



HAL
open science

Estimation non-paramétrique de la densité de variables aléatoires cachées

Charlotte Dion

► **To cite this version:**

Charlotte Dion. Estimation non-paramétrique de la densité de variables aléatoires cachées. Variables complexes [math.CV]. Université Grenoble Alpes, 2016. Français. NNT: 2016GREAM031 . tel-01685528

HAL Id: tel-01685528

<https://theses.hal.science/tel-01685528>

Submitted on 16 Jan 2018

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

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Mathématiques Appliquées**

Arrêté ministériel : 7 aout 2006

Présentée par

Charlotte Dion

Thèse dirigée par **Adeline Samson Leclercq**
et codirigée par **Fabienne Comte**

préparée au sein du **Laboratoire Jean Kuntzmann**
et de **École Doctorale Mathématiques, Sciences et Technologies de l'Information, Informatique**

Estimation non-paramétrique de la densité de variables aléatoires cachées

Thèse soutenue publiquement le **24 juin 2016**
devant le jury composé de :

M. Pascal Massart

Professeur, Université Paris-Sud, Président

M. Arnak Dalalyan

Professeur, ENSAE/CREST, Rapporteur

M. Reinhard Höpfner

Professeur, Université de Mainz, Rapporteur

M. Anatoli Juditsky

Professeur, Université Grenoble Alpes, Examinateur

Mme Clémentine Prieur

Professeure, Université Grenoble Alpes, Examinatrice

Mme Fabienne Comte

Professeure, Université Paris Descartes, Directrice de thèse

Mme Adeline Samson Leclercq

Professeure, Université Grenoble Alpes, Directrice de thèse



Estimation non-paramétrique de la densité de variables aléatoires cachées

Résumé : Cette thèse comporte plusieurs procédures d'estimation non-paramétrique de densité de probabilité. Dans chaque cas les variables d'intérêt ne sont pas observées directement, ce qui est une difficulté majeure. La première partie traite un modèle linéaire mixte où des observations répétées sont disponibles. La deuxième partie s'intéresse aux modèles d'équations différentielles stochastiques à effets aléatoires. Plusieurs trajectoires sont observées en temps continu sur un intervalle de temps commun. La troisième partie se place dans un contexte de bruit multiplicatif. Les différentes parties de cette thèse sont reliées par un contexte commun de problème inverse et par une problématique commune : l'estimation de la densité d'une variable cachée. Dans les deux premières parties la densité d'un ou plusieurs effets aléatoires est estimée. Dans la troisième partie il s'agit de reconstruire la densité de la variable d'origine à partir d'observations bruitées. Différentes méthodes d'estimation globale sont utilisées pour construire des estimateurs performants : estimateurs à noyau, estimateurs par projection ou estimateurs construits par déconvolution. La sélection de paramètres mène à des estimateurs adaptatifs et les risques quadratiques intégrés sont majorés grâce à une inégalité de concentration de Talagrand. Une étude sur simulations de chaque estimateur illustre leurs performances. Un jeu de données neuronales est étudié grâce aux procédures mises en place pour les équations différentielles stochastiques.

Mots clés : Estimation non-paramétrique. Estimateur à noyau. Méthode de sélection. Équations différentielles stochastiques. Effets aléatoires.

Abstract: This thesis contains several nonparametric estimation procedures of a probability density function. In each case, the main difficulty lies in the fact that the variables of interest are not directly observed. The first part deals with a mixed linear model for which repeated observations are available. The second part focuses on stochastic differential equations with random effects. Many trajectories are observed continuously on the same time interval. The third part is in a full multiplicative noise framework. The parts of the thesis are connected by the same context of inverse problems and by a common problematic: the estimation of the density function of a hidden variable. In the first two parts the density of one or two random effects is estimated. In the third part the goal is to rebuild the density of the original variable from the noisy observations. Different global methods are used and lead to well competitive estimators: kernel estimators, projection estimators or estimators built from deconvolution. Parameter selection gives adaptive estimators and the integrated risks are bounded using a Talagrand concentration inequality. A simulation study for each proposed estimator highlights their performances. A neuronal dataset is investigated with the new procedures for stochastic differential equations developed in this work.

Keywords: Nonparametric estimation. Kernel estimator. Selection method. Stochastic differential equations. Random effects.

Remerciements

Tout d'abord, je tiens à exprimer ma profonde gratitude envers mes directrices de thèse Fabienne Comte et Adeline Samson. L'aboutissement de cette thèse résulte de vos grandes qualités scientifiques et humaines. Depuis notre entretien pour un stage de master 2 vous avez toujours été présentes. Merci de m'avoir permis de vivre cette merveilleuse expérience qu'a été la thèse en me faisant confiance. J'ai tout appris auprès de vous et je serai toujours à l'écoute de vos conseils. Merci également pour votre patience et votre optimisme moteur au cours de ces trois années. Enfin, merci pour votre soutien infaillible lors de la recherche d'un contrat post-doctoral. Je voudrais de plus remercier Valentine Genon-Catalot avec qui j'ai été honorée de travailler. Notre collaboration a été très importante pour moi. J'espère avoir un peu appris de sa précision et sa rigueur mathématique. Merci également de m'avoir fait découvrir le réseau Dynstoch dont j'espère pouvoir continuer à suivre les rencontres dans les années à venir. Enfin j'aimerais remercier Simone Hermann ma collaboratrice dans la réalisation du package R.

Je tiens à remercier Arnak Dalalyan et Reinhard Höpfner pour l'intérêt qu'ils ont porté à mon travail en acceptant de rapporter cette thèse, ainsi que Anatoli Juditsky, Pascal Massart et Clémentine Prieur de me faire l'honneur de participer à mon jury de soutenance.

Puis je suis très reconnaissante envers mes professeurs d'Orsay qui ont su me guider dans mes choix. Tout d'abord Dominique Hulin lors des cours d'analyse pour l'agrégation et Grégory Miermont lors des cours de probabilité. Mon arrivée à Orsay a été très agréable grâce à vous. Merci de m'avoir soutenue même dans les moments les plus difficiles. Enfin merci à Christophe Giraud et Pascal Massart directeurs du master Mathématiques pour les Sciences du Vivant. Cette année a été décisive pour moi. Merci de m'avoir encouragée à continuer, encore actuellement.

Merci au directeur du Laboratoire Jean Kuntzmann (LJK) Eric Bonnetier pour son accueil. Merci à l'ex-directrice du laboratoire de Mathématiques Appliquées de Paris 5 (MAP5) Annie Raoult qui m'a permis d'avoir un bureau très agréable au MAP5. Je voudrais également remercier chaleureusement l'équipe des administrateurs, secrétaires et informaticiens pour avoir facilité mon quotidien. En particulier au MAP5 : merci à Marie-Hélène Gbaguidi pour ses réponses à toutes nos questions et pour son sourire quotidien, merci à Isabelle Valéro, Christophe Castellani, Marie Marduel. Merci à Azedine Mani et Arnaud Meunier pour tous les debugages ! Au LJK merci à Patrice Navarro, Frédéric Audra pour leurs conseils en informatique et leur réactivité inégalable, merci également à Catherine Laiolo et Juana Dos-Santos. Merci également aux membres permanents de l'équipe proba-stat du MAP5 et de l'équipe de statistiques de LJK.

J'aimerais remercier l'équipe de l'IUT Paris-Descartes avec laquelle j'ai passé 3 années lors de mon monitorat. Merci à François-Xavier Jollois puis Florence Muri de m'avoir intégrée à l'équipe. Un grand merci à Mohamed Mellouk pour m'avoir permis de donner le TD de probabilité associé à son cours, avec une liberté dans les choix d'exercices et de la notation des élèves.

Merci au groupe jeune de la SFDS de m'avoir fait découvrir un autre visage de la recherche, en particulier Thomas, Vincent, Maud et Benjamin.

Merci à tous les doctorants, ATER, post-doc, qui contribuent chaque jour à la bonne ambiance des laboratoires ! Au MAP5 : Alkeos, Loïc, Fabien, Maud, Jean, Gwen, Kevin, Anne-Claire, Julie, Thomas B, Thomas P, Pierre-Alexandre, Rémy, Pierre, Sonia, Ronan, Mélina, Laurent, Alasdair, Léo, Rébecca, Samuel. Merci pour les bonnes soirées (karaoké ?!) passées

ensemble, mais aussi plus simplement pour les pauses café et les déjeuners animés. Pierre R, on peut toujours compter sur toi, en toutes situations! Merci aussi d'avoir repris le GTTJD avec enthousiasme. Merci à Caroline P de me faire découvrir la musique de chambre; j'espère que l'on continuera nos déjeuners parisiens. Merci à Angelina d'être toujours disponible pour un café-discussion. Merci à Gaëlle pour sa gentillesse, ses conseils. Merci également à Christophe D.

Un merci spécial à Arthur, Charlotte L, Claire et Anne-So. Arthur merci pour toutes ces discussions sur le piano, les maths, la vie... Claire merci pour ta douceur et ton oreille si attentive. Anne-So merci pour d'être là, et de me suivre dans les expositions que j'ai envie de voir! Charlotte merci pour le réconfort que tu m'as apporté cette année. Votre amitié m'est très précieuse! Puis merci à Christèle E pour ces moments de rire, de danse, d'escapades, depuis le master 2 et tous ceux à venir!

Au LJK merci à Matthias, Romain, Nelson, Adrien, Cécile. Merci à Mélisande de m'avoir accueillie sans hésiter. Ester, merci pour ton amitié. Tu es un exemple par ton parcours brillant et ton soutien est une force pour moi. Enfin, merci à ma merveilleuse colocataire Chloé M. de partager mon quotidien. De Grenoble à Paris, what's next?

Je dois également beaucoup à mes amis de plus longue date sans qui la vie serait beaucoup moins intéressante! Mes amis de Cognac : Flo, quand je repense à tout ce chemin parcouru ensemble je suis vraiment heureuse que nous soyons toujours si proches. Merci à Polo pour son amitié que notre voyage au États-unis au lycée a soudé pour toujours! Merci à Marine et Charlotte T que j'ai toujours plaisir à voir et avec qui il est toujours tellement simple de se comprendre.

Merci à mes amis de Bordeaux : Marion, Gilles, Thibault. Marion, merci de venir me voir de temps en temps de Libourne en laissant pour un weekend ta petite famille si attachante, et merci pour toutes ces heures au téléphone depuis Ca-Ju! Gilles, inséparables/insupportables en classes prépa on se retrouve en master après quelques années à quelques salles d'écart. Que de souvenirs partagés, les cours, la « sonatina », puis Paris... Merci pour ton soutien. Enfin merci à Thibault pour tous ces moments de rire si importants, la plage, le ski, les férias!

Enfin mes amis de Paris. Il paraît que Paris est une ville où les gens sont désagréables, fermés et n'aiment pas bavarder. Allez dire ça à Christèle B et Céline! Cette année avec vous sur les bancs de l'université d'Orsay a été inoubliable. Merci pour tous ces moments ensemble! Merci aussi de m'avoir présentée Caroline L qui est devenue une de mes amies les plus chères. Merci à Chloé A avec qui l'année de l'agrégation fut (presque) une bonne année! J'espère que tu m'emmèneras bientôt sur ton bateau. Merci à Aurora pour tout ce qu'elle m'apporte. J'espère que l'on va continuer les visites ensemble, en France et ailleurs!

Mes derniers remerciements vont à ma famille. Merci à Agnès et François pour tous ces bons moments passés ensemble, souvent aux quatre coins du monde. Merci à Jacqueline et Francis avec qui je partage les meilleurs souvenirs de mon enfance. Merci à mon oncle Bernard d'être toujours présent et confiant en l'avenir. J'ai une pensée émue pour mes grands parents et ma grande tante. En particulier pour ma grand-mère Raymonde avec qui j'aurais tant aimé partager ce moment et beaucoup d'autres. Enfin, merci à mes parents. Il n'y a pas de mot assez fort pour vous exprimer ma gratitude et mon amour. Papa, merci d'avoir partagé avec moi ton goût pour les mathématiques; malgré mes efforts pour ne pas en faire trop au lycée, j'ai finalement suivi tes conseils. Maman, je suis tellement admirative et touchée que tu aies relu mon introduction, les étranges sigles qui suivent sont aussi les tiens!

Enfin merci à toi Pierre, tu emplis de joie mon quotidien un peu plus chaque jour, année après année.

Table des matières

1	Introduction	11
1.1	Notations	12
1.2	Estimation non-paramétrique	12
1.3	Modèles et approches	22
1.4	Étude de données neuronales	32
I	Modèle linéaire mixte	35
2	Nouvelles stratégies d'estimation non-paramétrique dans des modèles linéaires mixtes	37
2.1	Introduction	39
2.2	Construction of two estimators, risk bounds and adaptive results	40
2.3	Comparisons with previous results	46
2.4	Numerical illustration	47
2.5	Concluding remarks	52
2.6	Proofs	53
II	Modèles d'équations différentielles stochastiques à effets aléatoires	63
3	Estimation non-paramétrique dans un modèle d'Ornstein-Uhlenbeck à effets mixtes	65
3.1	Introduction	68
3.2	Presentation of the strategies	70
3.3	Study of the kernel estimator	72
3.4	Study of the deconvolution estimator	73
3.5	Simulation study	76
3.6	Application to neuronal data	82
3.7	Discussion	83
3.8	Proofs	84
	Appendix	90
4	Estimation bidimensionnelle dans un modèle d'équations différentielles stochastiques à deux effets aléatoires	93
4.1	Introduction	95
4.2	Notation and assumptions	96
4.3	Random effect estimation	98
4.4	Nonparametric estimation	100

4.5	Discrete data	103
4.6	Simulation study	104
4.7	Concluding remarks	110
4.8	Proofs	111
	Appendix 1	118
	Appendix 2: more about the MISE bounds	119
5	Estimation dans des modèles mixtes d'équations différentielles stochastiques avec le package R : mixedsde	123
5.1	Introduction	127
5.2	Density estimation in mixed stochastic differential models	128
5.3	Overview of the mixedsde functions	132
5.4	Package mixedsde through simulated examples	136
5.5	Package mixedsde through a real data example	142
	Appendix 1	148
	Appendix 2	148
III	Modèle de bruit multiplicatif	153
6	Estimation non-paramétrique dans un modèle de censure multiplicative avec un bruit symétrique	155
6.1	Introduction	157
6.2	Multiplicative denoising of density and survival function	158
6.3	Model transformation and deconvolution approach	164
6.4	Numerical study	166
6.5	Proofs	171
	Appendix: observations and outcomes	183
IV	Perspectives	189
7	Perspectives	191
7.1	Sur les modèles linéaires mixtes	192
7.2	Sur la modélisation neuronale	193
7.3	Sur le modèle de bruit multiplicatif	198
A	Liste de publications	203
B	Résultats d'analyse	205
	Bibliographie	209

Chapitre 1

Introduction

Sommaire

1.1	Notations	12
1.2	Estimation non-paramétrique	12
1.2.1	Méthode à noyau	12
1.2.2	Méthode par projection	13
1.2.3	Le principe de déconvolution	13
1.2.4	Qu'est-ce qui fait un bon estimateur ? Étude du risque quadratique intégré	15
1.2.5	Sélection de paramètre	18
1.3	Modèles et approches	22
1.3.1	Modèles linéaires mixtes	22
1.3.2	Modèles d'équations différentielles stochastiques à effets mixtes	25
1.3.3	Un modèle de censure	29
1.4	Étude de données neuronales	32

La densité de probabilité caractérise la répartition des données fournies par un échantillon. C'est une des premières quantités qu'un praticien peut étudier. Puis, l'estimation « non-paramétrique » d'une fonction fournit l'information nécessaire pour ensuite imposer un modèle paramétrique sur cette fonction si tel est le souhait du praticien. Cette approche statistique globale cherche à limiter le nombre d'hypothèses sur la forme éventuelle du modèle qui donne naissance aux données observées. Enfin, les méthodes développées pour l'estimation de densité peuvent se transposer pour l'estimation d'autres fonctions clés telles que la fonction de survie ou le taux de hasard.

Pour ces raisons nous nous intéressons dans ce travail à des problèmes d'estimation non-paramétrique, de fonctions densités. Nous nous concentrons sur la densité de variables cachées, c'est à dire qui ne sont pas observées directement. En effet, bien souvent, l'information principale est contenue dans ces variables. L'exemple de données contaminées par un bruit est le plus fréquent : l'enjeu du statisticien est de reconstruire l'information perdue. Nous proposons dans ce contexte de nouvelles procédures d'estimation globale et nous étudions les propriétés des estimateurs résultants. Cette introduction rappelle les principales méthodes d'estimation non-paramétrique de densité ainsi que les notions, les modèles et les résultats de cette thèse.

1.1 Notations

On note l'espace de Banach : $\mathbb{L}^p(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R}, \int_{\mathbb{R}} |f^p(x)| dx < \infty\}$ et la norme associée $\|f\|_p = (\int_{\mathbb{R}} |f^p(x)| dx)^{1/p}$, à l'exception de la norme de l'espace de Hilbert $\mathbb{L}^2(\mathbb{R})$ que l'on notera $\|f\|$ par simplicité. Pour deux fonctions f et g dans $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, le produit scalaire est défini par $\langle f, g \rangle = \int_{\mathbb{R}} f(x) \overline{g(x)} dx$. La transformée de Fourier de f est $f^*(x) = \int_{\mathbb{R}} e^{ixu} f(u) du$ pour tout $x \in \mathbb{R}$. Quand f est la densité d'une variable aléatoire X , $f^*(\cdot) = \mathbb{E}[e^{iX\cdot}]$ est appelée fonction caractéristique de X . Enfin le produit de convolution de f et g est pour tout $x \in \mathbb{R}$: $f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$. On rappelle le théorème de Plancherel-Parseval : pour toute fonction $f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $2\pi\|f\|^2 = \|f^*\|^2$. Enfin, si g est une fonction bornée sur \mathbb{R} on note : $\|g\|_{\infty} = \sup_{x \in \mathbb{R}} |g(x)|$.

1.2 Estimation non-paramétrique

Comment estimer la densité de probabilité f à partir de la connaissance d'une suite de variables aléatoires X_1, \dots, X_N indépendantes et identiquement distribuées (*i.i.d.*) issues de cette densité ? Lorsque le nombre d'observations N est assez grand, l'approche non-paramétrique peut être considérée. Dans cette section on se place dans le cadre où N observations directes sont disponibles, excepté dans le paragraphe 1.2.3. Des estimateurs classiques existent dans ce cas. Nous présentons ici seulement les méthodes à noyau et les méthodes par projection utilisées dans la suite de ce travail.

1.2.1 Méthode à noyau

Les estimateurs à noyau ont été introduits par Rosenblatt (1956) puis par Parzen (1962) pour estimer la fonction de répartition. Pour estimer une densité f sur \mathbb{R} , un estimateur à noyau (*kernel* en anglais) est donné, pour $h > 0$ et $x \in \mathbb{R}$, par :

$$\hat{f}_h(x) = \frac{1}{Nh} \sum_{j=1}^N K\left(\frac{x - X_j}{h}\right) = \frac{1}{N} \sum_{j=1}^N K_h(x - X_j) \quad (1.2.1)$$

où la fonction $K : \mathbb{R} \rightarrow \mathbb{R}$ est appelée noyau. Elle doit être intégrable telle que $\int_{\mathbb{R}} K(x)dx = 1$, et être positivement pour que l'estimateur soit une densité. Les noyaux usuels sont les noyaux gaussiens $K(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$ (spécialement pour l'estimation de densité), mais aussi le noyau d'Epanechnikov $K(x) = (3/4)(1 - x^2)\mathbb{1}_{-1 \leq x \leq 1}$ ou triangulaire $K(x) = (1 - |x|)\mathbb{1}_{|x| \leq 1}$. On note $K_h(\cdot) := K(\cdot/h)/h$ où le paramètre $h > 0$ est appelée fenêtre (*bandwidth*) du noyau, ou paramètre de lissage. Cet estimateur est un estimateur sans biais de la fonction $f_h = K_h \star f$, c'est à dire qu'il vérifie pour tout $x \in \mathbb{R}$, $\mathbb{E}[\hat{f}_h(x)] = f_h(x)$. De plus f_h converge vers f lorsque h tend vers 0 en norme \mathbb{L}^2 , ce qui justifie qu'en pratique on approche la fonction f par la fonction f_h .

1.2.2 Méthode par projection

Pour construire un estimateur par projection de f supposons que $f \in \mathbb{L}^2(\mathbb{R})$. Le principe est de développer f en séries de fonctions orthogonales et d'estimer les coefficients de ce développement à l'aide de moyennes empiriques. Soit $(\varphi_j)_{j \in \mathbb{N}}$ une base hilbertienne de $\mathbb{L}^2(\mathbb{R})$. Alors on définit la projection f_m de f sur le sous espace vectoriel engendré par les m premiers vecteurs de la base

$$\mathcal{S}_m = \text{Vect}\{\varphi_k, k = 1, \dots, m\} \quad (1.2.2)$$

avec la dimension $\dim \mathcal{S}_m = m$ (ou une fonction croissante de m), $f_m = \sum_{k=1}^m a_k \varphi_k$ et $a_k = \langle f, \varphi_k \rangle$. Les espaces \mathcal{S}_m sont dits emboîtés car $\mathcal{S}_m \subset \mathcal{S}_{m'} \Leftrightarrow m \leq m'$. Un estimateur par projection de f est donné par

$$\hat{f}_m = \sum_{k=1}^m \hat{a}_k \varphi_k \quad (1.2.3)$$

où \hat{a}_k est un estimateur de a_k . Dans le cadre d'observations directes on choisit les coefficients

$$\hat{a}_k = \frac{1}{N} \sum_{j=1}^N \varphi_k(X_j)$$

ainsi ils vérifient $\mathbb{E}[\hat{a}_k] = a_k$. L'estimateur \hat{f}_m est un estimateur sans biais du projeté f_m de f sur \mathcal{S}_m . L'estimation d'une fonction sur \mathbb{R} est alors ramenée à l'estimation de m coefficients réels. Plus la dimension m est grande plus le modèle est complexe. Il faudra optimiser le choix de m .

Nous avons présenté deux méthodes d'estimation de la densité f à partir d'un échantillon observé issu de cette loi. Cependant il est rare d'avoir des observations directes et c'est ce qui nous amène à décrire le cas d'observations bruitées dans le paragraphe suivant. Nous allons également préciser le lien entre la méthode à noyau et la méthode par projection.

1.2.3 Le principe de déconvolution

Les difficultés d'estimation viennent souvent du fait que la variable d'intérêt n'est pas la variable observée. Cela peut être dû par exemple à des erreurs de mesures lors d'une expérience. Le bruit qui masque l'échantillon d'origine est généralement modélisé de deux manières : par l'addition d'une variable indépendante (centrée et symétrique le plus souvent) ou par la multiplication par une variable indépendante. Le bruit multiplicatif est lié à la notion de censure de données que nous présentons dans le paragraphe 1.3.3. Le bruit additif conduit au problème dit « de déconvolution » et à des méthodes et des estimateurs spécifiques.

Notons les observations bruitées disponibles Y_j , telles que :

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, N$$

où les ε_j sont les variables aléatoires *i.i.d.* modélisant le bruit et sont indépendantes des X_j . Ici (et dès qu'il pourra y avoir une ambiguïté) nous précisons la notation : f_Z désigne la densité de la variable Z . Les observations $(Y_j)_j$ sont alors *i.i.d.* de densité f_Y . L'objectif est d'estimer la densité f_X des variables cachées X_j à partir des données Y_j , de façon non-paramétrique. La densité des observations s'écrit alors $f_Y = f_X \star f_\varepsilon$. En appliquant la transformée de Fourier à cette égalité, on obtient le produit simple $f_Y^* = f_X^* f_\varepsilon^*$. Si la densité du bruit ε est supposée connue et que f_ε^* ne s'annule pas, en estimant la fonction caractéristique f_Y^* on peut atteindre la fonction f_X^* par : $f_X^* = f_Y^* / f_\varepsilon^*$. Puis l'estimateur de f_X est construit en appliquant une formule d'inversion de Fourier. Cependant il reste à estimer la fonction caractéristique f_Y^* qui est inconnue. Un estimateur empirique est donné par :

$$\widehat{f}_Y^*(u) = \frac{1}{N} \sum_{j=1}^N e^{iuY_j}. \quad (1.2.4)$$

Mais, en substituant cet estimateur à la vraie fonction, l'intégrabilité du rapport $\widehat{f}_Y^* / f_\varepsilon^*$ n'est plus assurée. Une solution est d'utiliser un autre estimateur de la fonction caractéristique, noté $\widehat{f}_{Y,\alpha}^*$, pour obtenir un estimateur de la forme :

$$\widehat{f}_{X,\alpha}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{\widehat{f}_{Y,\alpha}^*(u)}{f_\varepsilon^*(u)} du. \quad (1.2.5)$$

Cet estimateur est obtenu en estimant la densité f_Y par un estimateur à noyau (Carroll & Hall, 1988; Fan, 1991; Meister & Neumann, 2010), ou par projection (Comte *et al.*, 2006; Comte & Lacour, 2011) puis en appliquant l'inversion de Fourier.

Une autre méthode consiste à inclure une fonction indicatrice dans l'intégrand, réduisant ainsi l'intervalle d'intégration sur un intervalle fermé. On dit que l'on introduit un *cut-off*. La collection d'estimateurs construite ainsi s'écrit :

$$\widehat{f}_{X,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{\widehat{f}_Y^*(u)}{f_\varepsilon^*(u)} du \quad (1.2.6)$$

avec par exemple $m = 1, \dots, N$. Mais cet estimateur est un estimateur par projection, en un certain sens. Soit la fonction sinus cardinal $\text{sinc}(x) = \frac{\sin \pi x}{\pi x}$, $\text{sinc}(0) = 1$ (cette fonction est dans $\mathbb{L}^2(\mathbb{R})$ mais pas dans $\mathbb{L}^1(\mathbb{R})$), à partir de cette fonction on construit la base : $\ell \in \mathbb{Z}$, $\phi_{m,\ell}(x) = \sqrt{m} \phi(mx - \ell)$ avec $\phi(x) = \frac{\sin \pi x}{\pi x}$. Le Théorème de Shannon (voir Leborgne, 2013) assure que c'est une base orthonormale de $\mathcal{S}_m = \{f \in \mathbb{L}^2(\mathbb{R}), \text{supp}(f^*) = [-\pi m, \pi m]\} = \text{Vect}(\phi_{m,\ell}, \ell \in \mathbb{Z})$ au sens $\mathbb{L}^2(\mathbb{R})$, où *supp* désigne le support.

Remarquons que cet espace \mathcal{S}_m est de dimension infinie ($m \neq \dim \mathcal{S}_m$). Nous notons la fonction $f_{X,m}$ définie par la formule $f_{X,m}^* := f_X^* \mathbf{1}_{[-\pi m, \pi m]}$, élément de \mathcal{S}_m et $\Pi_{\mathcal{S}_m}$ l'opérateur de projection de \mathcal{S}_m . Alors nous avons pour tout $x \in \mathbb{R}$

$$f_{X,m}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \mathbf{1}_{[-\pi m, \pi m]}(u) f_X^*(u) du = \sum_{\ell \in \mathbb{Z}} \langle f_X, \phi_{m,\ell} \rangle \phi_{m,\ell}(x) := \Pi_{\mathcal{S}_m}(f_X)(x).$$

Nous appelons les coefficients $a_{m,\ell} := \langle f_X, \phi_{m,\ell} \rangle$ donnés par

$$\langle f_X, \phi_{m,\ell} \rangle = \frac{1}{2\pi} \langle f_X^*, \phi_{m,\ell}^* \rangle = \frac{1}{2\pi} \left\langle \frac{f_Y^*}{f_\varepsilon^*}, \phi_{m,\ell}^* \right\rangle = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{e^{-(i\ell/m)u} f_Y^*(u)}{\sqrt{m} f_\varepsilon^*(u)} du.$$

Puis on définit l'estimateur par projection $\widehat{f}_{X,m}^P$ par $\widehat{f}_{X,m}^P = \sum_{\ell \in \mathbb{Z}} \widehat{a}_{m,\ell}^P \phi_{m,\ell}$ avec

$$\widehat{a}_{m,\ell}^P = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{e^{-(i\ell/m)u} \widehat{f}_Y^*(u)}{\sqrt{m} f_\varepsilon^*(u)} du$$

où \widehat{f}_Y^* est donné par la formule (1.2.4). Il reste à vérifier que $\widehat{f}_{X,m}^P = \widehat{f}_{X,m}$, en montrant que $\widehat{a}_{m,\ell}^P = \langle \widehat{f}_{X,m}, \phi_{m,\ell} \rangle$. En effet par la formule d'inversion de Fourier on trouve :

$$\langle \widehat{f}_{X,m}, \phi_{m,\ell} \rangle = \frac{1}{2\pi} \langle \widehat{f}_{X,m}^*, \phi_{m,\ell}^* \rangle = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{e^{-(i\ell/m)u} \widehat{f}_Y^*(u)}{\sqrt{m} f_\varepsilon^*(u)} du = \widehat{a}_{m,\ell}^P$$

ce qui donne l'égalité annoncée.

Mais l'estimateur (1.2.6) est aussi un estimateur à noyau. Posons : $m = 1/h, h > 0, 1 \leq m \leq N$

$$\widehat{f}_{X,m}(x) = \frac{1}{2\pi} \int \mathbf{1}_{[-\pi,\pi]}(uh) e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuY_j}}{f_\varepsilon^*(u)} du$$

puis $K^*(u) := \mathbf{1}_{[-\pi,\pi]}(u)$, $K(x) = \frac{\sin(\pi x)}{\pi x}$, $K_h^*(u) := K^*(uh)$; et $\widetilde{K}_h^*(u) := \frac{K^*(uh)}{f_\varepsilon^*(u)}$. Alors il vient :

$$\widehat{f}_{X,m}(x) = \frac{1}{2\pi} \frac{1}{N} \sum_{j=1}^N \int e^{-iu(x-Y_j)} \widetilde{K}_h^*(u) du = \frac{1}{N} \sum_{j=1}^N \widetilde{K}_h(x - Y_j)$$

Nous avons montré ici le fort lien qui existe entre noyau et projection. Ce lien sera renforcé dans la Section 1.2.4.

Cas de la loi du bruit inconnue

Il est possible de construire des estimateurs lorsque la loi du bruit est inconnue. Il faut alors estimer la fonction caractéristique du bruit f_ε^* ou dans certains cas une transformation de celle ci. Une méthode consiste à supposer que l'on dispose d'un échantillon du bruit seul (c'est le cas par exemple dans Comte & Lacour, 2011). En pratique cela correspond par exemple à l'étude de sorties d'une machine, à vide dans un premier temps. On possède alors un échantillon de l'erreur de mesure seule. D'autre part, on peut disposer de données répétées : pour une observation (un individu par exemple) j on a \mathcal{K} mesures bruitées pour $k = 1, \dots, \mathcal{K}$, $Y_{j,k} = X_j + \varepsilon_{j,k}$. Alors, une transformation des données peut permettre d'estimer la quantité voulue.

Cependant, cette quantité (par exemple f_ε^* ou $|f_\varepsilon^*|^2$) apparaît toujours au dénominateur de notre estimateur (1.2.6), il faudra donc inverser l'estimateur obtenu. Pour s'assurer que le dénominateur ne s'approche pas trop de 0, on fait appel à la technique de troncature due à Neumann (1997) (voir par exemple Comte & Lacour (2011), Comte & Samson (2012), Mabon (2015a) et le Chapitre 2).

1.2.4 Qu'est-ce qui fait un bon estimateur? Étude du risque quadratique intégré

Afin de juger la qualité d'un estimateur il faut choisir une fonction de perte et définir le risque associé. Pour les procédures d'estimation globales étudiées dans ce travail nous nous intéressons au risque quadratique intégré (MISE, pour *mean integrated squared error*). Pour cela la fonction

que l'on cherche à estimer est supposée de carré intégrable. La fonction de perte est la distance \mathbb{L}^2 et le MISE en est alors l'espérance. Pour un estimateur \widehat{f}_α de f , avec α dans la collection \mathcal{H} , le risque s'écrit $\mathbb{E}[\|\widehat{f}_\alpha - f\|^2]$. Il fournit une mesure moyenne de la distance entre l'estimateur et la vraie fonction.

Décomposition biais-variance

Pour calculer ce risque on le décompose en un terme qui mesure la distance entre la moyenne de l'estimateur et la vraie fonction, et de la variance intégrée de l'estimateur :

$$\mathbb{E}[\|\widehat{f}_\alpha - f\|^2] = \|\mathbb{E}[\widehat{f}_\alpha] - f\|^2 + \mathbb{E}[\|\mathbb{E}[\widehat{f}_\alpha] - \widehat{f}_\alpha\|^2]. \quad (1.2.7)$$

Le premier terme est appelé terme de biais et le second terme de variance. En calculant chaque terme, on obtient une borne supérieure de ce risque dépendant de N et de α . Le résultat est non-asymptotique, puisque valable pour toute valeur de N . Il est d'autant meilleur que N est grand.

L'estimateur à noyau dans le cadre des observations directes décrit Section 1.2.1, vérifie $\mathbb{E}[\widehat{f}_h] = f_h$, et si le noyau K est de carré intégrable, la borne obtenue est :

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq \|f_h - f\|^2 + \frac{\|K\|^2}{Nh}. \quad (1.2.8)$$

L'estimateur par projection dans le cadre des observations directes, décrit Section 1.2.2, vérifie $\mathbb{E}[\widehat{f}_m] = f_m$ et la borne obtenue est :

$$\mathbb{E}[\|\widehat{f}_m - f\|^2] \leq \|f_m - f\|^2 + \frac{C_\varphi m}{N} \quad (1.2.9)$$

si la base $(\varphi_k)_k$ choisie est telle que les vecteurs de base sont bornés : $\|\varphi_k\|_\infty \leq C_\varphi$ pour tout k , avec $C_\varphi \in \mathbb{R}_+^*$.

On peut faire ici une analogie entre les deux estimateurs grâce aux bornes de risques obtenues. En effet, dans la borne (1.2.8) on voit que le terme de biais décroît quand la fenêtre h tend vers 0 alors que dans ce cas le terme de variance croît. Dans la borne (1.2.9) on voit que le terme de biais décroît quand la dimension m tend vers l'infini alors que dans ce cas le terme de variance croît. Les paramètres h et m ont donc un rôle similaire.

Comment estimer l'ordre du terme de biais ?

Nous étudions l'estimation de fonctions densités, donc de fonctions intégrables. Les espaces intéressants pour cette étude sont les espaces de Sobolev et les espaces de Nikol'skii, qui sont des cas particuliers d'espaces plus généraux de Besov que nous ne présentons pas dans ce travail, (voir DeVore & Lorentz, 1993, pour les inclusions d'espaces). Supposer que la densité recherchée f appartient à un de ces espaces permet d'obtenir l'ordre du biais.

Tout d'abord, on s'intéresse aux classes de Hölder, définies par

$$\Sigma_I(\beta, L) = \{f : I \rightarrow \mathbb{R}, f^{[\beta]} \text{ existe}, \forall x, y \in I, |f^{[\beta]}(x) - f^{[\beta]}(y)| \leq L|x - y|^{\beta - [\beta]}\}$$

où $[\beta]$ est le plus grand entier plus petit que β . Le paramètre β est appelé indice de régularité de la fonction (voir Tsybakov, 2009), et $I = \mathbb{R}^+, \mathbb{R}$ le plus souvent. Par exemple, pour évaluer l'ordre du biais pour un estimateur à noyau \widehat{f}_h donné par l'équation (1.2.1), ponctuellement, c'est à dire $|f_h(x_0) - f(x_0)|$, on suppose que $f \in \Sigma_{\mathbb{R}}(\beta, L)$ mais alors il faut supposer que le noyau K est « d'ordre » $\ell = [\beta]$.

Définition 1.2.1. Soit $\ell \in \mathbb{N}^*$. On dit que $K : \mathbb{R} \rightarrow \mathbb{R}$ est un noyau d'ordre ℓ si les fonctions $u \mapsto u^j K(u)$ sont intégrables pour $j = 0, \dots, \ell$ et telles que : $\int_{\mathbb{R}} K(u) du = 1$, $\int_{\mathbb{R}} u^j K(u) du = 0$, pour $j = 1, \dots, \ell$.

Dans ces conditions il vient pour tout $x_0 \in \mathbb{R}$, $|f_h(x_0) - f(x_0)| \leq Ch^\beta$ (où C est une constante qui dépend de K et de β , voir Tsybakov (2009)).

Puis, plus généralement les classes de Nikol'skii sont définies par

$$\mathcal{N}(\beta, R) := \left\{ f \in \mathbb{L}^2(\mathbb{R}), \forall y \in \mathbb{R}, \left(\int |f^{\lfloor \beta \rfloor}(x+y) - f^{\lfloor \beta \rfloor}(x)|^2 dx \right)^{1/2} \leq R|y|^{\beta - \lfloor \beta \rfloor}, \quad \|f^{\lfloor \beta \rfloor}\| \leq R \right\}$$

(voir Nikol'skii, 1975). Elles sont adaptées aux estimateurs à noyau pour obtenir une majoration du biais au carré intégré. En effet si $f \in \mathcal{N}(\beta, R)$ et que K est un noyau d'ordre $\ell = \lfloor \beta \rfloor$, il vient : $\|\widehat{f}_h - f\| \leq Ch^\beta$. On réfère à Goldenshluger & Lepski (2011) et Comte & Lacour (2013) pour d'autres utilisations dans le cas multidimensionnel.

Enfin, les classes des Sobolev sont très utilisées. Nous nous intéressons à la définition des espaces de Sobolev en fonction de la transformée de Fourier des fonctions, où b n'est pas nécessairement entier,

$$H^b(\mathbb{R}) = W_2^b(\mathbb{R}) := \{f \in \mathbb{L}^2(\mathbb{R}), \int |f^*(x)|^2 (1+x^2)^b dx < \infty\}.$$

Rappelons que si $b > 1/2$ et k le plus grand entier tel que $k < b - 1/2$ alors $f \in H^b(\mathbb{R})$ est de classe \mathcal{C}^k (voir Lerner (2008)). Les deux classes classiquement utilisées pour qualifier la régularité des fonctions d'intérêt, sont :

- *Ordinary smooth* (OS) : cela s'écrit comme l'appartenance à une boule de Sobolev de $H^b(\mathbb{R})$ de régularité b de rayon L , pour la norme sous-jacente :

$$\mathcal{A}_b(L) := \{f \in \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1+x^2)^b dx \leq L\}$$

avec $b > 0$, $L > 0$. Par exemple les densités gamma, uniforme, Laplace sont dans un $\mathcal{A}_b(L)$. Si $b > 1/2$ les densités contenues dans $\mathcal{A}_b(L)$ sont continues et bornées.

- *Super smooth* (SS) : la fonction d'intérêt est dans l'espace suivant :

$$\mathcal{A}_{r,a} := \{f \in \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 \exp(2a|x|^r) dx \leq L\}.$$

Les fonctions de cette classe sont en particulier \mathcal{C}^∞ . Par exemple les densités Cauchy, gaussiennes ou les mélange de gaussiennes, appartiennent à un $\mathcal{A}_{r,a}$.

Ces espaces sont utilisés en déconvolution dans Butucea (2004) ou Lacour (2006). Par exemple, l'estimateur $\widehat{f}_{X,m}$ donné par l'équation (1.2.6) est construit comme un estimateur sans biais de $f_{X,m}$, si $f_X \in \mathcal{A}_b(L)$ il vient $\|f_{X,m} - f_X\| \leq Cm^{-b}$.

Enfin, on peut définir l'analogie des espaces de Sobolev dans $\mathbb{L}^2(\mathbb{R}^+)$. Dans ce cas, on considère une base sur cet espace, qui n'est plus la base trigonométrique mais la base de Laguerre $(\varphi_k)_k$ (définie dans la Partie 6). La construction et l'étude de l'espace suivant se trouvent dans Bongioanni & Torrea (2009). Une boule de Sobolev-Laguerre de régularité s de rayon L est définie par

$$W^s(\mathbb{R}^+, L) := \{f \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s \langle f, \varphi_k \rangle^2 \leq L < +\infty\},$$

avec $s > 0$. Comte & Genon-Catalot (2015) donnent un équivalent sur la régularité de la fonction f lorsque s est entier : si f est dans $\mathbb{L}^2(\mathbb{R}^+)$ alors cela équivaut à dire que f admet des dérivées

jusqu'à l'ordre $s - 1$ et $f^{(s-1)}$ est absolument continue et $\delta^m \circ \delta^{m-1} \circ \dots \circ \delta^1 \circ \delta^0 f \in \mathbb{L}^2(\mathbb{R}^+)$ avec $\delta^k f(x) = \sqrt{x} f'(x) + \frac{1}{2} \left(\sqrt{x} - \frac{k}{\sqrt{x}} f(x) \right)$. Dans ce cas, lorsque l'on estime une densité $f \in \mathbb{L}^2(\mathbb{R}^+)$ par un estimateur par projection \hat{f}_m , en supposant de plus que $f \in W^s(\mathbb{R}^+, L)$ on obtient : $\|f - f_m\|^2 \leq m^{-s}$.

