)dudv$ et $\sigma_K^2 = \int_{-1}^1 t^2 K(t)dt$.

Alors, lorsque $h \rightarrow 0$, nous avons, sous les hypothèses (i)-(ii),

$$\int_{-1}^1 \int_{-1}^1 K(u)K(v)R(t+uh, t+vh)dudv = R(t, t) + C_{K,R,t}h + o(h), \quad (2.15a)$$

et sous l'hypothèse (iii), il vient

$$\int_{-1}^1 \int_{-1}^1 K(u)K(v)R(t+uh, t+vh)dudv = R(t, t) + \sigma_K^2 R^{(2,0)}(t, t)h^2 + o(h^2). \quad (2.15b)$$

Ce résultat est obtenu essentiellement grâce au théorème de convergence dominée de Lebesgue qui permet d'établir la continuité d'une intégrale dépendant d'un paramètre. La démonstration de la relation (2.15a) apparaît dans Hart et Wehrly (1986) dans le cas où $R(s, t) = \sigma^2 \rho(s - t)$, avec ρ une fonction lipschitzienne et admettant des dérivées à gauche et à droite en 0). Dans ce cas, il vient $C_{K,R,t} = 2\sigma^2 \left(\int_{-1}^1 \int_v^1 (u-v)K(u)K(v)dudv \right) \rho'(0^+)$. La relation (2.15b) est quant à elle obtenue en intégrant le développement limité d'ordre 2 de R au point (t, t) puis en exploitant la symétrie de R et le fait que K est une densité symétrique par rapport à 0.

Exemple 2.2. Soit W un processus de Wiener standard sur $[0, 1]$. La fonction de covariance $R(s, t) = \min(s, t)$ de W vérifie les hypothèses (i)-(ii) précédentes, mais pas (iii). Pour tous réels $0 \leq t \leq 1$ et $-1 \leq u, v \leq 1$, on a $D_{(u,v)}^+ R(t, t) = R(u, v)$ et l'approximation (2.15a) est exacte : on peut omettre le terme $o(h)$.

Exemple 2.3. Soit X un processus du second ordre défini sur $[0, 1]$, ayant pour fonction de covariance $R(u, v) = \varphi(u)\varphi(v)$, où φ est une fonction continue admettant des dérivées à droite et à gauche au point $t \in]0, 1[$. Alors R vérifie les hypothèses (i)-(ii) du Lemme 2.1 et on peut montrer par le calcul que $C_{K,R,t} = 2\varphi(t)(\varphi'(t^+) - \varphi'(t^-)) \int_0^1 uK(u)du$. Cette quantité peut donc être rendue positive, négative, ou nulle par un choix adapté de φ .

Fortes des relations (2.13), (2.14), et du Lemme 2.1, nous pouvons déterminer l'équivalent asymptotique de la MSE de l'estimateur $\widehat{f}_{n,p}^{GM}$.

Théorème 2.6. Soit $t \in]0, 1[$. Considérons les hypothèses

- (i) $f \in C^2[0, 1]$;
- (ii) R vérifie les conditions (i)-(ii) du Lemme 2.1 au point t ;
- (ii') R vérifie la condition (iii) du Lemme 2.1 au point t ;
- (iii) n et p tendent vers l'infini de sorte que $n = \mathcal{O}(p)$;
- (iv) $h = h(n, p) \rightarrow 0$ lorsque $n, p \rightarrow \infty$.

Alors lorsque $n, p \rightarrow \infty$, sous les hypothèses (i)-(ii)-(iii)-(iv), nous avons

$$\mathbb{E} \left(f(t) - \widehat{f}_{np}^{GM}(t) \right)^2 = \frac{R(t, t) + C_{K,R,t}h}{n} + \frac{h^4 (f''(t))^2}{4} + o(h^4) + o\left(\frac{1}{n}\right), \quad (2.16a)$$

et sous les hypothèses (i)-(ii')-(iii)-(iv), il vient

$$\begin{aligned} \mathbb{E}\left(f(t) - \hat{f}_{np}^{GM}(t)\right)^2 &= \frac{R(t,t) + \sigma_K^2 R^{(2,0)}(t,t)h^2}{n} + \frac{h^4(f''(t))^2}{4} \\ &+ o(h^4) + o\left(\frac{1}{n}\right). \end{aligned} \quad (2.16b)$$

Le Théorème 2.6 revêt une grande importance car il fournit le signe de la correction $C_{K,R,t}h$ ou $\sigma_K^2 R^{(2,0)}(t,t)h^2$ apportée par le lissage dans le développement asymptotique de la variance. Ce signe dépend essentiellement des dérivées de R , et il indique le choix optimal de h qui minimise l'équivalent asymptotique de la MSE. Si par exemple $C_{K,R,t}$ ou $R^{(2,0)}(t,t)$ est positive, alors un lissage plus fort augmente simultanément le biais au carré et la variance asymptotiques, et on a intérêt à choisir h le plus petit possible. Si par contre $C_{K,R,t}$ ou $R^{(2,0)}(t,t)$ est négative, alors h détermine un équilibre entre le biais au carré et la variance asymptotiques, et le choix optimal de h est $h = (|C_{K,R,t}|/(4(f''(t))^2))^{1/3}n^{-1/3}$ pour (2.16a), et $h = (\sigma_K^2|R^{(2,0)}(t,t)|/(2(f''(t))^2))^{1/2}n^{-1/2}$ pour (2.16b), avec $f''(t) \neq 0$.

B. Estimateur spline de lissage.

Nous donnons à présent un équivalent asymptotique de la MSE de l'estimateur spline de lissage lorsque $n, p \rightarrow \infty$. Le résultat proposé fait appel aux travaux de Nychka (1995) sur l'approximation du noyau équivalent d'une spline de lissage par une fonction de Green et sur le biais et la variance asymptotiques d'un estimateur spline de lissage dans le cas de données i.i.d..

Théorème 2.7. *Soient un entier $m \geq 1$ et un réel $\Delta \in]0, 1/2[$. Soit \hat{f}_λ l'estimateur spline de lissage d'ordre $2m$ et de paramètre $\lambda = \lambda(n, p) > 0$ associé aux données $\bar{X}_n(t_{1p}), \dots, \bar{X}_n(t_{pp})$. Supposons que*

- (i) *f admet une dérivée $f^{(2m)}$ hölderienne sur $[0, 1]$;*
- (ii) *R admet une dérivée croisée $R^{(2m, 2m)}$ hölderienne d'indice β sur $[0, 1]^2$;*
- (iii) *il existe une famille de fonctions $(\tilde{R}_\tau)_{\tau \in [0, 1]}$ telle que pour tout τ , \tilde{R}_τ coïncide avec $R(\cdot, \tau)$ sur $[\Delta/2, 1 - \Delta/2]$, \tilde{R}_τ vérifie les conditions $\tilde{R}_\tau^{(k)}(0) = \tilde{R}_\tau^{(k)}(1) = 0$ pour $m \leq k \leq 2m - 1$, et telle que $\sup_{\tau \in [0, 1]} \|\tilde{R}_\tau\|_\mu < \infty$ pour un réel $\mu \in [2 + \frac{1}{2m}, 2 + \frac{1}{2m} + \frac{\beta}{2m}]$, où $\|\cdot\|_\mu$ est la norme d'interpolation définie par (2.60).*
- (iv) *la fonction de répartition F_p du dispositif expérimental converge uniformément sur $[0, 1]$ vers une fonction $F \in C^2[0, 1]$ de dérivée $F' = \psi > 0$;*
- (v) *$\lambda = o(1)$ et $\|F_p - F\|_\infty = o(\lambda^{2+1/2m})$ lorsque $n, p \rightarrow \infty$.*

Alors la MSE de \widehat{f}_λ vérifie

$$\begin{aligned} \mathbb{E} \left(f(t) - \widehat{f}_\lambda(t) \right)^2 &= \frac{R(t, t) + \lambda(-1)^{m+1} 2R^{(2m,0)}(t, t) + \lambda^2 R^{(2m,2m)}(t, t)}{n} \\ &+ \frac{\lambda^2 (f^{(2m)}(t))^2}{\psi^2(t)} + o(\lambda^2) + o\left(\frac{1}{n}\right). \end{aligned} \quad (2.17)$$

lorsque $n, p \rightarrow \infty$, uniformément par rapport à $t \in [\Delta, 1 - \Delta]$.

Remarque 2.7. Une conjecture de Nychka (1995) est utilisée pour établir le théorème précédent (voir la démonstration en page 95). Par ailleurs, l'hypothèse (ii) (continuité hölderienne de $R^{(2m,2m)}$) implique vraisemblablement l'hypothèse technique (iii), d'après une communication personnelle de D. Nychka.

2.4 Perspectives de développements

Dans ce chapitre, nous avons établi un ensemble de résultats asymptotiques concernant l'estimation non paramétrique de la fonction moyenne d'un processus aléatoire à partir de données observées dans un dispositif expérimental déterministe et équilibré.

Deux résultats fondamentaux sur la variance d'un estimateur linéaire général (Théorème 2.2) et sur sa normalité asymptotique (Théorème 2.3) ont été présentés. Ces résultats ont été exposés par commodité dans le cadre d'un processus défini sur $[0, 1]$, mais à l'instar du théorème de Mercer (et du TCL pour le second) qui servent à les établir, ils peuvent être généralisés aux processus définis sur un espace métrique compact (notamment les processus spatiaux ou spatio-temporels).

Une autre extension envisageable des Théorèmes 2.2 et 2.3 est le passage à un dispositif déterministe déséquilibré ou bien aléatoire. En fait, il est possible d'adapter la démonstration du Théorème 2.2 et d'obtenir, pour la variance de l'estimateur dans le cas déséquilibré et pour sa variance conditionnelle dans le cas aléatoire, le même résultat que dans le cas équilibré. Cependant, cette extension est obtenue au prix de conditions techniques difficilement interprétables sur les poids de l'estimateur. Dans ces deux cas, on peut toutefois montrer sous de faibles conditions de régularité du dispositif que la variance asymptotique de l'estimateur reste de l'ordre de n^{-1} , où n est la taille de l'échantillon (X_1, \dots, X_n) .

D'autre part, deux directions de recherche se dessinent à partir de la procédure d'intervalles de confiance simultanés (ICS) de la Section 2.2.3. En premier lieu, il est souhaitable de pouvoir construire des ICS par des méthodes moins conservatrices que celles de Scheffé ou de Bonferroni en s'appuyant sur le fait que la structure

de corrélation du processus observé se répercute sur les intervalles de confiance. En effet, si les valeurs de la trajectoire moyenne observée sont très (positivement) corrélées, alors les intervalles de confiance le sont aussi, et le problème de l'estimation simultanée se rapproche alors beaucoup de l'estimation par intervalles ponctuels. En second lieu, un travail intéressant mais difficile consisterait à obtenir de bandes de confiance simultanées en lieu et place des ICS de cette section qui portent seulement sur un nombre fini de points. Il existe une vaste littérature sur l'obtention de bandes de confiance simultanées approchées dans la régression avec variable explicative déterministe (voir par exemple Loader, 1992; Sun et Loader, 1994; Zhou *et al.*, 1996, et leur bibliographie). Au vu des techniques existantes et des résultats proposés, il semble cependant impossible de construire de telles bandes dans un cadre d'estimation aussi général que celui de ce chapitre. Certaines restrictions ou spécifications doivent être formulées au sujet du processus considéré, de sa fonction moyenne et de l'estimateur utilisé.

Enfin, les résultats de la Section 2.3.2 sur l'erreur quadratique ponctuelle MSE des estimateurs à noyau et des splines de lissage sont susceptibles d'être développés et appliqués à la sélection de paramètre de lissage par une méthode de type plug-in. En particulier, l'exploitation jointe du Théorème 2.6 et d'une bonne méthode d'estimation de la covariance du processus pourrait donner de bons résultats à l'instar de Hart et Wehrly (1986).

2.5 Démonstrations

Pour les besoins de cette section, nous introduisons de nouvelles notations. D'abord, pour simplifier l'écriture, les poids W_{nip} de l'estimateur linéaire général \widehat{f}_{np} sont notés simplement W_i . De même, les points d'observation t_{ip} sont notés t_i et la moyenne des processus \overline{X}_n est notée \overline{X} . On définit aussi le processus de bruit $\varepsilon = X - f$ associé à X , et la moyenne des bruits $\overline{\varepsilon} = \overline{X} - f$. On note $\mathbf{V} = (R(t_i, t_j))$ la matrice $p \times p$ de covariance du processus ε aux points d'observation. On rappelle enfin que pour toute fonction g définie sur $[0,1]$, la notation \mathbf{g} désigne le vecteur $(g(t_1), \dots, g(t_p))'$.

Commençons par rappeler le classique théorème de Mercer, qui interviendra dans les démonstrations des Théorèmes 2.1, 2.2 et 2.4.

Théorème 2.8. (Mercer, 1909) *Soit R un noyau symétrique, continu et semi-défini positif sur $[0,1] \times [0,1]$. Alors l'opérateur intégral associé à R admet des valeurs propres $\lambda_1, \lambda_2, \dots$ et des fonctions propres $\varphi_1, \varphi_2, \dots$ telles que*

- (i) $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, avec $\lambda_k \rightarrow 0$ lorsque $k \rightarrow \infty$;
- (ii) les fonctions $\varphi_1, \varphi_2, \dots$ sont continues et forment une b.o.n. de $L_2[0,1]$.

De plus, le noyau R admet la représentation

$$R(s, t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k(s) \varphi_k(t), \quad (2.18)$$

où la convergence de la série est absolue et uniforme sur $[0,1] \times [0,1]$.

Remarque 2.8. *La fonction $t \mapsto R(t, t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k^2(t)$, limite uniforme d'une série de fonction continues, peut être intégrée terme à terme sur $[0,1]$. Nous avons donc $\int_0^1 R(t, t) dt = \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_k^2(t) dt$, et comme les $\varphi_1, \varphi_2, \dots$ forment une b.o.n. de $L_2[0,1]$, il vient $\int_0^1 R(t, t) dt = \sum_{k=1}^{\infty} \lambda_k < \infty$.*

Démonstration du Théorème 2.1

Le principe de la démonstration est le suivant. On exprime d'abord la variance moyenne de l'estimateur \widehat{f}_{np} et la variance intégrée du processus X sous forme de séries grâce au théorème de Mercer. Le k -ème terme de la série associée à \widehat{f}_{np} correspond alors à la norme euclidienne des valeurs aux points t_1, \dots, t_p du lissage de la fonction propre φ_k par les poids W_1, \dots, W_p de \widehat{f}_{np} , tandis que le k -ème terme de la série associée à X correspond à la norme L_2 de φ_k . On

montre ensuite, à l'aide d'une décomposition en somme partielle plus reste, que l'écart entre ces séries devient arbitrairement petit pour p suffisamment grand uniformément par rapport à n . Les approximations numériques de fonctions et d'intégrales effectuées s'appuient essentiellement sur le biais asymptotique nul de \widehat{f}_{np} , sur le caractère borné de la somme de ses poids, et sur la continuité de R .

Soit $\epsilon > 0$. Commençons par écrire sous forme matricielle la variance moyenne de \widehat{f}_{np} aux points d'observation, à l'aide des matrices $\mathbf{A} = (W_j(t_i))$ et $\mathbf{V} = (R(t_i, t_j))$ de taille $p \times p$. En notant que $\widehat{\mathbf{f}}_{np} = (\widehat{f}_{np}(t_1), \dots, \widehat{f}_{np}(t_p))' = \mathbf{A}\overline{\mathbf{X}}$, on a

$$p^{-1} \sum_{j=1}^p \text{Var } \widehat{f}_{np}(t_j) = p^{-1} \text{tr } \text{Var } \widehat{\mathbf{f}}_{np} = p^{-1} \text{tr } \text{Var } \mathbf{A}\overline{\mathbf{X}} = (np)^{-1} \text{tr } \mathbf{A}\mathbf{V}\mathbf{A}'. \quad (2.19)$$

Avec le théorème de Mercer, nous pouvons exprimer \mathbf{V} sous forme d'une série convergente $\sum_{k=1}^{\infty} \lambda_k \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k'$ pour tout $p \geq 1$. Utilisant la linéarité et la commutativité de la trace matricielle, nous faisons le calcul suivant :

$$\begin{aligned} np^{-1} \sum_{j=1}^p \text{Var } \widehat{f}_{np}(t_j) &= p^{-1} \text{tr } \mathbf{A} \left(\sum_{k=1}^{\infty} \lambda_k \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k' \right) \mathbf{A}' \\ &= p^{-1} \sum_{k=1}^{\infty} \lambda_k \text{tr } \mathbf{A} \boldsymbol{\varphi}_k \boldsymbol{\varphi}_k' \mathbf{A}' \\ &= p^{-1} \sum_{k=1}^{\infty} \lambda_k \|\mathbf{A} \boldsymbol{\varphi}_k\|^2. \end{aligned} \quad (2.20)$$

D'un autre côté, intégrons la fonction $t \mapsto R(t, t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k^2(t)$ par rapport à la distribution limite F . D'après la Remarque 2.8, les signes somme et intégrale peuvent être permutés. Il vient

$$\int_0^1 R(t, t) dF(t) = \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_k^2(t) dF(t). \quad (2.21)$$

Ayant exprimé sous forme de séries la variance moyenne de \widehat{f}_{np} et la variance intégrée de X , nous rapprochons maintenant les formules (2.20) et (2.21). Nous constatons d'abord, d'après l'inégalité triangulaire et d'après la condition (i) du théorème (estimateur asymptotiquement sans biais), que pour tout $K \geq 1$, il existe un entier $p_1(K)$ tel que pour tous $n \geq 1$ et $p \geq p_1(K)$,

$$\left| p^{-1} \sum_{k=1}^K \lambda_k \|\mathbf{A} \boldsymbol{\varphi}_k\|^2 - p^{-1} \sum_{k=1}^K \lambda_k \|\boldsymbol{\varphi}_k\|^2 \right| < \epsilon. \quad (2.22)$$

Par ailleurs, les fonctions $\varphi_1, \varphi_2, \dots$ étant continues, la convergence de la fonction de répartition empirique F_p vers la mesure limite F assure que pour tout $K \geq 1$, il existe un entier $p_2(K)$ tel que pour tout $p \geq p_2(K)$,

$$\left| p^{-1} \sum_{k=1}^K \lambda_k \|\varphi_k\|^2 - \sum_{k=1}^K \lambda_k \int_0^1 \varphi_k^2(t) dF(t) \right| < \epsilon. \quad (2.23)$$

Ensuite, nous déduisons de (2.21) qu'il existe un entier K_1 tel que pour tout $K \geq K_1$,

$$\left| \sum_{k=1}^K \lambda_k \int_0^1 \varphi_k^2(t) dF(t) - \int_0^1 R(t, t) dF(t) \right| < \epsilon. \quad (2.24)$$

Nous allons à présent montrer que $p^{-1} \sum_{k=K+1}^{\infty} \lambda_k \|\mathbf{A}\varphi_k\|^2$, le reste d'ordre K de (2.20), devient négligeable lorsque $K \rightarrow \infty$, uniformément par rapport à n et p . En premier lieu, nous établissons, avec l'hypothèse (ii) et un argument classique de calcul matriciel (par exemple le Théorème 1.12 : disques de Gersgorin), que $\|\mathbf{A}\mathbf{A}'\| = \|\mathbf{A}\|^2 \leq (\max_{1 \leq j \leq p} \sum_{i=1}^p |W_i(t_j)|)^2 = \mathcal{O}(1)$ lorsque $n, p \rightarrow \infty$. Il existe donc une constante C finie telle que pour tous $n, p \geq 1$,

$$\|\mathbf{A}\| \leq C. \quad (2.25)$$

D'autre part, la convergence uniforme de (2.18) implique que pour tout $p \geq 1$,

$$\begin{aligned} \sum_{k=K+1}^{\infty} \lambda_k \|\varphi_k\|^2 &= \sum_{k=K+1}^{\infty} \lambda_k \left(\sum_{i=1}^p \varphi_k^2(t_i) \right) = \sum_{i=1}^p \left(\sum_{k=K+1}^{\infty} \lambda_k \varphi_k^2(t_i) \right) \\ &\leq p \sup_{t \in [0,1]} \left(\sum_{k=K+1}^{\infty} \lambda_k \varphi_k^2(t) \right) = p \cdot o(1) \end{aligned} \quad (2.26)$$

lorsque $K \rightarrow \infty$, où le $o(1)$ ne dépend ni de n ni de p . Nous déduisons alors de (2.25) et (2.26) qu'il existe un entier K_2 tel que pour tout $K \geq K_2$ et pour tous $n, p \geq 1$,

$$p^{-1} \sum_{k=K+1}^{\infty} \lambda_k \|\mathbf{A}\varphi_k\|^2 < \epsilon. \quad (2.27)$$

En posant enfin $K_0 = \max(K_1, K_2)$ et $p_0 = \max(p_1(K_0), p_2(K_0))$, nous pouvons exploiter les formules (2.20) à (2.24) et (2.27) pour former la majoration suivante,

valable pour tout $n \geq 1$ et pour tout $p \geq p_0$:

$$\begin{aligned}
& \left| np^{-1} \sum_{j=1}^p \text{Var} \widehat{f}_{np}(t_j) - \int_0^1 R(t, t) dF(t) \right| \\
&= \left| p^{-1} \sum_{k=1}^{\infty} \lambda_k \|\mathbf{A}\boldsymbol{\varphi}_k\|^2 - \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_k^2(t) dF(t) \right| \\
&\leq \left| p^{-1} \sum_{k=1}^{K_0} \lambda_k \|\mathbf{A}\boldsymbol{\varphi}_k\|^2 - p^{-1} \sum_{k=1}^{K_0} \lambda_k \|\boldsymbol{\varphi}_k\|^2 \right| \\
&\quad + \left| p^{-1} \sum_{k=1}^{K_0} \lambda_k \|\boldsymbol{\varphi}_k\|^2 - \sum_{k=1}^{K_0} \lambda_k \int_0^1 \varphi_k^2(t) dF(t) \right| \tag{2.28} \\
&\quad + \left| \sum_{k=1}^{K_0} \lambda_k \int_0^1 \varphi_k^2(t) dF(t) - \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_k^2(t) dF(t) \right| \\
&\quad + \left| p^{-1} \sum_{k=K_0+1}^{\infty} \lambda_k \|\mathbf{A}\boldsymbol{\varphi}_k\|^2 \right| \\
&< 4\epsilon.
\end{aligned}$$

Le théorème est alors démontré. ■

Démonstration du Théorème 2.2

Nous commençons par établir la relation (2.3) sous les conditions (ii)-(iii)-(iv) du théorème, par un raisonnement analogue à celui de la démonstration précédente. Dans un second temps, nous établirons cette même relation sous les conditions (i)-(iii)-(iv)-(v) par un argument indépendant.

Supposons pour le moment les conditions (ii)-(iii)-(iv) acquises, et montrons que l'écart $|n \text{Var} \widehat{f}_{np}(t) - R(t, t)|$ devient arbitrairement petit pour p suffisamment grand et pour tout $n \geq 1$. Soit $\epsilon > 0$. Écrivons d'abord

$$\text{Var} \widehat{f}_{np}(t) = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) R(t_i, t_j), \tag{2.29}$$

puis, avec le théorème de Mercer, établissons que

$$\begin{aligned}
n \operatorname{Var} \widehat{f}_{np}(t) &= \sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) \sum_{k=1}^{\infty} \lambda_k \varphi_k(t_i) \varphi_k(t_j) \\
&= \sum_{k=1}^{\infty} \lambda_k \sum_{i=1}^p \sum_{j=1}^p W_i(t) \varphi_k(t_i) W_j(t) \varphi_k(t_j), \\
&= \sum_{k=1}^{\infty} \lambda_k \left(\sum_{i=1}^p W_i(t) \varphi_k(t_i) \right)^2.
\end{aligned} \tag{2.30}$$

En appliquant à nouveau le théorème de Mercer à $R(t, t)$, nous avons

$$\left| n \operatorname{Var} \widehat{f}_{np}(t) - R(t, t) \right| = \left| \sum_{k=1}^{\infty} \lambda_k \left(\left(\sum_{i=1}^p W_i(t) \varphi_k(t_i) \right)^2 - \varphi_k^2(t) \right) \right|$$

et nous effectuons alors les majorations

$$\begin{aligned}
&\leq \sum_{k=1}^{\infty} \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| \left(\sum_{i=1}^p |W_i(t) \varphi_k(t_i)| + |\varphi_k(t)| \right) \\
&\leq \sum_{k=1}^{\infty} \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| \left(\sum_{i=1}^p |W_i(t)| + 1 \right) \|\varphi_k\|_{\infty}.
\end{aligned} \tag{2.31}$$

Nous observons maintenant que d'après les conditions (ii) et (iv), il existe une constante C finie telle que $(\sum_{i=1}^p |W_i(t)| + 1) \|\varphi_k\|_{\infty} \leq C$ pour tout $t \in [0, 1]$ et pour tous $n, p, k \geq 1$. Nous obtenons alors, pour tous entiers $n, p, K \geq 1$, les inégalités

$$\begin{aligned}
\left| n \operatorname{Var} \widehat{f}_{np}(t) - R(t, t) \right| &\leq C \sum_{k=1}^{\infty} \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| \\
&\leq C \sum_{k=1}^K \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| + C \sum_{k=K+1}^{\infty} \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| \\
&\leq C \sum_{k=1}^K \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right| + C^2 \sum_{k=K+1}^{\infty} \lambda_k.
\end{aligned} \tag{2.32}$$

Ensuite, comme $\sum_{k=1}^{\infty} \lambda_k = \int_0^1 R(t, t) dt < \infty$ (voir la Remarque 2.8), il existe un entier K_0 tel que $\sum_{k=K_0+1}^{\infty} \lambda_k < \epsilon$. La condition (iii) de biais asymptotique nul assure alors que pour p suffisamment grand, la somme $\sum_{k=1}^{K_0} \lambda_k \left| \sum_{i=1}^p W_i(t) \varphi_k(t_i) - \varphi_k(t) \right|$ est arbitrairement petite uniformément par rapport à n . En d'autres termes, il existe un entier p_0 tel que pour tous $n \geq 1$ et $p \geq p_0$,

$$\left| n \operatorname{Var} \widehat{f}_{np}(t) - R(t, t) \right| \leq C\epsilon + C^2\epsilon, \tag{2.33}$$

ce qui établit la relation (2.3) souhaitée. \square

Nous allons à présent montrer la validité de (2.3) sous les conditions (i)-(iii)-(iv)-(v), en abandonnant le formalisme précédent (ϵ, C, p_0 etc.) au profit des notations de Landau. Commençons par écrire

$$\begin{aligned} \left| n \operatorname{Var} \widehat{f}_{np}(t) - R(t, t) \right| &\leq \left| \sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) (R(t_i, t_j) - R(t, t)) \right| \\ &+ |R(t, t)| \left| \sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) - 1 \right|. \end{aligned} \quad (2.34)$$

Lorsque $p \rightarrow \infty$, le second terme de la somme tend vers zéro uniformément par rapport à n . En effet, en appliquant la condition (iii) à la fonction $g \equiv 1$ sur $[0, 1]$, il vient

$$\sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) = \left(\sum_{i=1}^p W_i(t) \right)^2 = (1 + o(1))^2 = 1 + o(1).$$

Il reste à majorer le premier terme. On obtient successivement

$$\begin{aligned} &\left| \sum_{i=1}^p \sum_{j=1}^p W_i(t) W_j(t) (R(t_i, t_j) - R(t, t)) \right| \\ &\leq \sum_{i=1}^p |W_i(t)| \sum_{j=1}^p |W_j(t)| |R(t_i, t_j) - R(t, t)| \end{aligned}$$

puis, avec la condition (i),

$$\begin{aligned} &\leq \sum_{i=1}^p |W_i(t)| \sum_{j=1}^p |W_j(t)| C (|t - t_i|^\gamma + |t - t_j|^\gamma) \\ &= C \sum_{i=1}^p |W_i(t)| \left(|t - t_i|^\gamma \sum_{j=1}^p |W_j(t)| + \sum_{j=1}^p |W_j(t)| |t - t_j|^\gamma \right), \end{aligned}$$

c'est-à-dire, compte tenu des conditions (iv) et (v),

$$\begin{aligned} &= \sum_{i=1}^p |W_i(t)| \left(|t - t_i|^\gamma \mathcal{O}(1) + o(1) \right) \\ &= \mathcal{O}(1) \sum_{i=1}^p |W_i(t)| |t - t_i|^\gamma + o(1) \sum_{i=1}^p |W_i(t)| = o(1) \end{aligned}$$

lorsque $p \rightarrow \infty$ uniformément par rapport à n . Étant assurés de la convergence vers zéro du majorant de (2.34), nous pouvons alors conclure à la validité de (2.3). \blacksquare

Démonstration du Théorème 2.4

Le principe de la démonstration est le suivant. On décompose d'abord la MISE de l'estimateur spline de lissage \widehat{f}_λ en biais au carré plus variance. On contrôle le terme de biais à l'aide des résultats de Ragozin (1983) énoncés dans le Chapitre 1. Pour l'étude de la variance, on fait appel au théorème de Mercer comme dans les Théorèmes 2.1 et 2.2 pour représenter les variances de \widehat{f}_λ et du processus X par des séries de fonctions. On intègre ensuite ces séries sur $[0, 1]$ et on majore leur écart à l'aide des propriétés des splines de lissage. Le majorant obtenu fait notamment intervenir la série numérique $\sum_{k=1}^{\infty} \lambda_k \int_0^1 (\varphi_k^{(m_2)}(t))^2 dt$, où les λ_k et les φ_k sont les valeurs propres et les fonctions propres associées à R . La convergence de cette série est obtenue grâce au résultat suivant qui permet de différentier terme à terme le développement de Mercer.

Théorème 2.9. (Kadota, 1967) *Si pour un entier $q \geq 0$, la dérivée croisée $R^{(q,q)}$ d'ordre $2q$ existe et est continue sur $[0, 1] \times [0, 1]$, alors les fonctions $\varphi_1, \varphi_2, \dots$ sont de classe C^q sur $[0, 1]$ et*

$$R^{(q,q)}(s, t) = \sum_{k=1}^{\infty} \lambda_k \varphi_k^{(q)}(s) \varphi_k^{(q)}(t), \quad (2.35)$$

où la convergence est uniforme sur $[0, 1] \times [0, 1]$. Réciproquement, si R est continue sur $[0, 1] \times [0, 1]$, si les fonctions $\varphi_1, \varphi_2, \dots$ sont de classe C^q sur $[0, 1]$ et si la série $\sum_{k=1}^{\infty} \lambda_k \varphi_k^{(q)}(s) \varphi_k^{(q)}(t)$ converge uniformément sur $[0, 1] \times [0, 1]$, alors $R^{q,q}$ existe, est continue et vérifie (2.35).

Les lignes directrices de la démonstration étant définies, nous entrons dans le détail des calculs. Décomposons d'abord la MISE de l'estimateur spline de lissage \widehat{f}_λ en biais au carré plus variance, selon la formule

$$\begin{aligned} \mathbb{E}|f - \widehat{f}_\lambda|_0^2 &= |f - \mathbb{E}\widehat{f}_\lambda|_0^2 + \mathbb{E}|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_0^2 \\ &= |f - \mathbb{E}\widehat{f}_\lambda|_0^2 + \int_0^1 \text{Var} \widehat{f}_\lambda(t) dt. \end{aligned} \quad (2.36)$$

Reprenons alors la majoration du biais au carré $|f - \mathbb{E}\widehat{f}_\lambda|_0^2$ du Chapitre 1. (Comme vu auparavant, le terme de biais ne dépend que de f , de λ , et du dispositif expérimental.) D'après le Théorème 1.11, nous avons

$$|f - \mathbb{E}\widehat{f}_\lambda|_0^2 = \mathcal{O}\left((\lambda + p^{-2m})^{m_1/m}\right) |f|_{m_1}^2. \quad (2.37)$$

lorsque $n, p \rightarrow \infty$, où la constante du \mathcal{O} ne dépend que de m, m_1 , et de la constante finie $r := \sup_{p \geq 1} \frac{\max_{0 \leq i \leq p} (t_{i+1} - t_i)}{\min_{1 \leq i \leq p-1} (t_{i+1} - t_i)}$ liée au dispositif expérimental quasi-uniforme.

Pour la majoration de la variance, notons $\varphi_{k,\lambda}(t) := \sum_{i=1}^p W_i(t)\varphi_k(t_i)$ la spline de lissage d'ordre $2m$ associée à λ , aux points t_1, \dots, t_p , et à la fonction propre φ_k de l'opérateur de covariance.

L'équation (2.30), valable pour tous les estimateurs linéaires, établit que pour tous $n \geq 1$ et $p \geq m$,

$$\int_0^1 \text{Var} \widehat{f}_\lambda(t) dt = \frac{1}{n} \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_{k,\lambda}^2(t) dt. \quad (2.38)$$

Nous savons aussi par la Remarque 2.8 que

$$\int_0^1 R(t, t) dt = \sum_{k=1}^{\infty} \lambda_k \int_0^1 \varphi_k^2(t) dt. \quad (2.39)$$

A partir des représentations (2.38) et (2.39), nous obtenons

$$\begin{aligned} \left| n \int_0^1 \text{Var} \widehat{f}_\lambda(t) dt - \int_0^1 R(t, t) dt \right| &\leq \sum_{k=1}^{\infty} \lambda_k \int_0^1 \left| \varphi_{k,\lambda}^2(t) - \varphi_k^2(t) \right| dt \\ &= \sum_{k=1}^{\infty} \lambda_k \int_0^1 \left| \varphi_{k,\lambda}(t) - \varphi_k(t) \right| \left| \varphi_{k,\lambda}(t) + \varphi_k(t) \right| dt, \end{aligned}$$

puis, avec l'inégalité de Cauchy-Schwartz et la majoration $(a+b)^2 \leq 2(a^2+b^2)$,

$$\leq \sum_{k=1}^{\infty} \lambda_k \left(\int_0^1 (\varphi_{k,\lambda}(t) - \varphi_k(t))^2 dt \right)^{1/2} \left(2 \int_0^1 \varphi_{k,\lambda}^2(t) dt + 2 \int_0^1 \varphi_k^2(t) dt \right)^{1/2}.$$

Avec la notation $|g|_j^2 = \int_0^1 (g^{(j)}(t))^2 dt$, il vient

$$\begin{aligned} \left| n \int_0^1 \text{Var} \widehat{f}_\lambda(t) dt - \int_0^1 R(t, t) dt \right| &\leq 2^{1/2} \sum_{k=1}^{\infty} \lambda_k |\varphi_{k,\lambda} - \varphi_k|_0 \left(|\varphi_{k,\lambda}|_0^2 + |\varphi_k|_0^2 \right)^{1/2} \\ &\leq 2^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k |\varphi_{k,\lambda} - \varphi_k|_0^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k (|\varphi_{k,\lambda}|_0^2 + |\varphi_k|_0^2) \right)^{1/2} \\ &= 2^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k |\varphi_{k,\lambda} - \varphi_k|_0^2 \right)^{1/2} \left(\sum_{k=1}^{\infty} \lambda_k |\varphi_{k,\lambda}|_0^2 + \int_0^1 R(t, t) dt \right)^{1/2} \quad (2.40) \end{aligned}$$

Nous étudions à présent le comportement des deux séries figurant dans (2.40) lorsque $n, p \rightarrow \infty$.

Du fait de l'existence et de la continuité de la dérivée partielle $R^{(m_2, m_2)}$, nous déduisons du Théorème 2.9 que les fonctions $\varphi_k, k \geq 1$, sont dans $C^{m_2}[0, 1]$. Nous appliquons alors le Théorème 1.11 et obtenons directement la majoration

$$|\varphi_{k,\lambda} - \varphi_k|_0^2 = \mathcal{O}\left((\lambda + p^{-2m})^{m_2/m}\right) |\varphi_k|_{m_2}^2$$

lorsque $n, p \rightarrow \infty$, où le \mathcal{O} est uniforme en $k \geq 1$, cf. (2.37).

Le Théorème 1.10 et la relation (1.15) (norme de la matrice de lissage \mathbf{S}_λ égale à 1) permettent respectivement d'obtenir les relations

$$\begin{cases} |\varphi_{k,\lambda}|_0^2 = \mathcal{O}\left(p^{-1} \|\boldsymbol{\varphi}_{k,\lambda}\|^2 + p^{-2m_2} |\varphi_{k,\lambda}|_{m_2}^2\right), \\ \|\boldsymbol{\varphi}_{k,\lambda}\|^2 = \|\mathbf{S}_\lambda \boldsymbol{\varphi}_k\|^2 \leq \|\boldsymbol{\varphi}_k\|^2. \end{cases}$$

Par ailleurs, on utilise le Lemme 1.4 avec $g = \varphi_k$ et $h = \varphi_k - \varphi_{k,\lambda}$ pour établir la positivité de l'intégrale $\int_0^1 (g - h)^{(m_2)}(t) h^{(m_2)}(t) dt = p^{-1} \|\mathbf{h}\|^2$. Il résulte alors de la décomposition $|g|_{m_2}^2 = |h|_{m_2}^2 + |g - h|_{m_2}^2 + 2 \int_0^1 (g - h)^{(m_2)}(t) h^{(m_2)}(t) dt$ que

$$|\varphi_{k,\lambda}|_{m_2}^2 \leq |\varphi_k|_{m_2}^2.$$

Enfin, nous employons la majoration suivante d'une norme discrète par des semi-normes continues (voir par exemple Ragozin, 1983, Théorème 3.7) :

$$p^{-1} \|\boldsymbol{\varphi}_k\|^2 = \mathcal{O}\left(|\varphi_k|_0^2 + p^{-2m_2} |\varphi_k|_{m_2}^2\right).$$

En notant que les \mathcal{O} précédents ne dépendent pas de k , mais seulement de m, m_2 , et de la constante r de (2.37), nous résumons les éléments précédents dans les formules

$$\begin{cases} |\varphi_{k,\lambda} - \varphi_k|_0^2 = \mathcal{O}\left((\lambda + p^{-2m})^{m_2/m}\right) |\varphi_k|_{m_2}^2, \\ |\varphi_{k,\lambda}|_0^2 = \mathcal{O}(1) |\varphi_k|_0^2 + \mathcal{O}(p^{-2m_2}) |\varphi_k|_{m_2}^2. \end{cases} \quad (2.41)$$

Substituons alors (2.41) dans (2.40) : nous obtenons, lorsque $n, p \rightarrow \infty$,

$$\begin{aligned} & \left| n \int_0^1 \text{Var} \widehat{f}_\lambda(t) dt - \int_0^1 R(t, t) dt \right| \\ &= \mathcal{O}\left((\lambda + p^{-2m})^{m_2/2m}\right) \times \left(\sum_{k=1}^{\infty} \lambda_k |\varphi_k|_{m_2}^2 \right)^{1/2} \\ & \times \mathcal{O}\left(\left(p^{-2m_2} \sum_{k=1}^{\infty} \lambda_k |\varphi_k|_{m_2}^2 + \int_0^1 R(t, t) dt \right)^{1/2} \right). \end{aligned} \quad (2.42)$$

Il reste enfin à prouver la convergence de la série $\sum_{k=1}^{\infty} \lambda_k |\varphi_k|_{m_2}^2$ figurant dans (2.42). Celle-ci découle du Théorème 2.9 dans lequel l'intégration du développement uniforme (2.35) donne $\sum_{k=1}^{\infty} \lambda_k |\varphi_k|_{m_2}^2 = \int_0^1 R^{(m_2, m_2)}(t, t) dt < \infty$. La démonstration se termine alors en regroupant les majorations asymptotiques (2.37) et (2.42). ■

Démonstration du Théorème 2.5

Soient deux entiers $0 \leq k < m$. Nous considérons l'estimateur spline de lissage \widehat{f}_λ d'ordre $2m$ associé aux points t_1, \dots, t_p , au processus moyen \overline{X} , et au paramètre $\lambda = \lambda(n, p)$. Nous allons étudier la convergence presque sûre de l'écart quadratique $|f - \widehat{f}_\lambda|_k^2 = \int_0^1 (f^{(k)}(t) - \widehat{f}_\lambda^{(k)}(t))^2 dt$ lorsque $n, p \rightarrow \infty$.

Majorons d'abord $|f - \widehat{f}_\lambda|_k^2$ à l'aide du biais et de l'estimateur centré $\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda$:

$$|f - \widehat{f}_\lambda|_k^2 \leq 2 \left(|f - \mathbb{E}\widehat{f}_\lambda|_k^2 + |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_k^2 \right). \quad (2.43)$$

D'après le Théorème 1.11, si le dispositif expérimental est quasi-uniforme, alors le biais au carré est majoré de la façon suivante lorsque $n, p \rightarrow \infty$:

$$|f - \mathbb{E}\widehat{f}_\lambda|_k^2 = \mathcal{O}\left(\lambda + p^{-2m}\right)^{(m-k)/m} |f|_k^2 \quad (2.44)$$

Par ailleurs, l'étude de $|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_k^2$ se déduit de celle de la norme discrète $\|\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2$ et de $|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2$ grâce au Théorème 1.10 qui affirme que

$$p^{-2k} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_k^2 \leq Cp^{-1} \|\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2 + Dp^{-2m} |\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2 \quad (2.45)$$

où C et D sont deux constantes finies qui ne dépendent pas de n et p .

D'après la relation (1.15) ($\|\mathbf{S}_\lambda\| = 1$), nous avons d'abord

$$p^{-1} \|\widehat{\mathbf{f}}_\lambda - \mathbb{E}\widehat{\mathbf{f}}_\lambda\|^2 = p^{-1} \|\mathbf{S}_\lambda \overline{\mathbf{X}} - \mathbf{S}_\lambda \mathbf{f}\|^2 = p^{-1} \|\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}}\|^2 \leq p^{-1} \|\overline{\boldsymbol{\varepsilon}}\|^2 \leq \|\overline{\boldsymbol{\varepsilon}}\|_\infty^2. \quad (2.46)$$

Intéressons nous à présent à la quantité $|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2$. Avec d'une part, la matrice d'énergie $\boldsymbol{\Gamma}$ du Théorème 1.1 et l'inégalité $\|\boldsymbol{\Gamma}\| \leq Dp^{2m-1}$ déduite de (1.22a) et (1.22b) (\tilde{D} est une constante), et d'autre part le Lemme 1.4 et la définie positivité de \mathbf{S}_λ , nous obtenons les expressions

$$|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_m^2 = \begin{cases} (\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}})' \boldsymbol{\Gamma} (\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}}) \leq \|\boldsymbol{\Gamma}\| \|\mathbf{S}_\lambda\|^2 \|\overline{\boldsymbol{\varepsilon}}\|^2 \leq \tilde{D} p^{2m} \|\overline{\boldsymbol{\varepsilon}}\|_\infty^2; \\ (p\lambda)^{-1} (\mathbf{I} - \mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}})' (\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}}) \leq (p\lambda)^{-1} (\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}})' (\mathbf{S}_\lambda \overline{\boldsymbol{\varepsilon}}) \leq \lambda^{-1} \|\overline{\boldsymbol{\varepsilon}}\|_\infty^2. \end{cases} \quad (2.47)$$

Exploitions maintenant la version bornée de la loi du logarithme itéré dans $C[0, 1]$ que vérifie le processus X . Il vient, avec une probabilité 1 lorsque $n \rightarrow \infty$,

$$\|\overline{\boldsymbol{\varepsilon}}\|_\infty^2 = \mathcal{O}(n^{-1} \log_2 n). \quad (2.48)$$

Nous déduisons finalement des relations (2.45)–(2.48) la majoration presque sûre

$$|\widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda|_k^2 = \mathcal{O}\left(p^{2k} \min(1, \lambda^{-1} p^{-2m}) n^{-1} \log_2 n\right) \quad (2.49)$$

lorsque $n, p \rightarrow \infty$.

La réunion des majorations (2.43), (2.44), et (2.49) achève alors la démonstration du Théorème 2.5. ■

Démonstration du Lemme 2.1

Soit $t \in]0, 1[$. Supposons les hypothèses (i) et (ii) du lemme vérifiées.

Soit φ la fonction définie sur $[0, \min(t, 1-t)] \times [-1, 1]^2$ par

$$\varphi(h, u, v) = \begin{cases} K(u)K(v) \frac{R(t+uh, t+vh) - R(t, t)}{h} & \text{si } h \neq 0, \\ K(u)K(v) D_{(u,v)}^+ R(t, t) & \text{si } h = 0. \end{cases}$$

Par construction de φ et d'après (ii), la fonction $h \mapsto \varphi(h, u, v)$ est continue au point $h = 0$ pour tout couple $(u, v) \in [-1, 1]^2$. De plus, d'après (i) nous avons $\varphi(h, u, v) \leq CK(u)K(v)$ pour tous h, u, v , le majorant précédent étant intégrable par continuité de K . Par conséquent, le théorème de convergence dominée de Lebesgue permet d'affirmer que

$$\iint_{[-1,1]^2} \varphi(h, u, v) dudv = \iint_{[-1,1]^2} \varphi(0, u, v) dudv + o(1)$$

lorsque $h \rightarrow 0$, ce qui fournit immédiatement (2.15a).

Supposons à présent l'hypothèse (iii) du lemme vérifiée. La fonction R étant de classe C^2 au voisinage de (t, t) , elle admet le développement de Taylor suivant :

$$\begin{aligned} R(t+uh, t+vh) &= R(t, t) + uhR^{(1,0)}(t, t) + vhR^{(0,1)}(t, t) \\ &\quad + \frac{(uh)^2}{2} R^{(2,0)}(t, t) + uvh^2 R^{(1,1)}(t, t) \\ &\quad + \frac{(vh)^2}{2} R^{(0,2)}(t, t) + o((u+v)^2h^2) \end{aligned} \quad (2.50)$$

lorsque $h \rightarrow 0$, où le $o((u+v)^2h^2) = o(h^2)$ est uniforme par rapport à u et v d'après la continuité des dérivées d'ordre 2 au voisinage de (t, t) , disons $]t \pm \lambda[\times]t \pm \lambda[$.

Soit alors ψ la fonction définie sur $[0, \lambda \times [0, 1]^2$ par

$$\psi(h, u, v) = \begin{cases} K(u)K(v) \frac{R(t+uh, t+vh) - R(t, t) - uhR^{(1,0)}(t, t) - vhR^{(0,1)}(t, t)}{h^2} \\ \text{si } h \neq 0, \\ K(u)K(v) \left(\frac{u^2}{2} R^{(2,0)}(t, t) + uv R^{(1,1)}(t, t) + \frac{v^2}{2} R^{(0,2)}(t, t) \right) \\ \text{si } h = 0. \end{cases}$$

Alors, d'après (2.50), il apparaît que $h \mapsto \psi(h, u, v)$ est continue au point $h = 0$ pour tout couple $(u, v) \in [0, 1]^2$. De plus, l'application de la formule de Taylor-Lagrange à cette fonction et la continuité des dérivées d'ordre 2 de R impliquent que

$$|\psi(h, u, v)| \leq u^2 \max_{|t \pm \lambda|^2} (|R^{(2,0)}|) + 2uv \max_{|t \pm \lambda|^2} (|R^{(1,1)}|) + v^2 \max_{|t \pm \lambda|^2} (|R^{(0,2)}|).$$

pour tous h, u, v dans $[0, \lambda \times [0, 1]^2]$. Par conséquent, nous pouvons à nouveau utiliser le théorème de convergence dominée et obtenir

$$\begin{aligned}
& \iint_{[0,1]^2} K(u)K(v)R(t+uh, t+vh)dudv \\
&= \iint_{[0,1]^2} K(u)K(v)R(t, t)dudv \\
&\quad - h \iint_{[0,1]^2} K(u)K(v)(uR^{(1,0)}(t, t) - vR^{(0,1)}(t, t))dudv \\
&\quad + \frac{h^2}{2} \iint_{[0,1]^2} K(u)K(v) (u^2 R^{(2,0)}(t, t) + 2uv R^{(1,1)}(t, t)) dudv \\
&\quad + \frac{h^2}{2} \iint_{[0,1]^2} K(u)K(v)v^2 R^{(0,2)}(t, t)dudv + o(h^2).
\end{aligned} \tag{2.51}$$

Il reste, pour établir la relation (2.15b), à exploiter le fait que $\int_0^1 uK(u)du = 0$ (car K est un noyau symétrique) et le fait que $R^{(2,0)}(t, t) = R^{(0,2)}(t, t)$ (car R est symétrique). ■

Démonstration du Théorème 2.6

Ce théorème est facile à démontrer car tous les résultats intermédiaires ont été établis au préalable : le biais de $\widehat{f}_{n,p}^{GM}(t)$ est majoré dans (2.13) et la variance, d'abord approchée dans (2.14) par une intégrale dépendant de la largeur de fenêtre h , est ensuite explicitée par le Lemme 2.1. Le seul point à vérifier est que dans l'approximation de la MSE, les restes sont effectivement négligeables devant les termes principaux.

Supposons les hypothèses (i)-(ii)-(iii)-(iv) vérifiées. Nous avons donc

$$\begin{aligned}
\mathbb{E}\left(f(t) - \widehat{f}_{n,p}^{GM}(t)\right)^2 &= \frac{R(t, t) + C_{K,R,t}h}{n} + o\left(\frac{h}{n}\right) \\
&\quad + \frac{h^4(f''(t))^2}{4} + o(h^4) + \mathcal{O}\left(\frac{h^2}{p}\right) + \mathcal{O}\left(\frac{1}{p^2}\right).
\end{aligned} \tag{2.52}$$

Avec les hypothèses (iii) ($n = \mathcal{O}(p)$) et (iv) ($h = o(1)$), il vient

$$h^2/p = o(h/n) = o(1/n) \quad \text{et} \quad 1/p^2 = o(1/n). \tag{2.53}$$

En joignant les formules (2.52) et (2.53), nous obtenons alors la relation (2.16a) attendue. La relation (2.16b) est démontrée de manière analogue. ■

Démonstration du Théorème 2.7

Afin d'établir le Théorème 2.7, nous allons commencer par écrire l'estimateur spline de lissage \widehat{f}_λ sous une forme adaptée. Nous introduirons ensuite quelques notations et ferons un rappel sur la fonction de Green associée à une équation différentielle. Nous rappellerons alors l'expression du biais asymptotique de \widehat{f}_λ (Nychka, 1995, Théorème 2.2), puis nous étudierons le terme de variance en étendant au cadre processus aléatoire la méthodologie de Nychka (1995) développée avec des erreurs de mesure décorréélées et de variance σ^2 .

L'estimateur spline de lissage est linéaire : il s'écrit sous la forme $\widehat{f}_\lambda(t) = \sum_{i=1}^p W_i(t) \overline{X}_n(t_i)$ comme dans la formule (2.1). De plus, les travaux de Cox (1983) et le Lemme 3.1 de Nychka (1995) ont montré que les poids $W_i(t)$ peuvent s'écrire, avec un changement d'échelle, sous la forme $p^{-1}W(t_i, t)$, où W est une fonction symétrique de deux variables qui dépend des points t_1, \dots, t_p et du paramètre λ . Pour la simplicité des calculs ultérieurs et notamment pour l'approximation de W par une fonction de Green, nous utiliserons cette dernière notation :

$$\widehat{f}_\lambda(t) = \frac{1}{p} \sum_{i=1}^p W(t_i, t) \overline{X}_n(t_i). \quad (2.54)$$

Notations.

On rappelle que la fonction de répartition du dispositif expérimental est notée F_p et que sa limite F est de classe C^2 sur $[0, 1]$, avec $F' = \psi$. Ensuite, pour toute fonction $g : [0, 1] \rightarrow \mathbb{R}$ intégrable par rapport à la mesure de Lebesgue, nous écrivons

$$\int_0^1 g(t) dF_p(t) = \frac{1}{p} \sum_{i=1}^p g(t_i), \quad \text{et} \quad \int_0^1 g(t) dF(t) = \int_0^1 g(t) \psi(t) dt.$$

Pour les fonctions de deux variables, nous notons $F_p \otimes F_p$ et $F \otimes F$ les mesures produits. En particulier,

$$\begin{aligned} \iint_{[0,1]^2} g(t, \tau) d(F_p \otimes F_p)(t, \tau) &= \frac{1}{p^2} \sum_{i=1}^p \sum_{j=1}^p g(t_i, t_j), \\ \iint_{[0,1]^2} g(t, \tau) d(F \otimes F)(t, \tau) &= \iint_{[0,1]^2} g(t, \tau) \psi(t) \psi(\tau) dt d\tau. \end{aligned}$$

Rappelons aussi la notation $g^{(m_1, m_2)}(s, \tau)$ pour la dérivée $(\partial t^{m_1} \partial \tau^{m_2})g(t, \tau)$ d'une fonction g de deux variables, et la notation $\|F_p - F\|_\infty = \sup_{t \in [0,1]} |F_p(t) - F(t)|$ pour la norme du supremum. On pose également $\rho = \lambda^{1/2m}$.