Il existe d'autres espaces associés à d'autres bases, par exemple les bases d'histogrammes ou de polynômes par morceaux sont associées à des espaces de Besov.

Les vitesses classiques

Une borne supérieure d'un risque \mathbb{L}^2 montre qu'il tend vers 0 lorsque N devient grand, on parle d'estimateur convergent. La vitesse de convergence ψ_N est la suite, décroissante avec N , telle que : $\mathbb{E}[\|\hat{f}_\alpha - f\|^2] \leq C\psi_N$ avec C une constante indépendante de N . La suite $(\psi_N)_N$ est dite « optimale au sens minimax » (ou vitesse minimax) si elle satisfait également la borne inférieure :

$$\inf_{\hat{f}} \sup_{\alpha \in \mathcal{H}} \mathbb{E}[\|\hat{f}_\alpha - f\|^2] \geq C'\psi_N.$$

Lorsque l'on a obtenu l'ordre du biais en fonction d'un paramètre de régularité de la fonction f , le choix du paramètre optimal α_{opt} qui minimise la borne du risque se déduit aisément. En effet le caractère non-asymptotique des bornes des MISE obtenues permettent de choisir α en fonction de N . Alors, on trouve la vitesse en fonction des paramètres de régularité. Par exemple pour l'estimateur à noyau de $f : \hat{f}_h$ donné par (1.2.1), si $f \in \mathcal{N}(\beta, R)$, et que K est un noyau d'ordre $\ell = \lfloor \beta \rfloor$, on minimise en h la fonction convexe $h \mapsto C_1 h^{2\beta} + C_2/(Nh)$. Il vient $h_{\text{opt}} = CN^{-1/(2\beta+1)}$ et la vitesse de convergence est en $N^{-2\beta/(2\beta+1)}$.

Pour l'estimateur par projection de $f : \hat{f}_m$ donné par (1.2.3), si $f \in W^s(\mathbb{R}^+, L)$, on minimise en m la fonction convexe $m \mapsto C'_1 m^{-s} + C'_2 m/N$. Il vient $m_{\text{opt}} = C'N^{1/(1+s)}$ et la vitesse de convergence est la vitesse de convergence est $N^{-s/(s+1)}$. Dans les deux cas le choix de α (h ou m) dépend de N et du paramètre de régularité de f .

Pour les méthodes de déconvolution les vitesses classiques sont répertoriées dans Lacour (2006). Elles font intervenir en plus un paramètre de régularité pour la densité du bruit, due à la forme de l'estimateur $\hat{f}_{X,\alpha}$ donnée par (1.2.6). Les régularités utilisées pour la densité du bruit f_ε sont données par :

$$k_0(x^2 + 1)^{-\gamma/2} \exp(-b|x|^s) \leq |f_\varepsilon^*(x)| \leq k_1(x^2 + 1)^{-\gamma/2} \exp(-b|x|^s)$$

avec $s \geq 0$, $b \geq 0$, $\gamma \in \mathbb{R}$ ($\gamma > 0$ si $s = 0$ et $k_0, k_1 > 0$). La densité f_ε est dite *super smooth* (SS) si $s > 0$ et *ordinary smooth* (OS) sinon. Lorsque f_X et f_ε sont OS, les vitesses sont alors dégradées du paramètre de régularité γ de f_ε on obtient la vitesse : $N^{-2b/(2b+2\gamma+1)}$. Lorsque le bruit est SS et que f_X est seulement OS on obtient des vitesses logarithmiques mauvaises. Ces vitesses sont optimales (voir Fan, 1991; Butucea, 2004). Dans le cas où les deux densités sont SS, les vitesses sont données dans Comte & Lacour (2010), elles sont meilleures et décroissent plus vite que n'importe quelle fonction logarithmique. Notre objectif sera d'atteindre les vitesses optimales.

1.2.5 Sélection de paramètre

Une fois la collection d'estimateurs construite, pour une collection de paramètres, il faut choisir le paramètre qui définit l'estimateur final. En pratique, la régularité de la fonction que l'on estime est inconnue. L'objectif est de construire des estimateurs qui s'adaptent à la régularité inconnue

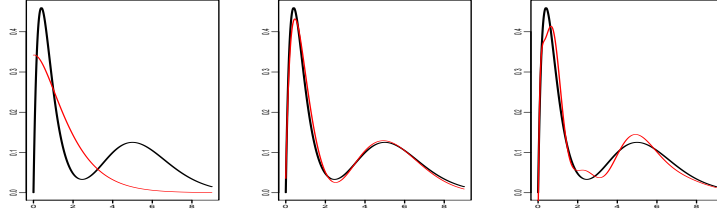


Figure 1.1 – Estimateur par projection \hat{f}_m , en rouge (gris), de f , en noir épais, densité d'un mélange de deux lois gamma. De gauche à droite : $m = 1$, $m = 7$, $m = 20$

de f , seulement à partir des données. On parlera d'estimateur « adaptatif » et de procédure *data-driven*.

Expliquons l'enjeu. Dans la décomposition biais-variance (1.2.7), les deux termes sont de monotonie opposée. Cela se voit sur les majorations (1.2.8) pour l'estimateur à noyau et (1.2.9) pour l'estimateur par projection. Mais le problème du choix de h ou m est le même : pour minimiser le risque, il faut faire un compromis entre ces deux termes, c'est à dire entre la qualité de l'approximation de la fonction d'intérêt et la complexité du modèle.

Sélection d'un paramètre de dimension

Considérons le cas de la collection d'estimateurs \hat{f}_m de f construits par projection et donnée en (1.2.3). La collection dénombrable finie des modèles, donc des paramètres m , est notée \mathcal{M}_N . L'estimateur par projection \hat{f}_m coïncide avec l'estimateur qui minimise le contraste $\gamma_N(t) = \|t\|^2 - (2/N) \sum_{j=1}^N t(X_j)$ sur \mathcal{S}_m (1.2.2). Les espaces \mathcal{S}_m sont emboîtés et il y a au plus un modèle par dimension. Plus la dimension sélectionnée est grande plus le modèle est précis mais complexe, un compromis est donc à faire. La Figure 1.1 illustre ce propos pour le choix de la base de Laguerre sur \mathbb{R}^+ . Trois estimateurs \hat{f}_m de la densité d'un échantillon issu d'un mélange de deux lois gamma sont représentés (trait gris/rouge) de gauche à droite pour $m = 1, 7, 20$. Seul le graphe du milieu représente une estimation satisfaisante. On recherche donc un critère pour choisir l'estimateur final $\hat{f}_{\hat{m}}$ de façon automatique. L'estimateur appelé « oracle » noté \hat{f}_{m^*} défini par : $m^* = \operatorname{argmin}_{m \in \mathcal{M}_N} \mathbb{E}[\|\hat{f}_m - f\|^2]$, est inaccessible en pratique, mais le but est de s'en approcher au mieux, c'est ce qui va définir le « meilleur » estimateur. D'après la majoration (1.2.9) et par orthogonalité de la base choisie, on a : $m^* = \operatorname{argmin}_{m \in \mathcal{M}_N} \left\{ \|f - f_m\|^2 + \frac{C_\varphi m}{N} \right\} = \operatorname{argmin}_{m \in \mathcal{M}_N} \left\{ -\|f_m\|^2 + \frac{C_\varphi m}{N} \right\}$. L'inégalité oracle est

$$\mathbb{E}[\|\hat{f}_{\hat{m}} - f\|^2] \leq C \mathbb{E}[\|\hat{f}_{m^*} - f\|^2] + \frac{C'}{N}$$

où C est une constante positive et C' est une constante positive dépendant de f et de C_φ . Les résultats que nous démontrons dans ce travail sont des résultats de « type oracle » c'est à dire de la forme

$$\mathbb{E}[\|\hat{f}_{\hat{m}} - f\|^2] \leq C \inf_{m \in \mathcal{M}_N} \mathbb{E}[\|\hat{f}_m - f\|^2] + \frac{C'}{N}.$$

L'inégalité sera d'autant meilleure que C est proche de 1. On dira que l'estimateur est « optimal au sens de l'oracle », s'il a un risque aussi petit, à une constante près et à un terme de reste près, que le meilleur des risques possibles dans une collection d'estimateurs.

Expliquons maintenant la méthode de sélection. Le contraste $\gamma_N(\hat{f}_m) = -\|\hat{f}_m\|^2$ décroît avec m par construction : par conséquent, en choisissant l'estimateur minimisant le contraste on sous-estimerait le vrai risque. C'est pourquoi on introduit une fonction de « pénalisation ». On définit une fonction $\text{pen} : \mathcal{M} \rightarrow \mathbb{R}^+$ telle m est choisi comme minimisant le critère pénalisé suivant

$$\text{crit}(m) = \gamma_N(\hat{f}_m) + \text{pen}(m).$$

Puis on choisit la fonction pénalité de l'ordre du majorant de la variance, évalué lors du calcul de MISE de l'estimateur, puisque l'on cherche à minimiser le MISE. Pour l'estimateur par projection on a donc $\text{pen}(m) = \kappa C_\varphi m/N$, avec κ une constante de calibration, à choisir en pratique. La collection de modèles \mathcal{M}_N est alors précisée, elle peut être choisie par exemple pour borner la pénalité : $\mathcal{M}_N = \{1, \dots, N\}$.

L'approche par minimisation de critère pénalisé date des années 70 avec les travaux de Akaike (1973) et Mallows (1973) qui proposent de pénaliser par la dimension puis elle a été généralisée par Barron *et al.* (1999), Massart (2007). D'autres critères tels que BIC (introduit par Schwarz en 1978) et AIC se basent sur une heuristique asymptotique et sont dépendants de la taille du modèle. C'est pourquoi des critères pénalisés basés sur des inégalités non-asymptotiques se sont développés.

Dans cette thèse nous choisissons de pénalités basées sur un majorant du terme de variance. La pénalité est déterministe le plus souvent et telle que

$$\text{pen}(\alpha) \geq \kappa \mathbb{E}[\|\mathbb{E}[\hat{f}_\alpha] - \hat{f}_\alpha\|^2]$$

avec κ une constante. Les inégalités de concentrations permettent de prouver les inégalités de type oracle pour ce choix de pénalité. En particulier on s'appuiera sur la forme intégrée du résultat de Talagrand (1996), donné en Annexe B. La pénalité est donnée à une constante près qu'il faudra calibrer en pratique. Par exemple Birgé & Massart (2006) ont proposé une méthode pour calibrer la constante des pénalités connues à une constante multiplicative près pour le modèle gaussien : l'heuristique de pente (on peut voir aussi Baudry *et al.*, 2012, pour l'implémentation de la méthode).

Sélection de fenêtre

Considérons maintenant la collection d'estimateurs à noyau \hat{f}_h de f donnée par (1.2.1). Notons \mathcal{H}_N la collection dénombrable finie de fenêtres h . Plus h est petit plus l'estimation est précise, mais plus la complexité du modèle est grande. Il faut donc faire un compromis entre précision et perte d'information. La Figure 1.2 montre trois estimateurs à noyau (en trait continu rouge/gris) de la densité d'un échantillon d'une variable X distribuée selon un mélange de deux lois gamma. De gauche à droite la fenêtre vaut $h = 0.01, 0.25, 1$. Seul le graphe du milieu montre un estimateur satisfaisant de la vraie densité (en trait noir épais), les deux autres choix ne sont pas acceptables. On voit ici la nécessité de définir une méthode de sélection de h . Pour minimiser le risque \mathbb{L}^2 de notre l'estimateur nous devons réaliser le compromis biais-variance, expliqué précédemment.

La méthode la plus ancienne appelée « règle du pouce » est basée sur des considérations empiriques. On trouve des détails dans Sheather (2004). Cette méthode s'appuie sur l'approximation lorsque h tend vers 0 du MISE pour le noyau gaussien et lorsque f est une gaussienne. Puis un méthode très utilisée est la *cross-validation*. Elle repose sur une minimisation du MISE approché par son homologue empirique. Il en existe différentes versions. On réfère aux travaux de Celisse (2008) et Celisse (2014) pour des résultats théoriques concernant cette méthode pour l'estimation de densité par projection.

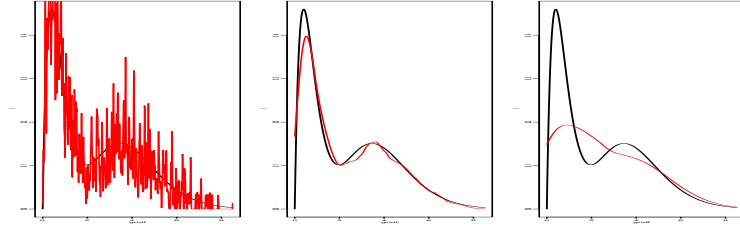


Figure 1.2 – Estimateur à noyau \hat{f}_h , en rouge (gris), de f , en noir épais, densité d'un mélange de deux lois gamma. De gauche à droite : $h = 0.01$, $h = 0.25$, $h = 1$

Dans le but d'obtenir des résultats théoriques sur les estimateurs à noyaux avec sélection de fenêtre, nous nous intéressons maintenant à la méthode de Goldenshluger et Lepski développée dans Goldenshluger & Lepski (2011). Soit \mathcal{H}_N une collection finie de fenêtres. Définissons la suite de noyaux itérés

$$\hat{f}_{h,h'}(x) = K_{h'} \star \hat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_{h'} \star K_h(x - X_j).$$

La fenêtre choisie est donnée par

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}_N} \{A(h) + V(h)\}$$

avec

$$A(h) = \sup_{h' \in \mathcal{H}_N} \left(\|\hat{f}_{h,h'} - \hat{f}_{h'}\|^2 - V(h') \right)_+, \quad \text{avec} \quad V(h) = \kappa \frac{\|K\|_1^2 \|K\|^2}{Nh}, \quad (1.2.10)$$

où κ est une constante de calibration, à choisir en pratique. Le terme $A(h)$ estime le terme de biais $\|f_h - f\|^2$ en remplaçant les quantités inconnues par des estimateurs. Mais le calcul montre qu'il faut encore retrancher un terme, de l'ordre du terme de variance ($V(h)$), pour que $A(h)$ soit de l'ordre du terme de biais, à un reste d'ordre $1/N$ près. La collection \mathcal{H}_N est choisie ici pour contrôler le terme $V(h)$ et est fournie par la démonstration de l'égalité de type oracle suivante, à l'aide de l'inégalité de Talagrand :

$$\mathbb{E}[\|\hat{f}_{\hat{h}} - f\|^2] \leq C \inf_{h \in \mathcal{H}_N} \{\|f_h - f\|^2 + V(h)\} + \frac{C'}{N}.$$

Par exemple $\mathcal{H}_N = \{1/k, k = 1, \dots, \lfloor \sqrt{N} \rfloor\}$. La preuve de ce résultat fournit une minoration de κ mais trop grande en général et en pratique il faudra choisir la constante par une étude sur simulations. Un travail récent de Lacour & Massart (2016) suggère de ne plus utiliser les estimateurs $\hat{f}_{h,h'}$ dans le terme $A(h)$ et de choisir une constante κ pour le terme $V(h)$ qui apparaît dans le critère $A(h)$ (1.2.10) puis 2κ dans le terme $V(h)$ qui définit \hat{h} .

Cas de la déconvolution

Nous avons vu dans la Section 1.2.3 que l'estimateur $\hat{f}_{X,m}$ construit pour le problème de déconvolution peut-être vu comme un estimateur à noyau ou un estimateur par projection (mais m n'est plus la dimension de l'espace \mathcal{S}_m). Dans ce cas, on peut alors soit utiliser une méthode de pénalisation classique (comme décrite ci dessus pour la projection), soit adapter la méthode de Goldenshluger et Lepski développée pour les noyaux pour la sélection de modèle (voir Comte & Johannes, 2012). En effet la méthode permet d'obtenir de façon simple des inégalités de type

oracle pour l'estimateur final et se transpose aisément à d'autres cadres. Nous adoptons cette seconde stratégie mixte dans les Chapitres 2 et 3 de cette thèse.

De plus, la méthode de Goldenshluger et Lepski s'adapte au choix de plusieurs fenêtres en dimension supérieure (nous utiliserons cette méthode dans le Chapitre 4). Nous nous sommes servis de cette remarque lors de la sélection de deux paramètres dans un cadre de déconvolution dans le Chapitre 3. En effet, dans ce cas, choisir le premier paramètre par pénalisation classique puis le second (avec un critère aléatoire cette fois) s'est avéré une mauvaise stratégie. La sélection de fenêtre par la méthode de Lepski donne généralement un paramètre qui annule le critère A (1.2.10) donc cela empêche une procédure en deux temps). Alors la méthode de Goldenshluger et Lepski, écrite dans ce contexte particulier, fournit une procédure adaptative satisfaisante.

Nous détaillons dans la section suivante les principaux résultats de cette thèse.

1.3 Les différents modèles étudiés et les approches proposées dans cette thèse

Différents modèles sont étudiés dans ce travail mais les parties sont liées par le même objectif : estimer la densité d'une variable aléatoire, non observée directement. Hormis dans le Chapitre 6, nous sommes dans le cadre de modèles mixtes. Nous rappelons rapidement pour chaque modèle les approches classiques puis nous présentons les nouvelles stratégies que nous proposons.

1.3.1 Modèles linéaires mixtes

Les modèles mixtes sont des outils très utiles pour décrire un groupe, cela peut être une population d'individus avec des traits en commun ou un ensemble de mesures issues d'une expérimentation. Leur particularité est de considérer des paramètres propres à chaque sujet et des paramètres communs à tout le groupe. Ceci permet de modéliser la variabilité inter-sujets et intra-sujets. L'intérêt principal de ces modèles est d'aboutir à une seule procédure d'estimation pour étudier la population entière. Certains paramètres peuvent être fixes et d'autres aléatoires, dépendants de chaque individu. Nous nous intéressons au cas de paramètres aléatoires.

Le premier modèle mixte que nous considérons est un modèle linéaire mixte. Il fait l'objet de la Partie I Chapitre 2. Ils sont étudiés en détails dans Pinheiro & Bates (2000). Pour $\mathcal{K} + 1$ temps d'observation $t_k = k\Delta$, la dynamique de N sujets Y_j s'écrit :

$$Y_{j,k} = \alpha_j + \beta_j t_k + \varepsilon_{j,k}, \quad j \in \{1, \dots, N\}, \quad k \in \{0, \dots, \mathcal{K}\}. \quad (1.3.1)$$

Les coefficients α_j, β_j sont deux effets aléatoires chacun supposé *i.i.d.* de densités respectives f_α et f_β , de carré intégrables. Les suites $(\varepsilon_{j,k})_{j,k}$ et $(\alpha_j, \beta_j)_j$ sont supposées indépendantes et les variables de bruit $\varepsilon_{j,k}$ *i.i.d.* de densité f_ε .

Remarquons que pour deux temps distincts deux variables $Y_{j,k}$ et $Y_{j,k'}$ ne sont pas indépendantes ce qui traduit la dépendance en temps des données. L'influence des effets aléatoires est illustrée par la Figure 1.3. Elle représente deux jeux de données simulées $(Y_{j,k})_{j,k}$ pour $N = 10$ et $\mathcal{K} = 20$, avec deux effets fixes (commun à chaque sujet) α et β sur le graphique de gauche et deux variables aléatoires α_j, β_j sur le graphique de droite. Lorsque l'effet aléatoire, par exemple α_j , est supposé Gaussien, on peut le réécrire $\alpha + \alpha_j$ avec $\alpha_j \sim \mathcal{N}(0, \sigma^2)$. Dans ce cas, on voit que la partie aléatoire est propre à chacun des individus et traduit l'écart à la valeur moyenne de chacun. Par exemple, les données de concentrations en anticorps pour une maladie au cours du temps pour un groupe d'individus peuvent être modélisées par le modèle (1.3.1). Alors les concentrations α_j et β_j pour l'individu j varient autour d'une valeur moyenne. Ces paramètres fournissent des informations supplémentaires sur chaque individu lorsqu'on les estime et sur leur

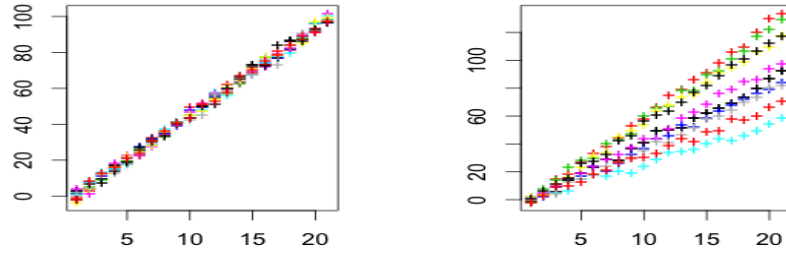


Figure 1.3 – Simulation de 10 trajectoires avec deux effets α et β constants à gauche et deux effets aléatoires à droite

répartition dans la population lorsqu'on estime leur densité (par exemple à $t_0 = 0$ on observe la concentration α_j initiale à un bruit près).

Les modèles mixtes sont le plus souvent étudiés avec une approche paramétrique où les effets aléatoires sont supposés gaussiens et indépendants. Le livre de Pinheiro & Bates (2000) fournit un répertoire des méthodes. Le but est d'estimer les paramètres dont dépendent les densités choisies. Pour cela on a généralement recouru à l'estimateur du maximum de vraisemblance lorsqu'il est accessible. Ces méthodes sont parfois appelées semi-paramétriques car c'est un paramètre de la densité qui est estimé, et non un paramètre du modèle. Mais une hypothèse gaussienne n'est pas toujours adaptée à la réalité. Un modèle de mélange de gaussiennes semble pouvoir décrire des données plus diverses. On peut citer Komárek & Lesaffre (2008) qui supposent que les effets aléatoires suivent un mélange de gaussiennes, et estiment les paramètres du modèle par une approche Bayésienne utilisant des techniques MCMC. Mais une approche complètement non-paramétrique semble indispensable pour prendre une décision avant d'imposer un modèle.

Pour notre étude, les deux effets aléatoires ne sont pas supposés indépendants et le bruit n'est pas supposé symétrique. L'objet du Chapitre 2 est d'estimer la fonction densité f_β à partir des observations $Y_{j,k}$, lorsque la densité f_ϵ est connue, dans un premier temps. L'estimation de la densité des f_α , peut être traitée comme un problème classique de déconvolution. En effet, lorsque $k = 0$, $t_0 = 0$ et alors on a N observations bruitées de la variable d'intérêt : $Y_{j,0} = \alpha_j + \epsilon_{j,0}$. On réfère par exemple à Comte *et al.* (2006) pour l'étude de cet estimateur.

Résultats

Nous proposons dans ce travail deux estimateurs non-paramétriques de la densité f_β des effets aléatoires β_j lorsque la densité du bruit est connue et que sa fonction caractéristique ne s'annule pas. Les procédures proposées s'appuient sur la méthode de déconvolution. En effet, il nous faut d'abord approcher la variable d'intérêt. Pour cela, comme nous disposons d'observations répétées pour chaque individu, une transformation des données nous donne notre estimateur :

$$Z_{j,k} := \frac{Y_{j,k} - Y_{j,0}}{k\Delta} = \beta_j + W_{j,k}$$

avec les variables $W_{j,k} = (\epsilon_{j,k} - \epsilon_{j,0})/(k\Delta)$ centrées et *i.i.d.* pour k fixé. Les $Z_{j,k}$ sont la somme de deux variables indépendantes pour k fixé. Des observations bruitées de la variable d'intérêt β sont donc disponibles. La densité des variables $Z_k = (Z_{j,k})_j$ est donnée par le produit de convolution entre la densité f_β et la densité de $W_k = (W_{j,k})_j$. Nous introduisons un *cut-off* pour la construction de l'estimateur par déconvolution, comme nous l'avons expliqué dans la Section

1.2.3. Le premier estimateur que nous considérons est :

$$\widehat{f}_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,m}}}{|f_{\varepsilon}^*\left(\frac{u}{m\Delta}\right)|^2} du .$$

La particularité de cette stratégie est donc de lier le paramètre de *cut-off* et le temps d'observation. La majoration de son risque quadratique intégré se décompose en un terme de biais et un terme de variance, qui sont :

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta}\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_{\beta}^*(u)|^2 du + F_{\varepsilon} \frac{m\Delta}{2\pi N}, \quad F_{\varepsilon} := \int_{-1}^1 (1/|f_{\varepsilon}^*(v)|^4) dv.$$

Ce résultat fait l'objet de la Proposition 2.2.1 du Chapitre 2. Cette borne dépend de la densité du bruit seulement à travers la constante F_{ε} . Numériquement on voit que cette constante est par exemple inférieure à 4 si le bruit est gaussien centré d'écart type inférieur à 1. Cette double dépendance en m nous conduit à utiliser une méthode de sélection inspirée de la méthode de Goldenshluger et Lepski, dans ce contexte. La procédure adaptative détaillée dans le Chapitre 2 mène à une inégalité de type oracle pour notre estimateur final $\widehat{f}_{\beta,\widehat{m}(1)}$ donnée dans le Théorème 2.2.3 et prouvée à la fin du chapitre. Dans la borne du risque ci-dessus, en considérant f_{β} dans la boule de Sobolev $\mathcal{A}_b(L)$ (donnée en Section 1.2.4), le terme de biais est d'ordre $(m\Delta)^{-2b}$. Alors la vitesse obtenue est en $N^{2b/(2b+1)}$ pour le m réalisant le compromis biais-variance. C'est la vitesse non-paramétrique obtenue lorsque des observations directes de la variable d'intérêt sont disponibles. Cette vitesse est meilleure que les vitesses de déconvolution classiques et que celle obtenue par Comte & Samson (2012) dans le même contexte.

Finalement nous nous intéressons ensuite à l'estimateur basé sur les variables $(Z_{j,\mathcal{K}})_j$ seulement afin de voir si les bonnes vitesses de convergences obtenues avec le précédent estimateur se confirment. Les variables utilisées sont donc celles du dernier temps et du premier temps d'observation seulement. L'estimateur s'écrit

$$\widehat{f}_{\beta,m}^{\mathcal{K}}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,\mathcal{K}}}}{|f_{\varepsilon}^*\left(\frac{u}{\mathcal{K}\Delta}\right)|^2} du .$$

Maintenant m est un paramètre de *cut-off* mais plus un temps d'observation. Alors pour $t \in S_m := \{t \in \mathbb{L}^2, \text{supp}(t^*) = [-m\Delta, m\Delta]\}$, on définit $\gamma(t) := \|t\|^2 - 2 \langle t, \widehat{f}_{\beta,m}^{\mathcal{K}} \rangle$, alors notre estimateur minimise ce contraste : $\widehat{f}_{\beta,m}^{\mathcal{K}} = \underset{t \in S_m}{\text{argmin}} \gamma(t)$. Si $m \leq \min(N, \mathcal{K})$ la borne de risque obtenue au Corollaire 2.2.6 est

$$\mathbb{E}[\|\widehat{f}_{\beta,m}^{\mathcal{K}} - f_{\beta}\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_{\beta}^*(u)|^2 du + B_{\varepsilon} \frac{m\Delta}{\pi N}, \quad B_{\varepsilon} := \sup_{v \in [-1,1]} \frac{1}{|f_{\varepsilon}^*(v)|^4}.$$

Nous obtenons la même vitesse de convergence pour cet estimateur. Puis la sélection du paramètre m se fait par une méthode de sélection classique et conduit à une inégalité de type oracle. Ces deux nouvelles procédures sont comparées sur simulations et donnent des résultats satisfaisants. On compare notamment les MISE par des simulations Monte Carlo pour différents choix de paramètres et de lois. La seconde procédure est plus rapide mais il est préférable d'utiliser la première lorsque le nombre d'observations temporelles par individu (\mathcal{K}) est petit, ce qui est souvent le cas en pratique. Nous proposons également une méthode dans le cas où la densité du bruit est inconnue en Section 2.6.7. Lors de l'étude sur simulations les résultats de cet estimateur sont très prometteurs. Cependant, dans ce travail nous ne menons pas la sélection adaptative du paramètre dont dépend l'estimateur.

1.3.2 Modèles d'équations différentielles stochastiques à effets mixtes

L'étude des équations différentielles stochastiques (EDS) et en particulier des processus de diffusion se développe. Cela s'explique par leur usage dans un grand nombre de domaines d'applications. Pour citer deux principaux exemples, ils sont utilisés en physique ou en biologie pour décrire les systèmes dynamiques perturbés par un bruit aléatoire, ainsi qu'en mathématiques financières. On les note le plus souvent :

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t), \quad X(0) = x \quad (1.3.2)$$

où W est un mouvement Brownien (indépendant de x s'il est aléatoire), la fonction b est appelée coefficient de dérive (*drift*) et σ coefficient de diffusion (*diffusion coefficient*). Pour assurer l'existence et l'unicité d'une solution, on suppose le plus souvent que b et σ sont localement lipschitziennes et σ strictement positive (voir par exemple Karatzas & Shreve, 1988, Chapitre 5).

Dans cette thèse nous nous plaçons dans le cadre de données continues. Pour pouvoir estimer le coefficient de dérive et le coefficient de diffusion il faut être en présence de données à haute fréquence, c'est à dire où le temps d'observation $T = \mathcal{K}\Delta$ tend vers l'infini avec \mathcal{K} et le pas de temps $\Delta = \Delta_{\mathcal{K}}$ vers 0. Il existe une littérature abondante sur l'estimation non-paramétrique de ces deux fonctions. Par exemple Hoffmann (1999), Dalalyan (2005) qui obtient un estimateur de la dérive convergeant à vitesse minimax ; Comte *et al.* (2006) estiment les deux coefficients sous hypothèses de stationnarité et d'ergodicité géométrique ; Höpfner & Brodda (2006), Höpfner (2007) étudie des estimateurs à noyau ; Schmisser (2013) étudie le cas de processus multidimensionnels. Ce n'est pas le propos de cette thèse donc nous ne développons pas davantage.

Les EDS à effets aléatoires généralisent les EDS classiques. Des applications sont : la dynamique des système thermaux (Bacher & Madsen, 2011), la prévision de force solaire et du vent (Iversen *et al.*, 2014), la modélisation de potentiels neuronaux (Ditlevsen & Samson, 2014), la pharmacocinétique/pharmacodynamique (PK/PD) (Hansen *et al.*, 2014) ou encore l'étude de la croissance de fissure (Hermann *et al.*, 2015). Selon les applications, on peut avoir accès à des mesures temporelles répétées et indépendantes. Par exemple, on a l'habitude d'observer la concentration dans le sang d'un médicament pour plusieurs sujets à des temps d'observations communs, ou la dynamique de plusieurs neurones au cours du temps. Chaque trajectoire représente alors le comportement d'un individu (ou d'une unité).

Dans la suite nous supposons observer de manière continue N trajectoires d'un processus sur un intervalle de temps $[0, T]$ assez grand. Les équations différentielles stochastiques à effets aléatoires (SDMEM pour *stochastic differential equation with mixed effects*) que nous considérons dans ce travail sont de la forme suivante :

$$dX_j(t) = b(X_j(t), \phi_j)dt + \sigma(X_j(t))dW_j(t), \quad X_j(0) = x_j, \quad j = 1, \dots, N, \quad t \in [0, T] \quad (1.3.3)$$

où ϕ_j est une variable aléatoire de \mathbb{R}^d et on considèrera dans cette thèse seulement les cas $d = 1, 2$. Sous de bonnes hypothèses, les couples $(\phi_j, X_j(t))$ sont markoviens. Si $X_j(0) = \gamma_j$ une variable aléatoire, on suppose que les $(W_j)_{j=1, \dots, N}$ et $(\phi_j, \gamma_j)_{j=1, \dots, N}$ sont définis sur le même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, les $(W_j)_{j=1, \dots, N}$ sont N processus de Wiener indépendants les $(\phi_j, \gamma_j)_{j=1, \dots, N}$ sont *i.i.d.*. De plus les suites $((\phi_1, \gamma_1), \dots, (\phi_N, \gamma_N))$ et (W_1, \dots, W_N) sont indépendantes. On suppose enfin qu'il existe un ouvert Φ de \mathbb{R}^d et un intervalle $(l, r) \subset \mathbb{R}$ tel que $\sigma(x) > 0$ pour tout $x \in (l, r)$. La Figure 1.4 montre la différence entre le faisceau de 10 trajectoires issues d'un processus d'Ornstein-Uhlenbeck (donné après) avec deux paramètres constants dans le coefficient de dérive (Figure de gauche), et avec deux paramètres aléatoires (Figure de droite). Quelques résultats théoriques sur ces modèles sont rappelés en Annexe B.

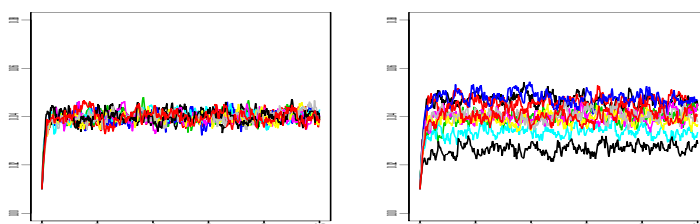


Figure 1.4 – Simulation de 10 trajectoires selon un processus d’Ornstein-Uhlenbeck avec deux paramètres constants dans le coefficient de dérive, à gauche, et deux effets aléatoires à droite

Les objectifs sont doubles : comme cela a été vu pour les modèles linéaires mixtes il faut tout d’abord approcher les effets aléatoires puis estimer leur densité. Nous sommes dans le cadre de problèmes inverses puisque les variables d’intérêt ne sont pas observées directement. Ces modèles seront l’objet des trois chapitres de la Partie II.

Principaux exemples

Nous allons donner deux cas particuliers de la classe plus générale des modèles de Pearson (voir Iacus, 2008) : le modèle d’Ornstein-Uhlenbeck et le modèle de Cox-Ingersoll-Ross.

Modèle d’Ornstein Uhlenbeck (OU)

Le modèle d’Ornstein-Uhlenbeck (sans effet aléatoire) a été introduit par Einstein en 1930 pour décrire le mouvement d’une particule libre dans un gaz ou un liquide qui développe l’effet de friction. Il est aussi utilisé en mathématiques financières sous le nom de modèle de Vasicek pour modéliser les taux d’intérêts (voir Vasicek & Gifford Fong, 1982) et pour décrire les données neuronales par exemple dans Burkitt (2006), Lansky *et al.* (2006), Höpfner (2007). Il décrit la trajectoire d’un processus attiré par une valeur (fixe ou non) mais constamment perturbé par un bruit. Le modèle avec effets aléatoires que l’on considère dans cette thèse pour $j = 1, \dots, N$, est

$$\begin{cases} dX_j(t) &= (\alpha_j - \beta_j X_j(t))dt + \sigma dW_j(t) \\ X_j(0) &= x_j \end{cases} \quad (1.3.4)$$

avec $\sigma > 0$. L’effet aléatoire est noté $\phi_j = (\alpha_j, \beta_j)^t$, soit seul α_j est aléatoire (et $\beta_j = \beta$ fixé) soit les deux sont aléatoires.

Ce modèle est un modèle de « retour à la moyenne » puisqu’il modélise un processus perturbé par un bruit au cours du temps et constamment attiré en même temps par une valeur d’équilibre. En fonction de nos notations, on peut réécrire la fonction de dérive comme : $\beta_j(\alpha_j/\beta_j - X_j(t))$; on voit alors que α_j/β_j est la valeur moyenne vers laquelle le processus est attiré, et β_j peut être interprété comme la vitesse avec laquelle cette valeur est atteinte.

En pratique, les données disponibles sont des données en temps discret. Il faudra donc discrétiser les procédures proposées dans cette thèse, qui supposent que le temps d’observation est continu. Lorsque nous avons besoin de simuler des trajectoires du modèle OU nous procédons à une simulation exacte car la solution de modèle est explicite

$$X_j(t) = X_j(0)e^{-\beta_j t} + \frac{\alpha_j}{\beta_j}(1 - e^{-\beta_j t}) + \sigma e^{-\beta_j t} \int_0^t e^{\beta_j s} dW_j(s). \quad (1.3.5)$$

Modèle de Cox-Ingersoll-Ross (CIR)

Le modèle de CIR est également très utilisé en finance (Cox *et al.*, 1985) et en neurosciences où il est appelé modèle de Feller (Höpfner, 2007). En effet le modèle précédent n'est pas borné ce qui peut ne pas être réaliste alors que le modèle CIR est borné inférieurement par 0. Le modèle avec effets aléatoires que nous allons étudier est donné par l'équation suivante :

$$\begin{cases} dX_j(t) &= (\alpha_j - \beta_j X_j(t))dt + \sigma \sqrt{X_j(t)} dW_j(t) \\ X_j(0) &= x_j > 0 \end{cases} \quad (1.3.6)$$

avec $\sigma > 0$, $\alpha_j > \sigma^2/2$ et $\beta_j > 0$, pour que le processus reste toujours assez loin de 0 (voir par exemple Lamberton & Lapeyre, 2012, page 133). Pour la discrétisation du modèle CIR nous utilisons un schéma de discrétisation implicite avec une solution explicite, dû à Alfonsi (2005).

De même que pour les modèles linéaires mixtes, pour les modèles SDMEM un pan de la littérature considère des modèles paramétriques. On s'intéresse à la densité des effets aléatoires, supposant que l'on connaît la forme de cette dernière et qu'elle dépend d'un paramètre que l'on va chercher à estimer. Dans ce cas l'estimation s'appuie sur la fonction de vraisemblance. Cette dernière (obtenue à l'aide de la propriété de Markov et de la formule de Bayes), le plus souvent inaccessible, peut être approchée ou des méthodes Bayésiennes peuvent être utilisées. Par exemple Picchini *et al.* (2008) étudient le modèle OU et la vraisemblance approchée par une méthode de Gauss-Hermite en supposant que l'effet aléatoire est gaussien. Picchini *et al.* (2010) étudient des modèles plus généraux avec un seul effet aléatoire dans la dérive admettant une densité continue dépendant de paramètres. Puis Delattre *et al.* (2013) donnent une expression exacte de la vraisemblance lorsque les effets aléatoires dans la dérive sont gaussiens et prouvent la consistance de l'estimateur du maximum de vraisemblance. De plus Delattre *et al.* (2015a) s'intéressent au cas d'un effet aléatoire dans le coefficient de diffusion, sans dérive, distribué selon une densité inverse Gamma et proposent deux méthodes estimation des paramètres (maximum de vraisemblance et approximation d'Euler). Enfin, Delattre *et al.* (2015b) supposent que les effets aléatoires suivent un modèle de mélange de gaussiennes et estiment les paramètres du modèle en approchant le maximum de vraisemblance exact avec l'algorithme EM lorsque le nombre de composantes est inconnu. Une approche non-paramétrique est donc novatrice dans ce domaine.

Résultats

Nous nous intéressons tout d'abord dans le Chapitre 3 au cas d'un seul effet aléatoire additif dans le modèle particulier OU (1.3.4). Pour approcher les effets aléatoires ϕ_j il faut considérer pour $j = 1, \dots, N$, $\tau \in [0, T]$, les variables

$$Z_{j,\tau} := \frac{X_j(\tau) - X_j(0) - \int_0^\tau (-\frac{X_j(s)}{\alpha} ds)}{\tau} = \phi_j + \frac{\sigma}{\tau} W_j(\tau).$$

Cet estimateur est l'estimateur du maximum de vraisemblance pour le modèle à $\phi_j = \varphi$ fixé. Ici encore, un problème de déconvolution apparaît : en effet les $Z_{j,\tau}$ sont la somme de deux variables indépendantes pour τ fixé. Et les variables $\varepsilon_{j,\tau} = \sigma W_j(\tau)/\tau$ sont centrées et *i.i.d.* pour τ fixé. Nous proposons deux solutions. Tout d'abord en remarquant que lorsque T tend vers l'infini, $Z_{j,T}$ tend vers ϕ_j p.s., nous étudions l'estimateur à noyau obtenu à partir des variables $(Z_{j,T})_j$.

$$\hat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - Z_{j,T}).$$

Il vérifie la borne de risque suivante

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2\sigma^4\|K''\|^2}{3T^2h^5}$$

donnée dans Comte *et al.* (2013) dans un cas plus général. Dans cette majoration, on voit le terme de biais en premier puis le terme usuel de variance et un terme inhabituel dû au fait que l'estimateur présente un biais supplémentaire par rapport aux estimateurs à noyau usuels : $\mathbb{E}[\widehat{f}_h] \neq f_h$. Cela correspond au terme $\|\mathbb{E}[\widehat{f}_h] - f_h\|^2 \leq \sigma^4\|K''\|^2/(3T^2h^5)$. Il vient donc de l'approximation des ϕ_j par les $Z_{j,T}$ et croît quand h tend vers 0. Nous adaptons la méthode de Goldenshluger et Lepski dans ce cadre pour proposer un estimateur adaptatif et obtenir une inégalité oracle qui fait l'objet du Théorème 3.3.2. Le terme de variance $V(h)$ contient maintenant deux termes, proportionnels aux deux derniers termes de la borne du MISE. Le théorème de Talagrand est appliqué plusieurs fois dans la preuve de ce résultat.