Fonction de Green.

Étant donné un réel $\lambda > 0$ et une fonction ψ continue sur $[0, 1]$, on note G_λ la fonction de Green associée à l'équation différentielle

$$\lambda(-1)^m h^{(2m)}(t) + \psi(t)h(t) = \psi(t)g(t), \quad (2.55)$$

sous les conditions limites $h^{(k)}(0) = h^{(k)}(1) = 0$ pour $m \leq k \leq 2m - 1$ ($g \in C[0, 1]$ est une fonction "source" dont ne dépend pas G_λ). Rappelons les faits suivants.

1. La fonction $G_\lambda(t, \tau)$ est symétrique en ses arguments.
2. Pour tout point $\tau \in [0, 1]$, $G_\lambda(\cdot, \tau)$ est de classe C^{2m-2} sur $[0, 1]$, de classe C^{2m} sur $([0, \tau] \cup [\tau, 1])$, et elle satisfait aux conditions limites de (2.55). De plus, $G^{(2m-1,0)}(\cdot, \tau)$ a un saut de discontinuité au point $t = \tau$.
3. La fonction $h : t \mapsto \int_0^1 G_\lambda(t, s)g(s)\psi(s)ds$, est solution de (2.55).
4. Pour toute fonction $h \in W_2^m[0, 1]$ et pour tout $t \in [0, 1]$, on a la relation

$$\int_0^1 G_\lambda(t, \tau)h(\tau)dF(\tau) + \lambda \int_0^1 G_\lambda^{(m,0)}(t, \tau)h^{(m)}(\tau)dF(\tau) = h(t), \quad (2.56)$$

ce qui revient à dire que G_λ est le noyau autoreproduisant de $W_2^m[0, 1]$ muni du produit scalaire $\langle h_1, h_2 \rangle = \int_0^1 h_1 h_2 \psi(\tau)d\tau + \lambda \int_0^1 h_1^{(m)}(\tau)h_2^{(m)}(\tau)d\tau$.

Hypothèses de Nychka (1995) dans le cas $m = 1$.

Nous présentons maintenant les hypothèses sur la fonction de Green permettant d'établir les deux résultats principaux de Nychka (1995) utilisés dans cette démonstration. Nous rappelons que ces hypothèses sont valides dans les cas où ψ est strictement positive et de classe C^1 sur $[0, 1]$ avec $m = 1$ (Nychka, 1995), et où $\psi \equiv 1$ (répartition limite uniforme) avec $m \geq 1$ (Messer et Goldstein, 1993).

(A) (condition d'enveloppe exponentielle). Il existe des constantes positives finies A_1, A_2, K telles que pour tous $0 \leq t, \tau \leq 1$,

$$|G_\lambda(t, \tau)| \leq \left(\frac{K}{\rho}\right) \exp\left(- (A_1 + A_2) \frac{|t - \tau|}{\rho}\right).$$

Si $G_\lambda^{(1,0)}$ et $G_\lambda^{(1,1)}$ existent pour $0 \leq t, \tau \leq 1$, alors

$$\begin{aligned} |G_\lambda^{(1,0)}(t, \tau)| &\leq \left(\frac{K}{\rho^2}\right) \exp\left(- (A_1 + A_2) \frac{|t - \tau|}{\rho}\right), \\ |G_\lambda^{(1,1)}(t, \tau)| &\leq \left(\frac{K}{\rho^3}\right) \exp\left(- (A_1 + A_2) \frac{|t - \tau|}{\rho}\right). \end{aligned}$$

Si $G_\lambda^{(1,0)}$ n'est pas continue pour $t = \tau$, alors

$$\lim_{\tau \rightarrow t^-} G_\lambda^{(1,0)}(t, \tau) - \lim_{\tau \rightarrow t^+} G_\lambda^{(1,0)}(t, \tau) = \frac{1}{\lambda}.$$

$$(B) \quad \delta_p := 2K(1/A_1 + 1/A_2)(\|F_p - F\|_\infty/\rho) < 1.$$

Remarque 2.9. Dans sa Section 7, Nychka (1995) indique de façon heuristique comment étendre ses résultats au cas où $m \geq 2$ avec $\psi \in C^{2m-1}[0, 1]$ vérifiant certaines conditions aux bornes. Il précise notamment une hypothèse sur $G_\lambda^{(k,0)}$, $k = 1, \dots, 2m-1$ et sur $G_\lambda^{(2m-1,1)}$ analogue à l'Hypothèse (A), ainsi que la stratégie pour établir la validité de cette hypothèse analogue.

Majoration du biais au carré.

Le biais de \widehat{f}_λ , qui ne dépend pas de la structure aléatoire des données, se traite comme dans le Théorème 2.2 de Nychka (1995). Avec les hypothèses (i) et (v) de notre théorème, il vient

$$f(t) - \mathbb{E}\widehat{f}_\lambda(t) = \frac{\lambda f^{(2m)}(t)}{\psi(t)} + \mathcal{O}(\lambda^2) \quad (2.57)$$

lorsque $n, p \rightarrow \infty$, uniformément par rapport à $t \in [\Delta, 1 - \Delta]$.

Majoration de la variance.

Concernant la variance, nous procédons en trois étapes : d'abord, nous décomposons la variance en une somme de termes notés V, N_1, N_2 , et N_3 ; ensuite, nous déterminons un équivalent asymptotique du terme principal V , sous la forme d'un polynôme du second degré en λ ; enfin, nous montrons que N_1, N_2 , et N_3 sont négligeables par rapport à λ^2 lorsque $p \rightarrow \infty$ uniformément par rapport à $n \geq 1$.

Soit $t \in [\Delta, 1 - \Delta]$. On décompose la variance de l'estimateur à l'aide des fonctions de poids W et de Green G_λ d'une part, et des fonctions de répartition F_p et F d'autre part :

$$\begin{aligned} n \operatorname{Var} \widehat{f}_\lambda(t) &= \iint_{[0,1]^2} W(t, s)W(t, \tau)R(s, \tau) d(F_p \otimes F_p)(s, \tau) \\ &= \iint_{[0,1]^2} G_\lambda(t, s)G_\lambda(t, \tau)R(s, \tau) d(F \otimes F)(s, \tau) \\ &\quad + \iint_{[0,1]^2} W(t, s)W(t, \tau)R(s, \tau) d((F_p - F) \otimes (F_p - F))(s, \tau) \\ &\quad + 2 \iint_{[0,1]^2} \{W(t, s) - G_\lambda(t, s)\}G_\lambda(t, \tau)R(s, \tau) d(F \otimes F)(s, \tau) \\ &\quad + \iint_{[0,1]^2} \{W(t, s) - G_\lambda(t, s)\}\{W(t, \tau) - G_\lambda(t, \tau)\}R(s, \tau) d(F \otimes F)(s, \tau) \\ &:= V + N_1 + N_2 + N_3. \end{aligned} \quad (2.58)$$

Nous abordons maintenant le calcul de V . En premier lieu, le théorème de Fubini permet d'écrire que

$$V = \int_0^1 \left(\int_0^1 G_\lambda(t, s) R(s, \tau) dF(s) \right) G_\lambda(t, \tau) dF(\tau). \quad (2.59)$$

Nous présentons ensuite un résultat nécessaire au calcul de l'intégrale intérieure de (2.59). Ce résultat, issu de la démonstration du Lemme 6.1 de Nychka (1995), fait appel à certaines normes d'interpolation définies ci-après. Soit $(\varphi_\nu)_{\nu \in \mathbb{N}^*}$ une base de $W_2^m[0, 1]$ telle que $\int_0^1 \varphi_k(t) \varphi_\nu(t) dF(t) = \delta_{k\nu}$ et $\int_0^1 \varphi_k^{(m)}(t) \varphi_\nu^{(m)}(t) dt = \gamma_\nu \delta_{k\nu}$ pour tout $\nu \geq 1$, avec $\gamma_1 \leq \gamma_2 \leq \dots$ ($\delta_{k\nu}$ désigne le delta de Kronecker). Soit également θ une fonction de classe C^{2m} sur $[0, 1]$ telle que $|\theta^{(2m)}(s) - \theta^{(2m)}(\tau)| \leq C|s - \tau|^\beta$ pour des constantes $\beta \in]0, 1]$ et $C > 0$, et vérifiant les conditions aux bornes $\theta^{(k)}(0) = \theta^{(k)}(1) = 0$ pour $m \leq k \leq 2m - 1$. Nychka (1995) montre par une adaptation des résultats de Cox (1988) que la série

$$\|\theta\|_\mu^2 = \sum_{\nu \geq 1} (1 + \gamma_\nu^\mu) \int_0^1 \theta(t) \varphi_\nu(t) dF(t) \quad (2.60)$$

est convergente pour tout réel $\mu < 2 + \frac{1}{2m} + \frac{\beta}{2m}$ et qu'elle définit alors une norme. Il montre aussi que si $\mu > 2 + \frac{1}{2m}$, alors

$$\left| \frac{d^{2m}}{dt^{2m}} \int_0^1 G_\lambda(s, t) \theta(s) dF(s) - \theta^{(2m)}(t) \right| \leq M (\|\theta\|_2^2 + \|\theta\|_\mu^2) \lambda^2 \quad (2.61)$$

pour tout $t \in [0, 1]$ et tout $\lambda > 0$, où M est une constante ne dépendant que de m . Revenons au calcul de $\int_0^1 G_\lambda(t, s) R(s, \tau) dF(s)$ dans (2.59). En suivant les arguments de (Nychka, 1995, p.1190) et en utilisant les hypothèses (ii) et (iii), nous pouvons montrer que

$$\int_0^1 G_\lambda(t, s) R(s, \tau) dF(s) = R(t, \tau) + \frac{(-1)^{m-1} \lambda}{\psi(t)} R^{(2m, 0)}(t, \tau) + \mathcal{O}\left(\|\tilde{R}_\tau\|_\mu^2 \lambda^3\right) \quad (2.62)$$

lorsque $\lambda \rightarrow 0$, uniformément par rapport à $\tau \in [0, 1]$, à $t \in [\Delta, 1 - \Delta]$ et à $\mu \in [2 + \frac{1}{2m}, 2 + \frac{1}{2m} + \frac{\beta}{2m}]$. Pour mémoire, \tilde{R}_τ est une fonction qui coïncide avec $R(\cdot, \tau)$ sur $[\Delta/2, 1 - \Delta/2]$ et qui vérifie les conditions aux bornes requises dans (2.60). Donnons ici un aperçu des arguments menant à (2.62). Il s'agit d'abord d'écrire $\int_0^1 G_\lambda(t, s) R(s, \tau) dF(s)$ comme la somme de $\int_0^1 G_\lambda(t, s) (R(s, \tau) - \tilde{R}_\tau(s)) dF(s)$ et de $\int_0^1 G_\lambda(t, s) \tilde{R}_\tau(s) dF(s)$. Pour tout $t \in [\Delta, 1 - \Delta]$, on montre avec la majoration $\sup_{s, \tau \in [0, 1]} |R(s, \tau) - \tilde{R}_\tau(s)| < \infty$ et avec l'hypothèse **(A)** que $\int_0^1 G_\lambda(t, s) (R(s, \tau) - \tilde{R}_\tau(s)) dF(s) = o\left(\frac{1}{\rho} \exp\left(-\frac{(A_1 + A_2)\Delta}{2\rho}\right)\right) = \mathcal{O}(\lambda^3)$. On utilise ensuite la propriété

3. de la fonction de Green (page 95) et la relation (2.61) avec l'autre intégrale $\int_0^1 G_\lambda(t, s) \tilde{R}_\tau(s) dF(s)$ pour établir (2.62).

Puisque le \mathcal{O} de (2.62) est uniforme en $\tau \in [0, 1]$, et d'après l'hypothèse (iii) ($\sup_{\tau \in [0, 1]} \|\tilde{R}_\tau\|_\mu < \infty$), l'équation (2.59) peut maintenant s'écrire

$$V = \int_0^1 G_\lambda(t, \tau) R(t, \tau) dF(\tau) + \frac{(-1)^{m-1} \lambda}{\psi(t)} \int_0^1 G_\lambda(t, \tau) R^{(2m, 0)}(t, \tau) dF(\tau) + \mathcal{O}(\lambda^3).$$

En appliquant le même raisonnement qui a mené à (2.62) aux deux intégrales précédentes, nous obtenons

$$V = R(t, t) + 2(-1)^{m-1} \lambda R^{(2m, 0)}(t, t) + \lambda^2 R^{(2m, 2m)}(t, t) + \mathcal{O}(\lambda^3). \quad (2.63)$$

Finalement, nous passons à la dernière étape qui consiste à montrer que N_1, N_2 et N_3 sont négligeables par rapport à λ^2 dans l'asymptotique. Dans ce but, les deux résultats suivants seront mis à contribution.

Lemme 2.2. (Nychka, 1995) *Soit $g : [0, 1] \rightarrow \mathbb{R}$ une fonction absolument continue. Alors pour tout $p \geq 1$,*

$$\left| \int_0^1 g(t) d(F_p - F)(t) \right| \leq \|F_p - F\|_\infty \int |g'(t)| dt.$$

Théorème 2.10. (Nychka, 1995) *Posons $\rho = \lambda^{1/2m}$. Sous les Hypothèses (A) et (B), nous avons, pour tout $p \geq 1$ et pour tous $0 \leq t, \tau \leq 1$,*

$$\begin{aligned} |W(t, \tau)| &\leq \frac{K}{(1 - \delta_p)\rho} \exp\left(-A_1 \frac{|t - \tau|}{\rho}\right), \\ |W(t, \tau) - G_\lambda(t, \tau)| &\leq \frac{\delta_p K}{(1 - \delta_p)\rho} \exp\left(-A_1 \frac{|t - \tau|}{\rho}\right), \\ |W^{(1, 0)}(t, \tau) - G_\lambda(t, \tau)| &\leq \frac{K}{(1 - \delta_p)\rho^3} \exp\left(-A_1 \frac{|t - \tau|}{\rho}\right). \end{aligned}$$

Observons d'abord que la covariance R , supposée continue sur le compact $[0, 1]^2$, est bornée. On peut donc l'ignorer dans la majoration des intégrales N_1, N_2 et N_3 définies par (2.58) puis séparer les intégrales doubles en produit d'intégrales simples.

D'après le Lemme 2.2 et le Théorème 2.10, il vient

$$\begin{aligned} N_1 &\leq C \left(\int_0^1 |W(t, \tau)| d(F_p - F)(\tau) \right)^2 \leq C \left(\int_0^1 \left| \frac{\partial}{\partial t} W(t, \tau) \right| d\tau D_p \right)^2 \\ &\leq C \left(\int_0^1 \frac{1}{\rho^2} \exp\left(-\frac{|t-\tau|}{\rho}\right) d\tau \right)^2 D_p^2 = \mathcal{O}\left(\frac{D_p^2}{\rho^2}\right), \end{aligned} \quad (2.64)$$

$$\begin{aligned} N_2 &\leq C \left(\int_0^1 |W(t, s) - G_\lambda(t, s)| ds \right) \left(\int_0^1 |G_\lambda(t, \tau)| d\tau \right) \\ &\leq C \left(\int_0^1 \frac{D_p}{\rho^2} \exp\left(-\frac{|t-s|}{\rho}\right) ds \right) \left(\int_0^1 \frac{1}{\rho} \exp\left(-\frac{|t-\tau|}{\rho}\right) d\tau \right) \\ &= \mathcal{O}\left(\frac{D_p}{\rho}\right), \end{aligned} \quad (2.65)$$

et aussi

$$\begin{aligned} N_3 &\leq C \left(\int_0^1 |W(t, \tau) - G_\lambda(t, \tau)| dt \right)^2 \\ &\leq C \left(\int_0^1 \frac{D_p}{\rho^2} \exp\left(-\frac{|x-t|}{\rho}\right) dt \right)^2 = \mathcal{O}\left(\frac{D_p^2}{\rho^2}\right). \end{aligned} \quad (2.66)$$

Il reste à constater que sous les hypothèses (iv) et (v), on a

$$D_p^2/\rho^2 = o(D_p/\rho) = o(\lambda^2)$$

lorsque $p \rightarrow \infty$, uniformément par rapport à $n \geq 1$. Rassemblant (2.57), (2.58), (2.63), et (2.64)–(2.66), on finit alors la démonstration du Théorème 2.7. ■

Chapitre 3

Pratique du lissage spline avec des mesures répétées

Ce chapitre porte sur la régression par splines de lissage dans un contexte de mesures répétées. L'accent est placé sur la pratique du lissage, et notamment sur le choix de la métrique et du paramètre de lissage utilisés pour la construction de l'estimateur. L'estimateur spline de lissage des moindres carrés généralisés [MCG], qui tient compte de la structure de covariance des erreurs du modèle, est défini. Son erreur quadratique MISE est comparée à celle de l'estimateur des moindres carrés ordinaires [MCO] du Chapitre 1 sur quelques exemples de fonction de régression et de structure de bruit. Les principales méthodes de construction d'estimateurs splines de lissage en présence de données corrélées sont présentées, et deux méthodes spécifiques aux mesures répétées sont comparées lors de simulations.

L'estimateur spline de lissage MCG est un estimateur qui semble particulièrement adapté au traitement semi/non paramétrique de données longitudinales ou clusterisées. Il est largement utilisé dans la pratique, d'autant plus que son lien avec les modèles linéaires mixtes (Brumback et Rice, 1998; Verbyla *et al.*, 1999) rend son implémentation facile. Ses propriétés théoriques ont fait l'objet de plusieurs travaux. Dans un modèle de régression longitudinale à points d'observation aléatoires, Welsh *et al.* (2002) ont montré que cet estimateur n'est pas local au sens où lorsque le nombre d'observations tend vers l'infini, l'estimation en un point x assigne un poids non négligeable aux observations distantes de x . Dans le même modèle, Lin *et al.* (2004) ont déterminé le biais et la variance asymptotiques ponctuels de cet estimateur à l'aide de la théorie des noyaux équivalents.

Dans le cas d'un échantillon corrélé et fini, Opsomer *et al.* (2001) ont observé que les méthodes usuelles de sélection du paramètre de lissage de l'estimateur spline MCO (voir Wahba, 1990, pour un exposé) réalisent de mauvaises performances car, développées sous l'hypothèse d'erreurs de mesure i.i.d., elles attribuent toute

la régularité des données à la fonction de régression et aucune à la corrélation des erreurs. L'adaptation de ces méthodes au cadre d'erreurs corrélées et l'estimation sous-jacente de la structure de covariance des erreurs ont été envisagées par de nombreux auteurs, dont Diggle et Hutchinson (1989), Kohn *et al.* (1992), Wang (1998), Zhang *et al.* (1998), et Lin et Zhang (1999). Souvent, cette adaptation nécessite le passage de la métrique MCO à la métrique MCG dans le critère de construction de l'estimateur, ce qui légitime l'usage de l'estimateur MCG. Cependant la littérature ne comporte pas, à notre connaissance, de comparaison rigoureuse des performances théoriques et/ou pratiques des deux estimateurs splines MCO et MCG construits avec le *même* paramètre du lissage.

Le présent chapitre offre une telle comparaison pour quelques exemples de fonctions de régression et de structures de bruit caractéristiques. Il apparaîtra dans ces exemples que les estimateurs splines MCO et MCG ont des performances similaires, et que la tâche véritablement décisive de l'estimation est le choix du paramètre de lissage. Un autre élément novateur de ce chapitre est la présentation d'une méthode originale de sélection du paramètre de lissage basée sur le calcul de la variance intégrée d'un estimateur spline cubique fait au Chapitre 1 et visant à rendre minimale l'erreur d'estimation MISE. Cette méthode est ensuite comparée par simulation à la validation croisée généralisée de Rice et Silverman (1991), dont l'efficacité pratique n'a pas été étudiée dans la littérature.

Le reste du chapitre est organisé de la manière suivante. L'estimateur spline de lissage MCG est présenté dans un modèle de mesures répétées équilibrées en Section 3.1. Son erreur quadratique moyenne aux points d'observation est étudiée en Section 3.1.1 à l'aide d'une extension des méthodes du Chapitre 1, et ses performances sur un échantillon fini sont comparées à celles de l'estimateur spline de lissage MCO dans une simulation qui figure en Section 3.1.2.

En Section 3.2 plusieurs méthodes de sélection du paramètre de lissage et/ou d'estimation de la structure de covariance des erreurs sont exposées. La Section 3.2.1 présente l'extension par Wang (1998) et Diggle et Hutchinson (1989) des méthodes usuelles du cadre d'erreurs i.i.d. au cadre corrélé. Ces extensions du maximum de vraisemblance généralisée [GML], de l'estimation sans biais du risque [UBR], et de la validation croisée généralisée [GCV] ont en commun le fait qu'elles adoptent une modélisation paramétrique de la covariance et qu'elles estiment simultanément les paramètres de lissage et de covariance. La méthode présentée en Section 3.2.2 est la validation croisée généralisée "GCV leave one curve out" de Rice et Silverman (1991). La méthode de la Section 3.2.3 vise à minimiser l'erreur quadratique MISE de l'estimateur spline cubique de lissage grâce au calcul explicite de sa variance intégrée effectué dans la Section 1.2.4 du Chapitre 1. Les deux dernières méthodes, qui sont propres au cadre de mesures répétées et qui n'utilisent pas de modélisation paramétrique de la covariance, sont comparées en Section 3.3 lors d'une simulation.

3.1 Estimateur spline de lissage des moindres carrés généralisés

Nous considérons à nouveau le modèle (1.4) de mesures répétées du Chapitre 1

$$y_{ij} = f(t_j) + \varepsilon_{ij}, \quad 1 \leq i \leq n, 1 \leq j \leq p,$$

où y_{ij} est la j -ème mesure effectuée sur la i -ème unité statistique au point d'observation t_j (avec $0 \leq t_1 < \dots < t_p \leq 1$), f est la fonction de régression, et ε_{ij} est une erreur aléatoire de mesure. On suppose que f est dans l'espace de Sobolev $W_2^m[0, 1]$ et que les vecteurs d'erreurs $(\varepsilon_{i1}, \dots, \varepsilon_{ip})'$, $1 \leq i \leq n$, sont indépendants, de moyenne nulle, et ont une matrice de covariance commune \mathbf{V} de taille $p \times p$.

Nous allons maintenant voir comment généraliser le critère (1.2) de construction de l'estimateur spline de lissage usuel, de manière à tenir explicitement compte de la structure de covariance des données. En termes mathématiques, cette généralisation est faite par l'introduction de poids dans la somme des carrés des résidus de l'ajustement, ce qui revient à spécifier une matrice de "métrique" \mathbf{M} de taille $p \times p$. Pour un réel $\lambda > 0$ donné, nous considérons alors le problème général suivant :

$$\min_{g \in W_2^m[0,1]} \frac{1}{np} \sum_{i=1}^n (\mathbf{g} - \mathbf{y}_i)' \mathbf{M} (\mathbf{g} - \mathbf{y}_i) + \lambda \int_0^1 (g^{(m)}(t))^2 dt, \quad (3.1)$$

avec les notations $\mathbf{g} = (g(t_1), \dots, g(t_p))'$ et $\mathbf{y}_i = (y_{i1}, \dots, y_{ip})'$ pour $1 \leq i \leq n$.

Il est immédiat, vu les propriétés de minimisation des splines naturelles (voir le Chapitre 1), que les solutions de (3.1), si elles existent, sont des splines naturelles de degré $2m - 1$ ayant leurs noeuds aux points t_1, \dots, t_p . On montre alors comme au Chapitre 1 (page 25), que le problème variationnel précédent se ramène à la minimisation du critère

$$\frac{1}{p} (\mathbf{g} - \bar{\mathbf{y}})' \mathbf{M} (\mathbf{g} - \bar{\mathbf{y}}) + \lambda \mathbf{g}' \mathbf{\Gamma} \mathbf{g},$$

où $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^n \mathbf{y}_i$ et $\mathbf{\Gamma}$ est la matrice d'énergie liée à la pénalité intégrale de (3.1) (voir le Théorème ??). Comme dans la Proposition 1.1, on en déduit qu'il existe une unique solution au problème initial, notée $\hat{f}_{\lambda, \mathbf{M}}$, et que cette spline de lissage est entièrement caractérisée par le vecteur de ses valeurs en ses noeuds

$$\hat{\mathbf{f}}_{\lambda, \mathbf{M}} := \left(\hat{f}_{\lambda, \mathbf{M}}(t_1), \dots, \hat{f}_{\lambda, \mathbf{M}}(t_p) \right)' = (\mathbf{M} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{M} \bar{\mathbf{y}}. \quad (3.2)$$

Notons que l'inversibilité de $\mathbf{M} + \lambda p \mathbf{\Gamma}$ est garantie par la définie positivité de la métrique \mathbf{M} et par la semi-définie positivité de $\mathbf{\Gamma}$ (voir le Théorème ??).

Nous abordons succinctement le problème de la spécification de la matrice \mathbf{M} . Observons d'abord que dans le cas où \mathbf{M} est la matrice identité, nous retrouvons l'estimateur spline de lissage MCO du Chapitre 1. D'autre part, si les erreurs ε_{ij} sont gaussiennes, alors le choix de la métrique $\mathbf{M} = \mathbf{V}^{-1}$ (estimateur MCG) pour la construction de $\widehat{f}_{\lambda, \mathbf{M}}$ correspond à une méthode de maximum de vraisemblance pénalisée. Notons par ailleurs que *le choix $\mathbf{M} = \mathbf{V}^{-1}$ ne fournit pas en général l'estimateur $\widehat{f}_{\lambda, \mathbf{M}}$ de variance minimale*, ni pour $\lambda > 0$ quelconque, ni pour λ suffisamment petit, contrairement au raisonnement heuristique de Welsh *et al.* (2002). Ce fait est propre aux splines de lissage. Par contraste, dans le cas d'estimateurs de projection satisfaisant à des contraintes linéaires sur le biais, on peut montrer que la variance de l'estimateur est rendue minimale par le choix de la métrique de projection $\mathbf{M} = \mathbf{V}^{-1}$ (voir le Chapitre 7 de Cornillon et Matzner-Løber (2007) pour le modèle linéaire, et Jallet (2007) pour un résultat nouveau sur les splines de régression). Revenant aux splines de lissage, une illustration du fait que le lissage MCG n'améliore pas le lissage MCO en termes d'erreur quadratique MISE sera fournie par l'étude de la Section 3.1.2.

3.1.1 Erreur quadratique moyenne de l'estimateur

Dans cette section, nous majorons l'erreur quadratique moyenne discrétisée (MASE) de l'estimateur $\widehat{f}_{\lambda, \mathbf{M}}$ de f . Ce résultat généralise le Théorème 1.2 du Chapitre 1 relatif à l'estimateur spline de lissage MCO. Les mêmes généralisations peuvent être obtenues pour l'erreur quadratique moyenne intégrée MISE et pour un dispositif déséquilibré par des calculs très semblables à ceux du Chapitre 1.

Nous commençons par rappeler certaines notations utiles et en introduire de nouvelles. On note $\mathbf{S}_{\lambda, \mathbf{M}} := (\mathbf{M} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{M}$ la matrice de lissage associée à l'estimateur $\widehat{f}_{\lambda, \mathbf{M}}$. Cette matrice n'est en général pas symétrique, mais en tant que produit des matrices définies positives $(\mathbf{M} + \lambda p \mathbf{\Gamma})^{-1}$ et \mathbf{M} , elle a des valeurs propres réelles strictement positives. Par ailleurs, on montre sans difficulté que $\|\mathbf{S}_{\lambda, \mathbf{M}}\| = 1$ (voir l'équation (1.15) du Chapitre 1).

On désigne par $h_j = t_{j+1} - t_j$, $0 \leq j \leq p$, les écarts inter-observations (avec $t_0 = 0$ et $t_{p+1} = 1$), et par $h_{max} = \max_{0 \leq j \leq p} h_j$ et $h_{min} = \min_{1 \leq j \leq p-1} h_j$ les écarts maximaux et minimaux du dispositif expérimental.

Concernant les normes vectorielles et matricielles, nous écrivons $\|\mathbf{x}\|^2 = \mathbf{x}'\mathbf{x}$ et $\|\mathbf{x}\|_{\mathbf{M}}^2 = \mathbf{x}'\mathbf{M}\mathbf{x}$ pour tout vecteur $\mathbf{x} \in \mathbb{R}^p$, et $\|\mathbf{M}\| = \sup_{\|\mathbf{x}\|=1} \|\mathbf{x}\|_{\mathbf{M}}$; nous notons aussi $|g|_m^2$ la semi-norme $\int_0^1 (g^{(m)}(t))^2 dt$ de toute fonction $g \in W_2^m[0, 1]$.

Nous décomposons d'abord la MASE en biais au carré plus variance :

$$\begin{aligned} \frac{1}{p} \sum_{j=1}^p \mathbb{E} \left(f(t_j) - \widehat{f}_{\lambda, \mathbf{M}}(t_j) \right)^2 &= \frac{1}{p} \left\| \mathbf{f} - \mathbb{E} \widehat{\mathbf{f}}_{\lambda, \mathbf{M}} \right\|^2 + \frac{1}{p} \mathbb{E} \left\| \widehat{\mathbf{f}}_{\lambda, \mathbf{M}} - \mathbb{E} \widehat{\mathbf{f}}_{\lambda, \mathbf{M}} \right\|^2 \\ &= \frac{1}{p} \left\| (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{M}}) \mathbf{f} \right\|^2 + \frac{1}{np} \operatorname{tr} (\mathbf{S}'_{\lambda, \mathbf{M}} \mathbf{V} \mathbf{S}_{\lambda, \mathbf{M}}). \end{aligned}$$

Pour majorer le biais au carré, nous avons besoin de l'inégalité classique de normes $\|\mathbf{x}\|^2 \leq \|\mathbf{M}^{-1}\| \cdot \|\mathbf{x}\|_{\mathbf{M}}^2$ valable pour tout $\mathbf{x} \in \mathbb{R}^p$, ainsi que du résultat suivant, qui est une adaptation facile du Lemme 4.3 de Craven et Wahba (1979).

Lemme 3.1. *Pour toute fonction $g \in W_2^m[0, 1]$,*

$$\frac{1}{p} \left\| (\mathbf{I} - \mathbf{S}_{\lambda, \mathbf{M}}) \mathbf{g} \right\|_{\mathbf{M}}^2 \leq \lambda |g|_m^2.$$

Concernant le terme de variance, nous utilisons une généralisation de la Proposition 1.2 obtenue par l'étude de l'équation propre généralisée $\mathbf{\Gamma} \mathbf{u} = \lambda \mathbf{M} \mathbf{u}$.

Proposition 3.1. *Il existe une constante $C = C(m, h_{\max}/h_{\min})$ telle que pour tout réel $\lambda \in]0, \|\mathbf{M}\|]$, la matrice de lissage $\mathbf{S}_{\lambda, \mathbf{M}}$ vérifie*

$$\operatorname{tr} (\mathbf{S}'_{\lambda, \mathbf{M}} \mathbf{S}_{\lambda, \mathbf{M}}) \leq C \left(\frac{\|\mathbf{M}\|}{\lambda} \right)^{1/2m}.$$

Démonstration.

Pour toute matrice \mathbf{A} diagonalisable, nous désignons par $\lambda_1(\mathbf{A}) \geq \lambda_2(\mathbf{A}) \geq \dots$ ses valeurs propres ordonnées.

Par un calcul d'algèbre matricielle standard, on montre que les valeurs propres de $\mathbf{S}_{\lambda, \mathbf{M}}$ sont exactement les racines carrées de celles de $\mathbf{S}'_{\lambda, \mathbf{M}} \mathbf{S}_{\lambda, \mathbf{M}}$. Ainsi,

$$\operatorname{tr} (\mathbf{S}'_{\lambda, \mathbf{M}} \mathbf{S}_{\lambda, \mathbf{M}}) = \sum_{j=1}^p \lambda_j^2 (\mathbf{S}_{\lambda, \mathbf{M}}).$$

D'autre part nous avons, pour tout $1 \leq j \leq p$,

$$\lambda_j (\mathbf{S}_{\lambda, \mathbf{M}}) = \lambda_j \left((\mathbf{I} + \lambda p \mathbf{M}^{-1} \mathbf{\Gamma})^{-1} \right) = \left(1 + \lambda p \lambda_{p+1-j} (\mathbf{M}^{-1} \mathbf{\Gamma}) \right)^{-1}.$$

Appliquons ensuite le Corollaire 2 de l'Annexe C (équations propres généralisées) aux matrices $\mathbf{A} = \mathbf{\Gamma}$ et $\mathbf{B} = \mathbf{M}$. Il vient $\lambda_{p+1-j} (\mathbf{M}^{-1} \mathbf{\Gamma}) \geq \lambda_{p-j} (\mathbf{\Gamma}) / \|\mathbf{M}\|$, puis on en déduit que

$$0 < \lambda_j (\mathbf{S}_{\lambda, \mathbf{M}}) \leq \left(1 + \lambda p \frac{\lambda_{p+1-j} (\mathbf{\Gamma})}{\|\mathbf{M}\|} \right)^{-1} \leq 1$$

et que

$$\text{tr } \mathbf{S}'_{\lambda, \mathbf{M}} \mathbf{S}_{\lambda, \mathbf{M}} \leq \sum_{j=1}^p \left(1 + \lambda p \frac{\lambda_j(\mathbf{\Gamma})}{\|\mathbf{M}\|} \right)^{-2}.$$

On termine alors la démonstration comme dans les Propositions 1.2 et 1.3. ■

En combinant les résultats précédents et en suivant à la lettre la démonstration du Théorème 1.2, nous obtenons la généralisation suivante dudit théorème.

Théorème 3.1. *Pour tout réel $\lambda \in]0, \|\mathbf{M}\|$, la MASE de l'estimateur $\widehat{f}_{\lambda, \mathbf{M}}$ vérifie*

$$\frac{1}{p} \mathbb{E} \left\| \mathbf{f} - \widehat{f}_{\lambda, \mathbf{M}} \right\|^2 \leq \lambda \|\mathbf{M}^{-1}\| \cdot |f|_m^2 + \frac{\text{tr}(\mathbf{V})}{np}, \quad (3.3a)$$

et

$$\frac{1}{p} \mathbb{E} \left\| \mathbf{f} - \widehat{f}_{\lambda, \mathbf{M}} \right\|^2 \leq \lambda \|\mathbf{M}^{-1}\| \cdot |f|_m^2 + \frac{C \|\mathbf{V}\|}{np} \left(\frac{\|\mathbf{M}\|}{\lambda} \right)^{1/2m}, \quad (3.3b)$$

où C est la constante de la Proposition 3.1.

Remarque 3.1. *Si on fixe $\|\mathbf{M}\| = \lambda_{max}$ dans (3.3b), la borne supérieure de la MASE diminue avec $\|\mathbf{M}^{-1}\| = \lambda_{min}^{-1}$, de sorte que ce majorant est rendu minimal pour $\lambda_{min} = \lambda_{max}$, i.e. pour le choix $\mathbf{M} \propto \mathbf{I}$ (et non pas $\mathbf{M} = \mathbf{V}^{-1}$).*

3.1.2 Comparaison des lissages splines MCO et MCG

Afin de savoir si l'estimateur spline MCG apporte une amélioration significative par rapport à l'estimateur MCO sur un échantillon fini, le logiciel R a été utilisé pour calculer l'erreur quadratique MISE de ces deux estimateurs dans le modèle (1.4) avec différentes fonctions de régression et différents types d'erreurs, tout en faisant varier les dimensions n et p du jeu de données.

Pour la simplicité des calculs, les points d'observation ont été pris équidistants ($t_j = (2j - 1)/2p$, $1 \leq j \leq p$) et on a utilisé des estimateurs splines cubiques ($m = 2$ dans (3.1)). Deux fonctions de régression ont été envisagées :

$$f(t) = 2t^2 - 0.5t + 1 + \sin(5\pi t), \quad t \in [0, 1], \quad (3.4)$$

et

$$f(t) = \exp(1 - 2t) \cos(30t) + 2 \ln(1 + t), \quad t \in [0, 1], \quad (3.5)$$

ainsi que deux types d'erreurs :

$$(\varepsilon_{1j}, \dots, \varepsilon_{nj}) \stackrel{i.i.d.}{\sim} \sigma W(t_j), \quad 1 \leq j \leq p, \quad (3.6)$$

où W est un processus de Wiener standard ($\mathbb{E}W(t) = 0$ et $\mathbb{E}W(s)W(t) = \min(s, t)$), avec $\sigma = 0.75$, et

$$(\varepsilon_{1j}, \dots, \varepsilon_{nj}) \stackrel{i.i.d.}{\sim} \sigma \varepsilon_j, \quad 1 \leq j \leq p, \quad (3.7)$$

où $(\varepsilon_j)_{j \in \mathbb{N}^*}$ est un processus AR(1) stationnaire gaussien de variance unité et de paramètre $\rho = 0.3$ ou $\rho = 0.9$, avec $\sigma = 0.75$.

La fonction définie par (3.4) est la somme d'une composante polynômiale de degré 2 et d'une composante sinusôidale de fréquence relativement rapide. La fonction définie par (3.5) représente une oscillation amortie à laquelle s'ajoute une composante logarithmique de faible amplitude. Les deux modèles de bruit illustrent pour l'un le processus aléatoire continu, et pour l'autre le processus AR(1) classique exhibant une corrélation à décroissance rapide. Pour ces deux modèles, l'intensité $\sigma = 0.75$ du bruit représente respectivement 25% de l'étendue des variations de la fonction définie par (3.4) et 16% des variations de la fonction définie par (3.5), ce qui représente un rapport signal/bruit modéré pour les deux fonctions (*cf.* Hart et Wehrly, 1986).

Dans un premier temps, nous nous sommes intéressé aux variations des composantes de la MISE (biais au carré et variance intégrés) lorsque le paramètre de lissage λ varie dans l'intervalle $[10^{-9}, 10^{-1}]$, en examinant en même temps l'effet de la discrétisation (nombre p de points d'observation) et de l'échantillonnage (nombre n d'unités statistiques).

La Figure 3.1 présente ces variations dans le cas où la fonction de régression f est (3.4) et où les erreurs sont modélisées à l'aide du processus de Wiener (3.6). Il apparaît de manière prévisible que la variance et le biais au carré sont respectivement des fonctions décroissante et croissante de λ pour les deux estimateurs. La valeur de n n'affecte pas le biais au carré mais elle modifie la variance en termes d'échelle. Elle joue donc sur l'équilibre entre biais et variance qui détermine la MISE optimale. La valeur de p influence peu la forme des courbes de biais de chaque estimateur, mais elle modifie en revanche les positions relatives de ces courbes sur certaines plages de valeurs de λ qui sont cependant éloignées de la région où la MISE est optimale. D'autre part, la MISE exhibe très peu de variations pour $\lambda \leq 10^{-6}$ et à partir de cette valeur elle croît très rapidement pour les deux estimateurs.

Un fait important est que dans la région $[10^{-9}, 10^{-6}]$ où la MISE prend sa valeur minimale, l'écart entre la MISE de l'estimateur MCO et celle de l'estimateur MCG ne semble pas significatif. Ce fait se retrouve dans l'examen de la Figure 3.2 dans laquelle f est définie par (3.4) et les erreurs proviennent du bruit AR(1) (3.7) avec deux niveaux de corrélation : $\rho = 0.3$ et $\rho = 0.9$. En effet, les deux courbes MISE

des estimateurs MCO et MCG sont très proches dans la région critique où la MISE prend sa valeur minimale.

On observe par ailleurs un minimum de la MISE plus marqué dans le cas du processus de bruit discret AR(1) que dans le cas du processus de Wiener. Cela est dû à la nature de ces deux bruits. Pour le processus de Wiener, quand λ devient petit, alors la variance intégrée des deux estimateurs s'approche de 0.28 dans le cas $n = 1$ et de 0.028 dans le cas $n = 10$. Ces valeurs sont à rapprocher de la variance intégrée sur $[0,1]$ du processus de moyenne empirique des erreurs $\bar{\varepsilon} \sim \sigma W/\sqrt{n}$, égale à $(\sigma^2/n) \int_0^1 t dt = (0.75^2/n) \cdot 0.5 = 0.28125/n$. On trouve ainsi une illustration du Théorème 2.2 du Chapitre 2 selon lequel lorsque $p \rightarrow \infty$, la variance de l'estimateur est équivalente à celle du processus de moyenne empirique uniformément en n . D'autre part, le lissage a peu d'influence sur la corrélation persistante du processus de Wiener (la corrélation reste forte entre des observations distantes), alors que dans le cas du processus AR(1), la corrélation s'atténue très rapidement pour des erreurs distantes et pour p grand, ce qui permet de réduire significativement la variance de l'estimateur par un lissage important. Pour cette raison, la courbe de variance des estimateurs décroît relativement vite dans la Figure 3.2 pour $\rho = 0.3$, de façon modérée pour $\rho = 0.9$, et de façon lente dans la Figure 3.1 (sauf pour l'estimateur MCG qui atténue l'effet de la corrélation du processus d'erreur).

Enfin, en comparant les six figures du haut de la Figure 3.2 (corrélation modérée, $\rho = 0.3$) et les six figures du bas (corrélation très forte, $\rho = 0.9$), on voit qu'en termes d'erreur d'estimation, la différence entre les estimateurs MCO et MCG est d'autant plus forte que la structure de corrélation est marquée. Dans le cas limite où les erreurs ne sont pas corrélées (bruit blanc), ces estimateurs sont d'ailleurs confondus.

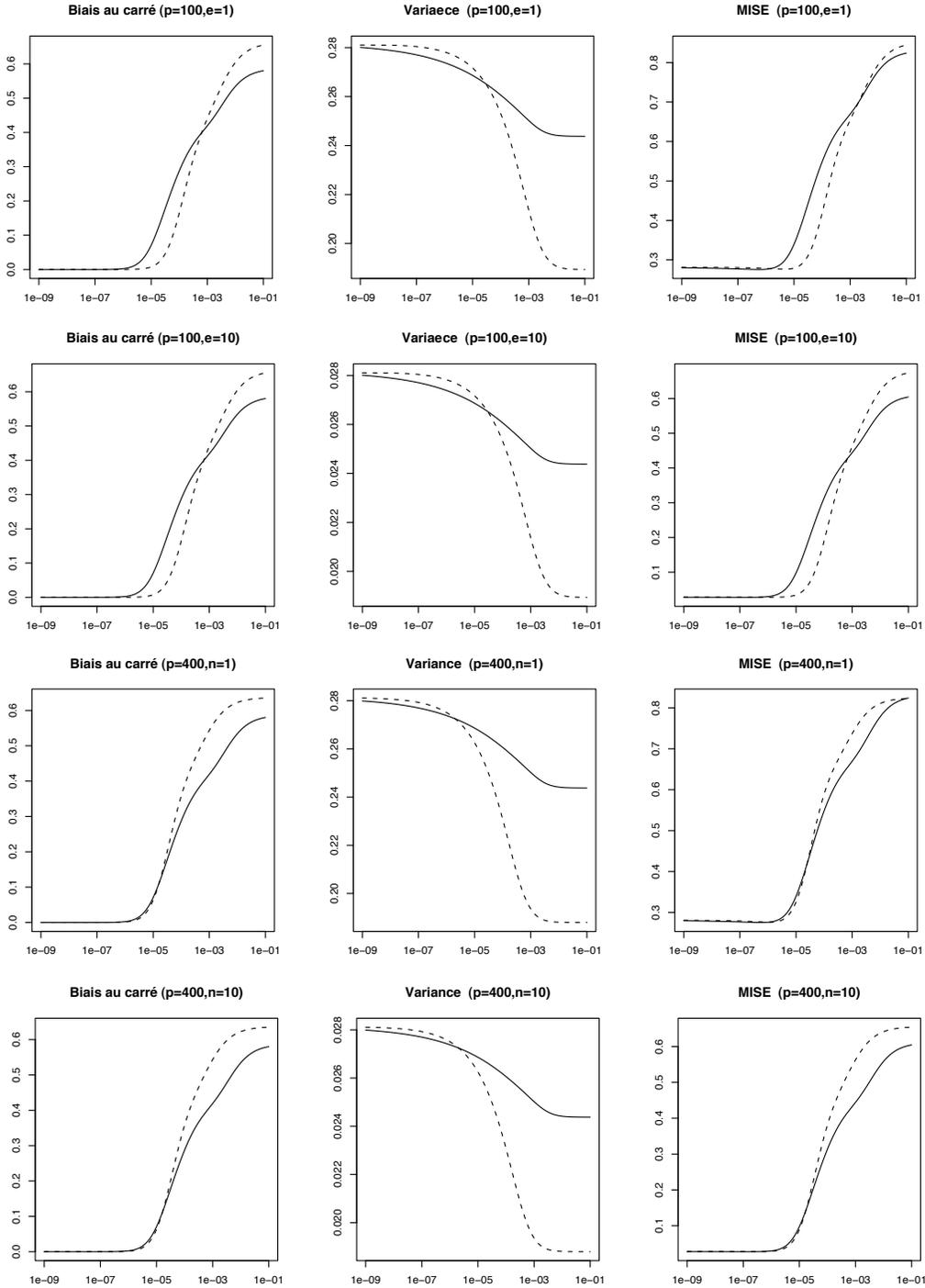


FIG. 3.1 – Biais au carré, variance, et MISE des estimateurs splines cubiques MCO et MCG en fonction du paramètre de lissage. La fonction de régression est $f(t) = 2t^2 - 0.5t + 1 + \sin(5\pi t)$ et le bruit est issu du processus de Wiener σW . Sur chaque graphe, la ligne pleine représente l'estimateur MCO et les tirets représentent l'estimateur MCG.

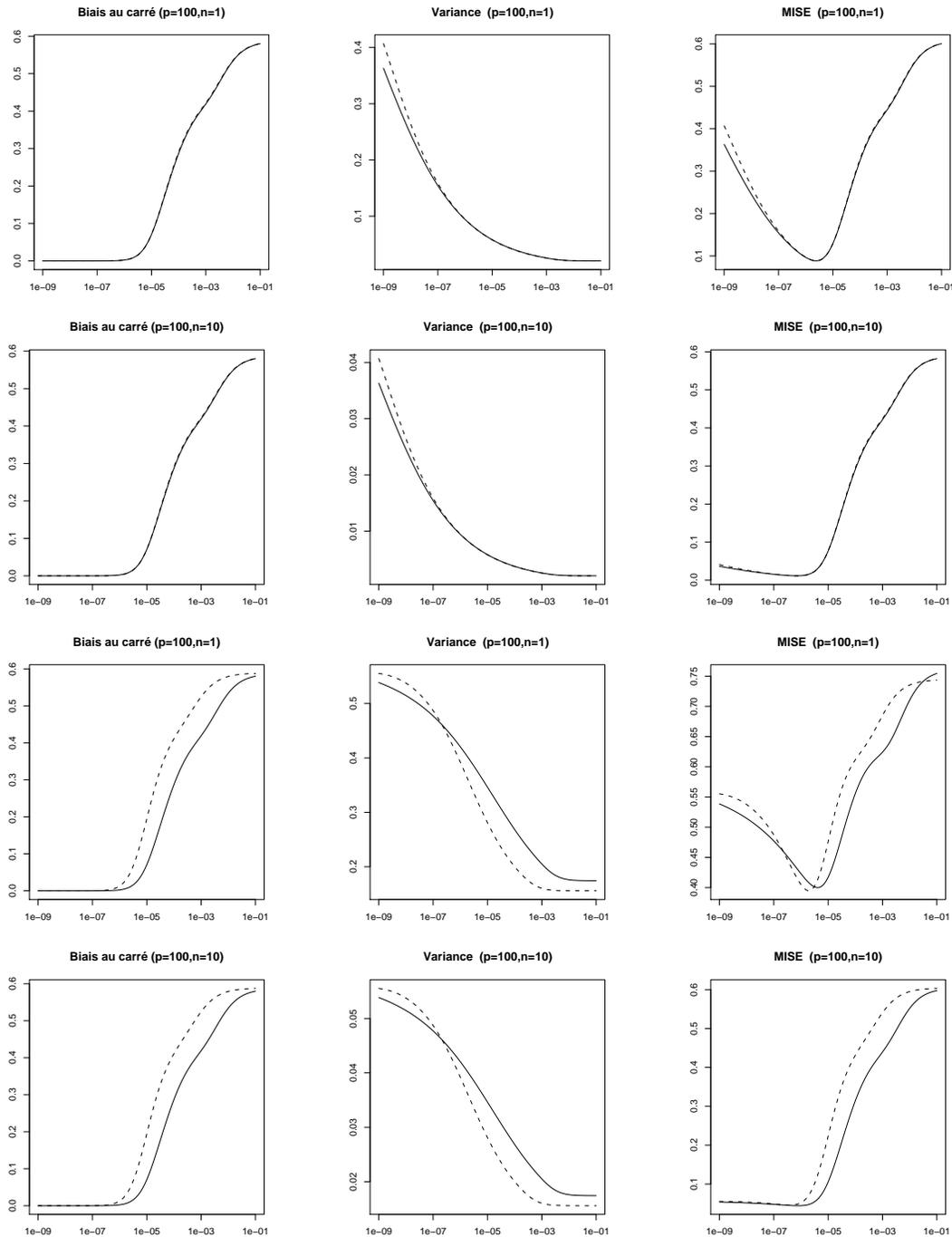


FIG. 3.2 – Biais au carré, variance, et MISE des estimateurs splines cubiques MCO et MCG en fonction du paramètre de lissage. La fonction de régression est $f(t) = 2t^2 - 0.5t + 1 + \sin(5\pi t)$ et le bruit est issu du processus AR(1) (3.7) avec $\rho = 0.3$ pour les six figures du haut et $\rho = 0.9$ pour les six figures du bas. Sur chaque graphe, la ligne pleine représente l'estimateur MCO et les tirets représentent l'estimateur MCG.

Au terme de l'examen des Figures 3.1 et 3.2, il apparaît assez nettement que la méthode de lissage MCG n'apporte pas d'amélioration significative par rapport à la méthode usuelle (MCO) au niveau de la MISE, au moins sur les exemples traités. Afin d'étayer cette affirmation, nous présentons deux tableaux qui comparent les performances optimales de chaque méthode dans plusieurs contextes.

Le Tableau 3.1 présente les valeurs du paramètre de lissage optimal et de la MISE correspondante pour les deux estimateurs splines MCO et MCG en fonction de n et de p . La fonction f considérée est définie par (3.4) et les erreurs sont issues d'un processus de Wiener comme en (3.6). De par la nature de processus aléatoire à covariance continue des erreurs, la MISE des estimateurs ne diminue pour ainsi dire pas quand p augmente pour un n fixé, et elle diminue proportionnellement à n pour un p fixé. Ces faits sont dus à ce que pour p grand, la variance de l'estimateur est équivalente à celle du processus de bruit divisée par n (voir le Théorème 2.2 et aussi Hart et Wehrly (1986)).

n	p	MCO		MCG	
		λ_{opt}	$MISE_{opt}$	λ_{opt}	$MISE_{opt}$
1	10	0	0.2895	0	0.2895
	20	5.59e-07	0.277531	2.05e-05	0.278109
	100	5.65e-07	0.275671	4.47e-06	0.276513
	200	5.65e-07	0.275620	2.24e-06	0.276440
	400	5.65e-07	0.275608	1.12e-06	0.276414
	800	5.65e-07	0.275605	5.60e-07	0.276403
10	10	0	0.036523	0	0.036523
	20	9.92e-08	0.028035	2.63e-06	0.028086
	100	1.34e-07	0.027733	8.99e-07	0.027915
	200	1.34e-07	0.027726	4.46e-07	0.027907
	400	1.34e-07	0.027725	2.23e-07	0.027904
	800	1.34e-07	0.027724	1.11e-07	0.027903

TAB. 3.1 – Paramètre de lissage optimal et MISE optimale des estimateurs splines cubiques MCO et MCG avec la fonction de régression $f(t) = 2t^2 - 0.5t + 1 + \sin(5\pi t)$ et le bruit processus de Wiener.

Pour un n fixé, quand p varie, le paramètre de lissage optimal reste à peu près constant pour l'estimateur MCO alors qu'il est beaucoup plus variable pour l'estimateur MCG. Pour la (faible) valeur $p = 10$ du nombre de points d'observation, le biais au carré des estimateurs domine leur variance, ce qui contraint à l'interpolation pour obtenir la MISE optimale. Enfin, il apparaît nettement que les deux estimateurs ont des performances optimales comparables pour toutes les valeurs de n et p , avec un très léger avantage pour l'estimateur MCO.