Dans un second temps, nous construisons un estimateur par déconvolution à partir des variables $(Z_{j,\tau})_{j,\tau}$, mais dépendant de deux paramètres. La collection s'écrit

$$\widetilde{f}_{m,s}(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m^2/s^2}} e^{\frac{u^2\sigma^2s^2}{2m^2}} du.$$

Le premier paramètre m joue le rôle de *cut-off* mais est aussi présent dans l'intégrant. La situation est similaire à celle de l'estimateur $\widehat{f}_{\beta,m}$ vu pour les modèles linéaires mixtes. Le second paramètre est choisi proportionnel à l'inverse de σ pour contrôler la variance de l'estimateur. La borne du risque obtenue dans ce cas Proposition 3.4.1 est

$$\mathbb{E} \left[\|\widetilde{f}_{m,s} - f\|^2 \right] \leq \|f_m - f\|^2 + \frac{m}{\pi N} \int_0^1 e^{\sigma^2 s^2 v^2} dv.$$

Finalement nous avons une collection d'estimateurs dépendants de deux paramètres de réglage que nous sélectionnons par une méthode de Goldenshluger et Lepski de nouveau mais bidimensionnelle cette fois. Nous obtenons un estimateur final adaptatif et nous démontrons une inégalité de type oracle donnée au Théorème 3.4.3. Ces estimateurs obtenus à l'aide d'une méthode de Goldenshluger et Lepski sont implémentés et étudiés sur des données simulées puis sur un jeu de données réelles (qui sera présenté par la suite dans cette introduction).

Puis nous étudions le cas de deux effets aléatoires dans le coefficient de dérive :

$$dX_j(t) = b(X_j(t), \phi_j)dt + \sigma(X_j(t))dW_j(t), \quad X_j(0) = \gamma_j, \quad j = 1, \dots, N, \quad t \in [0, T]$$

avec $b(X_j(t), \phi_j) = b^t(X_j(t))\phi_j = \phi_{j,1}b_1(X_j(t)) + \phi_{j,2}b_2(X_j(t))$, dans le Chapitre 4. Ce modèle est beaucoup plus général que le précédent. Une des motivations de ce travail est de pouvoir estimer la densité des effets aléatoires dans le cas du modèle CIR c'est à dire lorsque le coefficient de dérive n'est pas borné, ce qui sort du cadre du travail Comte *et al.* (2013). La stratégie que nous allons présenter ici s'inspire des résultats de Genon-Catalot & Larédo (2016) obtenus dans le cadre d'un seul effet aléatoire multiplicatif dans le coefficient de dérive $b(X_j(t), \phi_j) = b(X_j(t))\phi_j$. Les fonctions b_1, b_2, σ sont supposées continues sur \mathbb{R} , b_1, b_2 sont lipschitziennes et σ hôlderienne d'exposant dans $[1/2, 1]$. Ainsi l'existence et l'unicité d'une solution forte $X_j(\cdot)$ est assurée pour toute variables $(\phi_j, \gamma_j) \in \mathbb{R}^3$, adaptée à la filtration $(\mathcal{F}_t = \sigma((\phi_j, \gamma_j)_{j=1, \dots, N}, W_j(s), s \leq t, j = 1, \dots, N), t \geq 0)$.

Nous choisissons comme estimateur bidimensionnel des effets aléatoires $\phi_j = (\phi_{j,1}, \phi_{j,2})^t$ un estimateur du maximum de vraisemblance $A_j(T)$ (voir l'Annexe B pour des rappels sur la vraisemblance) tronqué, noté $\widehat{A}_j(T) = A_j(T)\mathbf{1}_{B_j(T)}$. Nous étudions le risque en norme euclidienne

de \mathbb{R}^2 , de cet estimateur et, sous des hypothèses techniques et des conditions de stationnarité du processus (détaillées dans le Chapitre 4), nous obtenons à la Proposition 4.3.1 la majoration du moment d'ordre 2 qui suit :

$$\mathbb{E} \left[\|\widehat{A}_j(T) - \phi_j\|_2^2 \right] \leq \frac{C}{T}. \quad (1.3.7)$$

La constante C est donnée dans les preuves comme $C = \mathbb{E}[S(\phi_j)]$ avec S une fonction explicite. Ce résultat est la clé du résultat de risque $\mathbb{L}^2(\mathbb{R}^2)$ de l'estimateur non-paramétrique de la densité donné ci-après. Nous construisons un estimateur à noyau de la densité bivariée des effets aléatoires :

$$\widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - \widehat{A}_j(T)).$$

La majoration du MISE de l'estimateur \widehat{f}_h obtenue Proposition 4.4.1 est

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh_1h_2} + 2 \max\left(\frac{1}{h_1^3h_2}, \frac{1}{h_1h_2^3}\right) \left(\left\| \frac{\partial K}{\partial u} \right\|^2 + \left\| \frac{\partial K}{\partial v} \right\|^2 \right) \frac{C}{T}.$$

On voit apparaître, comme dans le cas d'un seul effet aléatoire étudié précédemment, un terme supplémentaire de dû au biais de l'estimateur, mais contrôlable grâce à la majoration (1.3.7). Enfin, nous sélectionnons le couple de fenêtres par la méthode de Goldenshluger et Lepski et obtenons une inégalité de type oracle pour le l'estimateur final donnée au Théorème 4.4.5. Nous détaillons la procédure sur les exemples OU et CIR tout au long de ce second chapitre. Enfin nous présentons une étude numérique de notre estimateur (en sélectionnant le couple de fenêtres avec la fonction *kde2d* de R et non par la méthode de Lepski). Nous remarquons que les résultats numériques de MISE empiriques sont moins bons que pour un noyau unidimensionnel. Cela se voit en fait sur les vitesses de convergence dans un cadre classique. En effet, nous avons rappelé qu'un estimateur à noyau atteint la vitesse $N^{-2\beta/(2\beta+1)}$ lorsque $f \in \mathcal{N}(\beta, R)$ et que le noyau est d'ordre $\lfloor \beta \rfloor$. Ici avec $\beta = (\beta, \beta)^t$, si f est dans $\mathcal{N}(\beta, R)$ (la définition bidimensionnelle de cette espace est donnée au Chapitre 4 Section 4.4), la vitesse atteinte par l'estimateur est en $N^{-2\beta/(2\beta+2)}$, donc dégradée par la dimension 2.

Enfin, ces méthodes d'estimation sont implémentées dans le package `mixeddsde` pour le langage R disponible à partir du lien <https://cran.r-project.org/web/packages/mixeddsde/index.html>. Une méthode paramétrique y est développée par Adeline Samson et une approche Bayésienne paramétrique par Simone Hermann. Ce package met en valeur l'intérêt d'avoir différentes méthodes à disposition selon les cas. Précisons que ce code de recherche est toujours en développement. Il est perfectible et des améliorations sont à venir. Le Chapitre 5 est consacré à la présentation de ce package. Nous y proposons plusieurs exemples et un jeu de données réelles est étudié à l'aide des méthodes développées.

1.3.3 Un modèle de censure

Dans cette section nous nous intéressons à des variables positives. Elles peuvent modéliser des tailles, des poids ou encore des durées. Dans ce cas on les appelle données de survie. Cela peut être dans le domaine médical : la durée entre l'apparition du symptôme d'une maladie chez un patient et son décès suite à cette maladie ; ou dans le domaine de l'économie : la durée d'une période de chômage entre deux périodes de travail. Nous nous intéressons à deux fonctions clés pour ces variables (notons les X_j) : la densité f et la fonction de survie notée $\bar{F}(t) = \mathbb{P}(X_1 >$

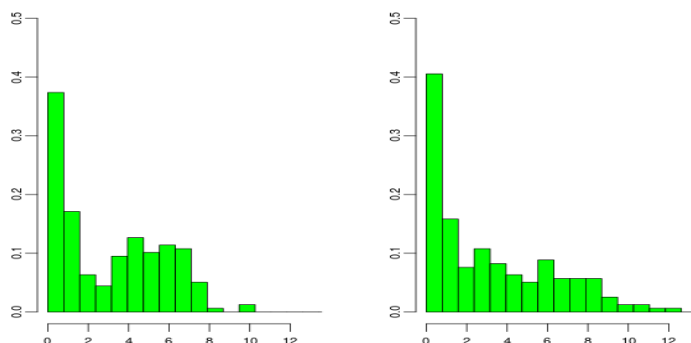


Figure 1.5 – Histogramme d'un échantillon X de taille 200 d'un mélange de deux lois gamma, à gauche, et l'échantillon bruité Y avec $a = 0.5$ à droite

$t) = \int_t^\infty f(x)dx$. La fonction de survie est très liée aux moments de la variable. En effet, pour toute fonction $s : \mathbb{R}^+ \rightarrow \mathbb{R}$ de classe \mathcal{C}^1 strictement croissante telle que $\mathbb{E}[s(X_1)] < +\infty$, on a $\mathbb{E}[s(X_1)] = s(0) + \int_0^{+\infty} \bar{F}(x)s'(x)dx$, en particulier $\mathbb{E}[X_1] = \int_0^{+\infty} \bar{F}(x)dx$. L'étude de données de survie s'intéresse généralement au taux de hasard $\lambda(x) = f(x)/\bar{F}(x)$, mais ce ne sera pas le cas ici.

Ces données sont souvent « censurées », ou tronquées, c'est à dire que l'on ne les observe pas complètement. Cela peut être dû à différentes causes. Lorsque l'on exclut toute une partie des valeurs d'une population (par exemple dans l'étude de manufacture), on parle de données tronquées. Lorsque l'on ne connaît pas quelles données sont exclues (par exemple dans l'étude de temps de réaction à une expérience chez des patients), on parle de censure. Pour la censure à droite on suppose observer $Y_i = \inf(X_i, C_i)$ et $\delta_i = \mathbb{1}_{X_i \leq C_i}$ c'est à dire les données censurées Y_i et les individus censurés à travers $\delta_i = 1$ si l'individu i (donc la variable X_i , le temps écoulé entre l'origine et le phénomène qui nous intéresse) est censuré, alors C représente le temps qui s'est écoulé entre l'origine et la censure (voir Alarcon *et al.*, 2009; Comte *et al.*, 2015b, pour des applications médicales).

Nous nous intéressons à une modélisation différente. En effet, nous supposons observer des données à un bruit multiplicatif près, uniforme centré en 1 s'écartant de 1 d'un paramètre $0 < a < 1$ (on notera $U \sim \mathcal{U}_{[1-a, 1+a]}$). Cela correspond à un phénomène d'atténuation/amplification de la variable d'origine décrit par le modèle

$$Y_i = X_i U_i, \quad i = 1 \dots n, \quad U_i \sim \mathcal{U}_{[1-a, 1+a]} \quad (1.3.8)$$

Ce modèle fait l'objet de la Partie III Chapitre 6. Les notations diffèrent quelque peu ici : le nombre d'observations est noté n (la notation N sera utilisée pour autre chose). Il découle naturellement du modèle classique de bruit additif. En effet, on a l'habitude d'écrire : $Y_i = X_i + \varepsilon_i$, $i = 1 \dots n$, où généralement ε_i est d'espérance nulle. Mais pour un grand nombre d'expériences, le bruit est en fait proportionnel au niveau du signal. Ce qui conduit à écrire le modèle sous la forme : $Y_i = X_i + X_i a \varepsilon_i = X_i(1 + a\varepsilon_i)$ où $U_i = (1 + a\varepsilon_i)$ est de moyenne 1. La Figure 1.5 illustre l'effet du bruit multiplicatif uniforme sur un échantillon issu d'un mélange de lois gamma représenté à gauche, et l'échantillon bruité à droite. Le caractère bimodal n'est plus très clair sur l'histogramme de droite. On peut donc penser qu'obtenir de l'information sur X , en considérant seulement l'échantillon bruité des Y , ne fournit pas une procédure satisfaisante.

Ce modèle peut représenter des résultats approximatifs obtenus lors d'un sondage sur le poids, ou la taille. Il modélise une transmission approximative de l'information : les enregistrements Y_i correspondent à la valeur d'intérêt X_i à un erreur $\pm 100a\%$ près. Mais il peut également

servir à la protection des données. En effet, lorsque l'on veut transmettre des données d'un groupe d'individus/de pays, sans communiquer chaque valeur précisément, on peut appliquer un bruit multiplicatif. Mais alors, avec la connaissance du niveau de bruit appliqué (et de la loi de la variable choisie) il faut savoir retrouver l'information principale contenue dans les données initiales. Cette question a été traitée dans Klein *et al.* (2013) pour différentes lois de U . Notre choix de modèle y est étudié.

Pour étudier la densité f de X , beaucoup de travaux existent dans le cas où les U_i sont de loi $\mathcal{U}_{[0,1]}$. Vardi (1989) puis Vardi & Zhang (1992) se placent dans ce cadre et supposent que l'on observe les Y_1, \dots, Y_n ainsi qu'un échantillon X_1, \dots, X_m antérieur. Ils étudient l'estimateur non-paramétrique du maximum de vraisemblance, dans le cas où $m, n \rightarrow \infty$. Ils montrent l'intérêt mathématique que comporte ce modèle, en le reliant à des problèmes classiquement étudiés. Par exemple le lien avec l'écriture classique de la censure ($Y_i = \inf X_i, C_i$ et $\delta_i = \mathbb{1}_{X_i \leq C_i}$ vue précédemment), ainsi qu'avec la déconvolution (lorsque l'on passe au logarithme dans le modèle multiplicatif, en particulier si $U \sim \mathcal{E}(1)$ alors $\exp(-U) \sim \mathcal{U}_{[0,1]}$). Enfin notons que dans ce cadre, le densité des données observées est décroissante et cette propriété est équivalente à l'écriture multiplicative du modèle (voir par exemple Gurland, 1948, pour la preuve). Ensuite, l'estimation de la fonction de répartition et de la densité est étudiée de façon non-paramétrique et asymptotique dans Asgharian *et al.* (2012), Abbaszadeh *et al.* (2013). Puis le travail Brunel *et al.* (2015) construit un estimateur à noyau de la densité et de la fonction de survie, avec une vitesse optimale au sens minimax. Enfin, Belomestny *et al.* (2016) s'intéressent au modèle de censure multiplicatif observé à un bruit additif près, avec $U_i \sim \beta(1, k)$.

Résultats

Pour le modèle (1.3.8) notre intérêt porte d'abord sur l'estimation de la fonction densité puis de la fonction de survie, de la variable masquée X . La procédure que nous proposons s'appuie sur le même principe pour les deux estimations : tout d'abord l'estimation d'une fonction auxiliaire puis l'estimation de la fonction d'intérêt.

Présentons l'estimation de la densité de X : f . La fonction auxiliaire dans ce cas est notée g . Elle apparait naturellement lorsque l'on calcule $\mathbb{E}[t(Y_1) + Y_1 t'(Y_1)] = \langle t, g \rangle$, où $\langle \cdot, \cdot \rangle$ est le produit scalaire sur $\mathbb{L}^2(\mathbb{R}^+)$ et $t \in \mathbb{L}^2(\mathbb{R}^+)$. Pour une base orthonormée de cet espace, on définit \mathcal{S}_m les espaces engendrés par les m premiers vecteurs de la base. Alors l'estimateur par projection de g s'écrit

$$\hat{g}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{1}{n} \sum_{i=1}^n [Y_i \varphi_j'(Y_i) + \varphi_j(Y_i)] = n^{-1} \sum_{i=1}^n \psi_{\varphi_j}(Y_i)$$

et c'est un estimateur sans biais de la fonction $g_m = \sum_{j=0}^{m-1} \langle g, \varphi_j \rangle \varphi_j$ de \mathcal{S}_m . Puis nous mettons en place la stratégie « d'inversion » qui donne une suite d'approximations de la fonction f à partir de g , en fonction de l'indice N de la suite. La collection d'estimateurs pour f est donnée finalement par

$$\hat{f}_{N,m}(x) = 2a \sum_{k=0}^{N-1} \hat{g}_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right).$$

Nous choisissons la base de Laguerre comme base orthonormée de $\mathbb{L}^2(\mathbb{R}^+)$ pour ses bonnes propriétés. Notamment les relations entre les vecteurs de base φ_j et leurs dérivées φ_j' sont simples et on dispose de majorations des normes infinies de ces fonctions. Nous étudions le MISE de l'estimateur de g en supposant que f est de carré intégrable et que $\mathbb{E}[X_1^2] < +\infty$, il s'en déduit le résultat suivant pour l'estimateur $\hat{f}_{N,m}$, en fonction de n le nombre d'observations et du pas

N de la suite d'approximation :

$$\mathbb{E}[\|\widehat{f}_{N,m} - f\|^2] \leq \frac{8a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} \left(\|g - g_m\|^2 + 4\frac{m}{n} + 16\mathbb{E}[Y_1^2] \frac{m^3}{n} \right) + 2 \left(\frac{1-a}{1+a} \right)^N \|f\|^2.$$

Ce résultat fait l'objet de la Proposition 6.2.9. Les constantes de cette borne ne sont pas très grandes, elles dépendent du paramètre a supposé connu. Les ordres du majorant obtenu sont ceux des modèles de bruit multiplicatif. Le terme supplémentaire dépendant de $\|f\|$ tend exponentiellement vite vers 0 lorsque N devient grand. Enfin, nous sélectionnons la dimension optimale de l'espace de projection par une procédure de pénalisation. La pénalité est de l'ordre du terme de variance. Mais ici elle est aléatoire car nous remplaçons $\mathbb{E}[Y_1^2]$ par son estimateur empirique. Nous obtenons une inégalité de type oracle pour notre estimateur sélectionné au Théorème 6.2.4, à l'aide de l'inégalité de Talagrand et de l'inégalité de Rosenthal (voir Annexe B). Dans un second temps nous estimons la fonction de survie \overline{F} en intégrant la relation entre f et g sur l'intervalle $(0, +\infty[$, ce qui donne une relation entre \overline{F} et \overline{G} . L'estimateur est le suivant :

$$\check{\overline{F}}_{N,m} = \frac{2a}{1+a} \sum_{k=0}^{N-1} \left(\frac{1-a}{1+a} \right)^k \check{\overline{G}}_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right).$$

En suivant les mêmes étapes et en supposant que $\mathbb{E}[X_1^2]$ est finie et que \overline{F} est de carré intégrable, il découle une inégalité de type oracle, non-asymptotique, pour l'estimateur de la fonction de survie construit, donnée au Théorème 6.2.7. Les vitesses obtenues sont les vitesses optimales au sens minimax dans le cadre multiplicatif pour un bruit uniforme sur $[0, 1]$. La vitesse d'estimation de f est en $n^{-s/(s+3)}$ lorsque $f \in W^s(\mathbb{R}^+, L)$. Pour la fonction de survie la vitesse est en $n^{-s/(s+1)}$ (vitesse qui reste moins bonne que la vitesse de l'estimateur empirique de la fonction de répartition, qui converge à vitesse $1/n$).

Lors de l'étude sur simulation, nous comparons l'estimateur construit par projection avec l'estimateur naturel construit par déconvolution. En effet, rappelons que si nous prenons le logarithme dans le modèle (1.3.8) nous obtenons un modèle avec bruit additif, cependant l'estimateur construit ainsi n'est pas défini au voisinage de 0. Cet estimateur fournit des MISE empiriques moins bons que notre nouvel estimateur dans tous les cas.

De plus, pour chacune des estimations (de la densité f et de la survie \overline{F}) nous comparons nos estimateurs aux estimateurs de référence obtenus sur les observations directes X_j , lorsqu'elles sont disponibles. Pour la densité on trace l'estimateur par projection issu de l'échantillon $(X_j)_j$, étudié par exemple dans Belomestny *et al.* (2016) où une borne inférieure est donnée lorsque $f \in W^s(\mathbb{R}^+, L)$. Pour la survie, notre estimateur de référence est l'estimateur empirique. L'étude montre que notre procédure d'estimation donne des résultats théoriques satisfaisants et que les résultats numériques le sont également. Il apparaît que lorsque le paramètre a est supérieur à 0.15, le bruit multiplicatif a un réel impact sur les données, alors estimer la densité f_Y en espérant approcher f n'est plus une stratégie possible (de même pour \overline{F}_Y et \overline{F}). Notre méthode d'estimation est alors efficace, tant pour la densité que pour la survie.

Enfin nous présentons un exemple d'application pratique de notre méthode pour la protection de données. Nous proposons de transmettre des données confidentielles bruitées par un bruit multiplicatif uniforme de paramètre a , transmis également. Alors, le receveur peut extraire l'information nécessaire de l'échantillon de départ grâce à notre méthode.

1.4 Étude de données neuronales

L'étude de données neuronales constitue une motivation au travail du Chapitre 3 de cette thèse. Elles seront étudiées plus précisément dans le Chapitre 5 notamment avec les méthodes développées au Chapitre 4. Mais détaillons d'abord ce que sont ces données.

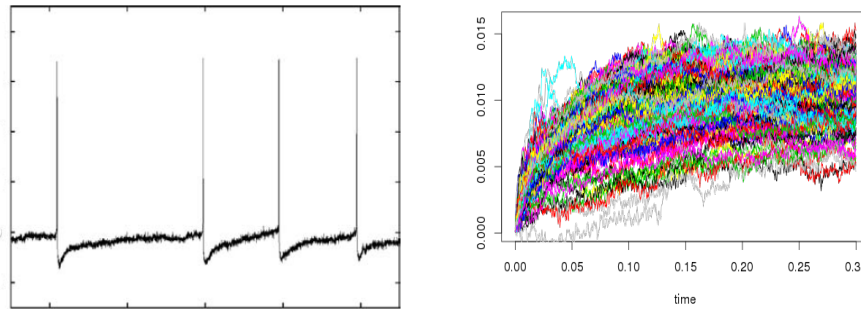


Figure 1.6 – Gauche : potentiel de membrane d’un neurone. Droite : 240 trajectoires entre les *spikes*

Les neurones sont des cellules nerveuses qui envoient constamment des signaux à notre cerveau, nos muscles et nos glandes. Quand un neurone reçoit un signal, il réagit en se chargeant en ions. Le potentiel de membrane est la différence entre le potentiel électrique intérieur et extérieur de la cellule, il est entre -40 mV et -80 mV. Le potentiel de membrane croît jusqu’à ce que le neurone envoie son signal (c’est la dépolarisation). Alors on voit un pic sur le potentiel (*spike*) puis il décroît en dessous de son potentiel de repos (c’est la repolarisation et l’hyperpolarisation) pour finalement y revenir lorsqu’il reçoit un nouveau signal. Ce que l’on vient de décrire s’appelle potentiel d’action. Un des buts de la modélisation mathématique de ce phénomène est de trouver un modèle pour décrire le potentiel de membrane à partir de mesures intracellulaires en dessous de seuil d’émission (les trajectoires *inter-spike*).

Un autre enjeu est de décrire la trajectoire complète. Ce second objectif fait appel à des modèles plus complexes, également non-homogène en temps, qui ne seront pas traités dans cette thèse mais quelques perspectives sont données Partie IV.

La famille des modèles LIF (*leaky-integrate and fire*) a été introduite par Lappicque (1907) pendant ces travaux sur l’excitabilité nerveuse humaine par un courant électrique, en faisant le parallèle entre la membrane du neurone et un circuit électrique. Ce sont des modèles à une équation différentielle ordinaire à une dimension (voir Izhikevich, 2003). Puis des versions stochastiques de ces modèles sont apparues pour intégrer le bruit induit par la présence d’autres neurones, en ajoutant un bruit stochastique. En effet des enregistrements de potentiels dans des expériences différentes suggèrent la présence de variables aléatoires dans le modèle prenant ainsi en compte le caractère stochastique de l’activité neuronale (voir Burkitt, 2006). Notamment les modèles OU et CIR qui sont simples et bien compris, permettent une étude mathématique approfondie. Cette modélisation est une modélisation naïve d’un phénomène complexe mais qui peut néanmoins apporter de nouvelles informations. Les *spikes* ne sont pas intrinsèques au modèle mais sont générés dès que le potentiel dépasse un certain seuil. Puis le processus est remis à son potentiel de repos. D’autre part, les temps entre deux impulsions ont été très étudiés dans la littérature (*e.g.* Tuckwell & Richter, 1978; Ditlevsen & Lansky, 2005) car ils contiennent de l’information importante.

Il y a actuellement une grande demande de modèles robustes et de nouvelles méthodes d’estimation des paramètres décrivant les données neuronales. On s’intéresse aux données intracellulaires qui sont à haute fréquence, et spécialement aux trajectoires entre deux *spikes* (elles sont délimitées par Lansky *et al.*, 2006). Elles sont classiquement décrites par un modèle OU (sans effet aléatoire) lorsque les impulsions sont spontanées et bien décrites par un modèle de Feller (sans effet aléatoire) lorsqu’il y a une stimulation extérieure électrique du neurone. On verra

rapidement dans la Partie IV que l'on peut coupler plusieurs modèles pour décrire les données. Le paramètre α_j des modèles (1.3.4) et (1.3.6) représente alors l'intensité du signal électrique reçu par le neurone. C'est naturellement ce paramètre que l'on suppose aléatoire en premier. Ensuite le paramètre β_j (strictement positif) est appelé constante de temps du neurone, et σ caractérise le niveau de bruit. Les coefficients de dérive des modèles OU et CIR expriment que le processus X_j oscille autour de la valeur $\alpha_j\beta_j$ avec une vitesse α_j .

Dans le Chapitre 3 nous proposons une première modélisation avec un modèle OU à un effet aléatoire additif. Nous comparons nos résultats à ceux de Picchini *et al.* (2008) et Picchini *et al.* (2010), qui supposent que cet effet aléatoire est gaussien. Nous concluons qu'il semble y avoir une queue à gauche sur la densité, non détectée par une modélisation paramétrique par une gaussienne et détectée par l'estimateur à noyau non-paramétrique. Puis, dans le Chapitre 5, nous étudions les données avec les modèles OU (1.3.4) et CIR (1.3.6) avec $d = 1$ ou $d = 2$. Il ressort de cette étude, et de la comparaison des méthodes paramétriques et non-paramétriques proposées dans le package `mixedSde`, que le modèle OU avec un effet additif est le meilleur pour modéliser ces données. L'estimation non-paramétrique par noyau semble être celle qui décrit le mieux la répartition des données et l'on trouve une valeur du paramètre β qui fournit un modèle final plus satisfaisant que les modèles passés. Pour poursuivre l'étude il serait intéressant de voir les résultats lorsque l'on suppose que l'effet aléatoire suit un mélange de gaussiennes, l'article Delattre *et al.* (2015b) fournit une méthode que l'on pourra ajouter au package.

Première partie

Modèle linéaire mixte

Chapitre 2

Nouvelles stratégies d'estimation non-paramétrique dans des modèles linéaires mixtes

Dans ce chapitre nous étudions un modèle linéaire mixte à deux effets aléatoires. Il permet de modéliser N individus ayant le même comportement, pour \mathcal{K} différents temps d'observations $t_k = k\Delta$, identiques pour tous les individus. Il s'écrit :

$$Y_{j,k} = \alpha_j + \beta_j t_k + \varepsilon_{j,k}, \quad j \in \{1, \dots, N\}, \quad k \in \{0, \dots, \mathcal{K}\}.$$

Les coefficients α_j, β_j sont deux paramètres aléatoires chacun supposé *i.i.d.* de densité respective f_α et f_β . Les suites $(\varepsilon_{j,k})_{j,k}$ et $(\alpha_j, \beta_j)_j$ sont indépendantes et les variables de bruit $\varepsilon_{j,k}$ sont *i.i.d.* de densité f_ε . L'objet de ce chapitre est d'estimer la fonction densité f_β à partir des observations $Y_{j,k}$, lorsque la densité f_ε est connue, dans un premier temps. Un travail précédent de Comte & Samson (2012) propose un estimateur non-paramétrique de la densité f_β construit par une méthode de déconvolution. Nous allons comparer nos résultats à cette méthode préexistante.

Nous proposons deux méthodes d'estimation par déconvolution, qui permettent d'obtenir de bonnes vitesses de convergence avec peu d'hypothèses sur la densité f_ε . Elles s'appuient sur les variables $Z_{j,k} = (Y_{j,k} - Y_{j,0})/(k\Delta)$ pour approcher β_j . Dans un premier temps nous construisons un estimateur $\hat{f}_{\beta,m}$ qui a la particularité de dépendre d'un paramètre de *cut-off* m qui est aussi l'indice de temps des observations. Nous calculons ensuite le risque quadratique intégré pour cet estimateur. Cette double dépendance en le paramètre à sélectionner nous mène à adapter une méthode de type Goldenshluger et Lepski dans ce cadre particulier de déconvolution.

Puis de ce premier estimateur nous déduisons un nouvel estimateur $\hat{f}_{\beta,m}^\mathcal{K}$ basé seulement sur les variables $(Z_{j,\mathcal{K}})_j$. Cette fois le *cut-off* m et le temps d'observation ne sont plus reliés. Le choix du paramètre m final est réalisé par minimisation d'un critère pénalisé (proche de celui utilisé pour les méthodes de projection).

Les deux estimateurs finaux utilisent $2N$ observations. Les méthodes originales que nous proposons fournissent des résultats théoriques et numériques satisfaisants. Il est possible de définir un cadre d'utilisation pour chaque estimateur. En effet, si le nombre d'observations \mathcal{K} est faible l'estimateur $\hat{f}_{\beta,m}$ semble conseillé. Dans le cas contraire l'estimateur $\hat{f}_{\beta,m}^\mathcal{K}$ est naturellement celui à utiliser. Ce second estimateur est souvent le plus performant.

Nous exposons ici ces méthodes d'estimation ainsi que leurs résultats théoriques, prouvés en détail à la fin du chapitre. Puis nous menons une étude approfondie sur simulation où nous comparons nos estimateurs et le précédent, pour différents choix de paramètres.

Ce chapitre est issu de l'article Dion (2014) publié dans *Journal of Statistical Planning and Inference*.

Sommaire

2.1	Introduction	39
2.2	Construction of two estimators, risk bounds and adaptive results	40
2.2.1	Estimator using all the observations : \hat{f}	40
2.2.2	Risk bound for the estimator \hat{f}	41
2.2.3	Adaptive selection of m and oracle risk bound for \hat{f}	43
2.2.4	Estimator built using the last observation time $\hat{f}^{\mathcal{K}}$	44
2.3	Comparisons with previous results	46
2.4	Numerical illustration	47
2.4.1	Implementation	47
2.4.2	Numerical simulation results	47
2.4.3	Case of unknown noise density	48
2.5	Concluding remarks	52
2.6	Proofs	53
2.6.1	Proof of Proposition 2.2.1	53
2.6.2	Notations	53
2.6.3	Proof of Theorem 2.2.3	54
2.6.4	Proof of Lemma 2.6.1	56
2.6.5	Proof of Theorem 2.2.7	58
2.6.6	Proof of the Lemma 2.6.2	59
2.6.7	Estimator with unknown noise density	59

2.1 Introduction

In order to analyse repeated measures in time for individuals with the same behaviour, we focus on a linear mixed-effects model. The observation value at time t_k with $k \in \{0, \dots, \mathcal{K}\}$ for individual j , $j \in \{1, \dots, N\}$, denoted $Y_{j,k}$, follows the model

$$Y_{j,k} = \alpha_j + \beta_j t_k + \varepsilon_{j,k}, \quad (2.1.1)$$

where the coefficients α_j and β_j are the random effects of regression and depend on the subject j . The times of observation t_k are known and equidistant with a time step Δ : $t_k = k\Delta$. The random variables $(\varepsilon_{j,k})_{1 \leq j \leq N, 0 \leq k \leq \mathcal{K}}$ are the measurement errors of zero mean. They are supposed independent and identically distributed (i.i.d.) with common density f_ε . We suppose α_j i.i.d. with density f_α and β_j i.i.d. with density f_β . Moreover we assume that $(\varepsilon_{j,k})_{1 \leq j \leq N, 0 \leq k \leq \mathcal{K}}$ and $(\alpha_j, \beta_j)_{1 \leq j \leq N}$ are independent sequences.

Mixed models with random effects are often used, for example in pharmacokinetics. They describe both individual behaviour and variability between individuals. The distribution of random effects is of special interest. It allows for example to describe the heterogeneity of the drug kinetics in the population of individuals. Mixed models have been widely studied, often with parametric strategies and Gaussian random effects and noise (see Pinheiro & Bates, 2000). However it is not clear that this normality assumption of the random effects is truly satisfied in practice. The aim of this work is to produce nonparametric estimation of the density of the random effects from the observations $Y_{j,k}$.

Several papers consider this problem. Wu & Zhu (2010) relax the Gaussian assumption proposing an orthogonality-based estimation of moments estimating the third and fourth moments of the random effects and errors. Komárek & Lesaffre (2008) suppose that the random effects are a mixture of Gaussian distributions, and estimate the weights of the mixture components using a penalized approach. Papageorgiou & Hinde (2012) propose semi-parametric density models of random effects. But it is the estimation of the complete random effects density which should be appropriate. This is the framework of nonparametric estimation that is adopted in this chapter.

Complete nonparametric estimation of f_β and f_α in model (2.1.1) has been little studied in the literature. The main reference is Comte & Samson (2012) who set up a nonparametric estimation with deconvolution ideas. Indeed, model (2.1.1) can be seen as a measurement error model. Deconvolution methods are various and have been initiated by the kernel study of Fan (1991), followed by adaptive strategies (Pensky & Vidakovic (1999), Comte *et al.* (2006), Delaigle & Gijbels (2004)) and optimality of rates (Butucea & Tsybakov, 2007). Except the penalized contrast estimator of Comte & Samson (2012), none of these previous estimators has been used in the mixed model framework. However, for all these methods, the noise density is usually assumed to be known with a certain regularity. The resulting rates are thus very specific and depend on the regularity of the noise and of the function under estimation. For example the noise is supposed to be ordinary smooth in Comte & Samson (2012).

This work focuses on improving existing nonparametric strategies for the estimation of f_β . We refer to Comte *et al.* (2006) for the estimation of f_α , noticing that for $t_0 = 0$ model (2.1.1) writes $Y_{j,0} = \alpha_j + \varepsilon_{j,0}$ which is a standard convolution equation. We first propose an estimator which uses all the available observations. A Mean Integrated Squared Error (MISE) bound is computed and it shows that a bias-variance compromise must be performed. The particularity here is the selection model procedure: we adapt a method set up in Goldenshluger & Lepski (2011) for kernel estimators. Then we establish a non-asymptotic oracle risk bound using Talagrand's inequality. In a second time, an other estimator is proposed, built using the last and the first time of observations. If two times of observation are separated by a quite long time,

enough information is available to obtain a good estimation. As in the first case, risk bound leads us to propose a model selection strategy. In that case a more classical and easy to implant strategy can be adopted, namely a penalized criterion (Birgé & Massart, 1998; Massart, 2007).

What is new is that our two strategies take advantage of both nonparametric and deconvolution ideas. With very mild assumption on the noise regularity (we only require that the characteristic function of the noise is nonzero), the two estimators recover standard rates in density estimation for f_β (Stone, 1980; Donoho *et al.*, 1996) while deconvolution rates were expected. Precisely logarithmic speed of convergence implied by deconvolution methods is avoided. Therefore our two estimators have better rates. Furthermore we do not assume that the noise is ordinary smooth, which was assumed in Comte & Samson (2012).

This chapter is organized as follows. We proceed with the construction of the estimators in Section 2.2.1 and 2.2.12. Bias-variance decompositions of the \mathbb{L}^2 -risk are proved (Propositions 2.2.1, 2.2.5) and lead us to propose adaptive strategies. The main results (Theorems 2.2.3, 2.2.7) prove that the resulting estimators are adaptive. A short Section 2.3 provides a comparison of our estimators with Comte & Samson (2012). Lastly, we illustrate the different methods by simulation experiments presented in Section 2.4. We briefly discuss therein the case of unknown noise density (see Section 2.4.3 and Section 2.6.7) which is also implemented for comparison. Proofs are gathered in Section 2.6.

2.2 Construction of two estimators, risk bounds and adaptive results

Let us introduce some notations. For two functions f and g in $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, the scalar product is defined by $\langle f, g \rangle = \int_{\mathbb{R}} f(x)g(x)dx$ and the associated norm is $\|f\|^2 = \int_{\mathbb{R}} |f(x)|^2 dx$. The Fourier transform of f is $f^*(x) = \int_{\mathbb{R}} e^{ixu} f(u) du$ for all $x \in \mathbb{R}$. When f is the density of a random variable X , $f^*(u) = \mathbb{E}[e^{iuX}]$ is called the characteristic function of X . Then the convolution product of f and g for all $x \in \mathbb{R}$, is $f \star g(x) = \int_{\mathbb{R}} f(x-y)g(y)dy$. Finally we remind the Plancherel-Parseval's formula: $\forall f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), 2\pi\|f\|^2 = \|f^*\|^2$.

2.2.1 Estimator using all the observations: \hat{f}

In this section we build an estimator of the density f_β using all the available observations. We consider the normalized variables $Z_{m,k}$ defined by the difference between two observations, for all $j \in \{1, \dots, N\}$, and for all $m \in \{1, \dots, \mathcal{K}\}$

$$Z_{j,m} := \frac{Y_{j,m} - Y_{j,0}}{m\Delta} = \beta_j + \frac{\varepsilon_{j,m} - \varepsilon_{j,0}}{m\Delta} =: \beta_j + W_{j,m}. \tag{2.2.1}$$

For a given m , the variables $W_{j,m}$, $j \in \{1, \dots, N\}$, are i.i.d. with density f_{W_m} . Due to the independence between the ε_j 's and the β_j 's, the $Z_{j,m}$ are i.i.d. with density denoted by f_{Z_m} , and definition (2.2.1) implies

$$f_{Z_m} = f_\beta \star f_{W_m}. \tag{2.2.2}$$

The Fourier transform of (2.2.2) yields

$$f_{Z_m}^* = f_\beta^* f_{W_m}^*.$$

By independence of the $\varepsilon_{k,m}$ we have for all $u \in \mathbb{R}$

$$\begin{aligned} f_{W_m}^*(u) &= \mathbb{E}[e^{iu(\varepsilon_{j,m} - \varepsilon_{j,0})/m\Delta}] = \mathbb{E}[e^{iu\varepsilon_{j,m}/m\Delta}] \mathbb{E}[e^{-iu\varepsilon_{j,0}/m\Delta}] \\ &= \left| f_\varepsilon^* \left(\frac{u}{m\Delta} \right) \right|^2. \end{aligned}$$

Let us consider the following assumptions:

- (A1) $f_\beta \in \mathbb{L}^2(\mathbb{R})$ and $f_\beta^* \in \mathbb{L}^1(\mathbb{R})$
- (A2) f_ε is known and $f_\varepsilon^* \neq 0$.

(A1) is needed to compute the $\mathbb{L}^2(\mathbb{R})$ -risk and apply the Fourier inversion formula $f_\beta(x) = 1/(2\pi) \int_{\mathbb{R}} e^{-iux} f_\beta^*(u) du$. Assumption (A2) is true for most densities (Gaussian, exponential, Gamma, Laplace) but excludes for example the uniform density. Under (A2), we obtain for all $u \in \mathbb{R}$

$$f_\beta^*(u) = \frac{f_{Z_m}^*(u)}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2},$$

and using the Fourier inversion, we deduce the following closed formula for f_β

$$\forall x \in \mathbb{R}, f_\beta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{f_{Z_m}^*(u)}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} du. \quad (2.2.3)$$

To propose an estimator of f_β , $f_{Z_m}^*$ is replaced by its empirical estimator

$$\forall u \in \mathbb{R}, \widehat{f}_{Z_m}^*(u) = \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m}}. \quad (2.2.4)$$

However, although formula (2.2.3) is well defined under (A1), the integrability of the ratio $\widehat{f}_{Z_m}^*/|f_\varepsilon^*|^2$ is not ensured, since f_ε^* tends to zero near infinity. Therefore, we do not only plug (2.2.4) in equation (2.2.3) but we also introduce a cut-off which avoids integrability problems. We choose $m\Delta$ as a cut-off where m is as before the time of observation subscript m . Finally the estimator is defined by:

$$\widehat{f}_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{\widehat{f}_{Z_m}^*(u)}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} du = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,m}}}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} du. \quad (2.2.5)$$

This original choice of the cut-off impacts the risk bound and leads to unexpectedly better rates for $\widehat{f}_{\beta,m}$, as seen below. In the following, we first compute the MISE of \widehat{f} .

2.2.2 Risk bound for the estimator \widehat{f}

We define the function $f_{\beta,m}$ by the relation

$$f_{\beta,m}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} f_\beta^*(u) du \quad (2.2.6)$$

which amounts to $\mathbb{E}[\widehat{f}_{\beta,m}] = f_{\beta,m}$. Thus $\widehat{f}_{\beta,m}$ is an unbiased estimator of $f_{\beta,m}$.

The bias-variance decomposition is the first step to study the MISE. Then we bound the two terms using Plancherel-Parseval's theorem, and we get the following bound.