Dans le Tableau 3.2, l'étude précédente est renouvelée avec la fonction de régression f de (3.5) et le bruit AR(1) de (3.7). Les valeurs de la MISE sont affichées avec une précision de 10^{-4} afin de permettre une visualisation claire de ses variations, qui sont beaucoup plus marquées ici que dans le cas du bruit processus (3.6).

Concernant les performances respectives des estimateurs MCO et MCG, on constate comme auparavant des valeurs très comparables de chaque erreur MISE optimale, avec un très léger avantage pour l'estimateur MCO qui n'excède pas 5%.

ρ	n	p	MCO		MCG		
			λ_{opt}	$MISE_{opt}$	λ_{opt}	$MISE_{opt}$	
0.3	1	10	0	1.1653	0	1.1653	
		20	3.83e-07	0.4506	8.67e-07	0.4549	
		100	2.23e-07	0.1493	2.33e-07	0.1510	
		200	1.55e-07	0.0844	1.52e-07	0.0848	
		400	1.06e-07	0.0470	1.02e-07	0.0469	
		800	7.20e-08	0.0259	6.90e-08	0.0259	
	10	10	0	0.6667	0	0.6667	
		20	7.80e-08	0.0600	1.50e-07	0.0608	
		100	5.62e-08	0.01995	6.19e-08	0.0202	
		200	4.10e-08	0.0115	4.14e-08	0.0115	
		400	2.83e-08	0.0064	2.78e-08	0.0064	
		800	1.91e-08	0.0039	1.86e-08	0.0036	
	0.9	1	10	0	1.1734	0	1.1734
			20	8.68e-08	0.5657	1.34e-06	0.5665
100			1.92e-07	0.4785	2.31e-07	0.4820	
200			2.65e-07	0.3932	9.72e-08	0.4000	
400			2.85e-07	0.2804	4.67e-08	0.2870	
800			2.38e-07	0.1762	2.65e-08	0.1787	
10		10	0	0.6675	0	0.6675	
		20	1.28e-08	0.0656	1.11e-07	0.0656	
		100	3.28e-08	0.0509	4.33e-08	0.0523	
		200	4.93e-08	0.0444	1.98e-08	0.0466	
		400	6.02e-08	0.0343	1.09e-08	0.0361	
		800	5.73e-08	0.0229	6.96e-09	0.0234	

TAB. 3.2 – Paramètre de lissage optimal et MISE optimale des estimateurs splines cubiques MCO et MCG avec la fonction de régression $f(t) = \exp(1 - 2t) \cos(30t) + 2 \ln(1 + t)$ et le bruit AR(1).

Un autre fait intéressant est la vitesse de décroissance de la MISE. Pour $\rho = 3$ et pour chaque valeur de n , on peut observer dans le tableau que la MISE optimale des estimateurs décroît proportionnellement à $p^{-4/5}$, à savoir la vitesse optimale de convergence dans la régression non paramétrique d'une fonction de classe C^2 avec des erreurs indépendantes (Stone, 1982). Il est normal que cette vitesse soit observée car l'estimation est effectuée par des splines cubiques, ce qui revient à supposer que f est de classe C^2 , même si cette fonction est en réalité de classe C^∞ . Par ailleurs, l'observation d'une telle convergence est en accord avec la Proposition 1.4 et avec les simulations de Wang (1998) qui montrent qu'en présence d'un bruit AR(1), la vitesse optimale du cas i.i.d. est préservée. Dans le cas $\rho = 0.9$ où la corrélation est très forte, le même ordre de décroissance devrait survenir pour de très grandes valeurs de p (supérieures à celles qui figurent dans le tableau).

En conclusion, l'étude effectuée dans cette section, qui couvre plusieurs types de fonction de régression, de bruit, et plusieurs tailles de dispositif expérimental, tend à montrer que l'estimateur spline MCG ne donne pas de meilleures performances que l'estimateur MCO en termes d'erreur quadratique MISE. Autrement dit, l'incorporation de la structure de covariance des données dans la métrique du lissage spline n'améliore pas l'estimation et ce d'autant plus qu'en pratique, la vraie matrice de covariance \mathbf{V} utilisée dans ces exemples est rarement connue et doit donc être estimée, ce qui risque d'augmenter encore la MISE. Le même constat a été fait par Jallet (2007) concernant les splines de régression. D'autre part les méthodes usuelles de sélection du paramètre de lissage développées dans le cadre d'erreurs i.i.d. sont peu efficaces en présence de données corrélées (voir entre autres Altman, 1990; Rice et Silverman, 1991; Wang, 1998; Opsomer *et al.*, 2001). Nous sommes donc amenés à penser que la tâche essentielle dans la construction d'un estimateur consiste à prendre en compte la structure de covariance des données de manière adéquate à travers le choix du paramètre de lissage.

3.2 Sélection du paramètre de lissage

3.2.1 Extension des méthodes usuelles

Wang (1998) a étendu trois méthodes usuelles de sélection du paramètre de lissage d'un estimateur spline en passant du cas d'observations indépendantes au cas d'observations corrélées.

Ces extensions sont réalisées dans le modèle (1.4) avec $n = 1$ (observations non répliquées), et sous l'hypothèse que la matrice de covariance \mathbf{V} s'écrit sous la forme $\sigma^2 \mathbf{W}^{-1}$, où \mathbf{W} est une matrice explicite dépendant d'un vecteur $\boldsymbol{\tau}$ de paramètres inconnus. La fonction de régression f est estimée par la spline de lissage $\hat{f}_{\lambda, \boldsymbol{\tau}}$ solution

de (3.1) avec la métrique $\mathbf{M} = \mathbf{W}$. Ainsi, $\widehat{f}_{\lambda, \tau}$ est l'unique solution de

$$\min_{g \in W_2^m[0,1]} \frac{1}{p} (\mathbf{g} - \mathbf{y})' \mathbf{W} (\mathbf{g} - \mathbf{y}) + \lambda \int_0^1 (g^{(m)}(t))^2 dt,$$

où $\mathbf{y} = (y_{11}, \dots, y_{1p})'$ et $\mathbf{g} = (g(t_1), \dots, g(t_p))'$.

Le formalisme des espaces de Hilbert à noyau autoreproduisant est bien adapté pour expliciter l'estimateur $\widehat{f}_{\lambda, \tau}$. Soit $\phi_\nu(t) = t^{\nu-1}/(\nu-1)!$ pour $\nu = 1, \dots, m$. Soit $R_1(s, t) = \int_0^1 (s-u)_+^{m-1} (t-u)_+^{m-1} du / ((m-1)!)^2$, avec la notation $(t)_+ = \max(t, 0)$. Soient les matrices $\mathbf{T} = (\phi_\nu(t_j))_{1 \leq j \leq p, 1 \leq \nu \leq m}$ de taille $p \times m$ et $\mathbf{\Sigma} = (R_1(t_i, t_j))_{1 \leq i, j \leq p}$ de taille $p \times p$. D'après Craven et Wahba (1979), $\widehat{f}_{\lambda, \tau}$ s'écrit sous la forme

$$\widehat{f}_{\lambda, \tau}(t) = \sum_{\nu=1}^m d_\nu \phi_\nu(t) + \sum_{j=1}^p c_j R_1(t, t_j), \quad (3.8)$$

où $\mathbf{c} = (c_1, \dots, c_p)'$ et $\mathbf{d} = (d_1, \dots, d_m)'$ sont les solutions du système linéaire

$$\begin{pmatrix} \mathbf{T}'\mathbf{W}\mathbf{T} & \mathbf{T}'\mathbf{W} \\ \mathbf{\Sigma}\mathbf{W}\mathbf{T} & \mathbf{\Sigma}\mathbf{W} + \lambda p \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{d} \\ \mathbf{\Sigma}\mathbf{c} \end{pmatrix} = \begin{pmatrix} \mathbf{T}'\mathbf{W}\mathbf{y} \\ \mathbf{\Sigma}\mathbf{W}\mathbf{y} \end{pmatrix}.$$

Par ailleurs, en notant

$$\mathbf{T} = (\mathbf{Q}_1 \quad \mathbf{Q}_2) \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix}$$

la décomposition QR de la matrice \mathbf{T} (la matrice $(\mathbf{Q}_1 \quad \mathbf{Q}_2)$ est orthogonale avec \mathbf{Q}_1 de taille $p \times (p-m)$ et \mathbf{Q}_2 de taille $p \times m$, et \mathbf{R} est triangulaire supérieure), la matrice de lissage de $\widehat{f}_{\lambda, \tau}$, définie formellement par $\mathbf{S}_{\lambda, \tau} = (\mathbf{W} + \lambda p \mathbf{\Gamma})^{-1} \mathbf{W}$, s'explique en

$$\mathbf{S}_{\lambda, \tau} = \mathbf{I} - \lambda p \mathbf{W}^{-1} \mathbf{Q}_2 (\mathbf{Q}_2' (\mathbf{\Sigma} + \lambda p \mathbf{W}^{-1}) \mathbf{Q}_2)^{-1} \mathbf{Q}_2'. \quad (3.9)$$

Le paramètre de lissage λ et le vecteur $\boldsymbol{\tau}$ de paramètres de corrélation sont alors estimés simultanément à l'aide de l'une des trois méthodes suivantes.

La méthode du maximum de vraisemblance généralisée (en anglais : Generalized Maximum Likelihood, abrégé en GML) est développée dans un cadre bayésien qui, en plus de l'estimation ponctuelle, permet de construire des intervalles de confiance. La distribution *a priori* suivante est spécifiée pour f :

$$F(t) = \sum_{\nu=1}^m \theta_\nu \phi_\nu(t) + b^{1/2} U(t), \quad t \in [0, 1],$$

où $\boldsymbol{\theta} = (\theta_1, \dots, \theta_m)'$ suit une loi normale $N(0, a\mathbf{I})$, avec a et b deux constantes positives, et $\{U(t); t \in [0, 1]\}$ est un processus gaussien centré, indépendant de $\boldsymbol{\theta}$,

et de covariance $\mathbb{E}U(s)U(t) = R_1(s, t)$. Les données du modèle (1.4) s'écrivent alors $y_j = F(t_j) + \varepsilon_j, 1 \leq j \leq p$, avec $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_p)' \sim N(0, \sigma^2 \mathbf{W}^{-1})$ indépendant de F .

Avec les mêmes arguments que Wahba (1990), en posant $\lambda = \sigma^2/pb$, on a

$$\lim_{a \rightarrow \infty} \mathbb{E}(F(t)|\mathbf{y}) = \hat{f}_{\lambda, \boldsymbol{\tau}}(t), \quad \text{et} \quad \lim_{a \rightarrow \infty} \text{Cov}(\mathbf{F}|\mathbf{y}) = \sigma^2 \mathbf{S}_{\lambda, \boldsymbol{\tau}} \mathbf{W}^{-1},$$

avec $\mathbf{F} = (F(t_1), \dots, F(t_p))'$. Dans les formules précédentes, la distribution *a priori* de la partie paramétrique de F (polynômes de degré inférieur à m) est rendue "impropre" par le fait que $a \rightarrow \infty$.

Les estimateurs $\hat{\lambda}$ et $\hat{\boldsymbol{\tau}}$ qui maximisent la vraisemblance a posteriori de $\lambda, \boldsymbol{\tau}$, et b sachant $\mathbf{Q}'_2 \mathbf{y}$ sont obtenus par la minimisation du score

$$M(\lambda, \boldsymbol{\tau}) = \frac{\mathbf{y}' \mathbf{W} (\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}}) \mathbf{y}}{[\det^+ (\mathbf{W} (\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}}))]^{1/(p-m)}}, \quad (3.10)$$

où \det^+ est le produit des valeurs propres non nulles d'une matrice.

La variance σ^2 du bruit peut être estimée par

$$\hat{\sigma}^2 = \frac{\mathbf{y}' \mathbf{W} (\mathbf{I} - \mathbf{S}_{\hat{\lambda}, \hat{\boldsymbol{\tau}}}) \mathbf{y}}{p - m}. \quad (3.11)$$

Remarque 3.2. *Zhang et al. (1998) et Lin et Zhang (1999) ont observé, dans le cadre de modèles semi-paramétriques, que la méthode d'estimation des paramètres de lissage de l'estimateur et des paramètres de corrélation du bruit par maximum de vraisemblance restreinte (REML) se ramène à la méthode GML dans le cas particulier du modèle de Wang (1998).*

La méthode d'estimation sans biais du risque (en anglais : Unbiased Risk Estimation, abrégé en UBR) consiste à minimiser le score

$$U_k(\lambda, \boldsymbol{\tau}) = \frac{1}{p} \mathbf{y}' (\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}}) \mathbf{W}^k (\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}}) \mathbf{y} - \frac{\sigma^2}{p} \text{tr} \mathbf{W}^{k-1} + \frac{2\sigma^2}{p} \text{tr} (\mathbf{W}^{k-1} \mathbf{S}_{\lambda, \boldsymbol{\tau}})$$

pour $k = 0, 1$, ou 2 . Comme son nom l'indique, le score U_k est un estimateur sans biais de l'erreur quadratique moyenne pondérée $\mathbb{E} \left[p^{-1} (\hat{\mathbf{f}}_{\lambda, \boldsymbol{\tau}} - \mathbf{f})' \mathbf{W}^k (\hat{\mathbf{f}}_{\lambda, \boldsymbol{\tau}} - \mathbf{f}) \right]$, avec $\hat{\mathbf{f}}_{\lambda, \boldsymbol{\tau}} = (\hat{f}_{\lambda, \boldsymbol{\tau}}(t_1), \dots, \hat{f}_{\lambda, \boldsymbol{\tau}}(t_p))' = \mathbf{S}_{\lambda, \boldsymbol{\tau}} \mathbf{y}$ et $\mathbf{f} = (f(t_1), \dots, f(t_p))'$. Le calcul de U_k nécessite la connaissance de σ^2 ou son estimation, à l'aide par exemple de (3.11). Par ailleurs, les simulations de Wang (1998) indiquent que le score U_k fournissant la meilleure estimation de f est U_2 .

La méthode de validation croisée généralisée (en anglais Generalized Cross-Validation, abrégé en GCV) est bien connue pour ses propriétés d'optimalité dans le cas de données i.i.d. (Wahba, 1990). Dans le cas i.i.d., la construction d'intervalles de confiance bayésiens basés sur l'estimateur spline de la méthode GCV a été discutée par Wahba (1983) et Nychka (1988, 1990). Par ailleurs, les simulations de Wahba (1985) comparant les méthodes GCV et GML indiquent une supériorité de la méthode GCV pour un échantillon de taille modérée ou grande.

L'extension au cadre corrélé de la méthode GCV proposée par Wang (1998) consiste à prendre pour estimateurs de λ et τ les valeurs $\hat{\lambda}$ et $\hat{\tau}$ qui minimisent la fonction

$$V_k(\lambda, \tau) = \frac{n^{-1} \|\mathbf{W}^{k/2}((\mathbf{I} - \mathbf{S}_{\lambda, \tau})\mathbf{y})\|^2}{[n^{-1} \text{tr}(\mathbf{W}^{k-1}(\mathbf{I} - \mathbf{S}_{\lambda, \tau}))]^2} \quad (3.12)$$

pour $k = 0, 1$, ou 2 . Les simulations de Wang (1998) incitent à utiliser le score V_2 plutôt que les deux autres.

Diggle et Hutchinson (1989) ont proposé deux autres adaptations de la méthode GCV dans un cas particulier du cadre de Wang (1998) (pour des erreurs issues d'un processus d'Ornstein-Uhlenbeck de covariance $R(x, y) = \sigma^2 \exp(-\rho|x - y|)$, avec σ^2 et ρ inconnus). La plus efficace de leur deux méthodes est celle qui consiste à minimiser le score

$$L(\lambda, \rho) = p \ln \left[\mathbf{y}'(\mathbf{I} - \mathbf{S}_{\lambda, \tau})' \mathbf{W}(\mathbf{I} - \mathbf{S}_{\lambda, \tau}) \mathbf{y} \right] - \ln |\mathbf{W}| + (\ln p) \text{tr} \mathbf{S}_{\lambda, \tau} . \quad (3.13)$$

Pour les petits échantillons ($p = 50$), cette méthode donne de moins bonnes performances que les trois autres proposées par Wang (1998), avec notamment une tendance à l'interpolation des données. Cependant, elle est compétitive par rapport aux autres pour les échantillons plus larges ($p \geq 100$).

3.2.2 La méthode GCV de Rice et Silverman (1991)

Dans un contexte d'analyse en composantes principales fonctionnelles, Rice et Silverman (1991) utilisent une méthode de validation croisée généralisée pour la sélection du paramètre de lissage d'un estimateur spline de f . Ils travaillent dans le cadre du modèle (1.4) avec $n > 1$, en supposant que les erreurs ε_{ij} , $1 \leq i \leq n$, $1 \leq j \leq p$, proviennent d'un processus aléatoire $\{\varepsilon(x); x \in [0, 1]\}$ de moyenne nulle et de covariance continue. Ils considèrent un estimateur spline de lissage \hat{f}_λ solution de (3.1) avec $\mathbf{M} = \mathbf{I}$ et $m = 2$, et construisent une fonction de score

$$S(\lambda) = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^p \left(\hat{f}_\lambda^{-i}(t_j) - y_{ij} \right)^2 \quad (3.14)$$

où \widehat{f}_λ^{-i} est l'estimateur spline de lissage solution de

$$\min_{g \in W_2^m[0,1]} \frac{1}{p} \sum_{j=1}^p \left(g(t_j) - (n-1)^{-1} \sum_{1 \leq k \neq i \leq n} y_{kj} \right)^2 + \lambda \int_0^1 (g''(x))^2 dx.$$

Cette fonction de score représente le pouvoir prédictif d'un estimateur spline usuel basé sur $(n-1)$ courbes. Définissant la quantité $M_n(\lambda)$ comme la somme des erreurs quadratiques $\mathbb{E} \sum_{j=1}^p (\widehat{f}_\lambda(t_j) - f(t_j))^2$ de l'estimateur \widehat{f}_λ (basé sur n courbes), Rice et Silverman (1991) montrent ensuite que

$$n^{-1} \mathbb{E} S(\lambda) = \text{tr}(\mathbf{V}) + M_{n-1}(\lambda). \quad (3.15)$$

Ainsi, la valeur de λ qui minimise $n^{-1} \mathbb{E} S(\lambda)$ est aussi celle qui minimise $M_{n-1}(\lambda)$, et ce quelle que soit la fonction de covariance $R(s, t) = \mathbb{E} \varepsilon(s) \varepsilon(t)$. Pourvu que n ne soit pas trop petit, la minimisation du score $S(\lambda)$ (en anglais : "GCV leave one curve out") devrait donc fournir une bonne valeur de λ , du point de vue de l'erreur quadratique moyenne discrétisée (MASE, dans notre terminologie).

Il existe deux différences majeures entre la méthode de Rice et Silverman (1991) et les méthodes de Wang (1998) et Diggle et Hutchinson (1989). D'une part, la première méthode n'est valide que dans le contexte de mesures répétées ($n > 1$), tandis que les secondes s'accommodent de l'observation d'une seule unité statistique. D'autre part, la première méthode fait l'économie de l'estimation de la structure de covariance \mathbf{V} des données, alors que dans les deux autres l'estimation paramétrique de cette matrice apparaît comme intéressante en elle-même.

3.2.3 Une nouvelle méthode pour les splines cubiques

Dans le cas où l'estimation de la régression f du modèle (1.4) est réalisée par une spline cubique de lissage de paramètre λ variant dans $[0, 1]$, il est possible de construire une fonction de score basée sur les données dont le minimum survient à peu près au même point λ qui réalise le minimum de la MISE de l'estimateur. La construction d'un tel score découle en fait du Théorème 1.7 (voir page 30) qui permet de calculer la norme L_2 d'une spline cubique naturelle à l'aide de ses valeurs en ses noeuds. Le détail de cette construction et la méthode de sélection du paramètre λ qui en résulte, intitulée temporairement ici SPCS, sont présentées ci-après.

Pour tout $\lambda \in [0, 1]$, notons \widehat{f}_λ la spline cubique de lissage solution de (3.1) pour $\mathbf{M} = \mathbf{I}$ (\widehat{f}_λ est un estimateur spline MCO pour $\lambda > 0$, et \widehat{f}_0 est la spline cubique naturelle d'interpolation associée à $\bar{y}_1, \dots, \bar{y}_p$). En utilisant la décomposition en biais

au carré plus variance de la MISE de \widehat{f}_λ et le Corollaire 1.1 (voir page 30), nous avons

$$\begin{aligned} \text{MISE}(\lambda) &= \left| f - \mathbb{E}\widehat{f}_\lambda \right|_0^2 + \mathbb{E} \left| \widehat{f}_\lambda - \mathbb{E}\widehat{f}_\lambda \right|_0^2 \\ &= \left| f - \mathbb{E}\widehat{f}_\lambda \right|_0^2 + \frac{1}{n} \text{tr}(\mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{S}_\lambda \mathbf{V}) \end{aligned} \quad (3.16)$$

(on rappelle que $|\cdot|_0$ et $\langle \cdot, \cdot \rangle_0$ désignent la norme et le produit scalaire de $L_2[0, 1]$).

Pour définir une fonction de score $S^*(\lambda)$ dont l'argument minimum est à peu près celui de la MISE, nous remplaçons dans (3.16) les quantités inconnues f et \mathbf{V} par \widehat{f}_0 et par la matrice $\widehat{\mathbf{V}} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})'$, respectivement. Remarquons qu'il est possible et parfois avantageux d'utiliser une autre estimation de \mathbf{V} que la covariance empirique, notamment si n est petit ou si \mathbf{V} a une forme paramétrique. En faisant appel au Corollaire 1.1, il vient

$$\begin{aligned} \mathbb{E} \left| \widehat{f}_0 - \widehat{f}_\lambda \right|_0^2 &= \left| \mathbb{E}(\widehat{f}_0 - \widehat{f}_\lambda) \right|_0^2 + \mathbb{E} \left| \widehat{f}_0 - \widehat{f}_\lambda - \mathbb{E}(\widehat{f}_0 - \widehat{f}_\lambda) \right|_0^2 \\ &= \left| \mathbb{E}\widehat{f}_0 - f \right|_0^2 + \left| f - \mathbb{E}\widehat{f}_\lambda \right|_0^2 + 2 \left\langle \mathbb{E}\widehat{f}_0 - f, f - \mathbb{E}\widehat{f}_\lambda \right\rangle_0 + \frac{1}{n} \text{tr}((\mathbf{I} - \mathbf{S}_\lambda) \boldsymbol{\Phi} (\mathbf{I} - \mathbf{S}_\lambda) \mathbf{V}) \\ &= \text{MISE}(\lambda) - \frac{1}{n} \text{tr}(\mathbf{S}_\lambda \boldsymbol{\Phi} \mathbf{V}) - \frac{1}{n} \text{tr}(\boldsymbol{\Phi} \mathbf{S}_\lambda \mathbf{V}) - 2 \left\langle f - \mathbb{E}\widehat{f}_0, f - \mathbb{E}\widehat{f}_\lambda \right\rangle_0 + cste, \end{aligned}$$

où le terme *cste* désigne une quantité indépendante de λ . Ceci nous amène à définir la fonction de score

$$S^*(\lambda) := \left| \widehat{f}_0 - \widehat{f}_\lambda \right|_0^2 + \frac{1}{n} \text{tr}(\mathbf{S}_\lambda \boldsymbol{\Phi} \widehat{\mathbf{V}}) + \frac{1}{n} \text{tr}(\boldsymbol{\Phi} \mathbf{S}_\lambda \widehat{\mathbf{V}}). \quad (3.17)$$

En prenant l'espérance de $S^*(\lambda)$, on obtient

$$\mathbb{E}S^*(\lambda) = \text{MISE}(\lambda) - 2 \left\langle f - \mathbb{E}\widehat{f}_0, f - \mathbb{E}\widehat{f}_\lambda \right\rangle_0 + cste. \quad (3.18)$$

Nous donnons ici quelques arguments qualitatifs indiquant pourquoi le produit scalaire de (3.18) n'influence que très peu la minimisation de $\mathbb{E}S^*(\lambda)$ (ce fait sera également illustré en pratique par les bonnes performances de cette méthode dans les simulations de la Section 3.3). D'une part, nous avons observé sur de nombreux exemples de fonctions f que la courbe de la fonction $\lambda \mapsto 2 \langle f - \mathbb{E}\widehat{f}_0, f - \mathbb{E}\widehat{f}_\lambda \rangle_0$ a la même allure que celle du biais au carré $\lambda \mapsto |f - \mathbb{E}\widehat{f}_\lambda|_0^2$, mais à une échelle plus petite. En particulier, ces courbes ont leur minimum presque au même point. Cela implique que l'équilibre entre le biais au carré et la variance pour la MISE, et celui entre la quantité $|f - \mathbb{E}\widehat{f}_\lambda|_0^2 - 2 \langle f - \mathbb{E}\widehat{f}_0, f - \mathbb{E}\widehat{f}_\lambda \rangle_0$ et la variance pour $\mathbb{E}S^*(\lambda)$, est réalisé à peu près pour la même valeur de λ . D'autre part, pour une fonction f ayant peu d'oscillations rapides et pour p assez grand, l'interpolation fournit un biais nettement

inférieur à celui du lissage, c'est-à-dire que $|f - \mathbb{E}\hat{f}_0|_0 \ll |f - \mathbb{E}\hat{f}_\lambda|_0$ même pour λ très petit de l'ordre de 10^{-7} ou 10^{-8} . Ensuite il vient $|\langle f - \mathbb{E}\hat{f}_0, f - \mathbb{E}\hat{f}_\lambda \rangle_0| \ll |f - \mathbb{E}\hat{f}_\lambda|_0^2$ avec l'inégalité de Cauchy-Schwartz. Par ailleurs, les valeurs minimales de $\mathbb{E}S^*(\lambda)$ et de la MISE ne sont pas obtenues pour λ plus petit que disons 10^{-8} , car sur cette plage de valeurs les quantités $\langle f - \mathbb{E}\hat{f}_0, f - \mathbb{E}\hat{f}_\lambda \rangle_0$ et $|f - \mathbb{E}\hat{f}_\lambda|_0^2$ sont négligeables devant la variance $\mathbb{E}|\hat{f}_\lambda - \mathbb{E}\hat{f}_\lambda|_0^2$ qui prend sa valeur maximale. Ainsi, la MISE de \hat{f}_λ et la quantité $\mathbb{E}S^*(\lambda)$ sont à peu près égales sur la plage de valeurs de λ où elles réalisent leur minimum, ce qui justifie la méthode de sélection du paramètre de lissage par minimisation du score S^* .

Remarque 3.3. À l'aide du Théorème 1.7 et de (3.17), on montre sans difficulté que $S^*(\lambda) = \bar{\mathbf{y}}' (\mathbf{I} - \mathbf{S}_\lambda) \Phi (\mathbf{I} - \mathbf{S}_\lambda) \bar{\mathbf{y}} + n^{-1} \text{tr} \left((\mathbf{S}_\lambda \Phi + \Phi \mathbf{S}_\lambda) \hat{\mathbf{V}} \right)$. Ceci rend la recherche du minimum de S^* relativement rapide, une fois effectuée la décomposition en éléments propres de la matrice d'énergie \mathbf{K} qui vérifie $\mathbf{S}_\lambda = (\mathbf{I} + \lambda p \mathbf{K})^{-1}$ (voir page 48).

3.3 Comparaison des méthodes précédentes

Deux des méthodes de la section précédente ont été comparées lors de simulations du modèle de mesures répétées (1.4). Ces méthodes sont la validation croisée de Rice et Silverman (1991) [GCV] et la minimisation approchée de la MISE de la Section 3.2.3 [SPSC]. Elles ont été choisies car elles partagent deux caractéristiques : d'une part, elles sont spécifiquement développées pour le traitement de mesures répétées (la condition $n > 1$ est nécessaire pour le calcul du score GCV et aussi pour le calcul de la matrice de covariance empirique utilisée dans la méthode SPSC). D'autre part, elles ne font pas d'hypothèse paramétrique sur le bruit, ce qui en l'absence d'information a priori évite le problème de la spécification d'un modèle. Une autre caractéristique commune est que les deux méthodes visent toutes deux à la minimisation d'une erreur quadratique (cf. (3.15) et (3.18)).

Remarque 3.4. Une caractéristique intéressante de la méthode SPSC tient en ce qu'elle permet soit d'estimer la covariance théorique \mathbf{V} du modèle (1.4) à l'aide de la covariance empirique, soit d'intégrer l'information disponible sur \mathbf{V} afin d'estimer la MISE de l'estimateur puis la fonction de régression elle-même plus efficacement.

Les simulations du modèle (1.4) ont été réalisées avec la fonction de régression

$$f(t) = 10t^3 - 15t^4 + 6t^5, \quad t \in [0, 1], \quad (3.19)$$

utilisée par Hart et Wehrly (1986), et des erreurs $\varepsilon_{ij} = \varepsilon_i(t_j)$, $1 \leq i \leq n$, $1 \leq j \leq p$, issues d'un échantillon i.i.d. $(\varepsilon_1, \dots, \varepsilon_n)$ d'un processus de Wiener σW tel que

$\sigma = 0.3$, $\mathbb{E}W(t) = 0$, et $\text{Cov}(W(s), W(t)) = \min(s, t)$. Pour chacune des différentes tailles (n, p) de dispositif étudiées, la simulation a été répétée 100 fois avec le logiciel R. Des ensembles typiques de trajectoires ainsi que les trajectoires moyennes correspondantes sont présentés dans la Figure 3.3. On constate qu'à cause de la corrélation très sévère de σW , qui est caractéristique des processus aléatoires à covariance continue, il peut arriver que la trajectoire moyenne soit systématiquement en dessous (ou au dessus) de la fonction f , ce qui rend une bonne estimation impossible.

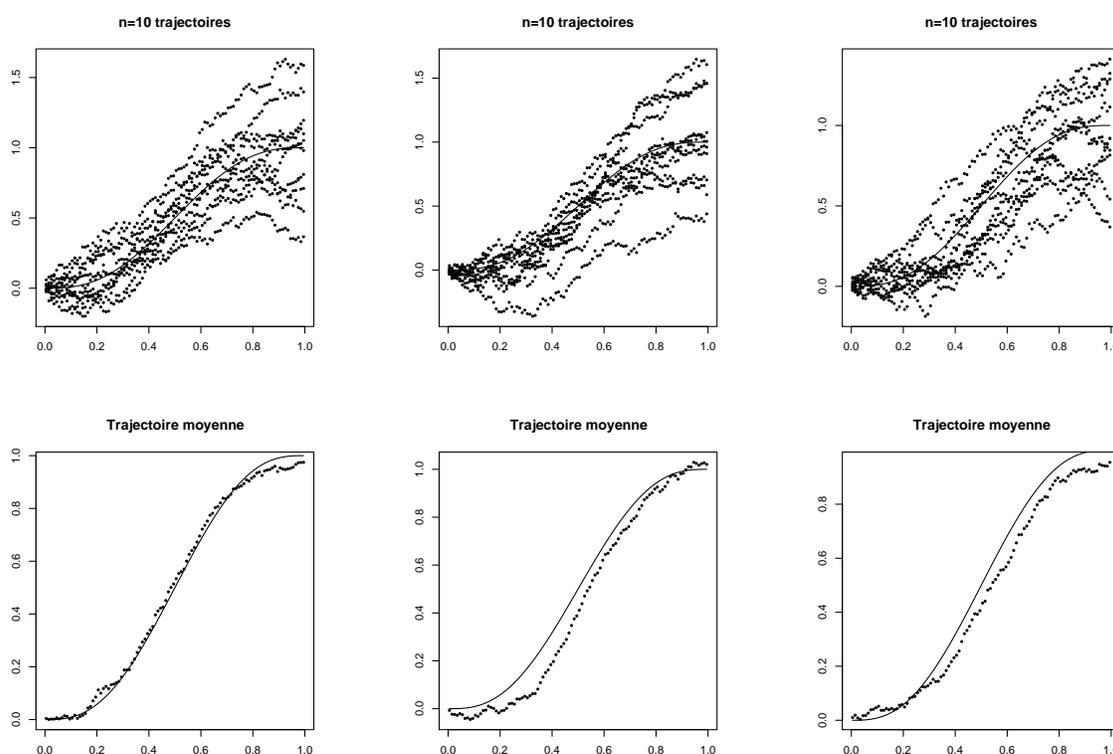


FIG. 3.3 – Exemples de trajectoires de $f + \sigma W$ observées en $p = 100$ points, avec $f(t) = 10t^3 - 15t^4 + 6t^5$, $\sigma = 0.3$, et W un processus de Wiener standard. En dessous de chaque figure du haut apparaît la trajectoire moyenne correspondante.

Avant de comparer les performances d'estimation de chacune des méthodes GCV et SPSC, nous observons dans la Figure 3.4 la ressemblance de leurs fonctions de scores, qui ont aussi la même allure que la courbe théorique de la MISE. Ceci n'est pas surprenant étant donné que ces scores sont construits de manière à reproduire approximativement les variations de la MISE, et en particulier à prendre leur valeur minimale au même point que la MISE.

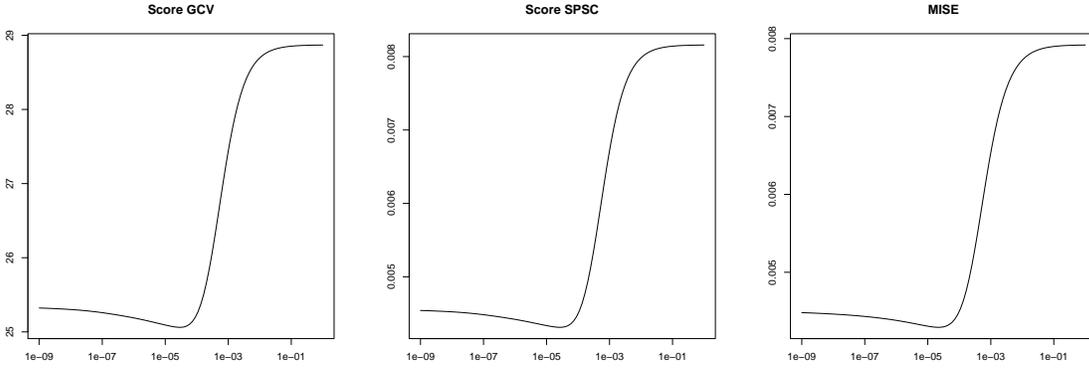


FIG. 3.4 – Scores GCV et SPSC en fonction du paramètre de lissage pour la simulation de $n = 10$ trajectoires observées en $p = 100$ points de $f + \sigma W$, et erreur théorique MISE correspondante ($f(t) = 10t^3 - 15t^4 + 6t^5$, $\sigma = 0.3$, et W processus de Wiener standard).

Nous passons maintenant à la comparaison proprement dite des performances de chaque méthode. Il est déjà ressorti de la Figure 3.4 que l'on pouvait s'attendre à ce que les deux méthodes choisissent des valeurs de paramètre de lissage très proches. Ceci est confirmé par le Tableau 3.3 qui indique l'égalité des erreurs d'estimation à 10^{-6} près. Sur l'ensemble des simulations réalisées, qui incluaient aussi l'étude d'autres fonctions de régression et d'un processus de bruit de type Ornstein-Uhlenbeck, nous avons constaté que les estimateurs fournis par les deux méthodes étaient indiscernables à l'oeil nu.

En dessous des erreurs L_2 moyennes de chaque méthode sur 100 répétitions figure la moyenne des erreurs L_2 optimales $\min_{\lambda > 0} \int_0^1 (\hat{f}_\lambda(t) - f(t))^2 dt$ [OPT], où \hat{f}_λ est un estimateur spline cubique de lissage MCO. On voit que l'efficacité des deux méthodes varie entre 97% et 99% sur l'ensemble des valeurs de (n, p) étudiées.

Remarque 3.5. *Face à de petits échantillons et/ou de forts niveaux de bruit, les méthodes SPSC et GCV fournissent parfois un paramètre de lissage infini qui mène à l'estimation de f par la droite de régression des moindres carrés usuelle. Ce phénomène rare survient lorsque les points de la trajectoire moyenne sont quasiment alignés. Dans ce cas, le score SPSC (3.17) est minimal pour λ très grand, du fait que la quantité $\text{tr}((\mathbf{S}_\lambda \Phi + \Phi \mathbf{S}_\lambda) \hat{\mathbf{V}})/n$ est minimisée pour λ très grand (\mathbf{S}_λ agit alors comme un filtre en ne laissant passer que les tendances affines), alors que la quantité $|\hat{f}_0 - \hat{f}_\lambda|_0^2$ reste très petite pour tout λ . Le score GCV (3.14) est également minimal pour λ très grand, car la prédiction optimale de chaque trajectoire observée est à peu près réalisée par la droite de régression associée aux $(n - 1)$ autres trajectoires. Dans une telle situation de quasi-alignement, l'erreur L_2 optimale est obtenue par interpolation des données, mais la différence entre interpolation et droite de régression n'étant alors pas flagrante, cela ne porte pas préjudice aux deux méthodes.*

		n= 10	n=20	n= 50	n= 100
$p = 10$	SPSC	0.004343	0.002000	0.000891	0.000423
	GCV	0.004351	0.002002	0.000891	0.000424
	OPT	0.004259	0.001960	0.000873	0.000416
$p = 20$	SPSC	0.004803	0.001882	0.000768	0.000423
	GCV	0.004805	0.001882	0.000769	0.000423
	OPT	0.004742	0.001849	0.000756	0.000418
$p = 100$	SPSC	0.004604	0.002004	0.000878	0.000342
	GCV	0.004606	0.002004	0.000878	0.000342
	OPT	0.004520	0.001979	0.000865	0.000337
$p = 200$	SPSC	0.004146	0.002387	0.000814	0.000562
	GCV	0.004149	0.002386	0.000814	0.000562
	OPT	0.004073	0.002349	0.000805	0.000557

TAB. 3.3 – Erreur L_2 dans l’estimation de $f(t) = 10t^3 - 15t^4 + 6t^5$ par des splines cubiques de lissage avec le bruit σW , où W est un processus de Wiener standard et $\sigma = 0.3$. Pour chaque valeur de (n, p) , 100 simulations sont réalisées au cours desquelles on construit les estimateurs par les méthodes SPSC, GCV, et par minimisation de l’erreur L_2 [OPT].

En définitive, il ressort des simulations précédentes que les méthodes GCV et SPSC sont très efficaces en présence de mesures répétées fortement corrélées. Leurs performances, qui sont comparables, réalisent en moyenne 98,3% de la performance L_2 optimale. L’excellence de ce résultat doit cependant être pondérée par le fait qu’en présence d’une corrélation très forte, même la meilleure estimation possible de f risque d’être mauvaise (voir la Figure 3.3). Dans d’autres simulations (non présentées ici) utilisant la même fonction f et des erreurs de type AR(1) avec le même écart-type $\sigma = 0.3$, les performances de ces méthodes restent comparables entre elles, et leur efficacité par rapport à la performance L_2 optimale varie entre 70% et 85% selon la force de la corrélation et la taille du jeu de données. Ces éléments font apparaître la méthode SPSC comme une alternative efficace à la méthode GCV “leave one curve out” de Rice et Silverman (1991).

Au terme de cette étude encourageante, des recherches plus élaborées peuvent être envisagées. D’abord, il serait intéressant d’examiner la performance de ces méthodes dans le cadre d’un dispositif expérimental déséquilibré. Ensuite, l’optimalité asymptotique de ces méthodes pourrait faire l’objet d’un travail du type de celui de Craven et Wahba (1979) et de Wahba (1983). Enfin une étude approfondie de la littérature sur l’estimation non paramétrique d’une fonction de covariance permettrait d’explorer des techniques plus sophistiquées que l’utilisation naïve de la covariance empirique effectuée ici.

Chapitre 4

Estimation adaptative en présence de corrélation

Dans ce chapitre, nous nous intéressons à un thème récent de la statistique qui est l'estimation adaptative dans la régression non paramétrique sur des données corrélées. Nous étendons de façon simple deux méthodes d'estimation qui sont le lissage spline adaptatif (Pintore *et al.*, 2006) et la méthode des P-splines adaptatives (Ruppert et Carroll, 2000). L'extension réalisée consiste à passer d'erreurs de type bruit blanc à des erreurs ayant une structure de covariance paramétrique. Nous étudions ensuite les deux estimateurs adaptatifs étendus lors de simulations et nous comparons leurs performances à celles d'estimateurs splines utilisant un paramètre de lissage global (*cf.* les méthodes GML, GCV, UBR pour données corrélées du Chapitre 3).

Le problème de la sélection du paramètre de lissage d'un estimateur dans la régression non paramétrique sur des données corrélées a largement été étudié. Parmi les nombreux travaux sur le sujet, citons Hart et Wehrly (1986), Diggle et Hutchinson (1989), et Rice et Silverman (1991) pour l'estimation avec des erreurs de mesure de type processus à temps continu ; Altman (1990) pour des suites de processus d'erreur stationnaires à temps discret ; Kohn *et al.* (1992) pour des erreurs de type processus ARMA ; Johnstone et Silverman (1997) pour des erreurs ayant une corrélation de longue portée ; Wang (1998) pour des erreurs ayant une structure de covariance paramétrique ; et enfin Hart (1996) et Opsomer *et al.* (2001) pour des revues de fond de la littérature à ce sujet.

Le champ statistique de l'estimation adaptative en présence d'erreurs décorréelées a aussi été exploré par de nombreux chercheurs. Müller et Stadtmüller (1987), Staniswalis (1989), et Fan et Gijbels (1995) ont étudié l'estimation par noyau avec une fenêtre locale. Dans Staniswalis (1989), cette fenêtre est obtenue par minimisation de l'erreur quadratique ponctuelle (MSE) qui est estimée à l'aide d'un estimateur pi-

lote ayant une fenêtre globale. Donoho et Johnstone (1995) ont établi des procédures de seuillage adaptatif des coefficients empiriques pour les estimateurs d'ondelettes. Friedman et Silverman (1989) et Zhou et Shen (2001) ont proposé des méthodes de sélection de noeuds pour les splines de régression. Eilers et Marx (1996) et Wand (1999) ont étudié les estimateurs P-splines avec une pénalité globale. À partir d'une approximation de l'erreur quadratique moyenne aux points d'observation (MASE), Wand (1999) a déterminé une formule explicite pour le paramètre de lissage optimal et a fourni une méthode plug-in simple pour estimer ce paramètre. L'estimation spatialement adaptative par P-splines a été discutée par Ruppert et Carroll (2000) et par Baladandayuthapani *et al.* (2005). Luo et Wahba (1997) ont développé une approche hybride entre les splines de régression et les splines de lissage qui utilise à la fois la sélection de noeuds et une pénalité de lissage. Heckman et Ramsay (2000) ont proposé une méthode d'estimation par L-splines dans laquelle l'opérateur différentiel linéaire L peut être déterminé adaptativement par les données. Cummins *et al.* (2001) ont construit une méthode GCV locale qui permet d'améliorer l'inférence par rapport au GCV global en diminuant la quantité de lissage (réduction du biais) dans les régions où la fonction de régression varie rapidement. Enfin, Pintore *et al.* (2006) ont donné la construction de l'estimateur spline de lissage adaptatif avec une pénalité variable localement constante.

Malgré l'importance des littératures relatives à chacun des deux thèmes de l'estimation adaptative et de l'estimation à partir de données corrélées, ces deux thèmes ont rarement été examinés *en même temps* (*cf.* Crainiceanu *et al.*, 2007). Une des raisons à cela est la difficulté de discerner ce qui, dans la variabilité observée d'un jeu de données, a trait à l'hétérogénéité spatiale de la fonction moyenne de ce qui a trait aux effets aléatoires systématiques (patterns des erreurs, corrélation et hétéroscédasticité). Pour faire de l'inférence statistique, une telle différenciation est nécessaire mais elle n'est possible que si l'on spécifie à la fois un modèle pour le signal et un modèle pour le bruit. Dans cette optique, Crainiceanu *et al.* (2007) ont étudié une modélisation bayésienne de la fonction de régression et de la fonction de variance des erreurs à l'aide de splines dans un modèle avec erreurs décorrélées et hétéroscédastiques.

Étant donné la nouveauté et la difficulté du sujet de l'estimation adaptative en présence de données corrélées, notre étude se place dans un cadre d'estimation relativement simple. Ainsi le modèle de régression considéré dans ce chapitre, la covariance des erreurs est modélisée de façon paramétrique, ce qui reste un cadre simple en regard d'une modélisation non paramétrique (éventuellement bayésienne) de ces erreurs. Par ailleurs, après avoir réalisé l'extension facile des méthodes d'estimation spline de Pintore *et al.* (2006) et de Ruppert et Carroll (2000), nous étudions par des simulations les performances des estimateurs obtenus sans examiner leurs propriétés théoriques. Nous tentons néanmoins de répondre à une question

importante du point de vue pratique : les estimateurs adaptatifs apportent-ils une amélioration significative par rapport aux estimateurs non adaptatifs dans l'estimation de la régression ? (nous savons déjà que dans le cadre d'erreurs de type bruit blanc et de fonctions de régression spatialement hétérogènes, la réponse est affirmative, cf. Staniswalis, 1989; Ruppert et Carroll, 2000; Pintore *et al.*, 2006).

Nous présentons maintenant le modèle étudié dans ce chapitre, avant d'étendre en Sections 4.1 et 4.2 les méthodes citées précédemment et de comparer les estimateurs étendus à des estimateurs non adaptatifs (c'est-à-dire avec un paramètre de lissage global) dans les simulations de la Section 4.3. Nous considérons comme Wang (1998) le modèle de régression non paramétrique

$$y_i = f(t_i) + \varepsilon_i, 1 \leq i \leq p,$$

où les points $0 \leq t_1 < \dots < t_p \leq 1$ sont fixés dans $[0, 1]$, f est la régression à estimer et les $\varepsilon_i, 1 \leq i \leq p$, sont des v.a. centrées ayant une matrice de covariance $\mathbf{V} = \sigma^2 \mathbf{W}^{-1}$ de forme paramétrique connue (on suppose que \mathbf{W} s'exprime en fonction d'un vecteur de paramètres $\boldsymbol{\tau}$ inconnu).

4.1 Lissage spline adaptatif

Nous proposons dans cette section une extension de la construction de Pintore *et al.* (2006), qui correspond à des erreurs de type bruit blanc ($\mathbf{W} = \mathbf{I}$), au cas où \mathbf{W} est (l'inverse d') une matrice de covariance quelconque. Dans ce but, nous utilisons les résultats de Wang (1998).

4.1.1 Estimateur

L'estimateur spline de lissage adaptatif $\hat{f}_{\lambda, \boldsymbol{\tau}}$ d'ordre $2m$ et de paramètre de lissage la fonction λ strictement positive est défini comme la solution du problème : Trouver $g \in W_2^m[0, 1]$ qui minimise

$$\frac{1}{p} (\mathbf{y} - \mathbf{g})' \mathbf{W} (\mathbf{y} - \mathbf{g}) + \int_0^1 \lambda(t) (g^{(m)}(t))^2 dt, \quad (4.1)$$

avec $\mathbf{y} = (y_1, \dots, y_p)'$ et $\mathbf{g} = (g(t_1), \dots, g(t_p))'$.

Pour caractériser cet estimateur de manière explicite, il est nécessaire de passer par le formalisme des espaces de Hilbert à noyau autoreproduisant. Considérons donc l'espace de Hilbert $\{g \in W_2^m[0, 1] : g^{(k)}(0) = 0, 0 \leq k \leq m\}$ muni du produit scalaire $\langle g, h \rangle = \int_0^1 \lambda(t) g^{(m)}(t) h^{(m)}(t) dt$. En notant $x_+ = \max(x, 0)$, le noyau

autoreproduisant de cet espace est (Pintore *et al.*, 2006) :

$$K_\lambda(s, t) = \int_0^1 \frac{1}{\lambda(u)} \frac{(s-u)_+^{m-1} (t-u)_+^{m-1}}{((m-1)!)^2} du \quad (4.2)$$

D'autre part, soit $(\phi_j)_{0 \leq j \leq m-1}$ la base de l'espace π_{m-1} des polynômes de degré inférieur ou égal à $m-1$ définie par $\phi_j(t) = t^j/j!$ pour $0 \leq j \leq m-1$. Nous montrons comme dans Craven et Wahba (1979) ou dans Gu (2002) que l'estimateur s'écrit sous la forme

$$\widehat{f}_{\lambda, \tau}(t) = \sum_{i=1}^p c_i K_\lambda(t, t_i) + \sum_{j=0}^{m-1} d_j \phi_j(t). \quad (4.3)$$

Introduisons alors les notations $\widehat{\mathbf{f}}_{\lambda, \tau} = (\widehat{f}_{\lambda, \tau}(t_1), \dots, \widehat{f}_{\lambda, \tau}(t_p))'$, $\mathbf{c} = (c_1, \dots, c_p)'$, $\mathbf{d} = (d_0, \dots, d_{m-1})'$, $\Sigma_\lambda = (K_\lambda(t_i, t_j))_{1 \leq i, j \leq p}$, et $\mathbf{T} = (\phi_j(t_i))_{1 \leq i \leq p, 1 \leq j \leq m}$. La formule (4.3) se retraduit en

$$\widehat{\mathbf{f}}_{\lambda, \tau} = \Sigma_\lambda \mathbf{c} + \mathbf{T} \mathbf{d},$$

et nous reformulons le problème (4.1) en : Trouver (\mathbf{c}, \mathbf{d}) qui minimise

$$\left\{ \frac{1}{p} (\mathbf{y} - \Sigma_\lambda \mathbf{c} - \mathbf{T} \mathbf{d})' \mathbf{V}^{-1} (\mathbf{y} - \Sigma_\lambda \mathbf{c} - \mathbf{T} \mathbf{d}) + \mathbf{c}' \Sigma_\lambda \mathbf{c} \right\}. \quad (4.4)$$

Posons $\mathbf{M}_{\lambda, \tau} = \Sigma_\lambda + p \mathbf{W}^{-1}$. En reprenant la décomposition QR de \mathbf{T} définie par (voir Section 3.2.1)

$$\mathbf{T} = (\mathbf{Q}_1 \quad \mathbf{Q}_2) \begin{pmatrix} \mathbf{R} \\ \mathbf{0} \end{pmatrix},$$

nous montrons aisément que la matrice de lissage associée à $\widehat{f}_{\lambda, \tau}$ s'écrit

$$\mathbf{S}_{\lambda, \tau} = \mathbf{I} - p \mathbf{W}^{-1} \mathbf{Q}_2 (\mathbf{Q}_2' \mathbf{M}_{\lambda, \tau} \mathbf{Q}_2)^{-1} \mathbf{Q}_2'. \quad (4.5)$$

4.1.2 Noyau autoreproduisant

Soient des noeuds $0 = \tau_0 < \tau_1 < \dots < \tau_S < \tau_{S+1} = 1$. Nous allons considérer une fonction de lissage λ en escalier :

$$\lambda(t) = \lambda_i \quad (\lambda_i > 0) \quad (t \in [\tau_i, \tau_{i+1}), \quad 1 \leq i \leq S). \quad (4.6)$$

Dans cette situation, il est possible de donner une forme explicite au noyau K_λ défini par (4.2). Fixons un t_i pour une valeur de i donnée ($1 \leq i \leq p$). Alors pour tout $v \in [t_i, 1]$, nous avons

$$K_\lambda(t_i, v) = \sum_{k=0}^S \sum_{j=1}^m \lambda_k^{-1} (-1)^j \left\{ \frac{(t_i - \tau_{k+1})_+^{m-1+j} (v - \tau_{k+1})_+^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i - \tau_k)_+^{m-1+j} (v - \tau_k)_+^{m-1+j}}{(m-1+j)!(m-j)!} \right\}, \quad (4.7)$$

et pour tout $v < t_i$ tel que $v \in [\tau_l, \tau_{l+1})$ pour un l donné, nous avons

$$K_\lambda(t_i, v) = \sum_{k=0}^S \sum_{j=1}^m \lambda_k^{-1} (-1)^j \left\{ \frac{(t_i - \tau_{k+1})_+^{m-1+j} (v - \tau_{k+1})^{m-j}}{(m-1+j)!(m-j)!} - \frac{(t_i - \tau_k)^{m-1+j} (v - \tau_k)_+^{m-1+j}}{(m-1+j)!(m-j)!} \right\} + \lambda_l^{-1} (-1)^m \frac{(t_j - v)^{2m-1}}{(2m-1)!}. \quad (4.8)$$

4.1.3 Estimation des paramètres de lissage et de corrélation

Nous utilisons un score GCV qui étend ceux de Pintore *et al.* (2006) et de Wang (1998) :

$$V(\lambda_1, \dots, \lambda_S, \boldsymbol{\tau}) = \frac{n^{-1} \|\mathbf{W}((\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}})\mathbf{y})\|^2}{[n^{-1} \text{tr}(\mathbf{W}(\mathbf{I} - \mathbf{S}_{\lambda, \boldsymbol{\tau}}))]^2}. \quad (4.9)$$

Dans la pratique, nous suivrons les recommandations de Pintore *et al.* (2006) en prenant S relativement petit (entre 5 et 20) et en choisissant les noeuds τ_1, \dots, τ_S parmi les points d'observation t_1, \dots, t_p .