Proposition 2.2.1. *Let assumptions (A1) and (A2) hold. Let estimator $\widehat{f}_{\beta,m}$ of f_β be given by (2.2.5). Let us denote $F_\varepsilon := \int_{-1}^1 (1/|f_\varepsilon^*(v)|^4) dv$. Then,*

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du + F_\varepsilon \frac{m\Delta}{2\pi N}. \quad (2.2.7)$$

The proofs are relegated to Section 2.6. The first term of the bound (2.2.7) is the integrated bias term $\|f_{\beta,m} - f_\beta\|^2$. It is due to the replacement of f_β as defined in (2.2.3) by $f_{\beta,m}$ in (2.2.5). This term is decreasing when m is increasing. The second term is the variance term which increases with m and comes from the estimation of $f_{\beta,m}$ by $\widehat{f}_{\beta,m}$. It has the order m/N : it is the order we would obtain in a direct density estimation context. This was unexpected here because we do not observe β . This gain is due to the choice of the cut-off linked to the observation index. Besides $F_\varepsilon = \int_{-1}^1 1/|f_\varepsilon^*(v)|^4 dv$ is a constant which depends on the model. Indeed, under **(A2)**, and because f_ε is a density thus its Fourier transform is continue, F_ε is a known numerical constant. Finally the bound (2.2.7) depends on the density of the noise only through a constant.

The next step is to select the cut-off m among the collection. It follows from Proposition 2.2.1 that we have to perform a bias-variance compromise to minimize the MISE. When f_β is in the Sobolev space with regularity parameter b , i.e. f_β belongs to the set $\mathcal{A}_b(L)$ defined by

$$\mathcal{A}_b(L) = \{f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1+x^2)^b dx \leq L\}$$

with $b > 0, L > 0$. In this particular case, the bias term satisfies:

$$\|f_{\beta,m} - f_\beta\|^2 = \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du \leq \frac{L}{2\pi} (m\Delta)^{-2b}.$$

Consequently, the \mathbb{L}^2 -risk of $\widehat{f}_{\beta,m}$ is bounded by,

$$\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] \leq C_1 (m\Delta)^{-2b} + C_2 \frac{m\Delta}{N}$$

with $C_1 = (L/(2\pi)) > 0$, and $C_2 = F_\varepsilon/(2\pi) > 0$. Therefore, the optimal theoretical choice of $m\Delta$ is $(m\Delta)_{opt} = C_b N^{\frac{1}{2b+1}}$ with

$$C_b = (2bC_1/C_2)^{1/(2b+1)}. \tag{2.2.8}$$

Then we can prove the following corollary of Proposition 2.2.1.

Corollary 2.2.2. *Assume that the assumptions of Proposition 2.2.1 hold, that $f_\beta \in \mathcal{A}_b(L)$, and that $C_b N^{\frac{1}{2b+1}} \leq \mathcal{K}\Delta$, with C_b given by (2.2.8). Then if we choose m_{opt} such that $m_{opt}\Delta = (m\Delta)_{opt} = C_b N^{1/(2b+1)}$, we have*

$$\mathbb{E}[\|\widehat{f}_{\beta,m_{opt}} - f_\beta\|^2] \leq KN^{-\frac{2b}{2b+1}}$$

for a constant K depending on b, L and F_ε .

The order of the risk is $N^{-2b/(2b+1)}$ for a large N , it is the nonparametric estimation rate of convergence obtained when the observation are realizations of the variable of interest. We refer to Lacour (2006) Table 1 for classical rates of convergence for nonparametric deconvolution. Comparing with these rates, the obtained speed of convergence is better in the present settings. This order is $N^{-b/(2b+1)}$ which is always larger than the parametric rate $1/\sqrt{N}$, but can get close to it when b increases. The condition $C_b N^{\frac{1}{2b+1}} \leq \mathcal{K}\Delta$ requires that \mathcal{K} is asymptotically large when N grows, but the larger b the weaker the condition. Nevertheless, this choice is theoretical because it depends on the regularity b of f_β which is unknown.

2.2.3 Adaptive selection of m and oracle risk bound for \widehat{f}

In this section we propose an adaptive selection of m which should outbalance the bias-variance compromise. We aim at finding a procedure that does not depend on the regularity of the density f_β , but only on the data. Suppose that m belongs to a finite set \mathcal{M} , the theoretical optimal choice for m is

$$m_{th} = \operatorname{argmin}_{m \in \mathcal{M}} \{\|f_{\beta,m} - f_\beta\|^2 + V(\widehat{f}_{\beta,m})\} = \operatorname{argmin}_{m \in \mathcal{M}} \{-\|f_{\beta,m}\|^2 + V(\widehat{f}_{\beta,m})\},$$

where we use the Plancherel-Parseval Theorem, and $V(\widehat{f}_{\beta,m}) := \int \mathbb{E}[(\widehat{f}_{\beta,m}(x) - \mathbb{E}[\widehat{f}_{\beta,m}(x)])^2] dx = \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2] \leq F_\varepsilon m \Delta / (2\pi N)$. The two terms appearing in this decomposition are unknown and thus replaced by estimators. The variance term is replaced by a penalty function proportional to the variance bound:

$$\operatorname{pen}^{(1)}(m) = \kappa^{(1)} F_\varepsilon \frac{m \Delta}{N}$$

with $m \in \mathcal{M} := \{1, \dots, N \wedge \mathcal{K}\}$ where $N \wedge \mathcal{K} = \min(N, \mathcal{K})$ and $\kappa^{(1)}$ a numerical constant. From the theoretical point of view $\kappa^{(1)} \geq 24/\pi$, but in practice, $\kappa^{(1)}$ is calibrated by intensive preliminary simulation experiments (see Section 2.4). This choice of \mathcal{M} implies $m \leq N$ and the penalty is bounded.

The dependence in m which holds in several aspects, prevents us from using usual selection method defined by penalized criterion. More precisely, Equation (2.2.5) depends on m in the bounds of the integral, and in the integrative term, thus estimator $\widehat{f}_{\beta,m}$ can not be written as the minimizer of a contrast (see Birgé & Massart, 1998; Massart, 2007). This is why we use a more sophisticated method of selection due to a recent work of Goldenshluger and Lepski (2011). The index m is adaptively chosen as the value which minimizes the sum of two terms as follows

$$\widehat{m}^{(1)} = \operatorname{argmin}_{m \in \mathcal{M}} \{\widehat{\Gamma}_m + \operatorname{pen}^{(1)}(m)\} \quad (2.2.9)$$

with $m \in \mathcal{M} = \{1, \dots, N \wedge \mathcal{K}\}$

$$\begin{aligned} \widehat{\Gamma}_m &:= \max_{1 \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_{\beta, m \wedge k} - \widehat{f}_{\beta, k}\|^2 - \operatorname{pen}^{(1)}(k) \right)_+ \\ &= \max_{m \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_{\beta, m} - \widehat{f}_{\beta, k}\|^2 - \operatorname{pen}^{(1)}(k) \right)_+ \end{aligned} \quad (2.2.10)$$

where $(x)_+ = \max(0, x)$. The contrast $\widehat{\Gamma}_m$ approximates the bias. The penalty term is subtracted in (2.2.10) to offset the variance effect implied by the estimation of the bias term. The final estimator $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ satisfies the following oracle inequality.

Theorem 2.2.3. *Construct an estimator of f_β of the form (2.2.5), and choose $\widehat{m}^{(1)}$ according to (2.2.9). Then, under assumptions (A1) and (A2), there exist two constants C and C' such that*

$$\mathbb{E}[\|\widehat{f}_{\beta, \widehat{m}^{(1)}} - f_\beta\|^2] \leq C \inf_{m \in \mathcal{M}} (\|f_{\beta,m} - f_\beta\|^2 + \operatorname{pen}^{(1)}(m)) + \frac{C'}{N}. \quad (2.2.11)$$

The constant C is numerical as soon as $\kappa^{(1)}$ is fixed, and C' depends on F_ε , Δ , $\|f_\beta\|$.

We prove this result with the same kind of tools as Comte *et al.* (2013) for a continuous time model. Theorem 2.2.3 provides a non-asymptotic result and bound (2.2.11) shows that $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ realizes automatically the bias-variance compromise.

This estimator fulfils the objective: it does not require any assumption on the regularity of f_β or f_ε . The bound we obtain depends on the density f_ε only through the constant F_ε . This

is new in the mixed model framework. Again, note that the variance is of order m/N which is better than a deconvolution order.

We establish the following asymptotic result in the particular case where f_β is in the Sobolev space with regularity b .

Corollary 2.2.4. *Assume that the assumptions of Theorem 2.2.3 hold, that $f_\beta \in \mathcal{A}_b(L)$, and that $C_b N^{\frac{1}{2b+1}} \leq \mathcal{K}\Delta$, with C_b given by (2.2.8). Then there exists a constant $C > 0$ such that*

$$\mathbb{E}[\|\widehat{f}_{\beta, \widehat{m}^{(1)}} - f_\beta\|^2] \leq CN^{-\frac{2b}{2b+1}}$$

Thus, if \mathcal{K} is large enough compared to \sqrt{N} , the estimator reaches the best possible rate for a function with regularity b (Tsybakov (2009) shows that $N^{-b/2b+1}$ is the optimal rate of convergence), without requiring the knowledge of b . This speed is the one we obtain when we observe directly the variable of interest. It is not the classical deconvolution speed which can be logarithmic in the Gaussian case for example (see Comte & Samson, 2012), or $N^{-b/(2b+2\gamma+1)}$ if the noise is ordinary smooth of order γ .

2.2.4 Estimator built using the last observation time $\widehat{f}^\mathcal{K}$

In order to simplify the previous estimator and its selection method, we propose a second estimator. This estimator is based only on the variables defined by (2.2.1) for $m = \mathcal{K}$, $\forall j \in \{1, \dots, N\}$,

$$Z_{j,\mathcal{K}} = \frac{Y_{j,\mathcal{K}} - Y_{j,0}}{\mathcal{K}\Delta} = \beta_j + \frac{\varepsilon_{j,\mathcal{K}} - \varepsilon_{j,0}}{\mathcal{K}\Delta} = \beta_j + W_{j,\mathcal{K}}.$$

The same arguments as before lead to, for any real-value x :

$$f_\beta(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{f_{Z_{\mathcal{K}}}^*(u)}{|f_\varepsilon^*(\frac{u}{\mathcal{K}\Delta})|^2} du.$$

Then we replace $f_{Z_{\mathcal{K}}}^*$ by its empirical estimator and we introduce for some m the cut-off $m\Delta$. The second estimator is defined by the following equation:

$$\widehat{f}_{\beta,m}^\mathcal{K}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,\mathcal{K}}}}{|f_\varepsilon^*(\frac{u}{\mathcal{K}\Delta})|^2} du. \quad (2.2.12)$$

We notice that m is only a cut-off here and no longer an observation time, then it does not need to be an integer. This estimator can also be seen as the argument minimum of a contrast type density. Indeed, for $t \in S_m := \{t \in \mathbb{L}^2, \text{supp}(t^*) = [-m\Delta, m\Delta]\}$, considering

$$\begin{aligned} \gamma(t) &:= \|t\|^2 - \frac{2}{N} \sum_{j=1}^N \frac{1}{2\pi} \int t^*(-u) \frac{e^{iuZ_{j,\mathcal{K}}}}{|f_\varepsilon^*(\frac{u}{\mathcal{K}\Delta})|^2} du \\ &= \|t\|^2 - 2\langle t, \widehat{f}_{\beta,m}^\mathcal{K} \rangle = \|t - \widehat{f}_{\beta,m}^\mathcal{K}\|^2 - \|\widehat{f}_{\beta,m}^\mathcal{K}\|^2, \end{aligned}$$

We have

$$\widehat{f}_{\beta,m}^\mathcal{K} = \underset{t \in S_m}{\operatorname{argmin}} \gamma(t).$$

This new estimator uses $2N$ observations like the previous estimator (2.2.5) when m is fixed. But as seen before the adaptive procedure of $\widehat{f}_{\beta,m}^\mathcal{K}$ uses all the $\mathcal{K}N$ observations while the adaptive procedure for $\widehat{f}_{\beta,m}^\mathcal{K}$ (see below) still uses only $2N$ observations. We have to check if this estimator is competitive. In order to lead a theoretical study, we compute the MISE of $\widehat{f}_{\beta,m}^\mathcal{K}$. Definition (2.2.6) still gives that $\widehat{f}_{\beta,m}^\mathcal{K}$ is an unbiased estimator of $f_{\beta,m}$. The bias-variance decomposition of the risk gives the following proposition.

Proposition 2.2.5. *Let assumptions (A1) and (A2) hold. Let estimator $\widehat{f}_{\beta,m}^{\mathcal{K}}$ of f_{β} be given by (2.2.12). Then, we have,*

$$\mathbb{E}[\|\widehat{f}_{\beta,m}^{\mathcal{K}} - f_{\beta}\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_{\beta}^*(u)|^2 du + \frac{1}{2\pi N} \int_{-m\Delta}^{m\Delta} \frac{du}{|f_{\varepsilon}^*\left(\frac{u}{\mathcal{K}\Delta}\right)|^4}. \quad (2.2.13)$$

The same comments we made after Proposition 2.2.1 apply here. We notice that if the observation time $\mathcal{K}\Delta$ is large enough, as the Fourier transform of a density evaluated in 0 is equal to 1, the variance term is under control. Under the condition $m \leq N \wedge \mathcal{K}$ we can prove the following result.

Corollary 2.2.6. *Let us denote $B_{\varepsilon} := \sup_{v \in [-1,1]} \frac{1}{|f_{\varepsilon}^*(v)|^4}$. If $m \leq N \wedge \mathcal{K}$, then under assumptions (A1) and (A2) one has*

$$\mathbb{E}[\|\widehat{f}_{\beta,m}^{\mathcal{K}} - f_{\beta}\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_{\beta}^*(u)|^2 du + B_{\varepsilon} \frac{m\Delta}{\pi N}. \quad (2.2.14)$$

Assumption (A2) assures that B_{ε} is finite. The dependence on the noise is contained in this constant B_{ε} . The estimator $\widehat{f}_{\beta,m}^{\mathcal{K}}$ is well defined when $m > \mathcal{K}$. Nevertheless, the variance in inequality (2.2.13) can be large in this case, that is the reason why we impose $m \leq N \wedge \mathcal{K}$ to obtain the bound (2.2.14). This assumption yields a variance order m/N , which is the best one can expect.

In the following we propose a penalization procedure. With the assumption of Corollary 2.2.6, the cut-off m is supposed to belong to the finite set $\mathcal{M} = \{1, \dots, N \wedge \mathcal{K}\}$ which is the same as in Section 2.2.3. The theoretical optimal choice of m is defined by

$$m_{th}^{(2)} = \operatorname{argmin}_{m \in \mathcal{M}} \{\|f_{\beta,m} - f_{\beta}\|^2 + V(\widehat{f}_{\beta,m}^{\mathcal{K}})\} = \operatorname{argmin}_{m \in \mathcal{M}} \{-\|f_{\beta,m}\|^2 + V(\widehat{f}_{\beta,m}^{\mathcal{K}})\}$$

where $V(\widehat{f}_{\beta,m}^{\mathcal{K}}) := \mathbb{E}[\|\widehat{f}_{\beta,m}^{\mathcal{K}} - f_{\beta,m}\|^2] \leq B_{\varepsilon} \frac{m\Delta}{\pi N}$. As f_{β} is unknown, so are $\|f_{\beta,m}\|^2$ and $V(\widehat{f}_{\beta,m}^{\mathcal{K}})$. Therefore the value $m_{th}^{(2)}$ must be estimated, and we consider

$$\widehat{m}^{(2)} = \operatorname{argmin}_{m \in \mathcal{M}} \{-\|\widehat{f}_{\beta,m}^{\mathcal{K}}\|^2 + \operatorname{pen}^{(2)}(m)\} = \operatorname{argmin}_{m \in \mathcal{M}} \{\gamma(\widehat{f}_{\beta,m}^{\mathcal{K}}) + \operatorname{pen}^{(2)}(m)\} \quad (2.2.15)$$

with $\operatorname{pen}^{(2)}$ a penalty function proportional to the variance term obtained in the bound (2.2.14):

$$\operatorname{pen}^{(2)}(m) = \kappa^{(2)} B_{\varepsilon} \frac{\Delta m}{N}$$

where $\kappa^{(2)}$ is a numerical constant calibrated in Section 2.4. As before the penalty function is bounded. In this case the data-driven choice of m uses less observations than the estimator defined by (2.2.5): $N\mathcal{K}$ observations are used in the first case, only $2N$ in the second. This method is more natural to realize the bias-variance trade off than the one set up in Section 2.2.1. But we can not set it up for the estimator $\widehat{f}_{\beta,m}$ in a theoretical point of view because of the dependence in m . It is easier to compute and the oracle risk bound we obtain is as good as the last one (2.2.11). In order to evaluate the order of the \mathbb{L}^2 -risk of the final estimator $\widehat{f}_{\beta,\widehat{m}^{(2)}}^{\mathcal{K}}$ we establish the following oracle inequality.

Theorem 2.2.7. *Let assumption (A1) and (A2) hold. Construct an estimator of f_{β} of the form (2.2.12) and choose $\widehat{m}^{(2)}$ according to (2.2.15). Then, there exist two constants C_1 and C_2 such that*

$$\mathbb{E}[\|\widehat{f}_{\beta,\widehat{m}^{(2)}}^{\mathcal{K}} - f_{\beta}\|^2] \leq C_2 \inf_{m \in \mathcal{M}} (\|f_{\beta,m} - f_{\beta}\|^2 + \operatorname{pen}^{(2)}(m)) + \frac{C_1}{N}.$$

The constant C_2 is a numerical constant as soon as $\kappa^{(2)}$ is fixed, and C_1 depends on B_{ε} , Δ , $\|f_{\beta}\|$.

Similarly to Theorem 2.2.3, this oracle inequality is a non-asymptotic result and is true for all N . The choice $\widehat{m}^{(2)}$ achieves automatically the bias penalty trade-off. It is worth stressing that the penalization procedure used to compute $\widehat{f}_{\beta, \widehat{m}^{(2)}}^{\mathcal{K}}$ is simpler and faster than the selection method used for $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ in Section 2.2.3. Thus, we shall check that its practical performances remain good even if the number of observations used here is much more limited.

2.3 Comparisons with previous results

In this section, estimators $\widehat{f}_{\beta, m}$ and $\widehat{f}_{\beta, m}^{\mathcal{K}}$ are compared with the estimator proposed in Comte & Samson (2012). Their study is led in the same context. Thus we can compare the construction, the penalty order and theoretical results. Let us introduce the estimator in Comte & Samson (2012) denoted by $\widetilde{f}_{\beta, m}$, for any real-value x

$$\widetilde{f}_{\beta, m}(x) = \frac{2}{N\mathcal{K}} \sum_{j=1}^N \sum_{k=1}^{\mathcal{K}/2} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{e^{iuU_{j,k}}}{|f_{\varepsilon}^*(\frac{u}{\Delta})|^2} du \quad (2.3.1)$$

with for $j \in \{1, \dots, N\}, k \in \{1, \dots, \mathcal{K}/2\}$ the random variables defined as

$$U_{j,k} := \frac{Y_{j,2k} - Y_{j,2k-1}}{\Delta}.$$

We first compare the construction. This estimator is computed by deconvolution method as the two estimators that we have presented. The main difference is the average over j appearing in the definition (2.3.1). Through this manipulation, this estimator uses all the observations even for a fixed m . The aim of the authors by averaging over j was to improve the variance. Then note that the parameter Δ appears (alone) in the denominator of the estimator. This forbids one to take a Δ too small. For example if the noise is Gaussian $1/|f_{\varepsilon}^*(\frac{u}{\Delta})|^2 = e^{\sigma_{\varepsilon}^2 u^2 / \Delta^2}$ can become very large as soon as $\Delta < 1$. Another remark is that the cut-off πm is independent of the observations as it is the case for $\widehat{f}_{\beta, m}^{\mathcal{K}}$. In our work the cut-off is $m\Delta$. We choose Δ instead of π in order to simplify the calculation.

In a second time we compare the adaptive procedures. The cut-off πm is selected as the minimizer of a penalized contrast, as it is the case for the estimator $\widehat{f}_{\beta, m}^{\mathcal{K}}$. The chosen penalty is composed of two terms, when we have only one. The first term is m/N and the second term is $1/(N\mathcal{K})$ multiplied with an integrative term depending on m , for which assumptions on f_{ε} are needed to determinate the order. On the contrary, for both of our estimators, the penalty functions have the order m/N and are explicit. They are thus simpler. Finally an oracle theorem is obtained in Comte & Samson (2012) assuming the noise is ordinary smooth. This assumption is strong and needed to bound the variance term, and to obtain a MISE order. However, they obtain the optimal rate of convergence of deconvolution cases (in minimax sense Fan (1991)), which is rather slow. In the present work we follow a different strategy aiming to reduce directly the effect of the noise instead of reducing the variance order. Thanks to this approach, no assumption on the noise regularity is needed to obtain Theorems 2.2.3, and 2.2.7. Furthermore, the obtained rates of convergence are similar to those obtained in the case of density estimation with direct observations.

Furthermore, as already said, estimator $\widetilde{f}_{\beta, m}$ given by equation (2.3.1) uses all the available observations while, the two new estimators $\widehat{f}_{\beta, m}$ (when m is fixed) and $\widehat{f}_{\beta, m}^{\mathcal{K}}$ use only $2N$ observations for their construction. However for the data-driven selection we need only $2N$ observations to compute $\widehat{m}^{(2)}$ whereas we need $N\mathcal{K}$ for $\widehat{m}^{(1)}$.

Nevertheless, we can define a field of design for each estimator. At first, note that it is possible to compute $\widehat{f}_{\beta, m}$ when $\Delta < 1$, while this does not seem reasonable with $\widetilde{f}_{\beta, m}$. Then,

estimator $\widehat{f}_{\beta,m}^{\mathcal{K}}$ given by (2.2.12) should benefit from a large total time of observations $\mathcal{K}\Delta$. Thus, when \mathcal{K} is small we should prefer using $\widehat{f}_{\beta,m}$ to $\widetilde{f}_{\beta,m}$ if all the observations are available. When \mathcal{K} is large and Δ is small both $\widehat{f}_{\beta,m}^{\mathcal{K}}$ and $\widetilde{f}_{\beta,m}$ should be efficient. And when only the first and the last time of observation are available we use $\widehat{f}_{\beta,m}^{\mathcal{K}}$. However if \mathcal{K} is large and Δ not too small we would prefer $\widetilde{f}_{\beta,m}$.

2.4 Numerical illustration

2.4.1 Implementation

In this part we implement the three estimators $\widehat{f}_{\beta,m}$, $\widehat{f}_{\beta,m}^{\mathcal{K}}$, and $\widetilde{f}_{\beta,m}$ in order to compare their performances on simulations. We try different scenarios: different designs for $\beta, \alpha, \varepsilon$ (ordinary smooth and super smooth) and different values of the parameters \mathcal{K}, Δ, N , to observe the behavior of each estimator. First we consider the case of known noise density. We consider simulated data according to model (2.1.1). For β we choose the distributions:

- Gaussian $\mathcal{N}(0, 1)$
- mixed Gaussian with $0.3\mathcal{N}(-1, 1/16) + 0.7\mathcal{N}(1, 1/16)$
- Gamma $\Gamma(25, 1/25) \times 5$
- mixed Gamma $(0.3\Gamma(2, 1/2) + 0.7\Gamma(20, 1/5)) \times (1/\sqrt{3})$

for α :

- Gaussian $\mathcal{N}(0, 1)$
- Gamma $\Gamma(25, 1/25) \times 5$

for ε :

- Gaussian $\mathcal{N}(0, \sigma_\varepsilon^2)$ with $\sigma_\varepsilon = 1/4$ and $1/2$
- Laplace $\varepsilon = \sigma_\varepsilon \times Y$ with Y standard Laplace distribution: $f_Y(y) = (1/\sqrt{2}) \exp(-\sqrt{2}|x|)$.

We choose the same distributions as in Comte & Samson (2012) in order to compare our results to theirs. These densities are all associated with variance one. Each estimation calls for Riemann Sums to compute the integrals. They appear in the estimators and in the quantities $F_\varepsilon, B_\varepsilon$. We choose a thin discretization of the interval by taking 200 points.

Then we need to calibrate the penalty constants $\widetilde{\kappa}, \kappa^{(1)}, \kappa^{(2)}$. We refer to Comte & Samson (2012) for the formula of the penalty of estimator \widehat{f} and we denote by $\widetilde{\kappa}$ the associated constant. Classically, the constants $\widetilde{\kappa}, \kappa^{(1)}, \kappa^{(2)}$ are fixed thanks to preliminary simulation experiments. Different cases of function f_β have been investigated with different parameter values, and a large number of repetitions. Comparing the MISE obtained as functions of the constants $\widetilde{\kappa}, \kappa^{(1)}, \kappa^{(2)}$ yields to select values making a good compromise over all experiences.

Finally for $\widehat{f}_{\beta,m}$ we choose $\kappa^{(1)} = 1$, for $\widehat{f}_{\beta,m}^{\mathcal{K}}$: $\kappa^{(2)} = 1$, and for $\widetilde{f}_{\beta,m}$ we choose $\widetilde{\kappa} = 6$.

In Section 2.4.3 we consider that the noise density is unknown, and the numerical results are presented. These new estimators are established in Section 2.6.7.

2.4.2 Numerical simulation results

All the designs discussed in Section 2.4.1 produce the same kind of results, thus we only present the case β Gaussian and mixed Gaussian, with α, ε Gaussian.

First we investigate the results on graphs. Figures 2.1, 2.2, and 2.3, represent two cases: on the left β is Gaussian, on the right β is mixed Gaussian. Figure 2.1 compares the estimator $\widehat{f}_{\beta, \widehat{m}^{(1)}}$ (the bold red line), and all the estimators of the collection $\widehat{f}_{\beta, m}$ (in coloured dashed lines), with the true density of β . We can see that the performance of the selection procedure is visually satisfying. Figure 2.2 represents 10 final estimators $\widehat{f}_{\beta, \widehat{m}^{(2)}}^{\mathcal{K}}$ (green dotted lines) versus the true

density of β . This figure ensures the stability of the estimation procedure. Indeed the beam of estimators is really narrow. Figures 2.3 and 2.4 compare the three estimators for different values of Δ . On Figure 2.3 we see that estimators are really close to each other and close to the true density f_β . Figure 2.4 illustrates the role of the parameter Δ . Indeed, due to its construction, estimator $\tilde{f}_{\beta,m}$ of Comte & Samson (2012) defined by (2.3.1) is more sensitive than the others to a small Δ , which is clear on Figure 2.4.

During this study we also observe differences between estimators when σ_ε increases, especially in the case of a bimodal density f_β . This highlights the role of the noise variance. Estimator $\hat{f}^\mathcal{K}$ is the most stable with respect to the variance effects.

Finally we compare empirical MISE ($\times 100$) of the three estimators. Results are presented in Table 2.1 as a function of \mathcal{K}, N and of the distribution of β . One can notice that the three estimators are really close and seem very competitive. For example, estimator $\hat{f}_{\beta,m}$ has the smallest MISE for $\mathcal{K} = 6, 20$ and β Gaussian with $\sigma_\varepsilon = 1/4$. Besides, for $\mathcal{K} = 6$ the two new estimators obtain smaller MISE than the previous estimator $\tilde{f}_{\beta,m}$. The influence of N is clear: the larger N the smaller the MISE. The MISE are improved when $N = 200$, or 2000 whereas for $N = 20$ especially when β is mixed Gaussian the MISE are high. Nevertheless, the role of \mathcal{K} is less explicit on this table. Indeed, when β is Gaussian, the MISE are smaller when $\mathcal{K} = 6$ than when $\mathcal{K} = 100$. However, when β is mixed Gaussian, the result are a bit better when \mathcal{K} increases. Besides, when $\mathcal{K} = 20$ or 100, estimator $\hat{f}_{\beta,\hat{m}}^\mathcal{K}$ produces the best MISE except for $N = 2000$. More precisely the results are better when $\mathcal{K} = N$ or if \mathcal{K} is of order \sqrt{N} , as we noticed in the theoretical study of $\hat{f}_{\beta,m}$, than when $\mathcal{K} > N$.

At the end, in order to highlight the role of Δ the cases $\Delta = 0.5$ and $\Delta = 1$ are surveyed (Table 2.2). The estimator $\tilde{f}_{\beta,m}$ was expected to be the most sensitive to this change. Table 2.2 shows that for $\Delta = 0.5$ and β Gaussian, $\hat{f}_{\beta,m}^\mathcal{K}$ outperforms, and when β is mixed Gaussian, $\hat{f}_{\beta,m}$ has the smallest MISE .

Then when $\Delta = 1$, if β is Gaussian the competition is between $\hat{f}_{\beta,m}^\mathcal{K}$ and $\hat{f}_{\beta,m}$, and if β is mixed Gaussian $\hat{f}_{\beta,m}^\mathcal{K}$ has the smallest MISE (except when $N = 2000$). Nevertheless, for these values of Δ , especially when β is mixed Gaussian, the results are not as good as before for the three estimators. For $\hat{f}_{\beta,m}^\mathcal{K}$ and $\hat{f}_{\beta,m}$ this may be the consequence of the reduction of the interval of integration $[-m\Delta, m\Delta]$ which becomes smaller than $[-m, m]$ when $\Delta < 1$ and thus the estimation may be affected.

2.4.3 Case of unknown noise density

We briefly consider the case of f_ε unknown. Several papers study the deconvolution method with unknown noise density from a theoretical point of view: Neumann (1997), Comte & Lacour (2011), Meister & Neumann (2010), Kappus & Mabon (2014). Here, we plug estimators of unknown quantities using a transformation of the data adapted from Comte & Samson (2012).

We generalize $\hat{f}_{\beta,m}$ and $\hat{f}_{\beta,m}^\mathcal{K}$ to unknown f_ε by computing new estimators $\widehat{\widehat{f}}_{\beta,m}$ and $\widehat{\widehat{f}}_{\beta,m}^\mathcal{K}$ (see Section 2.6.7 for details). The selection procedures presented in Section 2.2.3 and 2.2.4 are using the constants F_ε and B_ε . These quantities are estimated using relation (2.6.12) established in Section 2.6.7, and we obtain final estimators by applying the previous set up adaptive selection procedure with plugged estimated quantities. Of course, the new estimator $\widehat{\widehat{f}}_{\beta,m}^\mathcal{K}$ requires that intermediate observations between the first and the last are available.

In Table 2.3, we compare empirical MISE of $\widehat{\widehat{f}}_{\beta,m}$, $\widehat{\widehat{f}}_{\beta,m}^\mathcal{K}$ and $\widehat{\widehat{f}}_{\beta,m}$ built in Comte & Samson

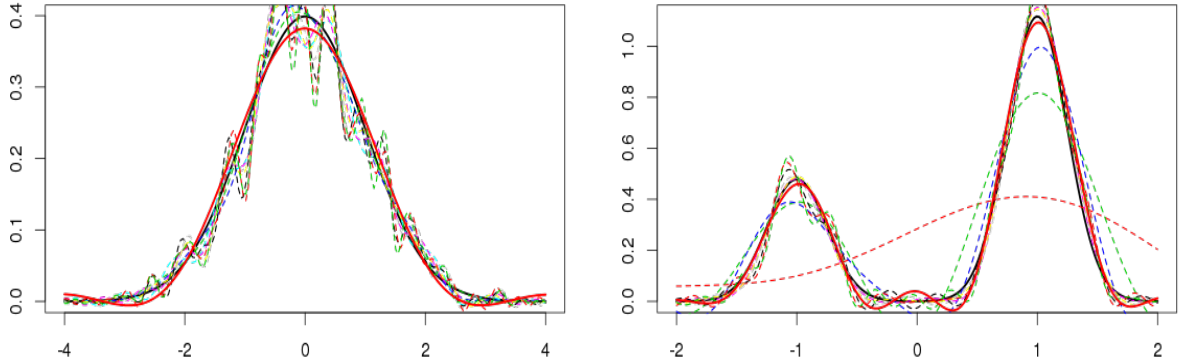


Figure 2.1 – The black bold line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The coloured dashed lines are the proposed estimators $\hat{f}_{\beta, m}$, the red (grey) bold line is the selected estimator $\hat{f}_{\beta, \hat{m}^{(1)}}$, with $N = 500$, $\mathcal{K} = 10$, $\Delta = 2$, $\sigma_\varepsilon = 0.25$

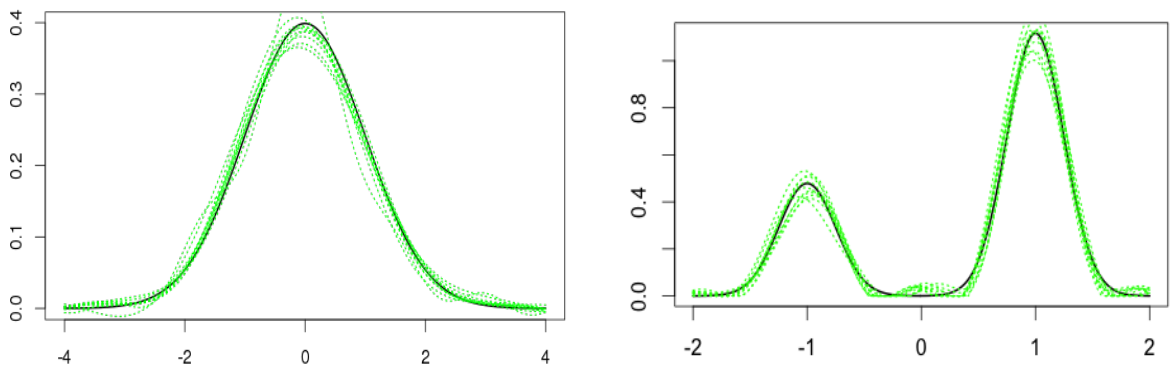


Figure 2.2 – The black bold line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The green (grey) dotted lines: 10 estimators $\hat{f}_{\beta, \hat{m}^{(2)}}^{\mathcal{K}}$ with $N = 500$, $\mathcal{K} = 6$, $\Delta = 2$, $\sigma_\varepsilon = 0.25$

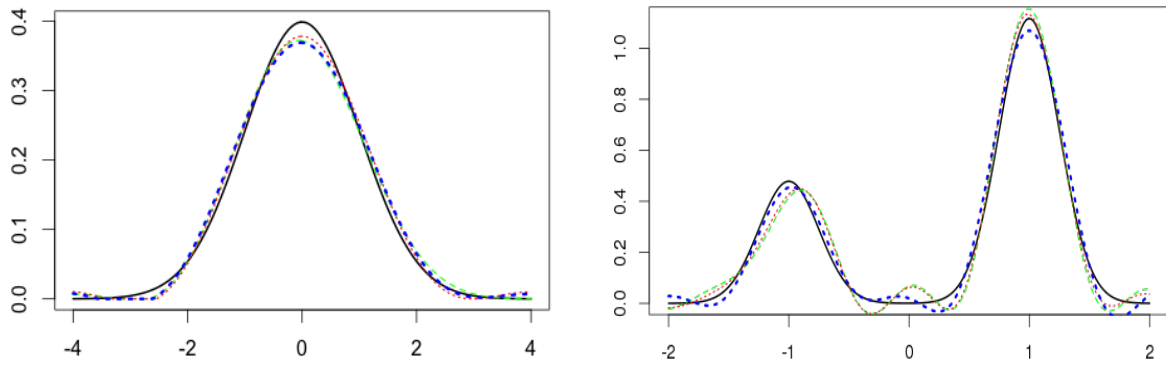


Figure 2.3 – The black line is the true density of β : Gaussian on the left, mixed Gaussian on the right. The green (grey) dashed line is $\tilde{f}_{\beta, \hat{m}}$, the red (green) thin dotted line is $\hat{f}_{\beta, \hat{m}^{(1)}}$, the thick blue (black) dotted line is $\hat{f}_{\beta, \hat{m}^{(2)}}^{\mathcal{K}}$, $N = 500$, $\mathcal{K} = 6$, $\Delta = 2$, $\sigma_{\varepsilon} = 0.25$

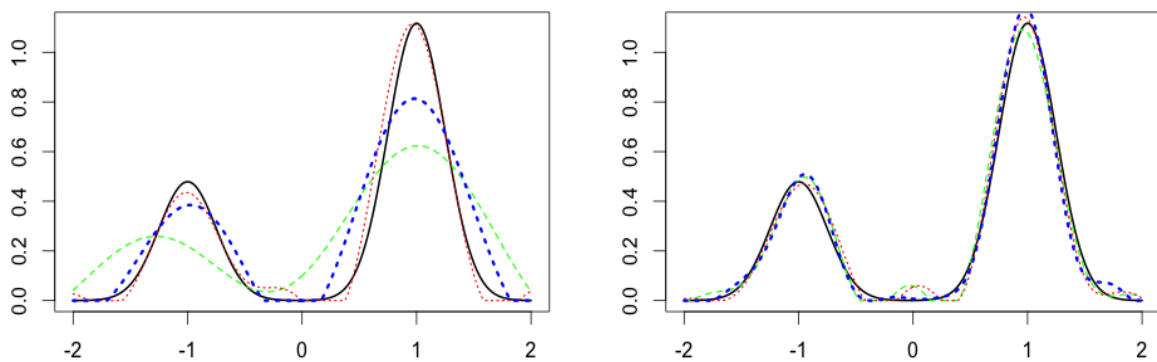


Figure 2.4 – The black bold line is the true density of β : mixed Gaussian. The green (grey) dashed line is $\tilde{f}_{\beta, \hat{m}}$, the red (grey) thin dotted line is $\hat{f}_{\beta, \hat{m}^{(1)}}$, the thick blue (black) dotted line is $\hat{f}_{\beta, \hat{m}^{(2)}}^{\mathcal{K}}$, $N = 500$, $\mathcal{K} = 6$, $\sigma_{\varepsilon} = 0.5$. On the left $\Delta = 1$, on the right $\Delta = 3$

Table 2.1 – Empirical MISE ($\times 100$) computed from 100 simulated data sets with α, ε Gaussian, and $\Delta = 2$ when f_ε is known

Estimator	\mathcal{K}	distribution of β	$\sigma_\varepsilon = 1/4$		$\sigma_\varepsilon = 1/2$		
			$N = 20$	$N = 200$	$N = 20$	$N = 200$	$N = 2000$
$\tilde{f}_{\beta,m}$	6	Gaussian	5.08	0.47	6.78	0.55	0.04
$\hat{f}_{\beta,m}$	6	Gaussian	1.94	0.24	4.21	0.30	0.11
$\hat{f}_{\beta,m}^\mathcal{K}$	6	Gaussian	2.27	0.29	2.88	0.32	0.03
$\tilde{f}_{\beta,m}$	6	Mixed Gaussian	30.0	2.40	34.9	6.30	3.04
$\hat{f}_{\beta,m}$	6	Mixed Gaussian	30.6	3.51	29.9	4.41	0.37
$\hat{f}_{\beta,m}^\mathcal{K}$	6	Mixed Gaussian	9.71	2.77	10.8	2.84	2.42
$\tilde{f}_{\beta,m}$	20	Gaussian	3.71	0.43	4.26	5.74	0.04
$\hat{f}_{\beta,m}$	20	Gaussian	1.78	0.29	2.33	4.14	0.11
$\hat{f}_{\beta,m}^\mathcal{K}$	20	Gaussian	2.44	0.44	2.54	2.76	0.03
$\tilde{f}_{\beta,m}$	20	Mixed Gaussian	16.1	1.51	21.2	5.75	1.39
$\hat{f}_{\beta,m}$	20	Mixed Gaussian	28.8	3.97	29.4	4.10	0.35
$\hat{f}_{\beta,m}^\mathcal{K}$	20	Mixed Gaussian	10.3	1.24	10.0	1.17	0.33
$\tilde{f}_{\beta,m}$	100	Gaussian	3.83	0.36	12.9	0.35	0.53
$\hat{f}_{\beta,m}$	100	Gaussian	4.62	2.61	31.3	0.91	0.51
$\hat{f}_{\beta,m}^\mathcal{K}$	100	Gaussian	2.73	0.27	9.97	0.30	2.22
$\tilde{f}_{\beta,m}$	100	Mixed Gaussian	12.4	1.30	14.2	3.28	0.56
$\hat{f}_{\beta,m}$	100	Mixed Gaussian	29.0	3.67	31.0	4.10	0.40
$\hat{f}_{\beta,m}^\mathcal{K}$	100	Mixed Gaussian	10.1	1.20	10.7	1.24	0.36

Table 2.2 – Empirical MISE ($\times 100$) computed from 100 simulated data sets with α, ε Gaussian, when f_ε is known, $\mathcal{K} = 6$

Estimator	distribution of β	Δ	$\sigma_\varepsilon = 1/4$		$\sigma_\varepsilon = 1/2$		
			$N = 20$	$N = 200$	$N = 20$	$N = 200$	$N = 2000$
$\tilde{f}_{\beta,m}$	Gaussian	0.5	4.60	0.48	6.48	4.11	0.61
$\hat{f}_{\beta,m}$	Gaussian	0.5	3.62	0.34	7.27	0.86	0.10
$\hat{f}_{\beta,m}^\mathcal{K}$	Gaussian	0.5	2.61	0.24	2.46	0.30	0.10
$\tilde{f}_{\beta,m}$	Mixed Gaussian	0.5	36.5	15.4	29.6	35.6	35.5
$\hat{f}_{\beta,m}$	Mixed Gaussian	0.5	29.8	37.1	29.8	4.03	0.50
$\hat{f}_{\beta,m}^\mathcal{K}$	Mixed Gaussian	0.5	35.6	32.6	35.9	32.8	32.75
$\tilde{f}_{\beta,m}$	Gaussian	1	5.10	0.52	5.88	0.53	0.06
$\hat{f}_{\beta,m}$	Gaussian	1	3.83	0.24	4.07	0.24	0.10
$\hat{f}_{\beta,m}^\mathcal{K}$	Gaussian	1	2.65	0.31	2.75	0.31	0.03
$\tilde{f}_{\beta,m}$	Mixed Gaussian	1	31.9	6.76	38.7	23.9	14.1
$\hat{f}_{\beta,m}$	Mixed Gaussian	1	37.6	6.81	37.6	6.96	2.24
$\hat{f}_{\beta,m}^\mathcal{K}$	Mixed Gaussian	1	9.80	6.80	10.97	6.89	6.48

Table 2.3 – Empirical MISE ($\times 100$) computed from 100 simulated data sets with α, ε Gaussian, and $\Delta = 2$, with f_ε unknown and estimated

Estimator	\mathcal{K}	distribution of β	$\sigma_\varepsilon = 1/4$		$\sigma_\varepsilon = 1/2$		
			$N = 20$	$N = 200$	$N = 20$	$N = 200$	$N = 2000$
$\widehat{f}_{\beta,m}$	6	Gaussian	6.37	0.66	6.37	0.54	0.04
$\widehat{\widehat{f}}_{\beta,m}$	6	Gaussian	2.22	0.28	2.32	0.30	0.11
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	6	Gaussian	3.14	0.31	3.53	0.29	0.04
$\widehat{f}_{\beta,m}$	6	mixed Gaussian	25.7	2.75	35.5	6.61	3.78
$\widehat{\widehat{f}}_{\beta,m}$	6	mixed Gaussian	30.4	3.98	29.8	3.60	0.38
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	6	mixed Gaussian	10.3	1.23	10.3	1.36	0.17
$\widehat{f}_{\beta,m}$	20	Gaussian	4.32	0.41	4.16	0.40	0.04
$\widehat{\widehat{f}}_{\beta,m}$	20	Gaussian	1.93	0.25	2.09	0.26	0.11
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	20	Gaussian	2.57	0.30	2.12	0.26	0.04
$\widehat{f}_{\beta,m}$	20	mixed Gaussian	17.5	1.70	27.2	6.12	3.50
$\widehat{\widehat{f}}_{\beta,m}$	20	mixed Gaussian	31.1	4.02	28.8	4.13	0.17
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	20	mixed Gaussian	9.68	1.29	10.2	1.26	0.15
$\widehat{f}_{\beta,m}$	100	Gaussian	3.62	0.38	3.27	0.41	0.38
$\widehat{\widehat{f}}_{\beta,m}$	100	Gaussian	6.62	2.91	5.46	0.74	0.73
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	100	Gaussian	2.136	0.32	2.314	0.33	0.28
$\widehat{f}_{\beta,m}$	100	mixed Gaussian	13.4	1.16	23.3	5.95	3.50
$\widehat{\widehat{f}}_{\beta,m}$	100	mixed Gaussian	27.3	3.51	30.2	4.00	0.23
$\widehat{f}_{\beta,m}^{\mathcal{K}}$	100	mixed Gaussian	10.6	1.10	10.7	1.26	0.14

(2012), as a function of \mathcal{K}, N and the distribution of β . The same comments as for Table 2.1 apply here. But we notice, more precisely, that $\widehat{\widehat{f}}_{\beta,m}$ has the smallest MISE most of the time when $\mathcal{K} = 6, 20$ and β is Gaussian except for $N = 2000$, and it is $\widehat{f}_{\beta,m}^{\mathcal{K}}$ when β mixed Gaussian. Similar plots than before were obtained in this case (not shown).