4.2 P-splines adaptatives

La méthode des P-splines spatialement adaptatives de Ruppert et Carroll (2000) fait appel au modèle de régression suivant :

$$f(t; \boldsymbol{\beta}) = \beta_0 + \beta_1 t + \dots + \beta_m t^m + \sum_{k=1}^K \beta_{p+k} (t - \tau_k)_+^m, \quad (4.10)$$

où $m \geq 1$ est un entier, $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p+K})'$ est un vecteur de coefficients de régression, et les $\tau_1 < \dots < \tau_K$ sont des noeuds fixés. Typiquement, le nombre K de noeuds peut être pris grand, mais il reste petit devant le nombre p d'observations. Par ailleurs on prend pour noeuds τ_k les quantiles de niveau $\lfloor k/K \rfloor$ ($1 \leq k \leq K$) des points d'observation du modèle, où $\lfloor \cdot \rfloor$ désigne la partie entière d'un réel.

Afin de construire l'estimateur de f dans le cadre d'erreurs de type bruit blanc, Ruppert et Carroll (2000) utilisent un critère de moindres carrés avec une pénalité sur les sauts de la dérivée d'ordre m de l'estimateur. C'est cette pénalité qui permet l'adaptation spatiale de l'estimateur aux variations plus ou moins rapides de f . Nous étendons la méthode de Ruppert et Carroll (2000) au cadre d'erreurs corrélées en remplaçant simplement dans leur critère les moindres carrés par des moindres

carrés généralisés qui incorporent la covariance des erreurs. Plus précisément, nous définissons $\widehat{\boldsymbol{\beta}}(\alpha, \boldsymbol{\tau})$ comme étant le vecteur $\boldsymbol{\beta} \in \mathbb{R}^{m+K}$ qui minimise la quantité

$$\frac{1}{p} (\mathbf{y} - \mathbf{f}_{\boldsymbol{\beta}})' \mathbf{W} (\mathbf{y} - \mathbf{f}_{\boldsymbol{\beta}}) + \sum_{k=1}^K \alpha(\tau_k) \beta_{m+k}^2, \quad (4.11)$$

où $\mathbf{y} = (y_1, \dots, y_p)'$, $\mathbf{f}_{\boldsymbol{\beta}} = (f(t_1; \boldsymbol{\beta}), \dots, f(t_p; \boldsymbol{\beta}))'$, et α est une fonction strictement positive (c'est la pénalité mentionnée).

Pour choisir la pénalité α , Ruppert et Carroll (2000) commencent par fixer un petit ensemble de noeuds $\{\tau_1 = \tau_1^* < \tau_2^* < \dots < \tau_{M-1}^* < \tau_M^* = \tau_K\}$, avec $M < K$. À chaque "sous-noeud" τ_k^* ils associent une pénalité $\alpha(\tau_k^*) = \alpha_k^*$, puis ils construisent la fonction α par interpolation linéaire des (τ_k^*, α_k^*) , $1 \leq k \leq M$. On voit alors que les pénalités $\alpha(\tau_k)$ utilisées dans la construction de l'estimateur ne dépendent que du vecteur $\boldsymbol{\alpha}^* = (\alpha_1^*, \dots, \alpha_M^*)'$. En notant \mathbf{X} la matrice de taille $p \times (m + K + 1)$ dont la i -ème ligne est $\mathbf{X}_i = (1, t_i, \dots, t_i^m, (t_i - \tau_1)_+^m, \dots, (t_i - \tau_K)_+^m)$ pour $1 \leq i \leq p$, et en notant $\mathbf{D}(\boldsymbol{\alpha}^*)$ la matrice diagonale dont les $(p + 1)$ premiers coefficients diagonaux sont nuls et dont les autres coefficients diagonaux sont $\alpha(\tau_1), \dots, \alpha(\tau_K)$, les auteurs obtiennent ensuite la solution de (4.11) dans le cas $\mathbf{W} = \mathbf{I}$ par des calculs standards. En suivant ces calculs, nous obtenons que le vecteur des coefficients de régression de l'estimateur P-spline s'écrit

$$\widehat{\boldsymbol{\beta}}(\alpha, \boldsymbol{\tau}) = (\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{D}(\boldsymbol{\alpha}^*))^{-1} \mathbf{X}'\mathbf{W}\mathbf{y}. \quad (4.12)$$

La méthode des P-splines est ainsi une ridge regression qui rapproche l'estimateur spline de régression d'un ajustement polynômial de degré m par la méthode des moindres carrés (généralisés).

Il reste enfin à déterminer le vecteur $\boldsymbol{\alpha}^*$ et aussi, dans le cadre corrélé, le vecteur $\boldsymbol{\tau}$ qui paramétrise la matrice \mathbf{W} . Dans ce but, Ruppert et Carroll (2000) font appel à une méthode GCV que nous généralisons ici par :

$$\text{GCV}(\boldsymbol{\alpha}^*, \boldsymbol{\tau}) = \frac{\|\mathbf{W}(\mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}}(\alpha, \boldsymbol{\tau}))\|^2}{(1 - \text{df}(\boldsymbol{\alpha}^*, \boldsymbol{\tau})/p)^2}, \quad (4.13)$$

où $\text{df}(\boldsymbol{\alpha}^*, \boldsymbol{\tau}) = \text{tr} \{(\mathbf{X}'\mathbf{W}\mathbf{X} + \mathbf{D}(\boldsymbol{\alpha}^*))^{-1} \mathbf{X}'\mathbf{W}\mathbf{X}\}$ est le nombre de degrés de liberté de l'estimateur qui est défini comme la trace de la matrice de lissage associée à cet estimateur (voir Wahba, 1990; Hastie et Tibshirani, 1990).

4.3 Lissage local vs lissage global : simulations

Nous avons conduit des simulations pour étudier les estimateurs adaptatifs splines de lissage et P-splines définis dans les sections précédentes. D'autres estimateurs splines ont été impliqués dans ces simulations dans un but de comparaison entre les performances du lissage global et du lissage local.

L'étude préliminaire comprenait cinq estimateurs splines dont trois "globaux", notés \hat{f}_{GCV} (spline cubique de lissage avec paramètre global choisi par GCV, cf. la Section 3.2.1), \hat{f}_{GML} (spline cubique de lissage avec paramètre global choisi par GML, cf. la Section 3.2.1), et $\hat{f}_{P,glob}$ (P-spline cubique avec un paramètre global choisi par GCV), et deux "locaux", notés $\hat{f}_{L,loc}$ (spline de lissage adaptative de la Section 4.1) et $\hat{f}_{P,loc}$ (P-spline adaptative de la Section 4.2).

L'ensemble des simulations a été réalisé dans l'environnement logiciel R. Les opérations de minimisation numérique pour l'estimation des paramètres de lissage et de corrélation du bruit ont été effectuées à l'aide des fonctions `optimize` et `optim` disponibles dans les packages standards. Les algorithmes de recherche utilisés étaient soit celui de Nelder-Mead, soit celui de Byrd *et al.* (1995). Cet algorithme de type quasi-Newton permet d'intégrer des contraintes de bornes supérieures et inférieures pour le domaine de recherche. Globalement, ces deux algorithmes se montrent robustes à l'initialisation de la recherche fournie par l'utilisateur. L'estimateur P-spline global $\hat{f}_{P,glob}$ a présenté de sérieux problèmes techniques (tendance à un lissage très fort, à une estimation de la corrélation ρ des observations supérieure à 1, et au plantage des algorithmes à cause de matrices mal conditionnées) et pour cette raison, nous l'avons écarté de l'étude.

Dans la première étude conduite, nous avons choisi comme régression la version suivante de la fonction de Heaviside :

$$f(t) = 0 \text{ si } t \leq 0.5, \quad \text{et} \quad f(t) = 1 \quad \text{si } t > 0.5, \quad (4.14)$$

qui est intéressante du fait de son hétérogénéité spatiale, et nous avons simulé des erreurs de type AR(1) stationnaire avec un paramètre de corrélation ρ prenant les valeurs 0.2, 0.5, et 0.8, et un écart-type $\sigma = 0.15$. Les données ont été générées dans un dispositif expérimental de $p = 100$ points équidistants.

Pour chaque valeur de ρ , nous avons répété 10 fois la simulation et nous avons noté l'erreur moyenne $ASE = p^{-1} \sum_{i=1}^p (\hat{f}(t_i) - f(t_i))^2$ pour chaque estimateur, ainsi que la valeur estimée de ρ . Le Tableau 4.1 présente les résultats de cette expérience moyennés à travers les 10 répétitions. On voit que pour toutes les valeurs de ρ étudiées, l'estimateur P-spline adaptatif est de loin le plus efficace dans l'estimation de f . L'estimateur qui semble-t-il estime le mieux la corrélation est celui de la méthode GCV globale. Cependant, les valeurs indiquées ici sont des moyennes et l'examen des estimations individuelles de ρ montre une importante variabilité de cette estimation. Cela est dû en partie au très petit nombre de répétitions de

la simulation. D'autre part, de façon intéressante l'erreur d'estimation de f ne semble pas augmenter avec ρ pour la P-spline adaptative, au contraire des autres estimateurs. Dans l'ensemble, nous avons observé que l'estimation de la structure de corrélation des erreurs (ici paramétrique) n'affecte que peu l'estimation parallèle de la fonction de régression. Ce constat a déjà été fait au Chapitre 3 lors de la comparaison des estimateurs des moindres carrés ordinaires et généralisés.

	$\rho = 0.2$		$\rho = 0.5$		$\rho = 0.8$	
	ASE	$\hat{\rho}$	ASE	$\hat{\rho}$	ASE	$\hat{\rho}$
\hat{f}_{GCV}	0.02569344	0.3413817	0.04852956	0.4824835	0.04836822	0.7734921
\hat{f}_{GML}	0.03016952	0.3735718	0.02931759	0.639774	0.06925794	0.8908166
$\hat{f}_{P,loc}$	0.01852055	0.6242201	0.01779877	0.8076486	0.01780423	0.8482969
$\hat{f}_{L,loc}$	0.02328520	0.2927062	0.03723562	0.6950397	0.04583518	0.7734902

TAB. 4.1 – Comparaison d'estimateurs splines ayant un paramètre de lissage soit global, soit local, dans l'estimation d'une fonction de Heaviside spatialement hétérogène observée avec un bruit AR(1) de paramètres ρ et $\sigma = 0.15$.

En définitive, il apparaît que les deux estimateurs adaptatifs apportent une réelle amélioration dans l'estimation. Une étude réalisée sur un échantillon de plus grande taille devrait conforter cette amélioration. Nous pouvons aussi nous donner une idée de l'apport des estimateurs adaptatifs en observant sur la Figure 4.1 la capacité qu'ont ces estimateurs à se rapprocher fortement du saut de la fonction de Heaviside au point 0.5. Il faut d'ailleurs remarquer que les estimateurs implémentés n'ont que 5 noeuds, et que leur adaptativité devrait aller croissant avec leur nombre de noeuds (voir à ce sujet Pintore *et al.*, 2006).

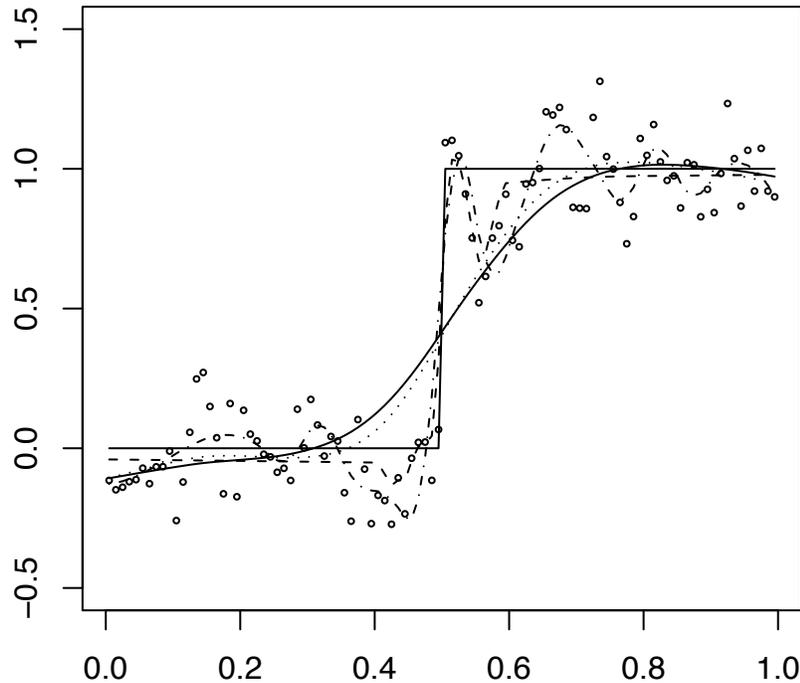


FIG. 4.1 – Estimation de la fonction de Heaviside par deux splines de paramètre global (méthodes GCV en ligne pleine et GML en pointillés) et par deux splines de paramètre local (méthode de lissage en tirets et P-splines en tirets alternés) avec 5 noeuds équidistants.

Conclusion et perspectives

L'étude succincte menée dans ce chapitre indique que l'implémentation d'une forme de lissage adaptatif sur des données corrélées améliore grandement l'estimation, comme dans le cas de la régression non paramétrique avec des erreurs de type bruit blanc. Le type d'estimation envisagé ici constitue en fait un lissage doublement adaptatif, en ce sens qu'il vise à s'ajuster simultanément au signal et au bruit des données. Cette idée a été expliquée en détail par Crainiceanu *et al.* (2007) qui, dans un cadre bayésien, ont montré qu'une estimation efficace en présence d'hétérogénéité du signal et d'hétéroscédasticité du processus d'erreur est reliée à la connaissance des variations du rapport signal/bruit. Par contre, dans une optique d'inférence statistique, il est nécessaire d'évaluer précisément la structure de ces deux composantes des données. Sans aller jusqu'à une modélisation purement bayésienne, une perspective bayésienne empirique (*cf.* Ruppert et Carroll, 2000) permet déjà de construire des intervalles de confiance pour la régression. Globalement, une méthode adaptative a le potentiel de capturer des variations très rapides d'une fonction d'intérêt, comme

dans l'exemple précédent du saut de la fonction de Heaviside. Pour aller plus loin sur le sujet de la construction de bandes de confiance pour les sauts d'une courbe, on pourra consulter Gijbels *et al.* (2004) et les références qu'il contient. Nous signalons enfin l'intérêt d'une étude pratique et/ou théorique qui pouvant être faite sur le nombre optimal (et éventuellement le positionnement) des noeuds d'un estimateur.

Annexe A

Rappels

Fonctions splines

Définition 1. Soit g une fonction à valeurs réelles définie sur \mathbb{R} , soit $[a, b]$ un intervalle borné et soit $m \geq 2$ un entier. On dit que g est une spline d'ordre m (ou de degré $(m - 1)$) sur $[a, b]$ s'il existe un ensemble fini $\{t_1 < \dots < t_p\}$ de points de $[a, b]$ tel que

- i) g est un polynôme d'ordre m sur chacun des intervalles $(-\infty, t_1[$, $]t_1, t_2[$, \dots , $]t_{p-1}, t_p[$ et $]t_p, \infty)$,
- ii) g est de classe C^{m-2} sur $[a, b]$.

Les points $t_1 < \dots < t_p$ sont alors appelés les noeuds de la spline g .

Définition 2. Avec les notations précédentes, on dit qu'une fonction g est une spline naturelle d'ordre $2m$ (ou de degré $(2m - 1)$) sur $[a, b]$ s'il existe un ensemble fini $\{t_1 < \dots < t_p\}$ de points de $[a, b]$ tel que

- (i) g est un polynôme d'ordre m sur $(-\infty, t_1[$ et sur $]t_p, \infty)$.
- (ii) g est un polynôme d'ordre $2m$ sur chacun des intervalles $]t_1, t_2[$, \dots , $]t_{p-1}, t_p[$.
- (iii) g est de classe C^{2m-2} sur $[a, b]$.

Remarque 1. Pour un entier $m \geq 1$ donné, l'ensemble des splines d'ordre m ainsi que l'ensemble des splines naturelles d'ordre $2m$ sont inclus dans l'espace de Sobolev $W_2^m[a, b]$.

Remarque 2. Dans les définitions précédentes, on peut associer à une même fonction spline g plusieurs ensembles de noeuds. Cependant, il existe un unique ensemble de noeuds (éventuellement vide) associé à g qui soit minimal au sens de l'inclusion. On peut donc parler des noeuds d'une fonction spline g sans ambiguïté.

Calcul matriciel

Une inégalité utile

Lemme 1.3. Soit $p \geq 1$ un entier et soient \mathbf{A} et \mathbf{B} deux matrices réelles de taille $p \times p$ telles que \mathbf{A} est symétrique et \mathbf{B} est symétrique et semi-définie positive. Alors, en notant $\lambda_{\min}(\mathbf{A})$ la plus petite valeur propre de \mathbf{A} , nous avons

$$\lambda_{\min}(\mathbf{A}) \operatorname{tr} \mathbf{B} \leq \operatorname{tr}(\mathbf{A}\mathbf{B}) \leq \|\mathbf{A}\| \operatorname{tr} \mathbf{B}.$$

Démonstration.

Nous ne montrerons que la majoration de la quantité $\operatorname{tr}(\mathbf{A}\mathbf{B})$, la minoration se faisant de manière similaire.

Puisque \mathbf{A} est symétrique, il existe une matrice \mathbf{P} orthogonale et une matrice \mathbf{D} diagonale, toutes deux de taille $p \times p$, telles que $\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}'$. De plus, les coefficients non nuls de la matrice \mathbf{D} , notés d_{11}, \dots, d_{pp} , sont les valeurs propres de \mathbf{A} et le plus grand d'entre eux (en valeur absolue) est égal à $\|\mathbf{A}\|$. Ensuite, par commutativité de la trace, nous avons $\operatorname{tr}(\mathbf{A}\mathbf{B}) = \operatorname{tr}(\mathbf{P}\mathbf{D}\mathbf{P}'\mathbf{B}) = \operatorname{tr}(\mathbf{D}\mathbf{P}'\mathbf{B}\mathbf{P})$. D'autre part, la matrice $\mathbf{D}\mathbf{P}'\mathbf{B}\mathbf{P}$ étant semi-définie positive comme \mathbf{B} , ses coefficients diagonaux, notés c_{11}, \dots, c_{pp} , sont positifs. Par conséquent, nous avons enfin

$$\operatorname{tr}(\mathbf{A}\mathbf{B}) = \sum_{i=1}^p d_{ii}c_{ii} \leq \max_{1 \leq i \leq p} d_{ii} \sum_{i=1}^p d_{ii} = \|\mathbf{A}\| \operatorname{tr} \mathbf{B}. \blacksquare$$

Valeurs propres généralisées

Nous commençons par rappeler le Théorème min-max de Courant-Fischer. Pour tout entier $p > 1$, nous identifions dans la suite un vecteurs-colonnes à p entrées et l'élément de \mathbb{R}^p associé.

Théorème 3. (Courant-Fischer) Soit \mathbf{A} une matrice $p \times p$ symétrique. On note $\lambda_1(\mathbf{A}) \geq \dots \geq \lambda_p(\mathbf{A})$ les valeurs propres de cette matrice. Alors pour tout $1 \leq k \leq p$, on a

$$\lambda_k(\mathbf{A}) = \min_{F_k \in \mathbf{E}_k} \max_{x \in F_k - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} = \max_{F_k \in \mathbf{E}_k} \min_{x \in F_k^\perp - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}},$$

où \mathbf{E}_k désigne l'ensemble des sous-espaces vectoriels de dimension k de \mathbb{R}^p .

Nous déduisons du Théorème de Courant-Fischer le résultat utile suivant.

Corollaire 1. Soient \mathbf{A} et \mathbf{B} deux matrices $p \times p$ symétriques. Alors en arrangeant dans l'ordre décroissant les valeurs propres d'une matrice, on a pour tout $1 \leq k \leq p$:

$$\lambda_k(\mathbf{A}) + \lambda_p(\mathbf{B}) \leq \lambda_k(\mathbf{A} + \mathbf{B}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}).$$

Démonstration.

Nous établirons seulement l'une des deux inégalités précédentes, l'autre s'obtenant de façon identique. Soit $1 \leq k \leq p$ un entier. Alors par le Théorème 3, on a

$$\begin{aligned}
\lambda_k(\mathbf{A} + \mathbf{B}) &= \min_{F_k \in \mathbf{E}_k} \max_{x \in F_k - \{0\}} \frac{\mathbf{x}'(\mathbf{A} + \mathbf{B})\mathbf{x}}{\mathbf{x}'\mathbf{x}} \\
&\leq \min_{F_k \in \mathbf{E}_k} \left(\max_{x \in F_k - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} + \max_{x \in F_k - \{0\}} \frac{\mathbf{x}'\mathbf{B}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right) \\
&\leq \min_{F_k \in \mathbf{E}_k} \left(\max_{x \in F_k - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} + \lambda_1(\mathbf{B}) \right) \\
&\leq \min_{F_k \in \mathbf{E}_k} \left(\max_{x \in F_k - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right) + \lambda_1(\mathbf{B}) \\
&= \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{B}). \blacksquare
\end{aligned}$$

Il existe aussi une version du Théorème 3 pour la caractérisation de valeurs propres généralisées. En effet, soient \mathbf{A} et \mathbf{B} deux matrices symétriques de taille $p \times p$, avec \mathbf{B} définie positive. On sait que les valeurs propres de l'équation propre généralisée $\mathbf{A}\mathbf{u} = \lambda\mathbf{B}\mathbf{u}$ sont celles de $\mathbf{B}^{-1}\mathbf{A}$, qui elles-mêmes sont celles de $\mathbf{B}^{-1/2}\mathbf{A}\mathbf{B}^{-1/2}$. On peut alors montrer que ces valeurs propres sont obtenues en maximisant (ou en minimisant suivant le cas) le quotient de Rayleigh $\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}}$ sur l'orthogonal dans la métrique \mathbf{B} des directions propres trouvées à chaque étape. Plus précisément, on a :

Théorème 4. Soient \mathbf{A} et \mathbf{B} deux matrices de taille $p \times p$, avec \mathbf{B} définie positive. On note $\lambda_1(\mathbf{B}^{-1}\mathbf{A}) \geq \dots \geq \lambda_p(\mathbf{B}^{-1}\mathbf{A})$ les valeurs propres de $\mathbf{B}^{-1}\mathbf{A}$. Soient $\mathbf{x}_1, \dots, \mathbf{x}_p$ des vecteurs propres linéairement indépendants associés à ces valeurs propres. Pour tout entier $1 \leq k \leq p$, on note $G_k = (\text{Vect}(\mathbf{x}_1, \dots, \mathbf{x}_k))_{\mathbf{B}}^{\perp}$, resp. $H_k = (\text{Vect}(\mathbf{x}_k, \dots, \mathbf{x}_p))_{\mathbf{B}}^{\perp}$, l'orthogonal pour la métrique \mathbf{B} de l'espace vectoriel engendré par les $\mathbf{x}_1, \dots, \mathbf{x}_k$, resp. par les $\mathbf{x}_{k+1}, \dots, \mathbf{x}_p$. Alors, avec la convention $G_0 = H_{p+1} = \mathbb{R}^p$, on a pour tout $1 \leq k \leq p$:

$$\lambda_k(\mathbf{B}^{-1}\mathbf{A}) = \min_{\mathbf{x} \in H_{k+1} - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} = \max_{\mathbf{x} \in G_{k-1} - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}},$$

Nous en déduisons l'encadrement suivant des valeurs propres de $\mathbf{B}^{-1}\mathbf{A}$:

Corollaire 2. Soient \mathbf{A} et \mathbf{B} deux matrices de taille $p \times p$, avec \mathbf{A} semi-définie positive et \mathbf{B} définie positive. Alors on a pour tout $1 \leq k \leq p$:

$$\frac{\lambda_k(\mathbf{A})}{\lambda_1(\mathbf{B})} \leq \lambda_k(\mathbf{B}^{-1}\mathbf{A}) \leq \frac{\lambda_k(\mathbf{A})}{\lambda_p(\mathbf{B})}.$$

Démonstration.

Nous établirons seulement l'une des deux inégalités précédentes, l'autre s'obtenant de façon identique. Soit $1 \leq k \leq p$ un entier. On commence par écrire, grâce au Théorème 4 la caractérisation

$$\begin{aligned} \lambda_k(\mathbf{B}^{-1}\mathbf{A}) &= \min_{\mathbf{x} \in H_{k+1} - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} \\ &= \min_{\mathbf{x} \in H_{k+1} - \{0\}} \left(\frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \frac{\mathbf{x}'\mathbf{x}}{\mathbf{x}'\mathbf{B}\mathbf{x}} \right) \end{aligned}$$

puis, étant donné que $\mathbf{x}'\mathbf{B}\mathbf{x} \geq \lambda_p(\mathbf{B})\mathbf{x}'\mathbf{x}$ pour tout $\mathbf{x} \in \mathbb{R}^p$, il vient

$$\lambda_k(\mathbf{B}^{-1}\mathbf{A}) \leq \left(\min_{\mathbf{x} \in H_{k+1} - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right) \frac{1}{\lambda_p(\mathbf{B})}.$$

En notant que le sous-espace vectoriel H_{k+1} est de dimension k , nous écrivons ensuite

$$\lambda_k(\mathbf{B}^{-1}\mathbf{A}) \leq \left(\max_{F_k \in \mathbf{E}_k} \min_{x \in F_k^\perp - \{0\}} \frac{\mathbf{x}'\mathbf{A}\mathbf{x}}{\mathbf{x}'\mathbf{x}} \right) \frac{1}{\lambda_p(\mathbf{B})}$$

et enfin nous pouvons conclure avec le Théorème 3 :

$$\lambda_k(\mathbf{B}^{-1}\mathbf{A}) = \frac{\lambda_h(\mathbf{A})}{\lambda_p(\mathbf{B})}. \blacksquare$$

Annexe B

Un résultat d'approximation spectrale

Soit R une fonction de covariance continue sur $[0, 1] \times [0, 1]$ et soit Γ l'opérateur de covariance associé à R défini par

$$\Gamma : \begin{cases} L_2[0, 1] & \rightarrow L_2[0, 1] \\ f & \mapsto \int_0^1 R(s, \cdot) f(s) ds \end{cases} .$$

Pas la continuité de R , l'opérateur Γ est compact et par le théorème spectral il admet des valeurs propres $\lambda_k, k \geq 1$ et des fonctions propres $\varphi_k, k \geq 1$ associées telles que la suite (λ_k) décroît vers zéro et les fonctions φ_k sont continues et forment une base orthonormée de $L_2[0, 1]$.