2.5 Concluding remarks

In summary, in this work we introduce two advanced estimators $\widehat{f}_{\beta,m}$ and $\widehat{f}_{\beta,m}^{\mathcal{K}}$ which are completely data driven. We obtained convincing theoretical results, validated by numerical study. In particular the original selection method set up in Section 2.2.1 leads to an adaptive estimator $\widehat{f}_{\beta,\widehat{m}}$. Furthermore it satisfies an oracle inequality which shows that it realizes automatically the bias-penalty trade-off for all N and all regularity of f_β, f_ε . The second estimator $\widehat{f}_{\beta,m}^{\mathcal{K}}$ seems simpler, nevertheless it produces really good theoretical results and also on simulations, in Section 2.4.2. It also has the advantage of using only the first and the last time of

observation by opposition with $\widehat{f}_{\beta,m}$.

The numerical study confirms the theoretical study: we can define a field of designs N, \mathcal{K} for each estimator. In particular, when \mathcal{K} is small and all the observations are available we will prefer $\widehat{f}_{\beta,m}$. But when $\mathcal{K}\Delta$ is large, and/or when only the first and the last observations are available, we will prefer $\widehat{f}_{\beta,m}^{\mathcal{K}}$.

We may wonder if the estimation of f_α could be improved. Moreover as no independence assumption is set between α and β , the joint distribution of the couple may be estimated over anisotropic regularity spaces. These two questions will be studied in forgoing works.

2.6 Proofs

2.6.1 Proof of Proposition 2.2.1

The MISE of $\widehat{f}_{\beta,m}$ is defined by $\mathbb{E}[\|\widehat{f}_{\beta,m} - f_\beta\|^2] = \|\mathbb{E}[\widehat{f}_{\beta,m}] - f_\beta\|^2 + \mathbb{E}[\|\widehat{f}_{\beta,m} - \mathbb{E}[\widehat{f}_{\beta,m}]\|^2]$. First, according to the definition of $f_{\beta,m}$, $\widehat{f}_{\beta,m}$ is an unbiased estimator of $f_{\beta,m}$, thus the squared integrated bias term is $\|\mathbb{E}[\widehat{f}_{\beta,m}] - f_\beta\|^2 = \|f_{\beta,m} - f_\beta\|^2$. But, for all $x \in \mathbb{R}$:

$$f_\beta(x) - \mathbb{E}(\widehat{f}_{\beta,m}(x)) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} (f_\beta^*(u) - f_{\beta,m}^*(u)) du = \frac{1}{2\pi} \int_{|u| \geq m\Delta} e^{-iux} f_\beta^*(u) du .$$

We obtain with Plancherel-Parserval's theorem

$$\|f_\beta - f_{\beta,m}\|^2 = \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f_\beta^*(u)|^2 du .$$

We study now the term of variance $V(\widehat{f}_{\beta,m}) = \|\mathbb{E}[\widehat{f}_{\beta,m} - \mathbb{E}[\widehat{f}_{\beta,m}]]\|^2 = \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2]$:

$$\begin{aligned} \|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2 &= \left\| \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{\widehat{f}_{Z_m}^*(u) - f_{Z_m}^*(u)}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} du \right\|^2 = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \left| \frac{\widehat{f}_{Z_m}^*(u) - f_{Z_m}^*(u)}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} \right|^2 du \\ &= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \left| \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,m}} - \mathbb{E}[e^{iuZ_{j,m}}]}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} \right|^2 du . \end{aligned}$$

By noticing that the random variables in the last sum are i.i.d, we have

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_{\beta,m} - f_{\beta,m}\|^2] &= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \text{Var} \left(\frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,m}}}{|f_\varepsilon^*(\frac{u}{m\Delta})|^2} \right) du = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \frac{1}{N} \frac{\text{Var}(e^{iuZ_{j,m}})}{|f_\varepsilon^*(\frac{u}{m\Delta})|^4} du \\ &\leq \frac{1}{2\pi N} \int_{-m\Delta}^{m\Delta} \frac{du}{|f_\varepsilon^*(\frac{u}{m\Delta})|^4} = F_\varepsilon \frac{m\Delta}{2\pi N} . \quad \square \end{aligned}$$

2.6.2 Notations

We define

$$\begin{aligned} S_m &= \{t \in \mathbb{L}^2, \text{supp}(t^*) = [-m\Delta, m\Delta]\} \\ B_{m,m'} &= \{t \in S_{m \vee m'}, \|t\| = 1\} \end{aligned}$$

and for $t \in S_m$:

$$\phi_t^k(x) = \frac{1}{2\pi} \int t^*(-u) \frac{e^{iux}}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} du. \quad (2.6.1)$$

Then we define the empirical contrast for an estimator \hat{f} of f , for $t \in S_m$

$$\nu(t) = \frac{1}{2\pi} \langle t^*, (\hat{f}_m - f_m)^* \rangle.$$

In order to make the reading easier we suppress the subscript β , and the exponents (1) in the following.

2.6.3 Proof of Theorem 2.2.3

Let us study the term $\|\hat{f}_{\hat{m}} - f\|^2$. For all $m \in \mathcal{M}$,

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq \left(\|\hat{f}_{\hat{m}} - \hat{f}_{\hat{m} \wedge m}\| + \|\hat{f}_{\hat{m} \wedge m} - \hat{f}_m\| + \|\hat{f}_m - f\| \right)^2 \\ &\leq 3 \left(\|\hat{f}_{\hat{m}} - \hat{f}_{\hat{m} \wedge m}\|^2 + \|\hat{f}_{\hat{m} \wedge m} - \hat{f}_m\|^2 + \|\hat{f}_m - f\|^2 \right). \end{aligned}$$

We are going to examine the three terms involved separately. For the first:

$$\begin{aligned} \|\hat{f}_{\hat{m}} - \hat{f}_{\hat{m} \wedge m}\|^2 &= \|\hat{f}_{\hat{m}} - \hat{f}_m\|^2 \mathbf{1}_{\{\hat{m} \geq m\}} \\ &= (\|\hat{f}_{\hat{m}} - \hat{f}_m\|^2 - \text{pen}(\hat{m})) \mathbf{1}_{\{\hat{m} \geq m\}} + \text{pen}(\hat{m}) \mathbf{1}_{\{\hat{m} \geq m\}} \\ &\leq \hat{\Gamma}_m + \text{pen}(\hat{m}). \end{aligned}$$

After,

$$\begin{aligned} \|\hat{f}_{\hat{m} \wedge m} - \hat{f}_m\|^2 &= \|\hat{f}_{\hat{m}} - \hat{f}_m\|^2 \mathbf{1}_{\{\hat{m} \leq m\}} \\ &= (\|\hat{f}_{\hat{m}} - \hat{f}_m\|^2 - \text{pen}(m)) \mathbf{1}_{\{\hat{m} \leq m\}} + \text{pen}(m) \mathbf{1}_{\{\hat{m} \leq m\}} \\ &\leq \hat{\Gamma}_{\hat{m}} + \text{pen}(m), \end{aligned}$$

what gives

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq 3 \left(\hat{\Gamma}_{\hat{m}} + \text{pen}(m) + \hat{\Gamma}_m + \text{pen}(\hat{m}) + \|\hat{f}_m - f\|^2 \right) \\ &\leq 6(\text{pen}(m) + \hat{\Gamma}_m) + 3\|\hat{f}_m - f\|^2. \end{aligned} \tag{2.6.2}$$

Now we study $\hat{\Gamma}_m$. For $k \geq m$, we have

$$\|\hat{f}_k - \hat{f}_m\|^2 \leq 3 \left(\|\hat{f}_k - f_k\|^2 + \|f_k - f_m\|^2 + \|f_m - \hat{f}_m\|^2 \right).$$

According to Plancherel-Parseval's theorem

$$\begin{aligned} \|f_k - f_m\|^2 &= \frac{1}{2\pi} \|f_k^* - f_m^*\|^2 = \frac{1}{2\pi} \int_{m\Delta \leq |u| \leq k\Delta} |f^*(u)|^2 du \\ &\leq \frac{1}{2\pi} \int_{|u| \geq m\Delta} |f^*(u)|^2 du = \|f - f_m\|^2. \end{aligned}$$

Besides

$$\begin{aligned} \|\hat{f}_m - f\|^2 &= \frac{1}{2\pi} \|\hat{f}_m^* - f^*\|^2 = \frac{1}{2\pi} \|(\hat{f}_m^* - f_m^*) + (f_m^* - f^*)\|^2 \\ &= \|\hat{f}_m - f_m\|^2 + \|f_m - f\|^2. \end{aligned}$$

Thus

$$\|\hat{f}_k - \hat{f}_m\|^2 \leq 3(\|\hat{f}_k - f_k\|^2 + \|\hat{f}_m - f\|^2).$$

Finally,

$$\begin{aligned}
 \widehat{\Gamma}_m &= \max_{m \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_m - \widehat{f}_k\|^2 - \text{pen}(k) \right)_+ \\
 &\leq \max_{m \leq k \leq N \wedge \mathcal{K}} \left(3\|\widehat{f}_k - f_k\|^2 + 3\|\widehat{f}_m - f\|^2 - \text{pen}(k) \right)_+ \\
 &\leq 3 \max_{m \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_k - f_k\|^2 - \frac{1}{3}\text{pen}(k) \right)_+ + 3\|\widehat{f}_m - f\|^2
 \end{aligned} \tag{2.6.3}$$

Finally, gathering (2.6.2) and (2.6.3) we obtain

$$\begin{aligned}
 \|\widehat{f}_{\widehat{m}} - f\|^2 &\leq 6 \left(3 \max_{m \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_k - f_k\|^2 - \frac{1}{3}\text{pen}(k) \right)_+ + 3\|\widehat{f}_m - f\|^2 + \text{pen}(m) \right) + 3\|\widehat{f}_m - f\|^2 \\
 &\leq 18 \max_{m \leq k \leq N \wedge \mathcal{K}} \left(\|\widehat{f}_k - f_k\|^2 - \frac{1}{3}\text{pen}(k) \right)_+ + 21\|\widehat{f}_m - f\|^2 + 6\text{pen}(m).
 \end{aligned}$$

The Proposition 2.2.1 implies for all $m \in \{1, \dots, N \wedge \mathcal{K}\}$,

$$\begin{aligned}
 \mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] &\leq 18 \sum_{k=m}^{\mathcal{K}} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - \frac{\text{pen}(k)}{3} \right)_+ \right] + 21\|f_m - f\|^2 + 6 \text{pen}(m) + \frac{21\Delta}{2\pi} F_\varepsilon \frac{m}{N} \\
 &\leq 18 \sum_{k=m}^{\mathcal{K}} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - \frac{\text{pen}(k)}{3} \right)_+ \right] + 21\|f_m - f\|^2 + \left(6 + \frac{21}{2\pi\kappa} \right) \text{pen}(m).
 \end{aligned} \tag{2.6.4}$$

It remains to bound from above the term $\sum_{k=m}^{N \wedge \mathcal{K}} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - \text{pen}(k)/3 \right)_+ \right]$, this is the aim of Lemma 2.6.1.

Lemma 2.6.1. *There exists a constant C'' such that for all $m \in \mathcal{M}$,*

$$\sum_{k=m}^{N \wedge \mathcal{K}} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - \text{pen}(k)/3 \right)_+ \right] \leq \frac{C''}{N}.$$

Lemma 2.6.1 and inequality (2.6.4) imply that, for all $m \in \mathcal{M}$

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leq 21\|f_m - f\|^2 + \left(6 + \frac{21}{2\pi\kappa} \right) \text{pen}(m) + 18 \frac{C''}{N}.$$

Thus, in particular

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leq C \inf_{m \in \mathcal{M}} (\|f_m - f\|^2 + \text{pen}(m)) + \frac{C'}{N}$$

with C and C' two constants. \square

2.6.4 Proof of Lemma 2.6.1

Consider $k \in \mathcal{M}$, and $t \in S_k$, define the empirical contrast:

$$\begin{aligned}
 \nu(t) &= \frac{1}{2\pi} \langle t^*, (\widehat{f}_k - f_k)^* \rangle \\
 &= \frac{1}{2\pi} \int_{\mathbb{R}} t^*(-u) \left(\frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,k}}}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} \mathbf{1}_{[-k\Delta, k\Delta]}(u) - f_k^*(u) \right) du \\
 &= \frac{1}{2\pi} \frac{1}{N} \sum_{j=1}^N \int_{-k\Delta}^{k\Delta} t^*(-u) \frac{e^{iuZ_{j,k}} - \mathbb{E}[e^{iuZ_{j,k}}]}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} du \\
 &= \frac{1}{N} \sum_{j=1}^N \phi_t^k(Z_{j,k}) - \mathbb{E}[\phi_t^k(Z_{j,k})].
 \end{aligned}$$

Let us define $A_k = \{t \in S_k, \|t\| = 1\}$, then $\|\widehat{f}_k - f_k\|^2 = \sup_{t \in A_k} |\nu(t)|^2$. We use Talagrand's inequality (see Annexe B) to prove the lemma. We study first the upper bound:

$$\begin{aligned}
 \sup_{t \in A_k} \|\phi^k\|_\infty &\leq \sup_{t \in A_k} \frac{1}{2\pi} \int_{\mathbb{R}} |t^*(u)| \frac{\mathbf{1}_{[-k\Delta, k\Delta]}(u)}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} du \\
 &\leq \sup_{t \in A_k} \frac{1}{2\pi} \left(\int_{\mathbb{R}} |t^*(u)|^2 du \right)^{1/2} \left(\int_{-k\Delta}^{k\Delta} \frac{1}{|f_\varepsilon^*(\frac{u}{k\Delta})|^4} du \right)^{1/2} \\
 &\leq \sup_{t \in A_k} \frac{1}{\sqrt{2\pi}} \|t\| \left(\int_{-k\Delta}^{k\Delta} \frac{du}{|f_\varepsilon^*(\frac{u}{k\Delta})|^4} \right)^{1/2} \\
 &\leq \frac{1}{\sqrt{2\pi}} \left(\int_{-k\Delta}^{k\Delta} \frac{du}{|f_\varepsilon^*(\frac{u}{k\Delta})|^4} \right)^{1/2} = \sqrt{\frac{\Delta}{2\pi}} \sqrt{kF_\varepsilon} := M.
 \end{aligned} \tag{2.6.5}$$

Then according to Proposition 2.2.1, for all $k \in \mathcal{M}$

$$\mathbb{E}[\|\widehat{f}_k - f_k\|^2] \leq F_\varepsilon \frac{k\Delta}{2\pi N},$$

thus

$$\mathbb{E}[\sup_{t \in A_k} |\nu(t)|^2] = \mathbb{E}[\|\widehat{f}_k - f_k\|^2] \leq F_\varepsilon \frac{k\Delta}{2\pi N} := H^2. \tag{2.6.6}$$

Finally we examine the term:

$$\begin{aligned}
 4\pi^2 \text{Var} \phi_t^k(Z_{j,k}) &\leq 4\pi^2 \mathbb{E} \left(\phi_t^k(Z_{j,k}) \overline{\phi_t^k(Z_{j,k})} \right) \\
 &\leq \mathbb{E} \left[\left(\int_{-k\Delta}^{k\Delta} t^*(-u) \frac{e^{iuZ_{j,k}}}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} du \right) \overline{\left(\int_{-k\Delta}^{k\Delta} t^*(-v) \frac{e^{ivZ_{j,k}}}{|f_\varepsilon^*(\frac{v}{k\Delta})|^2} dv \right)} \right] \\
 &= \mathbb{E} \left[\left(\int_{-k\Delta}^{k\Delta} t^*(-u) \frac{e^{iuZ_{j,k}}}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2} du \right) \left(\int_{-k\Delta}^{k\Delta} t^*(v) \frac{e^{-ivZ_{j,k}}}{|f_\varepsilon^*(\frac{v}{k\Delta})|^2} dv \right) \right] \\
 &= \mathbb{E} \left[\int_{-k\Delta}^{k\Delta} \int_{-k\Delta}^{k\Delta} t^*(-u) t^*(v) \frac{e^{i(u-v)Z_{j,k}}}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2 |f_\varepsilon^*(\frac{v}{k\Delta})|^2} dudv \right].
 \end{aligned}$$

Yet $\mathbb{E}(e^{i(u-v)Z_{j,k}}) = f_Z^*(u-v) = f^*(u-v) |f_\varepsilon^*(\frac{u-v}{k\Delta})|^2$, with Fubini's Theorem, it implies

$$4\pi^2 \text{Var}(\phi^k(Z_{j,k})) \leq \int_{-k\Delta}^{k\Delta} \int_{-k\Delta}^{k\Delta} t^*(-u) t^*(v) \frac{|f_\varepsilon^*(\frac{u-v}{k\Delta})|^2}{|f_\varepsilon^*(\frac{u}{k\Delta})|^2 |f_\varepsilon^*(\frac{v}{k\Delta})|^2} f^*(u-v) dudv.$$

Under assumption **(A2)** we obtain:

$$\begin{aligned} \sup_{u,v \in [-k\Delta, k\Delta]} \left| \frac{f_\varepsilon^*(\frac{u-v}{k\Delta})}{f_\varepsilon^*(\frac{u}{k\Delta})f_\varepsilon^*(\frac{v}{k\Delta})} \right|^2 &= \sup_{x,y \in [-1,1]} \left| \frac{f_\varepsilon^*(x-y)}{f_\varepsilon^*(x)f_\varepsilon^*(y)} \right|^2 \leq \sup_{x,y \in [-1,1]} \left| \frac{1}{f_\varepsilon^*(x)f_\varepsilon^*(y)} \right|^2 \\ &\leq \sup_{x \in [-1,1]} \left| \frac{1}{f_\varepsilon^*(x)} \right|^4 := C^2 < +\infty. \end{aligned}$$

thus

$$\begin{aligned} 4\pi^2 \sup_{t \in A_k} \frac{1}{N} \sum_{j=1}^N \text{Var}(\phi^k(Z_{j,k})) &\leq \sup_{t \in A_k} \int_{-k\Delta}^{k\Delta} \int_{-k\Delta}^{k\Delta} C^2 |t^*(-u)t^*(v)f^*(u-v)| dudv \\ &\stackrel{C.S}{\leq} \sup_{t \in A_k} 2\pi C^2 \|t\|^2 \left(\int_{-k\Delta}^{k\Delta} \int_{-k\Delta}^{k\Delta} |f^*(u-v)|^2 dudv \right)^{1/2} \\ &\leq 2\pi C^2 \left(\int_{-2k\Delta}^{2k\Delta} \int_{-k\Delta}^{k\Delta} |f^*(z)|^2 dudz \right)^{1/2} \\ &\leq C^2 4\pi \sqrt{\pi\Delta} \|f\| \sqrt{k} := 4\pi^2 v, \end{aligned}$$

with

$$v := \frac{C^2 \|f\|}{\sqrt{\pi}} \sqrt{k\Delta}. \quad (2.6.7)$$

We achieve with the Talagrand's theorem, gathering the bounds (2.6.5), (2.6.6), (2.6.7), for $\alpha = 1/2$,

$$\begin{aligned} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - 4H^2 \right)_+ \right] &\leq \frac{4}{K_1} \left(\frac{C^2 \|f\| \sqrt{k\Delta}}{\sqrt{\pi}} \frac{F_\varepsilon}{N} \exp \left(-\frac{K_1}{4} \frac{F_\varepsilon}{\sqrt{\pi} C^2 \|f\|} \sqrt{k\Delta} \right) \right. \\ &\quad \left. + \frac{49\Delta F_\varepsilon}{2\pi K_1 C^2 (1/2)} \frac{k}{N^2} \exp \left(-\frac{K_1 C(1/2)}{7\sqrt{2}} \sqrt{N} \right) \right) \end{aligned}$$

and

$$\mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - 4H^2 \right)_+ \right] \leq C_1 C^2 \|f\| \sqrt{\Delta} \frac{\sqrt{k}}{N} \exp \left(-C_2 \frac{F_\varepsilon \sqrt{\Delta}}{C^2 \|f\|} \sqrt{k} \right) + C_3 F_\varepsilon \Delta \frac{k}{N^2} \exp \left(-C_4 \sqrt{N} \right)$$

where C_i are positive numerical constants. Then we sum this inequality if $\kappa \geq 24/\pi$

$$\begin{aligned} \sum_{k=m}^{\kappa} \mathbb{E} \left(\|\widehat{f}_k - f_k\|^2 - \text{pen}(k)/3 \right)_+ &\leq \sum_{k=m}^{\kappa} \left(C_1 C^2 \|f\| \sqrt{\Delta} \frac{\sqrt{k}}{N} \exp \left(-C_2 \frac{F_\varepsilon \sqrt{\Delta}}{C^2 \|f\|} \sqrt{k} \right) \right. \\ &\quad \left. + C_3 F_\varepsilon \Delta \frac{k}{N^2} \exp \left(-C_4 \sqrt{N} \right) \right) \end{aligned}$$

We remark that $1 \leq k \leq N \wedge \mathcal{K}$ implies:

$$\begin{aligned} \sum_{k=m}^{N \wedge \mathcal{K}} \mathbb{E} \left[\left(\|\widehat{f}_k - f_k\|^2 - \text{pen}(k)/3 \right)_+ \right] &\leq \frac{C_1 C^2 \|f\| \sqrt{\Delta}}{N} \sum_{k=1}^{+\infty} \sqrt{k} e^{-\left(C_2 \frac{F_\varepsilon \sqrt{\Delta}}{C^2 \|f\|} \right) \sqrt{k}} \\ &\quad + \frac{C_3 F_\varepsilon \Delta}{N^2} \sum_{k=1}^{+\infty} k e^{-C_4 \sqrt{k}}. \end{aligned}$$

The sum $\sum_{k=0}^{+\infty} ke^{-C\sqrt{k}}$ converges $\forall C > 0$, there exists a constant C'' such as

$$18 \sum_{k=m}^{N \wedge K} \mathbb{E} \left(\|\widehat{f}_k - f_k\|^2 - \text{pen}(k)/3 \right)_+ \leq \frac{C''}{N} . \square$$

2.6.5 Proof of Theorem 2.2.7

We also omit the subscript β , and the exponents (2) in the following.

For $t \in S_m$, and $m \in \mathcal{M}$, we consider the contrast γ (independent on m)

$$\begin{aligned} \gamma(t) &= \|t\|^2 - \frac{2}{N} \sum_{j=1}^N \frac{1}{2\pi} \int t^*(-u) \frac{e^{iuZ_{j,\mathcal{K}}}}{|f_\varepsilon^*(\frac{u}{\mathcal{K}\Delta})|^2} du \\ &= \|t\|^2 - \frac{2}{N} \sum_{j=1}^N \phi_t^\mathcal{K}(Z_{j,\mathcal{K}}) \\ &= \|t\|^2 - 2 \langle t, \widehat{f}_m^\mathcal{K} \rangle = \|t - \widehat{f}_m^\mathcal{K}\|^2 - \|\widehat{f}_m^\mathcal{K}\|^2. \end{aligned}$$

Obviously $\gamma(\widehat{f}_m^\mathcal{K}) = -\|\widehat{f}_m^\mathcal{K}\|^2$. We define

$$\nu(t) := \frac{1}{2\pi} \langle t^*, (\widehat{f}_m^\mathcal{K} - f_m)^* \rangle = \frac{1}{N} \sum_{j=1}^N \phi_t^\mathcal{K}(Z_{j,\mathcal{K}}) - \mathbb{E}[\phi_t^\mathcal{K}(Z_{j,\mathcal{K}})]$$

then

$$(\gamma(t) - \|t - f\|^2) - (\gamma(s) - \|s - f\|^2) = -2\nu(t - s) \quad (2.6.8)$$

Thanks to the definition (2.2.15) of \widehat{m} : $\gamma(\widehat{f}_m^\mathcal{K}) + \text{pen}(\widehat{m}) \leq \gamma(f_m) + \text{pen}(m)$. Using the equation (2.6.8) we obtain

$$\|\widehat{f}_m^\mathcal{K} - f\|^2 \leq \|f_m - f\|^2 + 2\nu(\widehat{f}_m^\mathcal{K} - f_m) + \text{pen}(m) - \text{pen}(\widehat{m}).$$

Let us remark that

$$\nu(\widehat{f}_m^\mathcal{K} - f_m) = \|\widehat{f}_m^\mathcal{K} - f_m\| \nu \left(\frac{\widehat{f}_m^\mathcal{K} - f_m}{\|\widehat{f}_m^\mathcal{K} - f_m\|} \right).$$

This leads us to

$$\begin{aligned} \|\widehat{f}_m^\mathcal{K} - f\|^2 &\leq \|f_m - f\|^2 + 2\|\widehat{f}_m^\mathcal{K} - f_m\| \nu \left(\frac{\widehat{f}_m^\mathcal{K} - f_m}{\|\widehat{f}_m^\mathcal{K} - f_m\|} \right) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq \|f_m - f\|^2 + 2\|\widehat{f}_m^\mathcal{K} - f_m\| \sup_{t \in B_{m,\widehat{m}}} \nu(t) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq \|f_m - f\|^2 + \frac{1}{4}\|\widehat{f}_m^\mathcal{K} - f_m\|^2 + 4 \sup_{t \in B_{m,\widehat{m}}} \nu^2(t) + \text{pen}(m) - \text{pen}(\widehat{m}) \\ &\leq 3\|f_m - f\|^2 + 8 \sup_{t \in B_{m,\widehat{m}}} \nu^2(t) + 2\text{pen}(m) - 2\text{pen}(\widehat{m}) \\ &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 8 \left[\sup_{t \in B_{m,\widehat{m}}} \nu^2(t) - p(m, \widehat{m}) \right] \\ &\leq 3\|f_m - f\|^2 + 4\text{pen}(m) + 8 \sum_{m' \in \mathcal{M}} \left(\sup_{t \in B_{m,m'}} \nu^2(t) - p(m, m') \right)_+ \end{aligned}$$

with a function p such that $\forall m, m', 4p(m, m') \leq \text{pen}(m) + \text{pen}(m')$.

Lemma 2.6.2. *There exists a constant $C > 0$ such that*

$$\sum_{m' \in \mathcal{M}} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - p(m, m') \right)_+ \right] \leq \frac{C}{N}.$$

We conclude that there exist two constants $C_1, C_2 > 0$ such that

$$\mathbb{E}[\|\widehat{f}_{\widehat{m}} - f\|^2] \leq C_1 \inf_{m \in \mathcal{M}} \{\|f_m - f\|^2 + \text{pen}(m)\} + \frac{C_2}{N}.$$

2.6.6 Proof of the Lemma 2.6.2

We remind that $B_\varepsilon = \sup_{v \in [-1,1]} \frac{1}{|f_\varepsilon^*(v)|^4}$. For $m \in \mathcal{M}$ we consider $t \in S_m$. We use Talagrand's inequality to demonstrate Lemma 2.6.2 similarly to Lemma 2.6.1. We get:

$$M := \sqrt{\frac{\Delta B_\varepsilon m^*}{\pi}}, \tag{2.6.9}$$

$$H := \sqrt{\frac{\Delta B_\varepsilon m^*}{\pi N}}, \tag{2.6.10}$$

and

$$v = \frac{B_\varepsilon^2 \|f\|}{\sqrt{\pi}} \sqrt{m^* \Delta}. \tag{2.6.11}$$

According to Talagrand's inequality, for $\alpha = 1/2$ there exist four numerical constants A_1, A_2, A_3, A_4 such that:

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - 4H^2 \right)_+ \right] &\leq A_1 B_\varepsilon^2 \sqrt{\Delta} \frac{\sqrt{m^*}}{N} \exp \left(-A_2 \frac{\sqrt{\Delta}}{2B_\varepsilon \|f\|} \sqrt{m^*} \right) \\ &\quad + A_3 \Delta B_\varepsilon \frac{m^*}{N^2} \exp(-A_4 \sqrt{N}). \end{aligned}$$

Then we use that $1 \leq m^* \leq N \wedge \mathcal{K}$, thus for $\kappa \geq 4/\pi$ we obtain there exists a constant $C > 0$ such that

$$\begin{aligned} \sum_{m' \in \mathcal{M}} \mathbb{E} \left[\left(\sup_{t \in B_{m,m'}} \nu^2(t) - p(m, m') \right)_+ \right] &\leq \sum_{m=1}^{+\infty} A_1 B_\varepsilon^2 \frac{\sqrt{m\Delta}}{N} \exp \left(-A_2 \frac{\sqrt{m\Delta}}{2B_\varepsilon \|f\|} \right) \\ &\quad + \sum_{m=1}^{+\infty} A_3 B_\varepsilon \frac{m\Delta}{N^2} \exp(-A_4 \sqrt{m}) \\ &\leq \frac{C}{N} \end{aligned}$$

with $p(m, m') = 4 \frac{B_\varepsilon \Delta m^*}{\pi N}$. \square

2.6.7 Estimator with unknown noise density

We compute an estimator of the density f_β when f_ε is unknown, corresponding to (2.2.5). We use the same idea as Comte & Samson (2012). However it is $1/|f_\varepsilon^*(\frac{u}{m\Delta})|^2$ which has to be estimated and not $1/(f_\varepsilon^*(\frac{u}{\Delta}))^2$. This quantity is more reachable and the noise does not need

to be symmetric for the estimation. In this part we stand assumption **(A2)** for: **(A2')** f_ε is unknown and $f_\varepsilon^* \neq 0$.

We consider the random variables

$$A_j := \Delta(3Z_{j,3} - 2Z_{j,2} - Z_{j,1}) = \varepsilon_{j,3} - \varepsilon_{j,2} - \varepsilon_{j,1} + \varepsilon_{j,0}.$$

These new variables are i.i.d. with a density f_A . Then we have

$$\forall u \in \mathbb{R}, f_A^*(u) = \mathbb{E}[e^{iAu}] = |f_\varepsilon^*(u)|^4 = \mathbb{E}[\cos(Au)].$$

This leads to the unbiased empirical estimator of $|f_\varepsilon^*|^4$:

$$\widehat{|f_\varepsilon^*|^4}(u) = \frac{1}{N} \sum_{j=1}^N \cos(A_j u).$$

Nevertheless, this function can reach zero and has to be inverted. We define a truncated estimator following Comte & Lacour (2011) and we propose the estimator of $|1/f_\varepsilon^*|^2$ following:

$$\frac{1}{\widetilde{|f_\varepsilon^*|^2}(u)} = \frac{\mathbf{1}_{\widehat{|f_\varepsilon^*|^4}(u) \geq N^{-1/2}}}{[\widehat{|f_\varepsilon^*|^4}(u)]^{1/2}}. \quad (2.6.12)$$

The new estimator of f_β is

$$\forall x \in \mathbb{R}, \widehat{\widehat{f_{\beta,m}}}(x) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,m}}}{\widetilde{|f_\varepsilon^*|^2}\left(\frac{u}{m\Delta}\right)} du. \quad (2.6.13)$$

Proposition 2.6.3. *If $m > 3$, and denoting $K_\varepsilon := 5F_\varepsilon + 68G_\varepsilon$, where $G_\varepsilon := \int_{-1}^1 \frac{du}{|f_\varepsilon^*(u)|^8}$, under **(A1)**-**(A2')** we have*

$$\mathbb{E}[\|\widehat{\widehat{f_{\beta,m}}} - f_\beta\|^2] \leq \|f_{\beta,m} - f_\beta\|^2 + K_\varepsilon \frac{m\Delta}{\pi N}.$$

Proof

We denote

$$R(u) := \frac{1}{\widetilde{|f_\varepsilon^*|^2}(u)} - \frac{1}{|f_\varepsilon^*(u)|^2}$$

which is the subject of the following lemma. This lemma is an extension of the Neumann's Lemma (Neumann, 1997), demonstrated in Comte & Samson (2012) with a symmetric noise. The proof can be straightforwardly adapted and thus is omitted.

Lemma 2.6.4.

$$\mathbb{E} \left[\left(\frac{1}{\widetilde{|f_\varepsilon^*|^2}(u)} - \frac{1}{|f_\varepsilon^*(u)|^2} \right)^2 \right] \leq \frac{2}{|f_\varepsilon^*(u)|^4} \wedge \frac{3N^{-1/2}}{|f_\varepsilon^*(u)|^8} \wedge \frac{34N^{-1}}{|f_\varepsilon^*(u)|^{12}}.$$

We notice that $R(u)$ and $\widehat{f_{Z_m}^*} = 1/N \sum_{j=1}^N e^{iuZ_{j,m}}$ are independent as soon as $m > 3$. Let us decompose the norm: $\|\widehat{\widehat{f_{\beta,m}}} - f_\beta\|^2$.

$$\begin{aligned} \|\widehat{\widehat{f_{\beta,m}}} - f_\beta\|^2 &= \|\widehat{\widehat{f_{\beta,m}}} - f_{\beta,m}\|^2 + \|f_{\beta,m} - f_\beta\|^2 \\ &\leq 2\|\widehat{\widehat{f_{\beta,m}}} - \widehat{f_{\beta,m}}\|^2 + 2\|\widehat{f_{\beta,m}} - f_{\beta,m}\|^2 + \|f_{\beta,m} - f_\beta\|^2. \end{aligned}$$

The first member of the right hand side is simplified by Plancherel-Parseval's Theorem

$$\begin{aligned}
\|\widehat{f_{\beta,m}} - \widehat{f_{\beta_m}}\|^2 &= \left\| \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} e^{-iux} \widehat{f_{Z_m}^*}(u) R\left(\frac{u}{m\Delta}\right) du \right\|^2 \\
&= \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \left| \widehat{f_{Z_m}^*}(u) R\left(\frac{u}{m\Delta}\right) \right|^2 du \\
&\leq \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |\widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u)|^2 \left| R\left(\frac{u}{m\Delta}\right) \right|^2 du \\
&+ \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |f_{\beta}^*(u)|^2 \left| f_{\varepsilon}^*\left(\frac{u}{m\Delta}\right) \right|^4 \left| R\left(\frac{u}{m\Delta}\right) \right|^2 du
\end{aligned}$$

note that

$$\mathbb{E} \left[\left| \widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u) \right|^2 \right] = \text{Var} \left(\widehat{f_{Z_m}^*}(u) - f_{Z_m}^*(u) \right) = \text{Var} \left(\frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m}} \right) \leq \frac{1}{N}.$$

Then, using Lemma 2.6.4, we obtain

$$\begin{aligned}
\mathbb{E}[\|\widehat{f_{\beta,m}} - \widehat{f_{\beta,m}}\|^2] &\leq \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} \frac{1}{N} \mathbb{E} \left[\left| R\left(\frac{u}{m\Delta}\right) \right|^2 \right] du + \frac{1}{\pi} \int_{-m\Delta}^{m\Delta} |f_{\beta}^*(u)|^2 \frac{3N^{-1/2}}{|f_{\varepsilon}^*(\frac{u}{m\Delta})|^4} \wedge \frac{34N^{-1}}{|f_{\varepsilon}^*(\frac{u}{m\Delta})|^8} du \\
&\leq \frac{2}{\pi N} \int_{-m\Delta}^{m\Delta} \frac{1}{|f_{\varepsilon}^*(\frac{u}{m\Delta})|^4} du \\
&+ \frac{1}{\pi} \left(\frac{3}{\sqrt{N}} \int_{-m\Delta}^{m\Delta} \frac{|f_{\beta}^*(u)|^2}{|f_{\varepsilon}^*(\frac{u}{m\Delta})|^4} du \right) \wedge \left(\frac{34}{N} \int_{-m\Delta}^{m\Delta} \frac{|f_{\beta}^*(u)|^2}{|f_{\varepsilon}^*(\frac{u}{m\Delta})|^8} du \right) \\
&\leq \frac{2m\Delta}{\pi N} F_{\varepsilon} + 34 \frac{m\Delta}{\pi N} G_{\varepsilon}.
\end{aligned}$$

Finally, using Proposition 2.2.1, it yields

$$\mathbb{E}[\|\widehat{f_{\beta,m}} - f_{\beta}\|^2] \leq \|f_{\beta,m} - f_{\beta}\|^2 + \frac{m\Delta}{\pi N} F_{\varepsilon} + 4 \frac{m\Delta}{\pi N} F_{\varepsilon} + 68 \frac{m\Delta}{\pi N} G_{\varepsilon}. \quad \square$$

The same construction can be used to propose a version of $\widehat{f_{\beta,m}^{\mathcal{K}}}$ with estimated noise density. Adopting the same arguments we consider then the new estimator of f_{β}

$$\forall x \in \mathbb{R}, \quad \widehat{f_{\beta,m}^{\mathcal{K}}}(x) = \frac{1}{2\pi} \int_{-\Delta m}^{\Delta m} e^{-iux} \frac{1}{N} \sum_{j=1}^N \frac{e^{iuZ_{j,\mathcal{K}}}}{\widehat{|f_{\varepsilon}^*|^2}(\frac{u}{\mathcal{K}\Delta})} du. \quad (2.6.14)$$

Using Corollary 2.2.6 and Lemma 2.6.4 we obtain the following risk bound.

Proposition 2.6.5. *If $3 < m \leq N \wedge \mathcal{K}$, and denoting $D_{\varepsilon} := 4B_{\varepsilon} + 136C_{\varepsilon}$, where $C_{\varepsilon} := \sup_{u \in [-1,1]} \frac{1}{|f_{\varepsilon}^*(u)|^8}$, under **(A1)**-**(A2')** we have*

$$\mathbb{E}[\|\widehat{f_{\beta,m}^{\mathcal{K}}} - f_{\beta}\|^2] \leq \|f_{\beta,m}^{\mathcal{K}} - f_{\beta}\|^2 + D_{\varepsilon} \frac{m\Delta}{\pi N}.$$

Deuxième partie

Modèles d'équations différentielles stochastiques à effets aléatoires

Chapitre 3

Estimation non-paramétrique dans un modèle d'Ornstein-Uhlenbeck à effets mixtes

Dans ce chapitre nous nous intéressons au cas particulier du modèle d'Ornstein-Uhlenbeck (OU) à un effet aléatoire, additif. Nous supposons que N observations continues du processus suivant sont disponibles sur l'intervalle de temps $[0, T]$:

$$\begin{cases} dX_j(t) &= \left(\phi_j - \frac{X_j(t)}{\alpha} \right) dt + \sigma dW_j(t) \\ X_j(0) &= x_j \end{cases}$$

(c'est à dire avec les notations du modèle (1.3.4) donné au Chapitre 1 : $\phi_j = \alpha_j$ et $\beta_j = \beta$ constant, on notera ici $\beta = 1/\alpha$). Les valeurs initiales x_j sont supposées connues, ainsi que le paramètre α . Ce travail fait suite à l'article Comte *et al.* (2013). Dans cet article, les auteurs proposent deux estimateurs de la densité f des effets aléatoires ϕ_j . Chaque procédure commence par une phase d'estimation des effets aléatoires ϕ_j par une collection estimateurs $Z_{j,\tau} := (X_j(\tau) - X_j(0) - \int_0^\tau -(X_j(s)/\alpha)ds)/\tau$, pour $\tau \in [0, T]$, avant d'estimer la fonction densité f . Cette estimateur a la particularité de vérifier $\mathbb{E}[Z_{j,\tau}] = \mathbb{E}[\phi_j]$, et de plus $\mathbb{E}[|Z_{j,\tau} - \phi_j|^2] \leq \sigma^2/\tau$. Nous proposons tout d'abord un estimateur à noyau basé sur les variables $(Z_{j,T})_j$ pour lequel la fenêtre est choisie de manière adaptative par une méthode de Goldenshluger et Lepski, ce qui conduit à une inégalité de type oracle. Ensuite nous construisons un estimateur par déconvolution en rajoutant un paramètre dans l'estimateur par rapport à la procédure classique étudiée dans Comte *et al.* (2013). Cette méthode vise à réduire la variance des estimateurs $Z_{j,\tau}$. Le paramètre supplémentaire améliore les résultats de l'estimation et permet d'utiliser l'estimateur dans le cas de faibles valeurs de T (avec un grand nombre d'observations) ce qui n'était pas le cas avant. Notre collection d'estimateurs dépend alors d'un couple de paramètres que nous sélectionnons à l'aide d'une méthode de Goldenshluger et Lepski multidimensionnelle.

Une étude sur simulations atteste la bonne qualité de nos procédures. Enfin nous analysons une base de données neuronales, étudiée auparavant de manière paramétrique dans de nombreux travaux, notamment Lansky *et al.* (2006), Picchini *et al.* (2008), Picchini *et al.* (2010). La comparaison des résultats permet de mieux comprendre ces données.

Remarques

- Justifions la construction d'un nouvel estimateur par déconvolution. Lorsque nous discrétisons l'intervalle d'observation pour appliquer notre procédure sur des données réelles

(observées à des temps $0 = t_0 < t_1 < \dots < t_{\mathcal{K}} = T$ équidistants de δ), l'estimateur de ϕ_j pour $\tau = T$ s'écrit :

$$Z_{j,T} \approx \left(X_j(T) - X_j(0) + \frac{\delta}{\alpha} \sum_{l=1}^{\mathcal{K}} X_j((l-1)\delta) \right) \frac{1}{T}$$

Dans le cadre de données neuronales, on peut avoir les données en \mathbf{V}, \mathbf{s} ou en mV, ms . Les unités sont les suivantes : X_j le potentiel de membrane est en \mathbf{V} et T, δ, α en \mathbf{s} . Alors, on remarque que $Z_{j,T}$ est inchangé lorsque l'on change les unités (puisque'il s'exprime en \mathbf{V}/\mathbf{s}). Cependant, l'estimateur par déconvolution de Comte *et al.* (2013) pour $\tau = T$ change de valeur lorsque l'on change les unités. En effet,

$$\begin{aligned} \hat{f}_T(x) &= \frac{1}{2\pi} \int_{-\sqrt{T}}^{\sqrt{T}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,T}} e^{\frac{u^2\sigma^2}{2T}} du \\ &\approx \frac{1}{2\pi} \sum_{k=1}^{n_{\text{pas}}} \left((u_{k+1} - u_k) e^{-iu_k x} \frac{1}{N} \sum_{j=1}^N e^{iu_k Z_{j,T}} e^{\frac{u_k^2\sigma^2}{2T}} \right) \end{aligned}$$

Comme σ est en $\mathbf{V}/\sqrt{\mathbf{s}}$ le terme exponentiel $e^{u_k^2\sigma^2/(2T)}$ n'est pas homogène car il n'est pas sans unité. Les valeurs de u et donc l'intervalle d'intégration et les trois termes exponentiels changent. Finalement \hat{f}_T change avec les unités et lorsque T est petit l'estimation est mauvaise. Pour résoudre ce problème nous introduisons un nouveau paramètre proportionnel à $1/\sigma$, et on obtient un collection d'estimateurs $\tilde{f}_{m,s}$ homogènes (définis équation (3.2.6)).

- Pointons une difficulté numérique pour l'estimateur à noyau. Considérons un exemple de choix de paramètres pour des données simulées : $\phi_j \sim \mathcal{N}(0.5; 0.04^2)$ et $\alpha = 0.039$ avec $T = 0.3$, $\mathcal{K} = 2000$, $N = 200$ et $\sigma = 0.1$. Alors dans ce cas, l'estimateur à noyau est mauvais. Tout d'abord on constate que le rapport sigma sur bruit : $\sigma/\text{sd}(\phi) = 2$ où $\text{sd}(\phi)$ désigne l'écart type. Ce rapport, bien plus grand que 1 est en général à l'origine d'une estimation peu satisfaisante, le bruit pouvant masquer le signal. Nous avons détecté trois paramètres responsables : σ , T (et \mathcal{K}) et $\text{sd}(\phi)$. L'estimation s'améliore dans les cas suivants :

1. en changeant σ en $\sigma = 0.01$,
2. en changeant T en $T = 10$ ($\mathcal{K} = 4000$),
3. en changeant f en $\mathcal{N}(0.5; 0.4^2)$.

Donc c'est à la fois un problème d'asymptotique en T et N , et un problème de rapport signal sur bruit.

Ce chapitre est issu de l'article Dion (2016) publié dans *Metrika*.

Sommaire

3.1	Introduction	68
3.2	Presentation of the strategies	70
3.2.1	Notation and assumptions	70
3.2.2	Initial idea	70
3.2.3	Estimation strategies	70
3.3	Study of the kernel estimator	72
3.3.1	Risk bound	72
3.3.2	Adaptation of the bandwidth	72
3.4	Study of the deconvolution estimator	73
3.4.1	Risk bound	73
3.4.2	Selection of the final estimator	75
3.5	Simulation study	76
3.6	Application to neuronal data	82
3.6.1	Dataset	82
3.6.2	Comparison of estimators	82
3.7	Discussion	83
3.8	Proofs	84
3.8.1	Proof of Theorem 3.3.2	84
3.8.2	Proof of Proposition 3.4.1	87
3.8.3	Proof of Theorem 3.4.3	88
Appendix		90
3.8.4	Discretization	90

3.1 Introduction

Stochastic differential models have been intensively surveyed in the theoretical literature with either continuous observations (e.g. Kutoyants, 2004) or discrete observations, both in the parametric field (e.g. Genon-Catalot & Jacod, 1993) or in the nonparametric field (e.g. Hoffmann, 1999; Comte *et al.*, 2007). More recently, stochastic differential equations with random effects have been introduced with various applications such as neuronal modelling or pharmacokinetics (e.g. Picchini *et al.*, 2008; Delattre & Lavielle, 2013; Donnet & Samson, 2013). Mixed-effects models are used to analyse repeated measurements with similar functional form but with some variability between experiments (see Davidian & Giltinan, 1995; Pinheiro & Bates, 2000; Diggle *et al.*, 2002). The advantage is that a single estimation procedure is used to fit the overall data simultaneously.

Estimation methods in stochastic differential models with random effects have been proposed, especially in the parametric framework (e.g. Donnet & Samson, 2008; Donnet *et al.*, 2010; Picchini *et al.*, 2008, 2010; Picchini & Ditlevsen, 2011; Genon-Catalot & Larédo, 2016; Donnet & Samson, 2014; Delattre *et al.*, 2015a). All these parametric estimation methods of the density of the random effects are developed assuming a known model on the density, which is often Gaussian. However, one can wonder if this assumption is reasonable depending on the application context. We focus here on the nonparametric estimation of the density of the independent identically distributed random effects. To the best of our knowledge, the only references in this context are Comte *et al.* (2013) and the Chapter 4. The first one provides a nonparametric estimator of the density under restrictive assumptions on the drift and diffusion coefficients. The second one studies the more general case of two linear random effects in the drift. It provides a kernel estimator of the bivariate density of the couple of random parameters. Assuming that the process is at its stationary regime, the authors obtain \mathbb{L}^2 -convergence results.

The present work proposes two nonparametric estimation methods in the simpler model, *i.e.* an Ornstein-Uhlenbeck stochastic differential model with one additive random effect, X , the time scale parameter being assumed known. More precisely, we consider N real valued stochastic processes $(X_j(t), t \in [0, T])$, $j = 1, \dots, N$, with dynamics ruled by the following SDEs:

$$\begin{cases} dX_j(t) &= \left(\phi_j - \frac{X_j(t)}{\alpha} \right) dt + \sigma dW_j(t) \\ X_j(0) &= x_j \end{cases} \quad (3.1.1)$$

where $(W_j)_{1 \leq j \leq N}$ are N independent Wiener processes, and $(\phi_j)_{1 \leq j \leq N}$ are N unobserved independent and identically distributed (*i.i.d.*) random variables taking values in \mathbb{R} , with a common density f . The sequences $(\phi_j)_{1 \leq j \leq N}$ and $(W_j)_{1 \leq j \leq N}$ are independent. Here (x_1, \dots, x_N) are known values. The positive constants σ and α are supposed to be known; in practice they are estimated from experimental data. The estimation of σ can be done using the quadratic variation of the process. The constant α is a physical quantity. Picchini *et al.* (2008) give an estimator of α for the Ornstein-Uhlenbeck model (3.1.1) where the likelihood function is explicit and one can compute maximum likelihood estimators. Each process $(X_j(t), 0 \leq t \leq T)$ represents an individual and the variable ϕ_j is the random effect of individual j . Due to the independence of the ϕ_j and the W_j , the $X_j(t)$, for $j = 1, \dots, N$ are *i.i.d.* random variables when t is fixed, also the N trajectories $(X_j(t), 0 \leq t \leq T)$, $j = 1, \dots, N$ are *i.i.d.* Nevertheless, differences between observations are due to the realization of both W_j and ϕ_j . The Ornstein-Uhlenbeck model is very useful in practice: at first in physics to describe the movement of a particle, then in the econometric field, or for example in neuroscience to describe the membrane potential of a neuron.

The purpose of this work is to build nonparametric estimators of the random effect density

f , considering that only the processes are observed on $[0, T]$ with $T > 0$ given. In practice we consider discrete observations of the X_j 's with a very small time step δ . We are able to evaluate the error made by this discretization. The main difficulty is that we do not observe the ϕ_j 's but only the $X_j(k\delta)$'s. Thus the first step is to find an estimator of the random effects ϕ_j and then to estimate f , taking into account the approximation introduced by the estimation of the ϕ_j .

In the context of stochastic differential equations with random effects, Comte *et al.* (2013) propose different nonparametric estimators with good theoretical properties for large T and N . Here we adopt two different approaches. First we assume that T is large and we propose a direct estimation method of the density from the estimator of the ϕ_j 's. This is the kernel estimator. Then, assuming that T may be small (due to the chosen units for example but still with high frequency data) we focus on the deconvolution estimator.

The kernel estimator depends on a bandwidth to be chosen from the data. Several selection methods of the bandwidth of kernel estimators are known. The originality here is that we use a method, proposed by Goldenshluger & Lepski (2011), which provides an adaptive estimator. This kind of non-asymptotic result is new in this context.

Then we study an estimator built by a deconvolution method (see Butucea & Tsybakov, 2007; Comte *et al.*, 2013, for example). The novelty lies in the introduction of an additional tuning parameter to control the variance of the noise. The value of T is then allowed to be small but we still need high frequency observation meaning a small time step. We obtain a collection of estimators depending on two parameters. To select the final estimator among this collection, we extend the Goldenshluger and Lepski method for a two-dimensional model selection (Goldenshluger & Lepski, 2011). Finally we have a consistent estimator satisfying an oracle inequality, for any value of T . This estimator is likely to be applied to experimental data with small T .

We illustrate the properties of the proposed estimators with a simulation study. Especially, we compare them with standard bandwidth selection method of cross-validation type. Then, the estimators are applied to neuronal data. They are intracellular measurements of the neuronal membrane potential between two spikes which can be modelled with an Ornstein-Uhlenbeck model with one random effect as in (3.1.1). The potential being reset at a fixed initial value after a spike, we consider that the measure between two spikes is an independent experimental unit with a different realization of the random effect. This assumption has already been considered with parametric strategies in Picchini *et al.* (2008) and Picchini *et al.* (2010), where it is assumed that the random effect is Gaussian and proven that the Ornstein-Uhlenbeck model with one random effect fits better the data than the model without them. Our goal is to estimate nonparametrically the density of the random effect. This estimated density could be used in further works to model this phenomenon (instead of using the Gaussian density systematically.)

This Chapter is organized as follows. Section 3.2 is dedicated to giving definitions and presenting the estimators investigated in this work. Then in Section 3.3 we set up a method of bandwidth selection for the kernel estimator. In Section 3.4 we define and study the final data-driven estimator built by deconvolution. In Section 3.5 we calibrate the selection methods and illustrate the good performances of both estimators on simulated data. In Section 3.6 we experiment the procedures on real data. We conclude with a discussion in Section 3.7. All proofs are gathered in Section 3.8, and the computation of the error made by discretization is done in Appendix 3.8.3.

3.2 Presentation of the strategies

3.2.1 Notation and assumptions

Let us introduce some notations. For two functions g_1 and g_2 in $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, the convolution product of g_1 and g_2 for all $x \in \mathbb{R}$, is $g_1 \star g_2(x) = \int_{\mathbb{R}} g_1(x-y)g_2(y)dy$ and the scalar product is: $\langle g_1, g_2 \rangle = \int_{\mathbb{R}} g_1(x)\overline{g_2(x)}dx$. Then the Fourier transform of g_1 is $g_1^*(x) = \int_{\mathbb{R}} e^{iux}g_1(u)du$ for all $x \in \mathbb{R}$ and the \mathbb{L}^2 -norm is $\|g_1\|^2 = \int_{\mathbb{R}} |g_1(x)|^2dx$. Finally we recall the Plancherel-Parseval's formula: $2\pi\|g_1\|^2 = \|g_1^*\|^2$.

We assume **(A)** $f \in \mathbb{L}^2(\mathbb{R})$, $f^* \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$.

3.2.2 Initial idea

As previously mentioned, the first step of the procedure is to estimate the random effect ϕ_j which are not observed, in order to recover their density, in a second time.

For this purpose, we introduce the following random variables for $j = 1, \dots, N$ and $\tau \in]0, T]$,

$$Z_{j,\tau} := \frac{X_j(\tau) - X_j(0) - \int_0^\tau (-\frac{X_j(s)}{\alpha} ds)}{\tau} = \phi_j + \frac{\sigma}{\tau}W_j(\tau). \tag{3.2.1}$$

The $(Z_{j,\tau})_\tau$ are estimators of the ϕ_j based on the trajectory $(X_j(t))$. They correspond to the maximum in φ of the conditional likelihood of (3.1.1) given $\phi_j = \varphi$. Moreover random variables satisfy $\mathbb{E}[Z_{j,\tau}] = \mathbb{E}[\phi_j]$ and when τ goes to infinity, the noise $\sigma W_j(\tau)/\tau$ goes to zero. This attests the goodness of the estimator. Notice that the $(Z_{j,\tau})_{j=1,\dots,N}$ are *i.i.d.* when τ is fixed, with density f_{Z_τ} , due to the independence of $(\phi_j)_{j=1,\dots,N}$ and $(W_j)_{j=1,\dots,N}$. These new random variables are available, depending only on the observations and known parameters. Nevertheless, we only have discrete observations of the process. Thus we discretize the integral:

$\int_0^\tau X_s ds \approx \delta \sum_{k=1}^{\lfloor \tau/\delta \rfloor} X_{(k-1)\delta}$. The error due to this approximation is studied in Section 3.8.4. At this point, two strategies materialize which we explain in the following Section.

3.2.3 Estimation strategies

Let us present the two investigated methods.

Kernel strategy

The first idea is to reduce the noise which appears in formula (3.2.1). Indeed, $\text{Var}(\sigma W_j(\tau)/\tau) = \sigma^2/\tau$ leads to focus on the largest τ : $\tau = T$. Moreover, when T is large $Z_{j,T}$ clearly approximates ϕ_j without needing to remove the noise. Then we are able to build a kernel estimator of the density f of the ϕ_j 's based on the $Z_{j,T}$ using directly the $Z_{j,T}$ as an approximation of the non-observed random effects ϕ_j . These N random variables are *i.i.d.* and the resulting kernel estimator is given for all $x \in \mathbb{R}$, by

$$\widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - Z_{j,T}) \tag{3.2.2}$$

where $h > 0$ is a bandwidth, and $K : \mathbb{R} \rightarrow \mathbb{R}$ is a \mathcal{C}^2 kernel such that

$$\int K(u)du = 1, \quad \|K\|^2 = \int K^2(u)du < +\infty, \quad \int (K''(u))^2 du < +\infty, \quad K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right). \tag{3.2.3}$$

This natural estimator is studied in detail in Section 3.3.

Deconvolution strategy

The other idea is to build an estimator of f using all variables $Z_{j,\tau}$ for different $\tau \in [0, T]$. Recovering f from the observations $(X_1(t), \dots, X_N(t))_{t \in [0, T]}$ is called the deconvolution problem because the common density of $(Z_{j,\tau})_{j=1, \dots, N}$ is a convolution product between two densities. Indeed, the two members of the sum (3.2.1) are independent when τ is fixed, which implies for all $j = 1, \dots, N$,

$$f_{Z_\tau}(u) = f \star f_{\frac{\sigma}{\tau}W_1(\tau)}(u).$$

Then the characteristic function of ϕ_j is recoverable from that of Z_τ . Taking the Fourier transform under assumption **(A)** gives the simple product

$$f_{Z_\tau}^*(u) = f^*(u)f_{\frac{\sigma}{\tau}W_1(\tau)}^*(u),$$

with $f_{\frac{\sigma}{\tau}W_1(\tau)}^*(u) = e^{-\frac{u^2\sigma^2}{2\tau}}$. In this particular case the noise is Gaussian and this convolution problem has been investigated in the literature, see Fan (1991); Butucea & Tsybakov (2007) for example. However it has been proven in Carroll & Hall (1988) that the best rates of convergence obtained in this case are logarithmic. This suggests to improve the deconvolution procedure and this is the reason why we choose not to use previous estimators but to propose a new method, based on repeated observations and new parameters chosen carefully.

We have $f^*(u) = f_{Z_\tau}^*(u)e^{u^2\sigma^2/2\tau}$. Finally the Fourier inversion gives the closed formula, for all $x \in \mathbb{R}$,

$$f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} f_{Z_\tau}^*(u) e^{\frac{u^2\sigma^2}{2\tau}} du. \tag{3.2.4}$$

Then, we estimate $f_{Z_\tau}^*(u)$ by its empirical estimator $\widehat{f}_{Z_\tau}^*(u) = (1/N) \sum_{j=1}^N e^{iuZ_{j,\tau}}$. However, plugging this in formula (3.2.4) involves integrability problems. Indeed the integrability of $\widehat{f}_{Z_\tau}^*(u)e^{u^2\sigma^2/2\tau}$ is not ensured. Therefore, we have to introduce a cut-off. The nonparametric estimation using a deconvolution method in the Gaussian case commonly yields bad speeds of convergence. To improve the rates, an idea of Comte & Samson (2012), for linear mixed models, was to link this cut-off and the time of the process. Comte *et al.* (2013) link the time of the process τ and the cut-off as follows:

$$\widehat{f}_\tau(x) = \frac{1}{2\pi} \int_{-\sqrt{\tau}}^{\sqrt{\tau}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,\tau}} e^{\frac{u^2\sigma^2}{2\tau}} du. \tag{3.2.5}$$

Then the time τ is chosen by a Goldenshluger and Lepski's method and the final estimator is denoted $\widetilde{f}_{\widetilde{\tau}}$. Nevertheless, when τ is small (which is the case for the real dataset we investigate), the integration domain is not large enough, and the estimators of f are not satisfactory (cf an explicit example in Section 3.6). We adapt Comte *et al.* (2013) estimator to this small T framework. Indeed, to improve the previous estimator, we introduce a new parameter s in the cut-off:

$$\widehat{f}_{s,\tau}(x) = \frac{1}{2\pi} \int_{-s\sqrt{\tau}}^{s\sqrt{\tau}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,\tau}} e^{\frac{u^2\sigma^2}{2\tau}} du.$$

Then, in order to simplify the theoretical study, we replace $s\sqrt{\tau}$ in the integral by a new parameter m . The resulting estimator $\widetilde{f}_{m,s}$ is defined when $m^2/s^2 \in]0, T]$, by

$$\widetilde{f}_{m,s}(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m^2/s^2}} e^{\frac{u^2\sigma^2s^2}{2m^2}} du \tag{3.2.6}$$

with m and s in two finite sets \mathcal{M} and \mathcal{S} that we will precise later.

In the following we survey in detail the two strategies.

3.3 Study of the kernel estimator

The kernel estimator given by (3.2.2) has been investigated in Comte *et al.* (2013). First we recall the MISE bound that the kernel estimator \widehat{f}_h satisfies. Then we develop the bandwidth selection procedure we are interested in in this work.

3.3.1 Risk bound

Let us define $f_h := K_h \star f$, for $h > 0$. We denote for all $p \in \mathbb{R}$, $\|f\|_p = (\int |f(x)|^p dx)^{1/p}$ and for $p = 2$ we still use $\|f\|_2 = \|f\|$. Notice that $\|K_h\| = \|K\|/\sqrt{h}$ and $\|K_h\|_1 = \|K\|_1$. We recall the result proven in Comte *et al.* (2013) for the MISE.

Proposition 3.3.1. *Considering estimator \widehat{f}_h given by (3.2.2), we have*

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2\sigma^4\|K''\|^2}{3T^2h^5}. \quad (3.3.1)$$

The right-hand side of (3.3.1) involves three terms, and the middle one is the integrated variance. The integrated bias is $\|\mathbb{E}[\widehat{f}_h] - f\|^2 \leq 2\|f - f_h\|^2 + 2\|\mathbb{E}[\widehat{f}_h] - f_h\|^2$, with

$$\|\mathbb{E}[\widehat{f}_h] - f_h\|^2 \leq \frac{\sigma^4\|K''\|^2}{3T^2h^5}. \quad (3.3.2)$$

Therefore, the first term $\|f - f_h\|^2$ is a bias term, which decreases when h decreases. The second term is the term of variance which increases when h decreases. Finally, the third term, also given in (3.3.2), is an unusual error term due to the approximation of the ϕ_j 's by the $Z_{j,T}$ also increasing when h decreases. We see on this bound that the rate σ^2/T must be small to obtain a small risk.

3.3.2 Adaptation of the bandwidth

Now that we have at hand a collection of estimators depending on a bandwidth h , we focus on the crucial matter namely how to choose the bandwidth from the data. The best choice of h is the one which minimizes the sum of these three terms. The selection of the bandwidth can be done for example using cross validation, see *e.g.* the R-function *density* which is commonly used. However, the only theoretical results known for cross-validation procedure are asymptotic and to the best of our knowledge there is no adaptive result on the final estimator. In the present work, we propose to adapt a selection method due to Goldenshluger & Lepski (2011) mentioned before, which provides a data-driven bandwidth for which we provide non-asymptotic theoretical results.

We denote $\mathcal{H}_{N,T}$ the finite set of bandwidths h , to be defined later. The best theoretical choice of the bandwidth is the h which minimizes the bound on the MISE given by (3.3.1). Nevertheless, in practice, the bias term is unknown, and this bound has to be estimated.

To choose h adequately, we use a Goldenshluger and Lepski's criterion introduced in Goldenshluger & Lepski (2011). The idea is to estimate $\|f - f_h\|^2$ by the \mathbb{L}^2 -distance between two

estimators defined in (3.2.2). But this induces an error which has to be corrected by the variance term. Then the estimator of the bias term is

$$A(h) = \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 - V(h') \right)_+ \quad (3.3.3)$$

where

$$\widehat{f}_{h,h'}(x) := K_{h'} \star \widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_{h'} \star K_h(x - Z_{j,T})$$

and V correspond to the two terms of variance

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{T^2 h^5} \quad (3.3.4)$$

with κ_1 and κ_2 two numerical positive constants. We will prove that $A(h)$ has the order of the bias term (see equation (3.8.7)). Finally the bandwidth is selected as follows:

$$\widehat{h} = \operatorname{argmin}_{h \in \mathcal{H}_{N,T}} \{A(h) + V(h)\} \quad (3.3.5)$$

with $\mathcal{H}_{N,T}$ a finite discrete set of bandwidths h such that $h > 0$, $\frac{1}{Nh} \leq 1$, $\frac{1}{h^5 T^2} \leq 1$ and $\operatorname{Card}(\mathcal{H}_{N,T}) \leq N$. It must be chosen such that when N goes to infinity, for all $c \in \mathcal{C}$ $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}} \leq S(c)$ with $S(c)$ a positive constant depending on c . For example notice that taking $\mathcal{H}_{N,T} = \{1/k^2, k = 1, \dots, \sqrt{N}\}$, the sum $\sum_{h \in \mathcal{H}_{N,T}} h^{-1/2} e^{-c/\sqrt{h}} \leq \sum_{k \geq 1} k e^{-ck}$ converges, which is a necessary condition for the proof.

Then we can prove the following Theorem.

Theorem 3.3.2. *Consider estimator \widehat{f}_h given by (3.2.2) with $h \in \mathcal{H}_{N,T}$. Then, there exist two penalty constants κ_1, κ_2 such that*

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \{\|f - f_h\|^2 + V(h)\} + \frac{C_2}{N}$$

where C_1, C_2 are two positive constants such that $C_1 = \max(7, 30\|K\|_1^2 + 6)$ and C_2 depends on $\|f\|, \|K\|, \|K\|_1, \|K\|_{4/3}$.

The theoretical study gives $\kappa_1 \geq \max(40/\|K\|_1^2, 40)$ and $\kappa_2 \geq \max(10/3, 10/(3\|K\|_1^2))$. But in practice these two constants are calibrated from a simulation study (and always smaller than the theoretical ones). Theorem 3.3.2 is an oracle inequality: the bias variance compromise is automatically obtained and in a data-driven and non-asymptotic way.

This strategy requires large T as we assume $1/h^5 \leq T^2$. The error implied by the discrete observations and the use of Riemann sums to compute the $Z_{j,T}$ are detailed in Comte *et al.* (2013).

3.4 Study of the deconvolution estimator

3.4.1 Risk bound

Let us emphasize that the estimator $\widetilde{f}_{m,s}$ given by (3.2.6) depends on two parameters which have to be selected from the data. This is not usual in the deconvolution setting, where only one cut-off parameter is often introduced. The selection of these two parameters (m, s) among

the finite sets \mathcal{M} , \mathcal{S} is thus more difficult. It is even more challenging here because the cut-off m appears both in the integral and in the integrand. But this will induce gains in the rates of the estimators. Before proposing a selection method of (m, s) we start by evaluating the quality of the estimator with the mean integrated squared error (MISE):

$$\mathbb{E} \left[\|\tilde{f}_{m,s} - f\|^2 \right] = \|f - \mathbb{E}[\tilde{f}_{m,s}]\|^2 + \mathbb{E} \left[\|\tilde{f}_{m,s} - \mathbb{E}[\tilde{f}_{m,s}]\|^2 \right].$$

In Proposition 3.4.1 we prove that $\mathbb{E}[\tilde{f}_{m,s}] = f_m$ where f_m is defined by its Fourier transform

$$f_m^* := f^* \mathbf{1}_{[-m,m]}.$$

It means that the bias does not depend on s . We obtain the following bound on the MISE of $\tilde{f}_{m,s}$.

Proposition 3.4.1. *Under (A), the estimator $\tilde{f}_{m,s}$ given by (3.2.6) is an unbiased estimator of f_m and we have*

$$\mathbb{E} \left[\|\tilde{f}_{m,s} - f\|^2 \right] \leq \|f_m - f\|^2 + \frac{m}{\pi N} \int_0^1 e^{\sigma^2 s^2 v^2} dv. \quad (3.4.1)$$

The proofs are relegated to Section 3.8. Let us look at the risk bound. The first term of the bound (3.4.1) is the bias term. It represents the error resulting from estimating f by f_m and it decreases when m increases, indeed:

$$\|f_m - f\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du.$$

The second term is the variance term, and it increases with m and s . One can notice that it is bounded as soon as s is bounded and $m \leq N$.

We specify the two sets \mathcal{M} and \mathcal{S} . We notice that the quality of the estimate in the Fourier domain is good on an interval around zero with length related with σ . The chosen set for s is

$$\mathcal{S} := \left\{ s_l = \frac{1}{2^l} \frac{2}{\sigma}, l = 0, \dots, P \right\}.$$

Notice that for $s_l \in \mathcal{S}$, $1/2^{P-1} \leq \sigma s_l \leq 2$. Moreover with this chosen collection \mathcal{S} , the order of the variance term is m/N . With the idea that m^2/s^2 is homogeneous to a time, we choose m in the finite collection:

$$\mathcal{M} := \left\{ m = \frac{\sqrt{k\Delta}}{\sigma}, k \in \mathbb{N}^*, 0 < m \leq N \right\}$$

with $0 < \Delta < 1$ a small step to be fixed. The collection of couples of parameters is

$$\mathcal{C} := \{(m, s) \in \mathcal{M} \times \mathcal{S}, m^2/s^2 \leq T\}.$$

The final estimator is the estimator from the collection \mathcal{C} which achieves the bias-variance compromise. Choosing the final estimator is not an easy task except if we know the regularity of f . Indeed, let us assume that f is in the Sobolev ball with regularity parameter b , i.e. f belongs to the set defined by

$$\mathcal{A}_b(L) = \left\{ f \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \int_{\mathbb{R}} |f^*(x)|^2 (1+x^2)^b dx \leq L \right\}$$

with $b > 0$, $L > 0$. For example the standard normal distribution is in a space $\mathcal{A}_b(L)$ for some L and for all $b > 0$, an exponential distribution is in some $\mathcal{A}_b(L)$ for $b < 1/2$ or more generally a Gamma distribution with shape parameter k is in some $\mathcal{A}_b(L)$ for $b < (k - 1/2)$. Thus when $f \in \mathcal{A}_b(L)$, the bias term satisfies:

$$\|f_m - f\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du \leq \frac{L}{2\pi} m^{-2b}.$$

Consequently, the \mathbb{L}^2 -risk of $\tilde{f}_{m,s}$ is bounded by

$$\mathbb{E}[\|\tilde{f}_{m,s} - f\|^2] \leq \frac{L}{2\pi} m^{-2b} + \frac{m}{\pi N} e^{\sigma^2 s^2}.$$

Therefore, the best theoretical choice of s is s_P the smallest s in our collection, and

$$m = m^* = K_b N^{\frac{1}{2b+1}}$$

with $K_b = (bL \exp(-1/(2^{2(P-1)})))^{1/(2b+1)}$. Then we obtain the following asymptotic result.

Corollary 3.4.2. *If $f \in \mathcal{A}_b(L)$, and if we choose $s = s_P$ and $m = m^*$, there exists a constant K depending on b, L, P , such that*

$$\mathbb{E}[\|\tilde{f}_{m^*,s_P} - f\|^2] \leq KN^{-\frac{2b}{2b+1}}.$$

The order of the risk in this case is $N^{-2b/(2b+1)}$ for a large N , and it is the nonparametric estimation rate of convergence obtained when the observations are N realizations of the variable of interest. Nevertheless, it is not easy to see that $(m, s) \in \mathcal{C}$ and this choice is theoretical because it depends on the regularity b of f , which is unknown. The next section provides a data-driven method to select (m, s) .

3.4.2 Selection of the final estimator

In this Section we deal with the choice of the best estimator among the available collection of $\tilde{f}_{m,s}$. In the previous work Comte *et al.* (2013), s was fixed to $s = 1$ and m was selected. But we saw empirically that it did not work in the setting corresponding to the data. This is why we experimented different values for s . But then we did not find any reliable criterion to select m for any given s . On the contrary, if we look at the bound and try to select s first, we just get $s = 0$, which is not of interest if we are looking to improve the estimator through s in particular. This implies to select the couple (m, s) minimizing the MISE and realizing the compromise between the two terms, in a data-driven way. This is a crucial issue. Indeed, the role of the two parameters is not the same. Thus we propose a new criterion adapted from the Goldenshluger & Lepski (2011) method.

The idea is to select the couple which minimizes the MISE: $\mathbb{E}[\|\tilde{f}_{m,s} - f\|^2]$. As it is unknown, we have to find a computable approximation of this quantity. We define the best couple (m, s) as the one minimizing a criterion defined as the sum of a squared bias term and a variance term called penalty. We define the penalty function, which has the same order as the bound on the variance term:

$$\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2},$$

where κ is a numerical constant to be calibrated. Note that for $m \in \mathcal{M}$ and $s \in \mathcal{S}$, the penalty function is bounded.

To estimate the bias term, we generalize Goldenshluger and Lepski's criterion for a two-dimensional index. The method is inspired by the ideas developed for kernel estimators by Goldenshluger & Lepski (2011) and adapted to model selection in one dimension in Comte & Johannes (2012) and in two dimensions by Chagny (2013). The idea is to estimate $\|f - f_m\|^2$ by the \mathbb{L}^2 -distance between two estimators defined in (3.2.6). But this induces a bias which has to be corrected by the penalty function. We consider the following estimator of the bias, with $(m', s') \wedge (m, s) := (m' \wedge m, s' \wedge s)$,

$$\Gamma_{m,s} = \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{m',s'} - \tilde{f}_{(m',s') \wedge (m,s)}\|^2 - \text{pen}(m', s') \right)_+ \quad (3.4.2)$$

for $(m, s) \in \mathcal{C}$. Finally the selected couple is:

$$(\tilde{m}, \tilde{s}) = \arg \min_{(m,s) \in \mathcal{C}} \{\Gamma_{m,s} + \text{pen}(m, s)\}. \quad (3.4.3)$$

We are now able to obtain the following result.

Theorem 3.4.3. *Under (A), consider the estimator $\tilde{f}_{\tilde{m}, \tilde{s}}$ given by (3.2.6) and (3.4.3). There exists κ_0 a numerical constant such that, for all penalty constants $\kappa \geq \kappa_0$,*

$$\mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] \leq C \inf_{(m,s) \in \mathcal{C}} \{\|f - f_m\|^2 + \text{pen}(m, s)\} + \frac{C'(P+1)}{N} \quad (3.4.4)$$

where $C > 0$ is a numerical constant and C' is a constant depending on $\|f\|$, σ , Δ , and $P+1$ the cardinality of \mathcal{S} .

The key of the proof is to prove that

$$\mathbb{E}[\Gamma_{m,s}] \leq 18\|f - f_m\|^2 + \frac{C'(P+1)}{N}$$

(see the proof in Section 3.8.3.) Inequality (3.4.4) means that $\tilde{f}_{\tilde{m}, \tilde{s}}$ automatically makes the bias-variance trade-off. Moreover, our result is of non asymptotic nature w.r.t. N .

One should notice that this new parameter s generalizes the results of Comte *et al.* (2013) even if T is large. We choose the two parameters in an adaptive way, thus this gives more flexibility in the choice of the estimator.

It follows from the proof that $\kappa_0 = 24$ would suit. But in practice, values obtained from the theory are generally too large and the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments. Besides the cardinality P of the set \mathcal{S} is chosen small in practice ($P = 3$ or 10 for example).

In Appendix 3.8.4 we investigate the error implied by the discrete observations and thus of the discretization of $Z_{j,\tau}$ given by (3.2.1).

3.5 Simulation study

In the following section we compare on simulations the two procedures we compute $\hat{f}_{\hat{h}}$ and $\tilde{f}_{\tilde{m}, \tilde{s}}$, the estimator of Comte *et al.* (2013) $\tilde{f}_{\tilde{\tau}}$ and we compare our bandwidth selection method with the estimator from the R-function *density* with the cross-validation argument `bw=ucv`, on the $Z_{j,T}, j = 1, \dots, N$, we denote it \hat{f}_{cv} .

We simulate data by computing the exact solutions of (3.1.1) given by Itô's formula,

$$X_j(t) = X_j(0)e^{-t/\alpha} + \phi_j\alpha(1 - e^{-t/\alpha}) + \sigma e^{-t/\alpha} \int_0^t e^{s/\alpha} dW_j(s) \quad (3.5.1)$$

at discrete times $t_k \in \mathcal{T} := \{k\delta, k \in \{0, \dots, \mathcal{K}\}, \mathcal{K}\delta = T\}$. For the simulation study, we have to fix $N, \delta, T, \sigma, \alpha$, and the density f . We take $\sigma = 0.0135, 0.05, 1$ and $\sigma = 0.05, \alpha = 0.039, 1, 39$. For the time T , we choose $T = 0.3, 10, 50, 100, 300$ with different values of δ the discrete time step at which observations are recorded. The value of \mathcal{K} , the number of observations for one trajectory ranges from 150 to 5000 for Table 3.1 and is fixed to $\mathcal{K} = 2000$ for Table 3.2. All these parameter values are chosen in relation with the parameters of the real dataset. In this study we hope to highlight the influence of each one. For f , we investigate four different distributions:

- Gaussian distribution $\mathcal{N}(0.278, (0.041)^2)$
 - Gamma distribution $\Gamma(1.88, 0.148)$
 - mixed Gaussian distribution $0.3\mathcal{N}(0, (0.02)^2) + 0.7\mathcal{N}(1, (0.02)^2)$ (denoted M-Gaussian)
 - mixed Gamma distribution $0.4\Gamma(3, 0.08) + 0.6\Gamma(30, 0.035)$ (denoted M-Gamma)
- where we write $\Gamma(k, \theta)$ with k the scale parameter and θ the shape.

First, we implement the two collections of estimators: \hat{f}_h and $\tilde{f}_{\tilde{m}, \tilde{s}}$. We begin by computing the random variables used by both estimators: $Z_{j,\tau}$ given by (3.2.1), with Riemann sums approximations (see Appendix 3.8.4 for details). For the deconvolution estimator given by (3.2.6) we also use Riemann sums to compute the integral. For the collection of m , we choose $\Delta = 0.08$ and δ changes. Furthermore, for the kernel estimator given by (3.2.2), we choose a Gaussian kernel: $K(u) = (1/\sqrt{2\pi})e^{-u^2/2}$. In this case $\|K\|_1 = 1, \|K\|_2^2 = 1/(2\sqrt{\pi}), \|K''\|_2^2 = (1 + 1/\sqrt{2})/(\sqrt{2\pi})$. Then, the selected bandwidth \hat{h} is given by Equation (3.3.5). Note that for all $(h, h') \in \mathcal{H}^2$,

$$K_{h'} \star K_h(x) = \frac{1}{\sqrt{2\pi}\sqrt{h'^2 + h^2}} e^{-x^2/[2(h'^2+h^2)]}.$$

We use this relation to compute the $\hat{f}_{h,h'}$.

Secondly, we have to calibrate the penalty constants: κ_1, κ_2 for the kernel estimator and κ for the deconvolution estimator. Classically, the constants are fixed thanks to preliminary simulation experiments. Different functions f have been investigated with different parameter values, and a large number of replications. Comparing the MISE obtained as functions of the constants κ_1, κ_2 and κ yield to select values making a good compromise over all experiments. Finally we choose $\kappa_1 = 1, \kappa_2 = 0.0001$ and $\kappa = 0.3$. A recent work Lacour & Massart (2016) proposes to change the calibration constants in the variance term $V(h)$: taking κ in the term $\Gamma_{(m,s)}$ (3.4.2) and 2κ for the second $V(h)$ in the selection criterion (3.3.5). It has been done in practice for the kernel estimator. We notice that this strategy produces very good results in practice, better than choosing the same κ for the two apparitions of term $V(h)$. On Figure 3.1 25 estimators \hat{f}_h are plotted and on Figure 3.2 25 estimators $\tilde{f}_{\tilde{m}, \tilde{s}}$, for the 4 investigated densities f . The batch of estimators is close to the estimated density.