Soit $(t_{ip})_{p \in \mathbb{N}^*, 1 \leq i \leq p}$ un tableau triangulaire de points de $[0, 1]$ ayant une répartition quasi-uniforme. Pour tout entier $p \geq 1$, soit $\mathbf{V} = (R(t_{ip}, t_{jp}))_{1 \leq i, j \leq p}$. Alors les valeurs propres $\lambda_{1p} \geq \dots \geq \lambda_{pp} \geq 0$ de \mathbf{V} et les valeurs propres $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$ de l'opérateur Γ sont liées de la façon suivante.

Proposition 1. *Pour tout entier $k > 0$ fixé, $\lambda_{kp} = \lambda_k \cdot p(1 + o(1))$ quand $p \rightarrow \infty$.*

Démonstration.

Il suffit d'exploiter le résultat d'approximation spectrale suivant, tiré du Chapitre 5 de Chatelin (1983), qui affirme la convergence des valeurs propres, avec préservation de la multiplicité, de l'approximation "fortement stable" d'un opérateur linéaire fermé. On note $\sigma(T)$ le spectre d'un opérateur linéaire T et on note $C[0, 1]$ l'espace des fonctions continues sur $[0, 1]$ à valeur complexes ou réelles.

Théorème 5. (Chatelin, 1983) *Soit R une fonction appartenant à $C[0, 1]$. Soit $(t_{ip})_{p \in \mathbb{N}^*, 1 \leq i \leq p}$ un tableau triangulaire de points de $[0, 1]$ ayant une répartition quasi-uniforme. On considère l'approximation de l'opérateur intégral de Fredholm*

$$T : \begin{cases} C[0, 1] & \rightarrow C[0, 1] \\ f & \mapsto \int_0^1 R(s, \cdot) f(s) ds \end{cases}$$

par l'opérateur de Nyström

$$T_p : \begin{cases} C[0, 1] & \rightarrow C[0, 1] \\ f & \mapsto \frac{1}{p} \sum_{j=1}^p R(\cdot, t_{jp}) f(t_{jp}) \end{cases} .$$

Soit $\lambda \in \sigma(T) - \{0\}$ une valeur propre de multiplicité q (λ est nécessairement isolée), soit Δ un voisinage de λ dans \mathbb{C} tel que $\Delta \cap \sigma(T) = \{\lambda\}$, et soit Π_p l'opérateur de projection spectrale associé à $\sigma(T_p) \cap \Delta$. Alors

$$\lim_{p \rightarrow \infty} \{\sigma(T_p) \cap \Delta\} = \{\lambda\}, \quad \text{et} \quad \lim_{p \rightarrow \infty} \dim(\Pi_p(C[0, 1])) = q.$$

Nous voyons que par définition, les valeurs propres de l'opérateur de Nyström T_p sont celles de $p^{-1}\mathbf{V}$. Il suffit ensuite d'appliquer le théorème précédent à la fonction de covariance R qui est dans $C[0, 1]$ par hypothèse, et la Proposition 1 s'ensuit. ■

Une illustration avec le processus de Wiener.

Soit $\{W(t); t \in [0, 1]\}$ un processus de Wiener standard de fonction de covariance $R(s, t) := \text{Cov}(W(s), W(t)) = \min(s, t)$. Nous allons retrouver le résultat de la Proposition 1 par un calcul direct, dans le cas où les points de discrétisation t_{1p}, \dots, t_{pp} sont équidistants : $t_{jp} = j/p, 1 \leq j \leq p$.

Nous rappelons d'abord le développement classique en série de fonctions de R

$$R(s, t) = \sum_{k=1}^{\infty} \frac{4}{(2k-1)^2\pi^2} \left(\sqrt{2} \sin((k-1/2)\pi s) \right) \left(\sqrt{2} \sin((k-1/2)\pi t) \right),$$

où la convergence est absolue et uniforme en $s, t \in [0, 1]$.

Nous allons expliciter les valeurs propres $\lambda_{1p} \geq \dots \geq \lambda_{pp} > 0$ de la matrice $\mathbf{V} = \mathbf{V}(p) = (R(t_{ip}, t_{jp}))_{1 \leq i, j \leq p}$.

Tout d'abord on écrit

$$\mathbf{V} = \frac{1}{p} \begin{pmatrix} 1 & \cdots & \cdots & & 1 \\ & 2 & & \cdots & 2 \\ \vdots & & 3 & \cdots & \cdots & 3 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ & & \vdots & & \ddots & \\ 1 & 2 & 3 & \cdots & & p \end{pmatrix} .$$

On calcule sans difficulté l'inverse de \mathbf{V} en décomposant par exemple cette matrice comme le produit $p^{-1}\mathbf{T}\mathbf{T}'$ avec $\mathbf{T} = (\delta_{\{k \geq l\}})$ triangulaire inférieure. On note que le

Théorème 8.3.2 de Graybill (1969) permet aussi d'inverser la matrice \mathbf{V} dans le cas d'une répartition arbitraire des points t_{ip} . On obtient

$$\mathbf{V}^{-1} = p \begin{pmatrix} 2 & -1 & 0 & \cdots & 0 \\ -1 & 2 & \ddots & \ddots & \vdots \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 2 & -1 \\ 0 & \cdots & 0 & -1 & 1 \end{pmatrix}$$

La matrice \mathbf{V}^{-1} est tribande et ses valeurs propres sont relativement aisées à calculer. Supposons donnée une valeur propre $\lambda > 0$ de $p^{-1}\mathbf{V}^{-1}$ ($\lambda \leq 3$ par le théorème des disques de Gersgorin). Soit $(x_1, \dots, x_p)'$ un vecteur propre associé. Nous allons trouver des conditions nécessaires sur les coefficients x_1, \dots, x_p qui permettront de d'exhiber toutes les valeurs propres de \mathbf{V}^{-1} . En posant sans perte de généralité $x_0 = 0$ et $x_1 = 1$, on a nécessairement la relation de récurrence linéaire d'ordre 2

$$-x_k + 2x_{k+1} - x_{k+2} = \lambda x_{k+1}, \quad 0 \leq k \leq p - 1$$

avec comme condition finale $x_p = x_{p+1}$. L'équation caractéristique associée est

$$r^2 + (\lambda - 2)r + 1 = 0.$$

On fait alors les changements de variables $\lambda = 2a$ puis $1 - a = \cos t$ avec $t \in [0, \pi]$. L'équation devient alors

$$r^2 - 2r \cos t + 1 = 0,$$

de solutions complexes conjuguées e^{it} et e^{-it} . On a donc $x_k = \alpha e^{kit} + \beta e^{-kit}$ pour $k \geq 0$. Les conditions initiales $x_0 = 0$ et $x_1 = 1$ imposent que $x_0 = \alpha + \beta = 0$ puis $x_1 = \alpha(2i \sin t) = 1$. Il vient $x_k = \frac{\sin kt}{\sin t}$, $k \geq 0$. La condition finale $x_p = x_{p+1}$ est équivalente à $(p + 1)t \equiv \pi - pt \pmod{2\pi}$ car t est dans $[0, \pi]$. Il vient ainsi $t \equiv \frac{\pi}{2p+1} \left[\frac{2\pi}{2p+1} \right]$ donc les valeurs propres de $p\mathbf{V}^{-1}$ sont les $\lambda = 2a = 2(1 - \cos t) = 2 \left(1 - \cos \frac{(2k-1)\pi}{2p+1} \right)$, $1 \leq k \leq p$. On en déduit que les valeurs propres de \mathbf{V} sont

$$\lambda_{kp}(\mathbf{V}) = (2p)^{-1} \left(1 - \cos \frac{(2k-1)\pi}{2p+1} \right)^{-1}, \quad 1 \leq k \leq p. \quad (4.15)$$

On peut alors écrire quand $p \rightarrow \infty$, en utilisant le développement limité en zéro $\cos x = 1 - x^2/2 + o(x^2)$,

$$\lambda_{kp}(\mathbf{V}) = \|\mathbf{V}\| = \frac{4p}{(2k-1)^2\pi^2} (1 + o(1))$$

pour tout entier $k > 0$ fixé. Par ailleurs, $\lambda_{pp}^{-1} = \|\mathbf{V}^{-1}\| = 4p(1 + o(1))$. Ceci redémontre la Proposition 1 dans le cas particulier du processus de Wiener. On note que si les points d'observations t_{jp} vérifient $t_{jp} = \frac{j}{p} + o(\frac{1}{p})$, il est possible de montrer le même résultat asymptotique à l'aide de la théorie des perturbations.

Bibliographie

- Altman, N. S. (1990). Kernel smoothing of data with correlated errors. *J. Amer. Statist. Assoc.*, **85**(411), 749–759.
- Anilkumar, P. (1994). On estimating the mean function of a Gaussian process. *Statist. Probab. Lett.*, **19**, 77–84.
- Baladandayuthapani, V., Mallick, B. K., et Carroll, R. J. (2005). Spatially adaptive bayesian penalized regression splines (P-splines). *J. Comput. Graph. Statist.*, **14**(2), 378–394.
- Beder, J. H. (1987). A sieve estimator for the mean of a Gaussian process. *Ann. Statist.*, **15**(1), 59–78.
- Berger, J. et Wolpert, R. (1983). Estimating the mean function of a Gaussian process and the Stein effect. *J. Multivariate Anal.*, **13**(3), 401–424.
- Besse, P. (1991). Approximation spline de l’analyse en composantes principales d’une variable aléatoire hilbertienne. *Ann. Fac. Sci. Toulouse Math. (5)*, **12**(3), 329–349.
- Besse, P., Cardot, H., et Ferraty, F. (1997). Simultaneous nonparametric regressions of unbalanced longitudinal data. *Comput. Statist.*, **24**(3), 255–270.
- Besse, P., Cardot, H., Faivre, R., et Goulard, M. (2005). Statistical modelling of functional data. *Appl. Stochast. Models Busi. Ind.*, **21**, 165–173.
- Biritxinaga, E. (1987). *Estimation spline de la moyenne d’une fonction aléatoire*. Thèse d’État, Université de Pau. Directeur de thèse : Professeur Alain Pousse.
- Biritxinaga, E. (1990). Consistance forte de l’estimation de la moyenne d’un processus stochastique par interpolation spline. *Statistique et Analyse des Données*, **15**, 47–59.
- Bosq, D. et Nguyen, H. T. (1996). *A course in stochastic processes*. Theory and decision library. Series B : mathematical and statistical methods. Kluwer Academic Publishers, Dordrecht.

- Brockwell, P. J. et Davis, R. A. (1987). *Time series. Theory and methods*, volume 58 of *Springer series in statistics*. Springer-Verlag, New York.
- Brumback, B. et Rice, J. (1998). Smoothing spline models for the analysis of nested and crossed samples of curves. *J. Amer. Statist. Assoc.*, **93**(443), 961–976.
- Burman, P. (1991). Regression function estimation from dependent observations. *J. Multivariate Anal.*, **36**(2), 263–279.
- Byrd, R. H., Lu, P., Nocedal, J., et Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM J. Sc. Comput.*, **16**, 1190–1208.
- Cardot, H. (2000). Nonparametric estimation of smoothed principal components analysis of sampled noisy functions. *J. Nonparametr. Statist.*, **12**(4), 503–538.
- Cardot, H. et Diack, C. A. T. (1998). Convergence en moyenne quadratique de l'estimateur de la régression par splines hybrides. *C. R. Acad. Sci. Paris Sér. I Math.*, **326**(5), 615–618.
- Carmona, R. et Kôno, N. (1976). Convergence en loi et lois du logarithme itéré pour les vecteurs gaussiens. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, **36**(3), 241–267.
- Chatelin, F. (1983). *Spectral approximation of linear operators*. Computer Science and Applied Mathematics. Academic Press Inc., New York.
- Ciarlet, P. G. (1982). *Introduction à l'analyse numérique matricielle et à l'optimisation*. Collection Mathématiques Appliquées pour la Maîtrise. Masson, Paris.
- Cornillon, P. A. et Matzner-Løber, E. (2007). *Régression. Théorie et applications*. Collection Statistique et Probabilités Appliquées. Springer, France.
- Cox, D. (1983). Asymptotics for M -type smoothing splines. *Ann. Statist.*, **11**(2), 530–551.
- Cox, D. (1988). Approximation of method of regularization estimators. *Ann. Statist.*, **16**(2), 694–712.
- Crainiceanu, C. M., Ruppert, D., Carroll, R., Joshi, A., et Goodner, B. (2007). Spatially adaptive bayesian penalized splines with heteroscedastic errors. *J. Comput. Graph. Statist.*, **16**(2), 265–288.
- Craven, P. et Wahba, G. (1979). Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**(4), 377–403.

- Csörgő, S. et Mielniczuk, J. (1995). Nonparametric regression under long-range dependent normal errors. *Ann. Statist.*, **23**(3), 1000–1014.
- Cummins, D. J., Filloon, T. G., et Nychka, D. (2001). Confidence intervals for nonparametric curve estimates : toward more uniform pointwise coverage. *J. Amer. Statist. Assoc.*, **96**(453), 233–246.
- Dauxois, J., Pousse, A., et Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function : some applications to statistical inference. *J. Multivariate Anal.*, **12**(1), 136–154.
- Degras, D. et Jallet, R. (2005). Convergence de l'estimateur spline cubique de lissage dans un modèle de régression longitudinale avec erreur de type processus. *C. R. Math. Acad. Sci. Paris*, **340**(11), 851–854.
- Deheuvels, P., Del Barrio, E., et Van De Geer, S. (2007). *Lectures on Empirical Processes*. Series of Lectures in Mathematics. European Mathematical Society, New York.
- Diggle, P. J. et Hutchinson, M. F. (1989). On spline smoothing with autocorrelated errors. *Austral. J. Statist.*, **31**(1), 166–182.
- Donoho, D. L. et Johnstone, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, **90**(432), 1200–1224.
- Donoho, D. L. et Johnstone, I. M. (1998). Minimax estimation via wavelet shrinkage. *Ann. Statist.*, **26**(3), 879–921.
- Eilers, P. et Marx, B. (1996). Flexible smoothing with b-splines and penalties. *Statist. Sci.*, **89**, 89–121.
- Fan, J. et Gijbels, I. (1995). Data-driven bandwidth selection in local polynomial fitting : Variable bandwidth and spatial adaptation. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, **57**(2), 371–394.
- Ferraty, F. et Vieu, P. (2004). Nonparametric models for functional data with applications in regression, time series prediction and curve discrimination. *Nonparam. Stat.*, **16**, 111–125.
- Ferraty, F. et Vieu, P. (2006). *Nonparametric functional data analysis*. Springer Series in Statistics. Springer, New York.
- Friedman, J. et Silverman, B. W. (1989). Flexible parsimonious smoothing and additive modeling. *Technometrics*, **31**(1), 3–39.

- Gasser, T. et Müller, H.-G. (1979). Kernel estimation of regression functions. In *Smoothing techniques for curve estimation (Proc. Workshop, Heidelberg, 1979)*, volume 757 of *Lecture Notes in Math.*, pages 23–68. Springer, Berlin.
- Gijbels, I., Hall, P., et Kneip, A. (2004). Interval and band estimation for curves with jumps. *J. Appl. Probab.*, **41**, 65–79.
- Graybill, F. A. (1969). *Introduction to Matrices with Application in Statistics*. Wadsworth Publishing Company, Belmont, CA.
- Green, P. J. et Silverman, B. W. (1994). *Nonparametric regression and generalized linear models*, volume 58 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.
- Grenander, U. (1981). *Abstract inference*. Wiley, New York.
- Grizzle, J. E. et Allen, D. M. (1969). Analysis of growth and dose response curves. *Biometrics*, **25**, 357–381.
- Gu, C. (2002). *Smoothing spline ANOVA models*. Springer series in statistics. Springer-Verlag, New York.
- Györfi, L., Kohler, M., Krzyzak, A., et Walk (2002). *A distribution-free theory of nonparametric regression*. Springer-Verlag, New York.
- Hall, P. et Hart, J. D. (1990). Nonparametric regression with long-range dependence. *Stoch. Proc. Appl.*, **36**(2), 339–351.
- Härdle, W., Kerkycharian, G., Picard, D., et Tsybakov, A. (1998). *Wavelets, Approximation, and Statistical Applications*. Lecture Notes in Statistics. Springer-Verlag, New York.
- Hart, J. D. (1991). Kernel regression estimation with time series error. *J. Roy. Statist. Soc. Ser. B*, **53**(1), 173–187.
- Hart, J. D. (1996). Some automated methods of smoothing time-dependent data. *Nonparam. Stat.*, **6**, 115–142.
- Hart, J. D. et Wehrly, T. E. (1986). Kernel regression estimation using repeated measurements data. *J. Amer. Statist. Assoc.*, **81**(396), 1080–1088.
- Hastie, T. et Tibshirani, R. (1990). *Generalized additive models*. Chapman and Hall, London.
- Heckman, N. E. et Ramsay, J. O. (2000). Penalized regression with model-based penalties. *Canad. J. Statist.*, **28**(2), 24–258.

- Horn, R. A. et Johnson, C. R. (1990). *Matrix analysis*. Cambridge University Press, Cambridge. Corrected reprint of the 1985 original.
- Jallet, R. (2007). *Splines de régression et splines de lissage en régression non paramétrique avec bruit processus*. Thèse d'État, Université de Paris VI. Directeur de thèse : Professeur Daniel Pierre-Loti-Viaud.
- Johnstone, I. M. et Silverman, B. W. (1997). Wavelet threshold estimators for data with correlated noise. *J. Roy. Statist. Soc. Ser. B*, **59**(2), 319–351.
- Kadota, T. T. (1967). Term-by-term differentiability of mercer's expansion. *Proc. Amer. Math. Soc.*, **18**(1), 69–72.
- Kay, S. M. (1993). *Fundamentals of statistical signal processing : estimation theory*. Prentice-Hall Signal Processing Series. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.
- Kohn, R., Ansley, C. F., et H., W. C. (1992). Nonparametric spline regression with autoregressive moving average errors. *Biometrika*, **79**(2), 335–346.
- Kuelbs, J. (1976). The law of the iterated logarithm in $C[0, 1]$. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, **33**, 221–235.
- Künsch, H., Beran, J., et Hampel, F. (1993). Contrasts under long-range correlations. *Ann. Statist.*, **21**(2), 943–964.
- Ledoux, M. et Talagrand, M. (1991). *Probability in Banach spaces*. Lecture Notes in Statistics. Springer-Verlag, Berlin.
- Liebscher, E. (1999). Asymptotic normality of nonparametric estimators under α -mixing condition. *Statist. Probab. Lett.*, **43**(3), 243–250.
- Lin, X. et Zhang, D. (1999). Inference in generalized additive mixed models using smoothing splines. *J. Roy. Statist. Soc. Ser. B*, **61**, 381–400.
- Lin, X., Wang, N., Welsh, A. H., et Carroll, R. J. (2004). Equivalent kernels of smoothing splines in nonparametric regression for clustered/longitudinal data. *Biometrika*, **91**(1), 177–193.
- Loader, C. (1992). Simultaneous confidence bands for the rate of a nonhomogeneous Poisson process. Technical report, AT&T Bell Laboratories.
- Luo, Z. et Wahba, G. (1997). Hybrid adaptive splines. *J. Amer. Statist. Assoc.*, **92**(437), 107–116.

- Masry, E. (2005). Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stoch. Proc. Appl.*, **115**, 155–177.
- Messer, K. et Goldstein, L. (1993). A new class of kernels for nonparametric curve estimation. *Ann. Statist.*, **21**(1), 179–195.
- Müller, H. et Stadtmüller, U. (1987). Variable bandwidth kernel estimators of regression curves. *Ann. Statist.*, **15**, 182–201.
- Müller, H. G. (1988). *Nonparametric regression analysis of longitudinal data*. Lecture Notes in Statistics. Springer-Verlag, New York.
- Nussbaum, M. (1985). Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.*, **13**(3), 984–997.
- Nychka, D. (1988). Bayesian confidence intervals for smoothing splines. *J. Amer. Statist. Assoc.*, **83**(404), 1134–1143.
- Nychka, D. (1990). The average posterior variance of a smoothing spline and a consistent estimate of the average squared error. *Ann. Statist.*, **18**(1), 415–428.
- Nychka, D. (1995). Splines as local smoothers. *Ann. Statist.*, **23**(4), 1175–1197.
- Opsomer, J., Wang, Y., et Yang, Y. (2001). Nonparametric regression with correlated errors. *Statist. Sci.*, **16**(2), 134–153.
- Parzen, E. (1961). Regression analysis of continuous parameter time series. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, volume I of *Theory of Statistics*, pages 469–489. University of California Press, Berkeley, CA.
- Peele, L. et Kimeldorf, G. (1977). Prediction functions and mean-estimation functions for time series. *Ann. Statist.*, **5**(4), 709–721.
- Pintore, A., Speckman, P., et Holmes, C. (2006). Spatially adaptive smoothing splines. *Biometrika*, **93**(1), 113–125.
- Priestley, M. B. et Chao, M. T. (1972). Non-parametric function fitting. *J. Roy. Statist. Soc. Ser. B*, **34**, 385–392.
- R Development Core Team (2006). *R : A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
- Ragozin, D. L. (1983). Error bounds for derivative estimates based on spline smoothing of exact or noisy data. *J. Approx. Theory*, **37**(4), 335–355.

- Ramsay, J. O. et Silverman, B. W. (1997). *Functional data analysis*. Springer Series in Statistics. Springer, New York.
- Rice, J. A. et Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, **53**(1), 233–243.
- Roussas, G., Tran, L., et Ioannides, D. (1992). Fixed design regression for time series : asymptotic normality. *J. Multivariate Anal.*, **40**, 262–291.
- Roussas, G. G. et Ionides, D. (1987). Moment inequalities for mixing sequence of random variables. *Stochastic Anal. Appl.*, **5**(1), 61–120.
- Ruppert, D. et Carroll, R. (2000). Spatially-adaptive penalties for spline fitting. *Austr. New Z. J. Statist.*, **42**(2), 205–224.
- Sarda, P. et Vieu, P. (1986). Vitesse de convergence uniforme de l'estimateur à noyau de la régression pour des observations dépendantes. *C. R. Acad. Sci. Paris Sér. I Math.*, **302**(11), 423–426.
- Scheffé, H. (1959). *The Analysis of variance*. Wiley publications in statistics. Wiley, New York.
- Schoenberg, I. J. (1964a). Spline functions and the problem of graduation. *Proc. Nat. Acad. Sci. U.S.A.*, **52**, 947–950.
- Schoenberg, I. J. (1964b). Spline interpolation and the higher derivatives. *Proc. Nat. Acad. Sci. U.S.A.*, **51**, 24–28.
- Staniswalis, J. G. (1989). Local bandwidth selection for kernel estimates. *J. Amer. Statist. Assoc.*, **84**(405), 284–288.
- Stone, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.*, **10**(4), 1040–1053.
- Sun, J. et Loader, C. (1994). Simultaneous confidence bands for linear regression and smoothing. *Ann. Statist.*, **22**(3), 1328–1347.
- Truong-Van, B. et Bru, N. (2001). Asymptotic normality of spline estimator when the errors are a linear stationary process. *Nonparam. Stat.*, **13**(5), 741–761.
- Tsybakov, A. B. (2004). *Introduction à l'estimation non-paramétrique*, volume 41 of *Mathématiques & Applications*. Springer-Verlag, Berlin.
- Utreras, F. I. (1983). Natural spline functions, their associated eigenvalue problem. *Numer. Math.*, **42**(1), 107–117.

- Utreras, F. I. (1988). Convergence rates for multivariate smoothing spline functions. *J. Approx. Theory*, **52**(1), 1–27.
- Verbyla, A. P., Cullis, B. R., Kenward, M. G., et Welham, S. J. (1999). The analysis of designed experiments and longitudinal data by using smoothing splines (with discussion). *Appl. Statist.*, **48**(3), 269–311.
- Wahba, G. (1983). Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. Roy. Statist. Soc. Ser. B*, **45**(1), 133–150.
- Wahba, G. (1985). A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Statist.*, **13**(4), 1378–1402.
- Wahba, G. (1990). *Spline models for observational data*. Society for Industrial and Applied Mathematics.
- Wand, M. P. (1999). On the optimal amount of smoothing in penalised spline regression. *Biometrika*, **86**(4), 936–940.
- Wang, Y. (1996). Function estimation via wavelet shrinkage for long-memory data. *Ann. Statist.*, **24**(2), 466–484.
- Wang, Y. (1998). Smoothing spline models with correlated random errors. *J. Amer. Statist. Assoc.*, **93**(441), 341–348.
- Welsh, A. H., Lin, X., et Carroll, R. J. (2002). Marginal longitudinal nonparametric regression : locality and efficiency of spline and kernel methods. *J. Amer. Statist. Assoc.*, **97**(458), 482–493.
- Yang, Y. (1997). Nonparametric regression with dependent errors. *Bernoulli*, **7**(4), 633–655.
- Ycart, B., Woyczyński, W. A., Szulga, J., Reazor, S., et Mann, Jr., J. A. (1990). An interacting particle model of adsorption. *Zastos. Mat.*, **20**(3), 405–422.
- Zhang, D., Lin, X., Raz, J., et Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Amer. Statist. Assoc.*, **93**(442), 710–719.
- Zhou, S. et Shen, X. (2001). Spatially adaptive regression splines and accurate knot selection schemes. *J. Amer. Statist. Assoc.*, **96**(453), 247–259.
- Zhou, S., Shen, X., et Wolfe, D. (1996). Local asymptotics for regression splines and confidence regions. *Ann. Statist.*, **26**(5), 1760–1782.