In order to evaluate the performances of each estimator on the different designs, we compare their empirical MISE computed from 100 simulated data sets.

Table 3.1 summarises the results for different parameters values. It shows the bad performances of the estimator of Comte *et al.* (2013) $\tilde{f}_{\tilde{z}}$ when T is small compared to $\tilde{f}_{\tilde{m}, \tilde{s}}$. It performs clearly better when T is increasing. Besides we notice that both kernel estimators have good results. Nevertheless these results are satisfying because it appears that our estimator \hat{f}_h fits slightly better the true density than \hat{f}_{cv} . The computation time is close for both selection

method. We show the results for different values of α do not seem to influence the quality of estimators (while the selected h, m, s are very different). During the simulation study, we noticed that the parameter α is important in the sense that when the value of α does not have the same order as the values of ϕ , the estimation is harder. Except when $T = 300$ the ratio signal noise which is the standard deviation of the random effect divided by σ is larger than one thus the settings are favourable. But for both Gamma and mixed Gamma cases, the standard deviations are respectively 0.2 and 0.15, which is not small compared to σ or to their mean. The mixed Gamma case is difficult for the nonparametric estimation: Figure 3.3 illustrates the performances of estimators for this choice. Finally, it is interesting to note that when the standard deviation of the random effect of interest has a larger variance, the density estimation is easier, which is the case of the chosen Gamma density for example compared with the Gaussian case.

In the following as the two kernel estimators seem very close, we only show the results for \hat{f}_h which is of interest here. Besides we no longer investigate the previous estimator $\tilde{f}_{\tilde{m}, \tilde{s}}$ in light of Table 3.1 for $T < N$.

We further compare \hat{f}_h and $\tilde{f}_{\tilde{m}, \tilde{s}}$. The two estimators seem close to the true density on graphs, see Figure 3.1 and 3.2. In Table 3.2 on the MISE it is clear that the kernel estimator is the best. Furthermore, we can point out some differences. The first row of the Table corresponds to simulations with the parameters of the real dataset. In the first column, the Gaussian case, the MISE are 10 times larger than the ones for other cases. This can be easily explained: the values of the estimated density are 10 times larger than others. Nevertheless, on lines 3 and 4 for the Gaussian case, the MISE are very large. This is due to the bad estimation of the ϕ_j by the $Z_{j,T}$ with $\sigma = 0.05$ and $T = 0.3$ ¹. The quality of the estimation is significantly better when we try a $\mathcal{N}(0.278, 0.2)$ (0.2 is the variance of the mixed Gaussian density we implement for example). In general one can notice that when σ is larger than the standard deviation of the density of the random effects f , the estimation is less precise, which is coherent in term of ratio signal noise.

Table 3.2 shows that if T increases, it improves the results for $\sigma = 0.05$, compare cases 2 and 5 with 4 and 7 for example. If \mathcal{K} is large enough, meaning if δ is small enough (which is the case even for $\mathcal{K} = 150$ when $T = 0.3$), the deconvolution estimator fits well the density. In practice, when T increases, the selected value of s decreases, which could have been predicted. The results are still satisfying for large T . For the kernel estimator, although the theoretical condition $1/h^5 < T^2$ is not satisfied, the numerical results are good.

Another point is, as expected, that the larger N is, the better the estimators \hat{f}_h and $\tilde{f}_{\tilde{m}, \tilde{s}}$. We can refer to Comte *et al.* (2013) for a study with different values for N . It highlights the influence of N when the estimated density has two modes; for example with $N = 50$ the estimation is clearly less precise than for $N = 200$.

A main difference between our two estimators \hat{f}_h and $\tilde{f}_{\tilde{m}, \tilde{s}}$ is the computation time: a few seconds for the first one and ten minutes for the second one. Thus the kernel estimator with the method of bandwidth selection is very efficient, especially in the case of multi-modal densities, and performs often better than the deconvolution one.

1. We insist that this bad estimation is not due to the fact that the noise is Gaussian. Indeed even if Fan (1991) proves the rates to be logarithmic in that case, the rates are improved and can be polynomial when the density under estimation is of the same type as the noise (see Lacour (2006), Comte *et al.* (2006)).

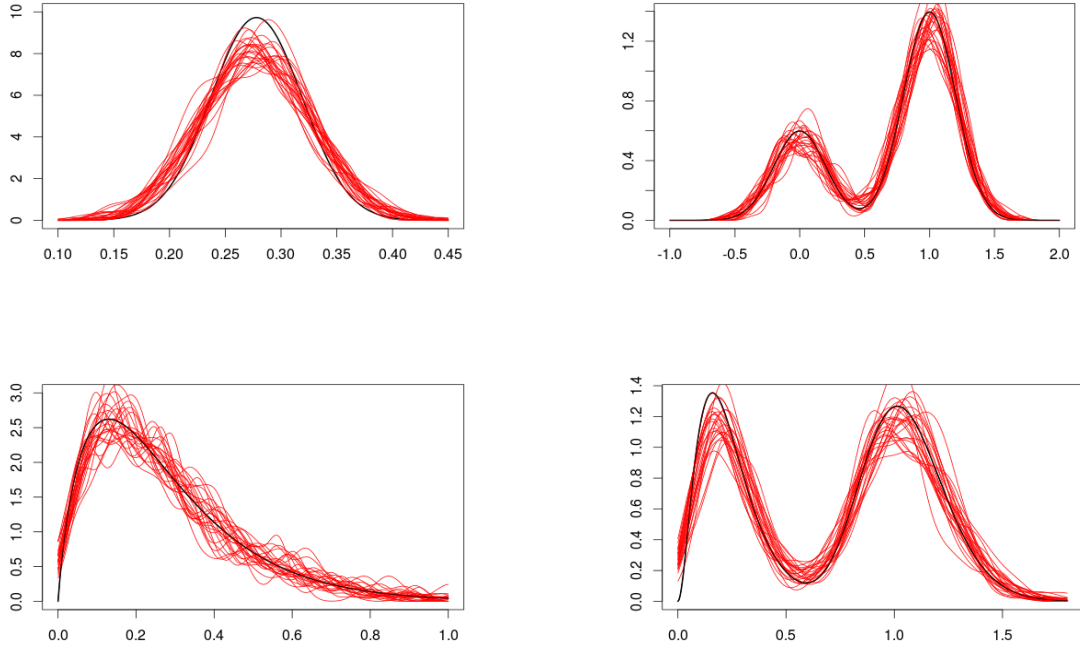


Figure 3.1 – Simulated data. In plain red (grey) 25 estimators \hat{f}_h with parameters: $N = 240$, $T = 0.3, \delta = 0.00015, \sigma = 0.0135, \alpha = 0.039$ and the true density f in plain bold black line. Top left: f Gaussian, top right f mixed Gaussian, bottom left f Gamma, bottom right f mixed Gamma

Table 3.1 – Empirical MISE computed from 100 simulated data sets, with $N = 200$, with various T, δ, σ and α for two kernel estimators \hat{f}_{cv}, \hat{f}_h and two deconvolution estimators $\tilde{f}_{\tilde{\tau}}$ and $\tilde{f}_{\tilde{m}, \tilde{s}}$

	\hat{f}_{cv}	\hat{f}_h	$\tilde{f}_{\tilde{\tau}}$	$\tilde{f}_{\tilde{m}, \tilde{s}}$
Case f Gamma				
$T = 0.3 \delta = 0.0005 \sigma = 0.0135 \alpha = 0.039$	0.043	0.037	1.547	0.071
$T = 300 \delta = 0.5 \sigma = 0.5 \alpha = 39$	0.048	0.039	0.049	0.055
$T = 50 \delta = 0.1 \sigma = 0.05 \alpha = 1$	0.042	0.039	0.218	0.050
Case f mixed Gamma				
$T = 0.3 \delta = 0.0005 \sigma = 0.0135 \alpha = 0.039$	0.033	0.030	0.712	0.035
$T = 300 \delta = 0.5 \sigma = 0.5 \alpha = 39$	0.032	0.030	0.035	0.043
$T = 50 \delta = 0.1 \sigma = 0.05 \alpha = 1$	0.033	0.031	0.145	0.043

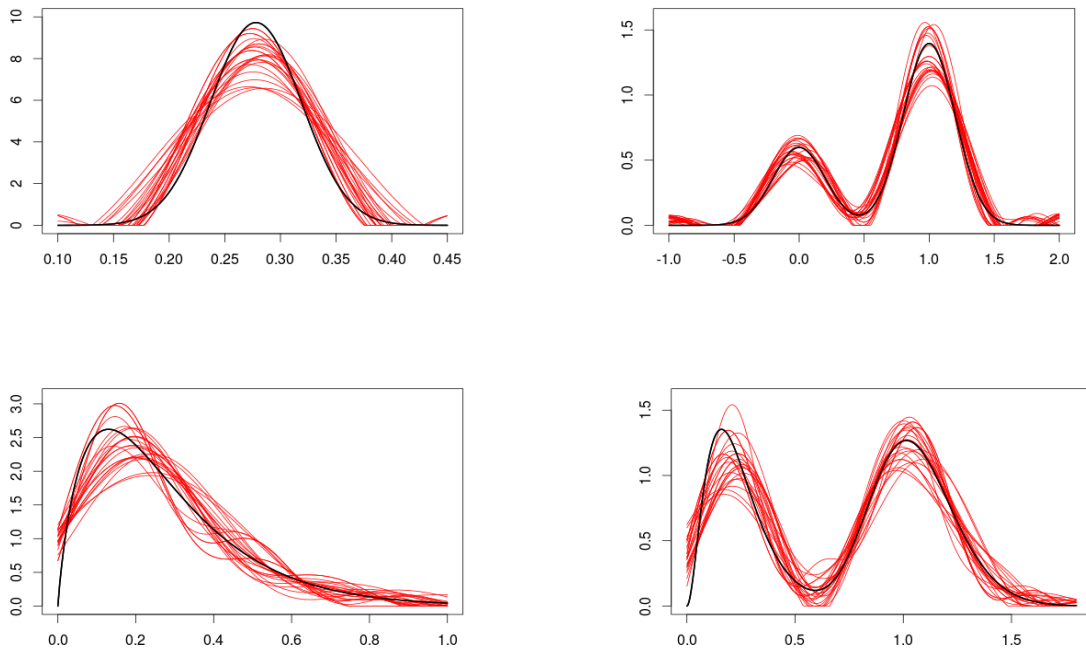


Figure 3.2 – Simulated data. In plain red (grey) 25 estimators $\tilde{f}_{\tilde{m},\tilde{s}}$ with parameters: $N = 240$, $T = 0.3$, $\delta = 0.00015$, $\sigma = 0.0135$, $\alpha = 0.039$ and the true density f in bold plain black line. Top left: f Gaussian, top right f mixed Gaussian, bottom left f Gamma, bottom right f mixed Gamma

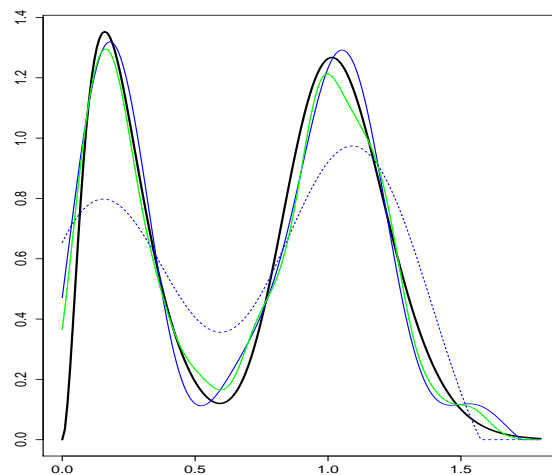


Figure 3.3 – Simulated data. The bold plain black curve is the true density f mixed Gamma, the estimator \hat{f}_{cv} and estimator \hat{f}_h are superposed in plain green (grey), estimator $\tilde{f}_{\tilde{z}}$ is dotted blue (black) and estimator $\tilde{f}_{\tilde{m},\tilde{s}}$ is plain blue (black), with $N = 200$, $T = 50$, $\delta = 0.05$, $\sigma = 0.05$, $\alpha = 1$

Table 3.2 – Empirical MISE computed from 100 simulated data sets, with $N = 240$, $\alpha = 0.039$ and various T, δ, σ for the kernel estimator \widehat{f}_h and the deconvolution estimator $\widetilde{f}_{\widetilde{m},\widetilde{s}}$

Case	T	δ	σ	Estimator	f			
					Gaussian	Gamma	M-Gaussian	M-Gamma
1	0.3	0.0002	0.0135	\widehat{f}_h	0.254	0.037	0.016	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.381	0.072	0.021	0.038
2	0.3	0.002	0.0135	\widehat{f}_h	0.270	0.032	0.016	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	1.584	0.071	0.019	0.038
3	0.3	0.0002	0.05	\widehat{f}_h	2.544	0.098	0.026	0.054
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	3.402	0.231	0.078	0.226
4	0.3	0.002	0.05	\widehat{f}_h	2.524	0.092	0.028	0.053
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	3.269	0.154	0.029	0.232
5	10	0.002	0.0135	\widehat{f}_h	0.091	0.031	0.017	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.080	0.037	0.035	0.052
6	10	0.02	0.0135	\widehat{f}_h	0.085	0.033	0.014	0.023
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.151	0.048	0.033	0.034
7	10	0.002	0.05	\widehat{f}_h	0.111	0.031	0.016	0.025
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.687	0.061	0.016	0.034
8	10	0.02	0.05	\widehat{f}_h	0.121	0.035	0.014	0.025
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.181	0.053	0.023	0.040
9	100	0.2	0.0135	\widehat{f}_h	0.106	0.032	0.014	0.024
				$\widetilde{f}_{\widetilde{m},\widetilde{s}}$	0.123	0.062	0.091	0.046

3.6 Application to neuronal data

3.6.1 Dataset

We describe quickly the data but we refer to Yu *et al.* (2004); Lansky *et al.* (2006) for example for details on data acquisition. The data are intracellular measurements of the membrane potential in volts along time, for one single neuron of a pig between the spikes. This is the depolarization phase. In this neuronal context, between the $(j - 1)^{\text{th}}$ and the j^{th} spike, the depolarization of the membrane potential receiving a random input can be described by the Ornstein-Uhlenbeck model with one random effect (3.1.1). The spikes are not intrinsic to the model but are generated when the voltage reaches for the first time a certain threshold S , then the process is reset to a fixed initial voltage. Thus each trajectory is observed on an interval $[0, T_j]$ where $T_j = \inf\{t > 0, X_j(t) \geq S\}$. The initial voltage (the value following a spike) is assumed to be equal to the resting potential. The present dataset has been normalised to obtain N trajectories which begin in zero: $x_j = 0$.

The positive constant parameter α is called the time constant of the neuron (the coefficient of decay in the exponential, when there is no noise), which is intrinsic to the neuron and fixed to $\alpha = 0.039$ [s] (Lansky *et al.*, 2006). The diffusion coefficient σ [V/ $\sqrt{\text{s}}$] has been estimated using the estimator $\hat{\sigma}^2 = (1/N) \sum_{j=1}^N \left((1/\mathcal{K}) \sum_{k=1}^{\mathcal{K}} ((X_j(\delta(k+1)) - X_j(\delta k))^2 / \delta) \right)$. We obtain $\hat{\sigma} = 0.0135$, which is the same value as that used in Picchini *et al.* (2008). The ϕ_j represents the local average input that the neuron receives during the j^{th} inter-spike interval. We assume that ϕ_j changes from one trajectory to another because of other neurons or the influence of the environment, for example. So parameters ϕ and σ characterize the input, while α , x_j (the resting potential) and S (the firing threshold) describe the neuron irrespectively of the incoming signal (Picchini *et al.*, 2008).

Data are composed of $N = 312$ inter-spike trajectories. For each interval $[0, T_j]$ the time step is the same: $\delta = 0.00015$ [s]. We decide to keep only realizations with more than 2000 observations ($T_j/\delta \geq 2000$). Finally we have $N = 240$ realizations with $\mathcal{K} = 2000$ observations and for $j = 1, \dots, N$, $T = T_j = 0.3$ [s]. Also the data are normalized in order to begin with zero at the initial time. The study of the units of measurement can highlight the collections given in Section 3.4. One can notice that the unit of measurement of v in the integrand must be [s/V] (same unit as $1/Z_{j,\tau}$) such that the exponential terms are without unit. The unit of s is [$\sqrt{\text{s}}/\text{V}$], and the choice of \mathcal{M} with the same unit as v seems natural.

It is interesting to note that the normality of the $Z_{j,T}$ is rejected by Shapiro and Wilk test (p-value 10^{-7}) and Kolmogorov-Smirnov test (p-value 10^{-3}). This suggests that the ϕ_j 's are not Gaussian. Thus we want to estimate nonparametrically their density. In the following we compare our results to the estimation obtained in Picchini *et al.* (2008) under the parametric Gaussian assumption.

3.6.2 Comparison of estimators

The estimation of the density f obtained by Picchini *et al.* (2008) under the Gaussian assumption on these data are $\mathcal{N}(\mu = 0.278, \eta^2 = (0.041)^2)$. Using a maximum-likelihood estimator on the $(Z_{j,T})$'s we obtain for the mean 0.270 and for the standard deviation 0.046. We notice that these two estimations are close to that of Picchini *et al.* (2008). We use our two nonparametric estimators to see how close to a Gaussian density they are.

On Figure 3.4 we represent both estimators $\hat{f}_{\hat{\eta}}$ and $\hat{f}_{\tilde{m}, \tilde{s}}$ applied on the real data and the density $\mathcal{N}(\mu, \eta^2)$. The two estimations are close, and close to the estimation of Picchini *et al.* (2008). However, it is also legitimate to think about a Gamma distribution to model the random parameters ϕ_j 's because, as the data have been normalised, the estimated random effects

are positive. Then it seems reasonable to use a non-negative random variable to model this local average input. Thus, a Gamma distribution may seem more appropriate than a Gaussian distribution, even if the chosen Gaussian has small probability to be negative. We look for the Gamma distribution which has for mean $\mu = 0.278$ and for standard deviation $\eta = 0.041$. This corresponds to a Gamma distribution with the shape parameter 46.3 and the scale parameter 0.006. We notice the similarity between the previous Gaussian curve and the new one. Thus this distribution seems also suitable to fit the distribution of the ϕ_j 's, as Figure 3.4 shows.

The Gaussian assumption is strong and leads to parametric tractable models. The present work confirms that this approximation is acceptable. However, the nonparametric estimation gives a density for the ϕ_j 's that can be used to simulate the random effect and could be closer to the true one.

Notice that, as mentioned in introduction, Comte *et al.* (2013)'s estimator cannot handle small values of T while our new proposals are successful in such case. Let us precise this point. If the number of points is large enough, the variable $Z_{j,T}$ can be approximate

$$Z_{j,T} \approx \left(X_j(T) - X_j(0) + \frac{\delta}{\alpha} \sum_{l=1}^{\mathcal{K}} X((l-1)\delta) \right) \frac{1}{T}.$$

It is unchanged when we change the units: \mathbf{V}, \mathbf{s} to \mathbf{mV}, \mathbf{ms} thus $T = 0.3$ to $T = 300$. However, the deconvolution estimator of Comte *et al.* (2013) for $\tau = T$ is changing when the value of T is changing. In fact, the estimator is

$$\begin{aligned} \widehat{f}_T(x) &= \frac{1}{2\pi} \int_{-\sqrt{T}}^{\sqrt{T}} e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,T}} e^{\frac{u^2\sigma^2}{2T}} du \\ &\approx \frac{1}{2\pi} \sum_{k=1}^{n_{\text{pas}}} \left((u_{k+1} - u_k) e^{-iu_k x} \frac{1}{N} \sum_{j=1}^N e^{iu_k Z_{j,T}} e^{\frac{u_k^2\sigma^2}{2T}} \right) \end{aligned}$$

and according to $T = 0.3$ or $T = 300$ the values of u and thus the interval of integration and the three exponential terms are changing. Finally estimator \widehat{f}_T is changing with the units. And the interval of integration is not large enough in the case $T = 0.3$ to give a good estimation.

To solve this problem we have proposed a new estimator $\widetilde{f}_{\widehat{m}, \widehat{s}}$ to allow the user to deal with data with the units he/she wants and to not oblige him/her to change it.

But one can wonder if the new estimators are robust when increasing T . Indeed, our method works for larger T . Precisely changing volts in millivolts and seconds in milliseconds implies $T = 300$, $\sigma = 0.426$, and on simulated data we adequately reconstruct the shape of the density.

3.7 Discussion

In this work we study a stochastic differential Ornstein-Uhlenbeck mixed-effects model. We propose two estimators of the density of the random effect. Both estimators are not very sensitive to the effect of the time of observation T . Indeed the kernel strategy corresponds to a context with large T while we built a deconvolution estimator especially for small values of T . Both are data-driven and satisfy an oracle-type inequality. According to the numerical study, the kernel estimator seems to be the efficient one: the numerical results are convincing and close to the ones obtained by cross-validation. Besides we provide non-asymptotic theoretical results. Furthermore we study neuronal data with nonparametric estimation strategy. Instead of making any parametric assumptions for the random effect distribution, we build an estimator of its density. Future work based on this estimation could be more precise and closer to the

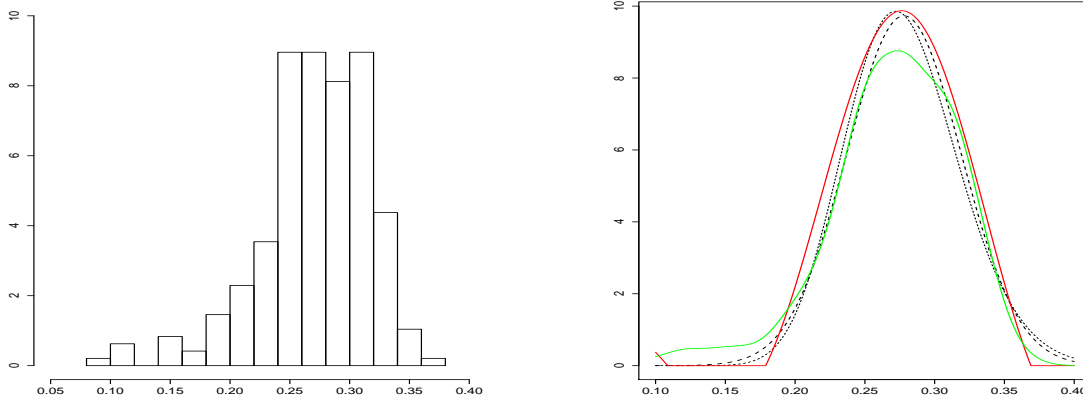


Figure 3.4 – Real data. In bold green (grey) estimator \hat{f}_h , in bold red (black) $\tilde{f}_{\tilde{m},\tilde{s}}$, the black dotted and bold line the density $\mathcal{N}(\mu, \eta^2)$ from Picchini *et al.* (2008) and the black dotted thin line the density $\Gamma(46.3, 0.006)$

real neuronal data. To complete the study, the method for different times of observation T_j could be settled up. Besides, some goodness-of-fit tests could be produced; we refer to Bissantz *et al.* (2007) who construct confidence bands for an estimator of f in the ordinary smooth deconvolution problem.

The model can be completed by adding another random effect: the time constant of the neuron. Picchini & Ditlevsen (2011) have investigated this model in a parametric way and Chapter 4 in a nonparametric way. A recent work Delattre *et al.* (2015b) assumes that the density of the random effect is a Gaussian mixture and uses data clustering method, which is an interesting approach for the data that we described. The question of a random effect in the diffusion coefficient also is open (see Delattre *et al.* (2015a)). Moreover, a model with a drift $b(X_j(t)) + \phi_j$, where b is a known function, can be treated with the same method. However, dealing with a diffusion $\sigma(X_j(t))$ where σ is a known function is a more complex problem.

3.8 Proofs

3.8.1 Proof of Theorem 3.3.2

Given $h \in \mathcal{H}_{N,T}$, we denote:

$$V(h) = \kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + \kappa_2 \frac{\sigma^4 \|K\|_1^2 \|K''\|^2}{T^2 h^5} =: V_1(h) + V_2(h).$$

Using the definition of $A(h)$ and of \hat{h} we obtain

$$\begin{aligned} \|\hat{f}_{\hat{h}} - f\|^2 &\leq 3\|\hat{f}_{\hat{h}} - \hat{f}_{\hat{h},\hat{h}}\|^2 + 3\|\hat{f}_{\hat{h},\hat{h}} - \hat{f}_{\hat{h}}\|^2 + 3\|\hat{f}_{\hat{h}} - f\|^2 \\ &\leq 3\left(A(\hat{h}) + V(\hat{h})\right) + 3\left(A(\hat{h}) + V(\hat{h})\right) + 3\|\hat{f}_{\hat{h}} - f\|^2 \\ &\leq 6A(\hat{h}) + 6V(\hat{h}) + 3\|\hat{f}_{\hat{h}} - f\|^2. \end{aligned}$$

Thus,

$$\mathbb{E}[\|\hat{f}_{\hat{h}} - f\|^2] \leq 6\mathbb{E}[A(\hat{h})] + 6V(\hat{h}) + 3\mathbb{E}[\|\hat{f}_{\hat{h}} - f\|^2],$$

hence, we only have to study the term $\mathbb{E}[A(\hat{h})]$. We can decompose $\|\hat{f}_{\hat{h},h'} - \hat{f}_{\hat{h}}\|^2$ as follows:

$$\|\widehat{f}_{h,h'} - \widehat{f}_{h'}\|^2 \leq 5\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 + 5\|f_{h,h'} - f_{h'}\|^2 + 5\|f_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 + 5\|\mathbb{E}[\widehat{f}_{h'}] - \widehat{f}_{h'}\|^2$$

thus

$$A(h) \leq 5(D_1 + D_2 + D_3 + D_4 + D_5)$$

with:

$$\begin{aligned} D_1 &:= \sup_{h' \in \mathcal{H}_{N,T}} \|f_{h,h'} - f_{h'}\|^2, \\ D_2 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h'} - \mathbb{E}[\widehat{f}_{h'}]\|^2 - \frac{V_1(h')}{10} \right)_+, \quad D_3 := \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\widehat{f}_{h,h'} - \mathbb{E}[\widehat{f}_{h,h'}]\|^2 - \frac{V_1(h')}{10} \right)_+, \\ D_4 &:= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10} \right)_+, \quad D_5 := \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10} \right)_+. \end{aligned}$$

According to Young's inequality (see Theorem B.0.1 Annexe B), we obtain

$$\|f_{h,h'} - f_{h'}\|^2 = \|K_{h'} \star (f_h - f)\|^2 \leq \|K_{h'}\|_1^2 \|f_h - f\|^2 = \|K\|_1^2 \|f_h - f\|^2$$

thus

$$D_1 \leq \|K\|_1^2 \|f_h - f\|^2. \quad (3.8.1)$$

Let us study the term D_2 . We denote $\mathcal{B}(1) = \{g \in \mathbb{L}^2(\mathbb{R}), \|g\| = 1\}$. We define

$$\nu_{N,h}(g) := \langle g, \widehat{f}_h - \mathbb{E}[\widehat{f}_h] \rangle$$

then $|\nu_{N,h}(g)| \leq \|g\| \|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|$ thus, the estimator \widehat{f}_h satisfies:

$$\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2 = \sup_{g \in \mathcal{B}(1)} (\nu_{N,h}(g))^2.$$

We can also compute the scalar product which defines $\nu_{N,h}$ and we obtain

$$\nu_{N,h}(g) = \frac{1}{N} \sum_{j=1}^N (g \star K_h^-(Z_{j,T}) - \mathbb{E}[g \star K_h^-(Z_{j,T})]) \quad (3.8.2)$$

with $K_h^-(x) := K_h(-x)$. This finally conducts to:

$$\mathbb{E}[D_2] \leq \sum_{h' \in \mathcal{H}_{N,T}} \mathbb{E} \left[\sup_{g \in \mathcal{B}(1)} (\nu_{N,h'}(g))^2 - \frac{V_1(h')}{10} \right]_+.$$

This bound and Equation (3.8.2) leads to apply Talagrand's Theorem (see Annexe B). We have to compute 3 quantities: M , H^2 and v .

First:

$$\begin{aligned} \sup_{g \in \mathcal{B}(1)} \|g \star K_{h'}^-\|_\infty &= \sup_{g \in \mathcal{B}(1)} \sup_{x \in \mathbb{R}} \left| \int g(y) K_{h'}^-(x-y) dy \right| = \sup_{g \in \mathcal{B}(1)} \sup_{x \in \mathbb{R}} |\langle g, K_{h'}^-(\cdot - x) \rangle| \\ &\leq \sup_{g \in \mathcal{B}(1)} \|g\| \|K_{h'}\| = \frac{\|K\|}{\sqrt{h'}} := M. \end{aligned} \quad (3.8.3)$$

Secondly, the bound of Proposition 3.3.1 gives

$$\mathbb{E} \left[\sup_{g \in \mathcal{B}(1)} (\nu_{N,h}(g))^2 \right] = \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2] \leq \frac{\|K\|_5^2}{Nh} := H^2. \quad (3.8.4)$$

Thirdly:

$$\begin{aligned} \sup_{g \in \mathcal{B}(1)} (\text{Var}(g \star K_{h'}^-(Z_{1,T}))) &\leq \sup_{g \in \mathcal{B}(1)} \mathbb{E}[(g \star K_{h'}^-(Z_{1,T}))^2] \\ &\leq 2 \sup_{g \in \mathcal{B}(1)} \mathbb{E}[(g \star K_{h'}^-(\phi_1))^2] + 2 \sup_{g \in \mathcal{B}(1)} \mathbb{E}[(g \star (K_{h'}^-(Z_{1,T}) - K_{h'}^-(\phi_1)))^2]. \end{aligned}$$

Let us investigate the two terms separately. Young's inequality gives:

$$\mathbb{E}[(g \star K_{h'}^-(\phi_1))^2] = \int (g \star K_{h'}^-(x))^2 f(x) dx \leq \|f\| \|g \star K_{h'}^-\|_4^2 = \frac{\|f\| \|K\|_{4/3}^2}{\sqrt{h'}} := v_1. \quad (3.8.5)$$

Then, one can write: $K_{h'}(x - Z_{1,T}) - K_{h'}(x - \phi_1) = (\phi_1 - Z_{1,T}) \int_0^1 (K_{h'}')'(x - \phi_1 + u(\phi_1 - Z_{1,T})) du$, thus

$$\begin{aligned} (g \star K_{h'}^-(Z_{1,T}) - g \star K_{h'}^-(\phi_1))^2 &= (\phi_1 - Z_{1,T})^2 \left(\int g(x) \int_0^1 (K_{h'}')'(x - \phi_1 + u(\phi_1 - Z_{1,T})) du dx \right)^2 \\ &\leq (\phi_1 - Z_{1,T})^2 \int g^2(x) \left(\int_0^1 (K_{h'}')'^2(x - \phi_1 + u(\phi_1 - Z_{1,T})) du \right) dx \\ &\leq (\phi_1 - Z_{1,T})^2 \|g\|^2 \int (K_{h'}')'^2(y) dy = (\phi_1 - Z_{1,T})^2 \|(K_{h'}')'\|^2. \end{aligned}$$

With $\mathbb{E}[(\phi_1 - Z_{1,T})^2] = \frac{\sigma^2}{T^2} \mathbb{E}[W_1(T)^2] = \frac{\sigma^2}{T}$, the assumption $T^{-1} \leq h^{5/2}$ leads to

$$\mathbb{E}[(g \star K_{h'}^-(Z_{1,T}) - g \star K_{h'}^-(\phi_1))^2] \leq \frac{\|K'\|^2 \sigma^2}{h^3 T} \leq \frac{\|K'\|^2 \sigma^2}{\sqrt{h'}} := v_2. \quad (3.8.6)$$

Finally $v = v_1 + v_2 = A_0/\sqrt{h'}$ with $A_0 = \|f\| \|K\|_{4/3}^2 + \|K'\|^2 \sigma^2$.

If $\kappa_1 \|K\|_1^2 \geq 40$, with the assumption $1/(Nh) \leq 1$, Talagrand's inequality (under the assumptions of the Theorem 3.3.2) gives

$$\begin{aligned} \mathbb{E} \left(\sup_{g \in \mathcal{B}(1)} (\nu_{N,h'}(g))^2 - \frac{V_1(h')}{10} \right)_+ &\leq \frac{C_1}{N\sqrt{h'}} e^{-C_2/\sqrt{h'}} + C_3 \frac{1}{h'N^2} e^{-C_4\sqrt{N}} \\ &\leq \frac{C_5}{N} \sum_{h' \in \mathcal{H}_{N,T}} \frac{1}{\sqrt{h'}} e^{-C_6/\sqrt{h'}} \leq \frac{C_5 S(C_6)}{N}. \end{aligned}$$

One can lead the study of D_3 as we have done for D_2 , using the same steps and tools. However $K_h \star K_{h'}$ instead of $K_{h'}$, adds $\|K\|_1$ in M and $\|K\|_1^2$ in H^2 and v .

Then, let us study the term D_4 . If $\kappa_2 \geq 10/(3\|K\|_1^2)$, the bound (3.3.2) leads us to

$$\begin{aligned} D_4 &= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h'}] - f_{h'}\|^2 - \frac{V_2(h')}{10} \right)_+ \\ &\leq \sup_{h' \in \mathcal{H}_{N,T}} \left(\frac{\|K''\|^2 \sigma^4}{3h'^5 T^2} - \frac{\kappa_2 \|K\|_1^2 \|K''\|^2 \sigma^4}{10T^2 h'^5} \right)_+ = 0 \end{aligned}$$

thus $D_4 = 0$. Finally, similarly, if $\kappa_2 \geq 10/3$, we obtain

$$\begin{aligned} D_5 &= \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\mathbb{E}[\widehat{f}_{h,h'}] - f_{h,h'}\|^2 - \frac{V_2(h')}{10} \right)_+ \\ &\leq \sup_{h' \in \mathcal{H}_{N,T}} \left(\frac{\|K''\|^2 \|K\|_1^2 \sigma^4}{3h^5 T^2} - \frac{\kappa_2 \|K\|_1^2 \|K''\|^2 \sigma^4}{10T^2 h'^5} \right)_+ = 0. \end{aligned}$$

Thus finally we obtained that:

$$\mathbb{E}[A(h)] \leq 5 \left(\|K\|_1^2 \|f_h - f\|^2 + \frac{c}{N} \right) \quad (3.8.7)$$

with c a constant depending on $\|f\|, \|K\|_1, \|K\|, \|K\|_{4/3}$. Finally we have shown that for all $h \in \mathcal{H}_{N,T}$:

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_h - f\|^2] &\leq 6\kappa_1 \frac{\|K\|_1^2 \|K\|^2}{Nh} + 6\kappa_2 \frac{\|K\|_1^2 \|K''\|^2 \sigma^4}{T^2 h^5} + 3 \left(2\|f - f_h\|^2 + \frac{\|K\|^2}{Nh} + \frac{2\|K''\|^2 \sigma^4}{3T^2 h^5} \right) \\ &\quad + 30 \left(\|K\|_1^2 \|f - f_h\|^2 + \frac{c}{N} \right) \\ &\leq \left(6 + \frac{3}{\|K\|_1^2 \kappa_1} \right) V_1(h) + \left(6 + \frac{9}{2\|K\|_1^2 \kappa_2} \right) V_2(h) + (30\|K\|_1^2 + 6)\|f_h - f\|^2 + \frac{C}{N} \\ &\leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \{ \|f - f_h\|^2 + V(h) \} + \frac{C_2}{N}. \end{aligned}$$

where $C_1 = \max(7, 30\|K\|_1^2 + 6)$ and C_2 depends on $\|f\|, \|K\|_1, \|K\|, \|K\|_{4/3}$. \square

3.8.2 Proof of Proposition 3.4.1

The bias term is $\|f - \mathbb{E}[\widetilde{f}_{m,s}]\|^2$. Let us compute $\mathbb{E}[\widetilde{f}_{m,s}]$. As the $Z_{j,\tau}$ are *i.i.d.* when τ is fixed and due to the independence of ϕ_1 and W_1 , we obtain:

$$\begin{aligned} \mathbb{E}[\widetilde{f}_{m,s}(x)] &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} \mathbb{E} \left[e^{iuZ_{1,m^2/s^2} + u^2 \sigma^2 s^2 / (2m^2)} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} \mathbb{E} \left[e^{iu\phi_1 + iu\sigma W_1(m^2/s^2)s^2/m^2 + u^2 \sigma^2 s^2 / (2m^2)} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux + u^2 \sigma^2 s^2 / (2m^2)} f^*(u) \mathbb{E} \left[e^{iu\sigma W_1(m^2/s^2)s^2/m^2} \right] du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux + u^2 \sigma^2 s^2 / (2m^2)} f^*(u) e^{-u^2 \sigma^2 s^2 / (2m^2)} du \\ &= \frac{1}{2\pi} \int_{-m}^m e^{-iux} f^*(u) du =: f_m(x). \end{aligned}$$

Therefore this gives $\mathbb{E}[\widetilde{f}_{m,s}(x)] = f_m(x)$, and $\|f - \mathbb{E}[\widetilde{f}_{m,s}]\|^2 = \|f - f_m\|^2 = \frac{1}{2\pi} \int_{|u| \geq m} |f^*(u)|^2 du$. The variance term is:

$$\begin{aligned} \mathbb{E} \left[\|\widetilde{f}_{m,s} - f_m\|^2 \right] &= \frac{1}{2\pi} \mathbb{E} \left[\int_{-m}^m \left| \frac{1}{N} \sum_{j=1}^N e^{iuZ_{j,m^2/s^2}} e^{\frac{u^2 \sigma^2 s^2}{2m^2}} - f^*(u) \right|^2 du \right] \\ &= \frac{1}{2\pi N} \int_{-m}^m e^{\frac{u^2 \sigma^2 s^2}{m^2}} \text{Var} \left(e^{iuZ_{1,m^2/s^2}} \right) du \\ &\leq \frac{1}{2\pi N} \int_{-m}^m e^{\frac{u^2 \sigma^2 s^2}{m^2}} du = \frac{m}{\pi N} \int_0^1 e^{s^2 \sigma^2 v^2} dv. \quad \square \end{aligned}$$

3.8.3 Proof of Theorem 3.4.3

Let us study the term $\|\tilde{f}_{\tilde{m},\tilde{s}} - f\|^2$. We decompose it into a sum of three terms and the definition of (\tilde{m}, \tilde{s}) (3.4.3) implies for all $(m, s) \in \mathcal{C}$

$$\begin{aligned} \|\tilde{f}_{\tilde{m},\tilde{s}} - f\|^2 &\leq 3 \left(\|\tilde{f}_{\tilde{m},\tilde{s}} - \tilde{f}_{(\tilde{m},\tilde{s})\wedge(m,s)}\|^2 + \|\tilde{f}_{(\tilde{m},\tilde{s})\wedge(m,s)} - \tilde{f}_{m,s}\|^2 + \|\tilde{f}_{m,s} - f\|^2 \right) \\ &\leq 3 (\Gamma_{m,s} + \text{pen}(\tilde{m}, \tilde{s})) + 3 (\Gamma_{\tilde{m},\tilde{s}} + \text{pen}(m, s)) + 3 \|\tilde{f}_{m,s} - f\|^2 \\ &\leq 6\Gamma_{m,s} + 6\text{pen}(m, s) + 3\|\tilde{f}_{m,s} - f\|^2 \end{aligned} \quad (3.8.8)$$

Now we study $\Gamma_{m,s}$. First:

$$\|\tilde{f}_{(m,s)\wedge(m',s')} - \tilde{f}_{m',s'}\|^2 \leq 3 \left(\|\tilde{f}_{m',s'} - f_{m'}\|^2 + \|f_{m'} - f_{m\wedge m'}\|^2 + \|f_{m\wedge m'} - \tilde{f}_{(m',s')\wedge(m,s)}\|^2 \right).$$

Thus:

$$\begin{aligned} \Gamma_{m,s} &\leq \max_{(m',s') \in \mathcal{C}} \left(3\|\tilde{f}_{m',s'} - f_{m'}\|^2 + 3\|f_{m'} - f_{m\wedge m'}\|^2 + 3\|f_{m\wedge m'} - \tilde{f}_{(m',s')\wedge(m,s)}\|^2 - \text{pen}(m', s') \right)_+ \\ &\leq 3 \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &\quad + 3 \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &\quad + 3 \max_{m' \in \mathcal{M}} \|f_{m'} - f_{m\wedge m'}\|^2. \end{aligned}$$

The last maximum can be explicit. If $m' \leq m$, then $\|f_{m'} - f_{m\wedge m'}\|^2 = \|f_{m'} - f_{m'}\|^2 = 0$. Otherwise,

$$\|f_{m'} - f_{m\wedge m'}\|^2 = \|f_{m'} - f_m\|^2 = \int_{m \leq |u| \leq m'} |f^*(u)|^2 du \leq \|f - f_m\|^2.$$

Finally:

$$\max_{m' \in \mathcal{M}} \|f_{m'} - f_{m\wedge m'}\|^2 \leq \|f - f_m\|^2.$$

We get the following bound for $\Gamma_{m,s}$:

$$\begin{aligned} \Gamma_{m,s} &\leq 3 \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &\quad + 3 \max_{(m',s') \in \mathcal{C}} \left(\|\tilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &\quad + 3\|f - f_m\|^2. \end{aligned} \quad (3.8.9)$$

Then we gather Equations (3.8.8) and (3.8.9):

$$\begin{aligned} \|\tilde{f}_{\tilde{m},\tilde{s}} - f\|^2 &\leq 6\text{pen}(m, s) + 3\|\tilde{f}_{m,s} - f\|^2 + 18\|f - f_m\|^2 + \max_{(m',s') \in \mathcal{C}} 18 \left(\|\tilde{f}_{m',s'} - f_{m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+ \\ &\quad + \max_{(m',s') \in \mathcal{C}} 18 \left(\|\tilde{f}_{(m',s')\wedge(m,s)} - f_{m\wedge m'}\|^2 - \frac{1}{6}\text{pen}(m', s') \right)_+. \end{aligned}$$

We first notice that our penalty function is increasing in s and m , thus we get the following bound for the last term:

$$\begin{aligned}
 & \mathbb{E} \left[\max_{(m', s') \in \mathcal{C}} \left(\|\tilde{f}_{(m', s') \wedge (m, s)} - f_{m \wedge m'}\|^2 - \frac{1}{6} \text{pen}((m', s') \wedge (m, s)) \right)_+ \right] \\
 & \leq \mathbb{E} \left[\max_{m' \leq m, s' \leq s} \left(\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right)_+ \right] + \mathbb{E} \left[\max_{m \leq m', s \leq s'} \left(\|\tilde{f}_{m, s} - f_m\|^2 - \frac{1}{6} \text{pen}(m, s) \right)_+ \right] \\
 & \quad + \mathbb{E} \left[\max_{m \leq m', s' \leq s} \left(\|\tilde{f}_{m, s'} - f_m\|^2 - \frac{1}{6} \text{pen}(m, s') \right)_+ \right] + \mathbb{E} \left[\max_{m' \leq m, s \leq s'} \left(\|\tilde{f}_{m', s} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s) \right)_+ \right] \\
 & \leq 4 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ .
 \end{aligned}$$

Moreover, according to Proposition 3.4.1 and using the inequality $\int_0^1 e^{\sigma^2 s^2 v^2} dv \leq e^{\sigma^2 s^2}$, we obtain, for all $(m, s) \in \mathcal{C}$,

$$\begin{aligned}
 \mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] & \leq 5 \times 18 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ + 6 \text{pen}(m, s) \\
 & \quad + 3 \frac{m}{\pi N} e^{\sigma^2 s^2} + 21 \|f - f_m\|^2 .
 \end{aligned}$$

Then we obtain the announced result with the following Lemma.

Lemma 3.8.1. *There exists a constant $C' > 0$ such that for $\text{pen}(m, s)$ defined by $\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2}$,*

$$\sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ \leq \frac{C'(P+1)}{N} .$$

According to Lemma 3.8.1, to be proved next, we choose $\text{pen}(m, s) = \kappa \frac{m}{N} e^{\sigma^2 s^2}$, thus, there exist two constants $C = 145$, $C' > 0$ such that,

$$\begin{aligned}
 \mathbb{E}[\|\tilde{f}_{\tilde{m}, \tilde{s}} - f\|^2] & \leq 5 \times 18 \sum_{m' \in \mathcal{M}} \sum_{s' \in \mathcal{S}} \mathbb{E} \left[\|\tilde{f}_{m', s'} - f_{m'}\|^2 - \frac{1}{6} \text{pen}(m', s') \right]_+ + \left(6\kappa + \frac{3}{\pi} \right) \frac{m}{N} e^{\sigma^2 s^2} + 21 \|f - f_m\|^2 \\
 & \leq C \inf_{(m, s) \in \mathcal{C}} \{ \|f - f_m\|^2 + \frac{m}{N} e^{\sigma^2 s^2} \} + \frac{C'}{N} . \quad \square
 \end{aligned}$$

Proof of Lemma 3.8.1

For a couple $(m, s) \in \mathcal{C}$ fixed, let us consider the subset $S_m := \{t \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R}), \text{supp}(t^*) = [-m, m]\}$. For $t \in S_m$,

$$\nu_N(t) = \frac{1}{N} \sum_{j=1}^N (\varphi_t(Z_{j, m^2/s^2}) - \mathbb{E}[\varphi_t(Z_{j, m^2/s^2})])$$

with $\varphi_t(x) := \frac{1}{2\pi} \int \overline{t^*(u)} e^{iux + \sigma^2 u^2 s^2 / (2m^2)} du$, then $\nu_N(t) = \frac{1}{2\pi} \langle t^*, (\tilde{f}_{m, s} - f_m)^* \rangle$. This leads to

$$\|\tilde{f}_{m, s} - f_m\|^2 = \sup_{t \in S_m, \|t\|=1} |\nu_N(t)|^2 . \tag{3.8.10}$$

We also have by Cauchy-Schwarz inequality

$$\begin{aligned}
 \|\varphi_t\|_\infty & \leq \frac{1}{2\pi} \int |t^*(u)| e^{\sigma^2 u^2 s^2 / (2m^2)} du \leq \frac{1}{2\pi} \left(\int_{-m}^m |t^*(u)|^2 du \right)^{1/2} \left(\int_{-m}^m e^{\sigma^2 u^2 s^2 / m^2} du \right)^{1/2} \\
 & \leq \frac{\sqrt{2m}}{\sqrt{2\pi}} e^{\sigma^2 s^2 / 2}
 \end{aligned}$$

thus

$$\sup_{t \in S_m, \|t\|=1} \|\varphi_t\|_\infty \leq \frac{\sqrt{m}}{\sqrt{\pi}} e^{\sigma^2 s^2 / 2} := M.$$

Then, by Proposition 3.4.1,

$$\mathbb{E} \left[\sup_{t \in S_m, \|t\|=1} |\nu_N(t)|^2 \right] = \mathbb{E} \left[\|\tilde{f}_{m,s} - f_m\|^2 \right] \leq \frac{m}{\pi N} \int_0^1 e^{\sigma^2 s^2 v^2} dv \leq \frac{m}{\pi N} e^{\sigma^2 s^2} := H^2.$$

Using Fubini and Cauchy-Schwarz inequalities we obtain for all $(m, s) \in \mathcal{C}$:

$$\begin{aligned} 4\pi \sup_{t \in S_m, \|t\|=1} \text{Var}(\varphi_t(Z_{j,m^2/s^2})) &\leq \sup_{t \in S_m, \|t\|=1} \iint t^*(u)t^*(-v) \mathbb{E} \left[e^{i(u-v)Z_{j,m^2/s^2}} \right] e^{(u^2+v^2)\sigma^2 s^2 / (2m^2)} dudv \\ &\leq 2\pi \left(\iint_{[-m,m]^2} |f^*(u-v)|^2 e^{(u^2+v^2)\sigma^2 s^2 / m^2} dudv \right)^{1/2} \\ &\leq 2\pi \left(e^{2\sigma^2 s^2} \iint_{[-m,m]^2} |f^*(u-v)|^2 dudv \right)^{1/2} \\ &\leq 2\pi e^{\sigma^2 s^2} \sqrt{2m} \left(\int_{-2m}^{2m} |f^*(z)|^2 dz \right)^{1/2} \leq 2\sqrt{2m} \sqrt{2\pi} \sqrt{\pi} e^{\sigma^2 s^2} \|f\| =: 4\pi^2 v, \\ v &:= \frac{\sqrt{m} e^{\sigma^2 s^2} \|f\|}{\sqrt{\pi}}. \end{aligned}$$

Finally using that $m \leq N$, $s \leq 2/\sigma$ and $\sum_{s \in \mathcal{S}} s = (4/\sigma)(1 - (1/2)^{P+1}) < 4/\sigma$, the Talagrand's inequality with $\alpha = 1/2$ if $4H^2 \leq \text{pen}(m, s)/6$ implies,

$$\begin{aligned} \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} \mathbb{E} \left[\|\tilde{f}_{m,s} - f_m\|^2 - \frac{1}{6} \text{pen}(s, m) \right]_+ &\leq \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} \left(\frac{C_1 \|f\|}{N} e^{\sigma^2 s^2} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} + C_3 \frac{m}{N^2} e^{\sigma^2 s^2} e^{-C_4 \sqrt{N}} \right) \\ &\leq \sum_{s \in \mathcal{S}} \frac{C_1 \|f\|}{N} e^{\sigma^2 s^2} \left(\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + \sum_{s \in \mathcal{S}} \sum_{m \in \mathcal{M}} C_3 e^4 \frac{1}{N} e^{-C_4 \sqrt{m}} \\ &\leq \frac{C_1 \|f\| (P+1) e^4}{N} \left(\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \right) + C_3 e^4 \frac{P+1}{N} \sum_{m \in \mathcal{M}} e^{-C_4 \sqrt{m}} \\ &\leq \frac{C'(P+1)}{N} \end{aligned}$$

because with the definition of \mathcal{M} , $\sum_{m \in \mathcal{M}} \sqrt{m} e^{-C_2 \frac{\sqrt{m}}{\|f\|}} \leq a_1 \sum_{k \in \mathbb{N}} k^{1/4} e^{-a_2 k^{1/4}} < +\infty$, and $\sum_{m \in \mathcal{M}} e^{-C_4 m^{1/2}} \leq \sum_{k \in \mathbb{N}} e^{-a_3 k^{1/4}} < +\infty$, with a_1, a_2, a_3 three positive constants. Notice that $C' > 0$ depends on $\sigma, \|f\|, \Delta$.

We choose $\text{pen}(m, s) = \kappa m e^{\sigma^2 s^2} / N$ with $\kappa \geq 24$. \square

Appendix

3.8.4 Discretization

Indeed, if we assume that the times of observations are the $t_k = k\delta, k = 1, \dots, N$ and $0 < \delta < 1$, we must study the error applied by discretization of the $Z_{j,\tau}$. Then, for any $0 < m^2/s^2 \leq T$ we use:

$$\hat{Z}_{j,m^2/s^2} = \frac{s^2}{m^2} \left[X_j(\delta[m^2/(s^2\delta)]) - X_j(0) + \frac{\delta}{\alpha} \sum_{k=1}^{\lfloor m^2/(s^2\delta) \rfloor} X_j((k-1)\delta) \right] \quad (3.8.11)$$

to approximate $Z_{j,m^2/s^2}$ given by (3.2.1). The corresponding estimator of f is

$$\widehat{f}_{m,s}(x) = \frac{1}{2\pi} \int_{-m}^m e^{-iux} \frac{1}{N} \sum_{j=1}^N e^{iu\widehat{Z}_{j,m^2/s^2}} e^{\frac{u^2\sigma^2 s^2}{2m^2}} du. \quad (3.8.12)$$

We investigate the error:

$$\mathbb{E}[\|\widehat{f}_{m,s} - f\|^2] \leq 2\mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] + 2\mathbb{E}[\|\widetilde{f}_{m,s} - f\|^2]$$

where the second term of the right hand side is bounded by Proposition 3.4.1. Then, Plancherel-Parseval's Theorem implies:

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq \frac{1}{2\pi} \mathbb{E} \left[\int_{-m}^m \frac{1}{N} \sum_{j=1}^N e^{u^2\sigma^2 s^2/m^2} \left| e^{iu\widehat{Z}_{j,m^2/s^2}} - e^{iuZ_{j,m^2/s^2}} \right|^2 du \right] \\ &\leq \frac{1}{2\pi} \int_{-m}^m e^{u^2\sigma^2 s^2/m^2} \mathbb{E} \left[\left| e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}} \right|^2 \right] du \end{aligned}$$

and

$$\mathbb{E} \left[\left| e^{iu\widehat{Z}_{1,m^2/s^2}} - e^{iuZ_{1,m^2/s^2}} \right|^2 \right] \leq |u|^2 \mathbb{E} \left[\left| \widehat{Z}_{1,m^2/s^2} - Z_{1,m^2/s^2} \right|^2 \right]$$

thus we study the last term. For all $(m, s) \in \mathcal{C}$, $m^2/s^2 \leq T$,

$$\begin{aligned} Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2} &= \frac{s^2}{m^2} (X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)])) \\ &\quad + \frac{s^2}{\alpha m^2} \sum_{k=1}^{\lfloor m^2/(s^2\delta) \rfloor} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \end{aligned}$$

then by Cauchy-Schwarz's inequality we obtain

$$\begin{aligned} (Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2 &\leq \frac{2s^4}{m^4} (X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2 \\ &\quad + \frac{2s^4}{\alpha^2 m^4} \left[\sum_{k=1}^{\lfloor m^2/(s^2\delta) \rfloor} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2. \end{aligned}$$

Hölder's inequality yields

$$\begin{aligned} \left[\sum_{k=1}^{\lfloor \frac{m^2}{s^2\delta} \rfloor} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2 &\leq \sum_{k=1}^{\lfloor \frac{m^2}{s^2\delta} \rfloor} \left[\int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta)) ds \right]^2 \left[\frac{m^2}{s^2\delta} \right] \\ &\leq \left[\frac{m^2}{s^2\delta} \right] \delta \sum_{k=1}^{\lfloor \frac{m^2}{s^2\delta} \rfloor} \int_{(k-1)\delta}^{k\delta} (X_j(s) - X_j((k-1)\delta))^2 ds. \end{aligned}$$

Let us study $\mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2]$, for $(k-1)\delta \leq s \leq k\delta$:

$$X_j(s) - X_j((k-1)\delta) = \int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha} \right) du + \int_{(k-1)\delta}^s \sigma dW_j(u)$$

and Cauchy-Schwarz's inequality gives

$$\begin{aligned} \mathbb{E}[(X_j(s) - X_j((k-1)\delta))^2] &\leq 2\mathbb{E}\left[\left(\int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha}\right) du\right)^2\right] + 2\mathbb{E}\left[\left(\int_{(k-1)\delta}^s \sigma dW_j(u)\right)^2\right] \\ &\leq 2\mathbb{E}\left[\int_{(k-1)\delta}^s \left(\phi_j - \frac{X_j(u)}{\alpha}\right)^2 du\right] + 2\delta\sigma^2 \\ &\leq 4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2. \end{aligned} \quad (3.8.13)$$

Finally, after simplification and using for all $x \in \mathbb{R}^+$, $[x] \leq x$,

$$\begin{aligned} \mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] &\leq \frac{2s^4}{m^4}\mathbb{E}[(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2] \\ &\quad + \frac{2}{\alpha^2}\left(4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2\right) \end{aligned}$$

and we can deal with the term $\mathbb{E}[(X_j(m^2/s^2) - X_j(\delta[m^2/(s^2\delta)]))^2]$ using formula (3.8.13) and $m^2/s^2 - \delta[m^2/(s^2\delta)] \leq \delta$. Thus:

$$\mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] \leq \left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)\left(4\delta^2\left(\mathbb{E}(\phi_j^2) + \frac{1}{\alpha^2}\sup_{s\geq 0}\mathbb{E}[X_j(s)^2]\right) + 2\delta\sigma^2\right).$$

Besides, for model (3.1.1), Equation (3.5.1) implies $\mathbb{E}[X_j(s)^2] \leq 3x_j^2 + 3\alpha^2\mathbb{E}[\phi_j^2] + 3\sigma^2$, and $0 < \delta < 1$ implies

$$\mathbb{E}\left[(Z_{1,m^2/s^2} - \widehat{Z}_{1,m^2/s^2})^2\right] \leq C\delta\left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)$$

with C a positive constant which does not depend on δ or m^2/s^2 . Finally,

$$\begin{aligned} \mathbb{E}[\|\widetilde{f}_{m,s} - \widetilde{f}_{m,s}\|^2] &\leq C\delta\left(\frac{2s^4}{m^4} + \frac{2}{\alpha^2}\right)\frac{1}{2\pi}\int_{-m}^m u^2 e^{u^2\sigma^2 s^2/m^2} du \\ &\leq C'\delta\left(\int_0^1 v^2 e^{v^2\sigma^2 s^2} dv\right)\left(\frac{s^4}{m} + \frac{m^3}{\alpha^2}\right). \end{aligned}$$

But $s \leq 2/\sigma$ and $m = \sqrt{k\Delta}/\sigma$, with $k \in \mathbb{N}^*$ and $0 < \Delta < 1$, thus we obtain

$$\mathbb{E}[\|\widetilde{f}_{m,s} - \widetilde{f}_{m,s}\|^2] \leq \frac{C'}{\sigma^3}\left(\int_0^1 v^2 e^{v^2\sigma^2 s^2} dv\right)\left(2^4\sqrt{k}\left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2}\left(\delta\Delta^{3/2}\right)\right).$$

Proposition 3.8.2. *Under (A), assuming $\mathbb{E}[\phi_j^2] < +\infty$, the estimator $\widehat{f}_{m,s}$ given by (3.8.12) satisfies*

$$\mathbb{E}\left[\|\widetilde{f}_{m,s} - f\|^2\right] \leq \|f_m - f\|^2 + \frac{\sqrt{k\Delta}}{\sigma\pi N}e^{\sigma^2 s^2} + \frac{C'}{\sigma^3}\frac{e^{\sigma^2 s^2}}{2\sigma^2 s^2}\left(2^4\sqrt{k}\left(\frac{\delta}{\sqrt{\Delta}}\right) + \frac{k^{3/2}}{\alpha^2}\left(\delta\Delta^{3/2}\right)\right).$$

Finally if Δ is fixed and δ is small, the error is acceptable. For example if $\delta = \Delta$ the error is of order $\sqrt{\delta}$.

For study on the kernel estimator we refer to Comte *et al.* (2013).

Chapitre 4

Estimation bidimensionnelle dans un modèle d'équations différentielles stochastiques à deux effets aléatoires

Dans ce chapitre nous étudions l'estimation d'effets aléatoires $\phi_j \in \mathbb{R}^2$ et de leur densité commune, dans le cas $d = 2$. Le modèle s'écrit

$$\begin{cases} dX_j(t) &= [\phi_{j,1}b_1(X_j(t)) + \phi_{j,2}b_2(X_j(t))]dt + \sigma(X_j(t))dW_j(t) \\ X_j(0) &= \gamma_j \end{cases}$$

c'est à dire $\phi_j = (\phi_{j,1}, \phi_{j,2})^t \in \mathbb{R}^2$, où u^t est la transposée du vecteur u , et le coefficient de dérive se réécrit $b^t(X_j(t))\phi_j$. Les fonctions b_1, b_2, σ sont connues. Les hypothèses sur ce modèle sont précisées dans le chapitre. Les deux effets aléatoires ne sont pas supposés indépendants et nous devons estimer une densité bivariable de \mathbb{R}^2 dans \mathbb{R} . Ce contexte bidimensionnel rend l'estimation plus difficile, en particulier à cause de l'utilisation d'opérateurs et de normes matriciels. Afin d'obtenir des résultats sur le risque quadratique intégré, nous supposons que le processus est à l'état stationnaire. Cette hypothèse importante simplifie la procédure et permet d'obtenir des résultats théoriques satisfaisants. La méthode d'estimation s'appuie sur l'écriture multiplicative du modèle. Comme dans le chapitre précédent, N trajectoires *i.i.d.* sont disponibles sur un intervalle commun $[0, T]$ donné, avec T et N grands. Pour chaque trajectoire $(X_j(t))_t$ nous regardons l'estimateur du maximum de vraisemblance $A_j(T)$ de ϕ_j . Puis, dans le but d'obtenir la convergence de cet estimateur, on étudie le risque $\mathbb{E}[\|A_j(T) - \phi_j\|_2^2]$ (où $\|\cdot\|_2$ désigne la norme euclidienne de \mathbb{R}^2). Or ces moments ne sont pas accessibles par le calcul. Nous introduisons alors un nouvel estimateur $\hat{A}_j(T)$, tronqué sur un domaine aléatoire bien choisi. Dans ce cas, nous obtenons une borne d'ordre $1/T$ de la distance moyenne dans \mathbb{R}^2 (comme dans le cas d'un seul effet aléatoire vu au Chapitre 3). Puis nous estimons la densité f avec un estimateur à noyau sur les estimateurs $\hat{A}_j(T)$ possédant de bonnes propriétés. Ce cas multidimensionnel n'avait pas encore été traité de façon non-paramétrique avant ce travail, à notre connaissance. Nous obtenons une borne du risque $\mathbb{L}^2(\mathbb{R}^2)$ de cet estimateur non-paramétrique. Puis nous menons une procédure de sélection des fenêtres $h = (h_1, h_2)$ inspirée d'une méthode de Goldenshluger et Lepski qui conduit à un choix adaptatif et à une inégalité de type oracle.

Tout le long de ce chapitre nous illustrerons nos propos sur les deux exemples OU et CIR, sur lesquels les hypothèses imposées par le cadre théorique se vérifient par le calcul. Après avoir présenté nos résultats nous proposons une analyse numérique sur simulations.

Ce chapitre est issu de l'article Dion & Genon-Catalot (2015) écrit avec Valentine Genon-Catalot et publié dans *Statistical Inference for Stochastic Processes*. La section Appendix 2 a été rajoutée pour figurer dans cette thèse.

Sommaire

4.1	Introduction	95
4.2	Notation and assumptions	96
4.2.1	General assumptions on the model	96
4.2.2	Key examples	97
4.2.3	Specific assumptions and notations for estimation	97
4.3	Random effect estimation	98
4.3.1	Definition of the estimator of the random effects	98
4.3.2	Main result	99
4.3.3	Key examples continued	99
4.4	Nonparametric estimation	100
4.4.1	Nonparametric estimator of the density of the random effects	100
4.4.2	Rates of convergence	101
4.4.3	Data-driven bandwidth and adaptation	102
4.5	Discrete data	103
4.6	Simulation study	104
4.7	Concluding remarks	110
4.8	Proofs	111
4.8.1	An auxiliary result and its proof	111
4.8.2	Proof of Proposition 4.3.1	113
4.8.3	Proof of proposition 4.4.1	116
4.8.4	Proof of Proposition 4.4.3	117
Appendix 1		118
4.8.5	Useful results of algebra	118
4.8.6	Details on examples	118
Appendix 2	: more about the MISE bounds	119

4.1 Introduction

Mixed-effects models are used to analyse repeated measurements with similar functional form but with some variability between experiments (see *e.g.* Davidian & Giltinan, 1995; Pinheiro & Bates, 2000; Diggle *et al.*, 2002). The advantage is that a single estimation procedure is used to fit the overall data simultaneously. In the present work the model of interest is a mixed-effects stochastic differential equation (SDE). Each equation represents the behaviour of one subject and corresponds to one realization of the random effects. Hence the random effects represent the particularity of each process. Mixed-effects SDEs have various applications such as neuronal or pharmacokinetic modelling (see *e.g.* Picchini *et al.*, 2008; Donnet & Samson, 2014). Estimation methods in SDEs with random effects have been proposed in literature. The main purposes are the estimation of the common distribution of the random effects in a parametric or nonparametric way. The estimation of the common density of the random effects is mainly parametric. Most methods assume normality of the random effects and estimate the population parameters (see *e.g.* Picchini *et al.*, 2008; Delattre *et al.*, 2013; DeVore & Lorentz, 1993). However, one can ask if this assumption is reasonable in some application contexts. Nonparametric estimation can allow us to get around this problem. To the best of our knowledge, the first reference in this context is Comte *et al.* (2013) who propose different nonparametric estimators and then the Chapter 3 develops two adaptive nonparametric estimators for the Orsntein-Uhlenbeck model with an application to a neuronal database. But these two references focus on a one-dimensional random effect.

In the present work we study the case of two random effects or, in other words, of one bidimensional random effect. We want to deal with two points: the estimation of the random effects and the nonparametric estimation of their common density. This bivariate context makes the study more complex. In fact the estimation of the random effects is done using matrix norm and operator, and anisotropy appears in the density estimation part. We consider N trajectories, observed on the interval $[0, T]$ where T is given. We denote u^t the transpose of a vector or a matrix u . For $j = 1, \dots, N$ the dynamics of each process is described by the stochastic differential equation

$$\begin{cases} dX_j(t) &= b^t(X_j(t))\phi_j dt + \sigma(X_j(t))dW_j(t) \\ X_j(0) &= \gamma_j \end{cases} \quad (4.1.1)$$

where $\phi_j = (\phi_{j,1}, \phi_{j,2})^t \in \mathbb{R}^2$ is the bidimensional random effect, $b(\cdot) = (b_1(\cdot), b_2(\cdot))^t$, $\sigma(\cdot)$ are known functions defined on \mathbb{R} , $(W_j)_{1 \leq j \leq N}$ are N independent Wiener processes and γ_j is a real valued random variable. The random variables $((\phi_1, \gamma_1), \dots, (\phi_N, \gamma_N))$ are *i.i.d.* and the sequences $((\phi_1, \gamma_1), \dots, (\phi_N, \gamma_N))$ and (W_1, \dots, W_N) are independent. The two random effects are not assumed independent. This model is more general than the ones investigated in Comte *et al.* (2013) where the drift has the form $b(\cdot)\phi$ or $b(\cdot) + \phi$ with ϕ a real valued random effect. Notice that the N trajectories $(X_j(t), 0 \leq t \leq T)_{1 \leq j \leq N}$ are *i.i.d.*.

We assume that the $\phi_j = (\phi_{j,1}, \phi_{j,2})^t$'s have a common bivariate density f . Our goal is twofold: first estimate the random effects ϕ_j 's and then their density f , from the observations $(X_j(t), 0 \leq t \leq T)_{1 \leq j \leq N}$, with large T and N . The estimation of the random variables ϕ_j 's follows the steps of Genon-Catalot & Larédo (2016) where only one multiplicative random effect in the drift is considered. We build an estimator $\hat{A}_j(T)$ of ϕ_j based on the trajectory $(X_j(t), 0 \leq t \leq T)$ and study its L^2 -risk. This leads to a bound of order $1/T$. Then we propose a kernel estimator of the density f which uses the sample $(\hat{A}_j(T))_j$. When f is in a Nikol'ski space a bound of the mean integrated squared error is established and the rate of convergence is evaluated. Finally a data-driven choice of the bandwidth based on a Goldenshluger and Lepski's criterion

for anisotropic multi-index is proposed (see *e.g.* Kerkyacharian *et al.*, 2007; Goldenshluger & Lepski, 2011) and leads to an adaptive estimator.

Section 4.2 is dedicated to assumptions and definitions of some useful quantities for the estimation of ϕ_j . In Section 4.3 the estimator of the random effects is built and its L^2 -risk is bounded. In particular we deal with two main examples: the Ornstein-Uhlenbeck model and the Cox-Ingersoll-Ross model. In Section 4.4 the estimator of the density f is studied. In Section 4.5, we explain how the estimation procedure can be applied when only discrete data are available. Finally, Section 4.6 is devoted to numerical simulations to illustrate estimators. Section 4.7 gives some concluding remarks. Proofs are relegated in Section 4.8.

4.2 Notation and assumptions

4.2.1 General assumptions on the model

Consider real valued processes $(X_j(t))_{j=1,\dots,N}$ given by (4.1.1). We assume that $(W_j)_{j=1,\dots,N}$ and $(\phi_j, \gamma_j)_{j=1,\dots,N}$ are defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Consider the following assumptions (see *e.g.* Kutoyants, 2004; Kessler *et al.*, 2012, for details).

(A1) The real valued functions $x \mapsto b_1(x)$, $x \mapsto b_2(x)$ and $x \mapsto \sigma(x)$ are continuous on \mathbb{R} and b_1, b_2 satisfy a Lipschitz condition and σ satisfies a Hölder condition with exponent belonging to $[1/2, 1]$.

(A2) There exists an open set Φ of \mathbb{R}^2 and an interval $(l, r) \subset \mathbb{R}$ such that $\sigma^2(x) > 0$ for $x \in (l, r)$, and for all $\varphi \in \Phi$ the function $s_\varphi : x \mapsto \exp(-2 \int_{x_0}^x \frac{b^t(u)\varphi}{\sigma^2(u)} du)$ with $x_0 \in (l, r)$, satisfies

$$\int_l s_\varphi(y) dy = +\infty, \quad \int^r s_\varphi(y) dy = +\infty.$$

The function $m_\varphi : x \mapsto 1/(\sigma^2(x)s_\varphi(x))$ satisfies $M(\varphi) := \int_l^r m_\varphi(x) dx < +\infty$. We set

$$\pi_\varphi(x) := \mathbf{1}_{(l,r)}(x) \frac{m_\varphi(x)}{M(\varphi)}. \quad (4.2.1)$$

(A3) The random variable ϕ_j takes values in Φ , has distribution $f(\varphi)d\varphi$ and (ϕ_j, γ_j) has distribution on $\Phi \times (l, r)$

$$\pi(d\varphi, dx) = f(\varphi)d\varphi \otimes \pi_\varphi(x)dx.$$

Assumption **(A1)** ensures the existence and uniqueness of a strong solution $X_j(\cdot)$ to (4.1.1) for all random variables $(\phi_j, \gamma_j) \in \mathbb{R}^3$, adapted to the filtration $(\mathcal{F}_t = \sigma((\phi_j, \gamma_j)_{j=1,\dots,N}, W_j(s), s \leq t, j = 1 \dots, N), t \geq 0)$. Then (ϕ_j, γ_j) is \mathcal{F}_0 -measurable and W_j is a $(\mathcal{F}_t, t \geq 0)$ Brownian motion.

In the following, for $\varphi = (\varphi_1, \varphi_2)^t$ a fixed value in Φ we denote by X^φ the process solution of the stochastic differential equation with fixed φ :

$$\begin{cases} dX_j^\varphi(t) &= b^t(X_j^\varphi(t))\varphi dt + \sigma(X_j^\varphi(t))dW_j(t) \\ X_j^\varphi(0) &\sim \pi_\varphi, \quad X_j^\varphi(0) \text{ independent of } W_j. \end{cases} \quad (4.2.2)$$

Under **(A1)**-**(A2)**, for $\varphi \in \Phi$, the process defined by (4.2.2) is strictly stationary and ergodic with scale density s_φ , speed density m_φ and marginal distribution $\pi_\varphi(x)dx$. Under **(A1)**-**(A3)**, according to Genon-Catalot & Larédo (2016), the conditional distribution of X_j given $\phi_j = \varphi$ is identical to the distribution of (4.2.2), the process $((\phi_j, X_j(t)), t \geq 0)$ is strictly stationary

with marginal distribution π and $\mathbb{P}(X_j(t) \in (l, r), \forall t > 0) = 1$. Finally under **(A1)**-**(A3)** as for $j = 1, \dots, N$, for all $t \geq 0$, $X_j(t) \in (l, r)$, $\forall T > 0$:

$$\int_0^T \frac{b_1^2}{\sigma^2}(X_j(s))ds < +\infty, \quad \int_0^T \frac{b_2^2}{\sigma^2}(X_j(s))ds < +\infty, \quad a.s.$$

4.2.2 Key examples

We investigate two classical examples: the mixed Ornstein-Uhlenbeck model and the mixed Cox-Ingersoll-Ross model.

Example 4.2.1. [1] *The Ornstein Uhlenbeck model (OU) with two random effects is defined as*

$$\begin{cases} dX_j(t) &= (\phi_{j,1} - \phi_{j,2}X_j(t))dt + \sigma dW_j(t) \\ X_j(0) &= \gamma_j \end{cases} \quad (4.2.3)$$

with $\sigma(x) = \sigma > 0$, $(l, r) = \mathbb{R}$, and $b(x) = (1, -x)^t$. Assumption **(A2)** requires $\Phi = \mathbb{R} \times (0, +\infty)$ and leads to $\pi_\varphi = \mathcal{N}(\varphi_1/\varphi_2, \sigma^2/(2\varphi_2))$ as invariant distribution for fixed φ .

Example 4.2.2. [2] *The Cox-Ingersoll-Ross model (CIR) with two random effects is defined as*

$$\begin{cases} dX_j(t) &= (\phi_{j,1} - \phi_{j,2}X_j(t))dt + \sigma\sqrt{X_j(t)}dW_j(t) \\ X_j(0) &= \gamma_j \end{cases} \quad (4.2.4)$$

with $\sigma(x) = \sigma\sqrt{x}$ with $\sigma > 0$, and $b(x) = (1, -x)^t$. Assumption **(A2)** requires $2\varphi_1/\sigma^2 \geq 1$ and $2\varphi_2/\sigma^2 > 0$ (in particular the process is always positive). This leads to consider $\Phi = (\sigma^2/2, +\infty) \times (0, +\infty)$. The invariant distribution, for fixed φ , $\pi_\varphi = \Gamma(2\varphi_1/\sigma^2, \sigma^2/(2\varphi_2))$ is the Gamma distribution with shape parameter $2\varphi_1/\sigma^2$ and scale parameter $\sigma^2/(2\varphi_2)$.

4.2.3 Specific assumptions and notations for estimation

Our estimation is based on the following quantities. As in Delattre *et al.* (2013) we define for $j = 1, \dots, N$,

$$U_j(T) := \int_0^T \frac{b}{\sigma^2}(X_j(s))dX_j(s), \quad (4.2.5)$$

which is a column vector with size 2×1 and the 2×2 symmetric matrix:

$$V_j(T) := \int_0^T \frac{bb^t}{\sigma^2}(X_j(s))ds = \begin{pmatrix} \int_0^T \frac{b_1^2}{\sigma^2}(X_j(s))ds & \int_0^T \frac{b_1b_2}{\sigma^2}(X_j(s))ds \\ \int_0^T \frac{b_1b_2}{\sigma^2}(X_j(s))ds & \int_0^T \frac{b_2^2}{\sigma^2}(X_j(s))ds \end{pmatrix}. \quad (4.2.6)$$

Using (4.1.1), $U_j(T)$ can be decomposed as the sum of two terms:

$$U_j(T) = V_j(T)\phi_j + M_j(T) \text{ with } M_j(T) := \int_0^T \frac{b}{\sigma}(X_j(s))dW_j(s). \quad (4.2.7)$$

Note that

$$V_j(T) = \langle M_j \rangle_T \quad (4.2.8)$$

where $\langle . \rangle$ is the quadratic variation of a continuous local martingale. For all measurable function $h : (l, r) \rightarrow \mathbb{R}$, π_φ -integrable for all $\varphi \in \Phi$, we define the random variable $\pi_{\phi_j}(h) :=$

$\int_l^r h(x)\pi_{\phi_j}(x)dx$.

(A4) For $\varphi \in \Phi$ and $i \in \{1, 2\}$, we assume

$$\pi_{\varphi} \left(\frac{b_i^2}{\sigma^2} \right) = \int_l^r \frac{b_i^2}{\sigma^2}(x)\pi_{\varphi}(x)dx < +\infty.$$

We define the random matrix

$$L_j := \pi_{\phi_j} \left(\frac{bb^t}{\sigma^2} \right).$$

(A5) For $j = 1, \dots, N$, $\mathbb{P}(L_j \text{ invertible}) = 1$.

Under (A1)-(A4), Theorem 3.1. of Genon-Catalot & Larédo (2016) gives:

$$\frac{\langle M_j \rangle_t}{t} = \frac{V_j(t)}{t} \xrightarrow[t \rightarrow +\infty]{} L_j, \text{ a.s.} \tag{4.2.9}$$

One can notice that assumption (A5) implies that $V_j(t)$ is invertible for t large enough.

In the study we denote $\|\cdot\|$ the $L^2(\mathbb{R})$ -norm, $\|\cdot\|_p$ the $L^p(\mathbb{R})$ -norm when $p \neq 2$, $\|\cdot\|_2$ the euclidean norm of \mathbb{R}^2 and $\|\cdot\|_F$ the Frobenius norm of matrices defined for $A \in M_2(\mathbb{R})$ by $\|A\|_F^2 = \sum_{i,j} A_{i,j}^2 = \text{Tr}(A^t A)$. We denote $S_2(\mathbb{R})$ the subset of symmetric matrices of $M_2(\mathbb{R})$.

4.3 Random effect estimation

We define first for each j an estimator of the random variable ϕ_j based on the trajectory $X_j(t)$, $t \in [0, T]$. Then these estimators are used to build an estimator of the density f .

4.3.1 Definition of the estimator of the random effects

Let us define N bidimensional random variables

$$A_j(T) := (A_{j,1}(T), A_{j,2}(T))^t = V_j(T)^{-1}U_j(T), \tag{4.3.1}$$

which corresponds to the maximum likelihood estimator of ϕ_j when $\phi_j = \varphi$ is deterministic (see *e.g.* Lipster & Shiryaev (2001)). Note that

$$A_j(T) = \phi_j + V_j(T)^{-1}M_j(T). \tag{4.3.2}$$

Thus $A_j(T)$ is a consistent estimator of ϕ_j as T tends to infinity according to convergence (4.2.9). Because of the presence of the inverse matrix $V_j(T)^{-1}$ in formulae (4.3.1)-(4.3.2), it is difficult to prove that $A_j(T)$ has finite moments and to compute any of them. To overcome this theoretical difficulty, we consider a truncated estimator of ϕ_j . Let

$$\widehat{A}_j(T) := A_j(T)\mathbf{1}_{B_j(T)}, \quad B_j(T) := \{V_j(T) \succeq \kappa\sqrt{T}I_2\} = \{\min(\lambda_{1,j}(T), \lambda_{2,j}(T)) \geq \kappa\sqrt{T}\} \tag{4.3.3}$$

where $\lambda_{i,j}(T)$, $i = 1, 2$ are the eigenvalues of $V_j(T)$, I_2 is the identity matrix of $M_2(\mathbb{R})$ and κ is a calibration constant in the threshold. The inequality in the definition of $B_j(T)$ has a matrix sense: for two matrices $(A, B) \in S_2(\mathbb{R})$, $A \preceq B$ if and only if $B - A$ is a non negative matrix (see Appendix 4.8.5).

Relations (4.2.6) and (4.2.8) show that $V_j(T)$ is a non negative symmetric matrix, thus its eigenvalues are non negative. We are able to bound the L^2 -risk of the estimator $\widehat{A}_j(T)$ of ϕ_j . This bound is needed to evaluate the mean integrated squared error of the nonparametric estimator of the density f .

4.3.2 Main result

We denote \mathcal{L}_φ the infinitesimal generator of the process (4.2.2), given for $F \in \mathcal{C}^2((l, r))$, by $\mathcal{L}_\varphi F(x) := (\sigma^2(x)/2)F''(x) + (b^t(x)\varphi)F'(x)$. Its domain is included in $L^2_{\pi_\varphi}$ which is the space of functions f such that $\int_l^r f^2 d\pi_\varphi < \infty$ (for details see *e.g.* Genon-Catalot *et al.* (2000)). When F is a matrix, the notation $\mathcal{L}_\varphi F$ indicates that we apply the operator on each coefficient of the matrix. We also define analogously F' as the matrix of derivatives. For all $g = (g_{i,k})_{1 \leq i,k \leq 2} \in M_2(\mathbb{R})$ such that $\pi_\varphi(g_{i,k}^2) < +\infty$ for all $\varphi \in \Phi$, we associate the matrix

$$F_\varphi^g = \begin{pmatrix} F_\varphi^{g_{1,1}} & F_\varphi^{g_{1,2}} \\ F_\varphi^{g_{2,1}} & F_\varphi^{g_{2,2}} \end{pmatrix}$$

satisfying $-\mathcal{L}_\varphi F_\varphi^g = g - \pi_\varphi g$ for all $\varphi \in \Phi$. We denote for simplicity

$$H_{\phi_j} := F_{\phi_j}^{bb^t/\sigma^2}. \quad (4.3.4)$$

Examples below show how H_{ϕ_j} can be constructed. We are now able to announce the main result on the estimator on the random effects.

Proposition 4.3.1. *Consider the processes $(X_j(t), 0 \leq t \leq T)_{1 \leq j \leq N}$ given by (4.1.1) under (A1)-(A5). Assume that for $j = 1, \dots, N$,*

$$\mathbb{E} \left[\frac{\|L_j\|_F^2}{[\det(L_j)]^2} \text{Tr}(L_j) \right] < \infty, \quad \mathbb{E} \left[\frac{\|L_j\|_F^4}{[\det(L_j)]^4} \text{Tr} \left(\pi_{\phi_j} \left(\left(\frac{bb^t}{\sigma^2} \right)^2 \right) \right) \right] < \infty,$$

$$\mathbb{E} \left[\|\phi_j\|_2^2 \left(\frac{1}{L_{j,1,1}^2} + \frac{1}{L_{j,2,2}^2} \right) \pi_{\phi_j}(\|H_{\phi_j}\|_F^2) \right] + \mathbb{E} \left[\|\phi_j\|_2^2 \left(\frac{1}{L_{j,1,1}^2} + \frac{1}{L_{j,2,2}^2} \right) \pi_{\phi_j}(\|H'_{\phi_j}\|_F^2 \sigma^2) \right] < \infty$$

and

$$\mathbb{E} \left[\pi_{\phi_j}(\|H_{\phi_j}\|_F^4) \right] + \mathbb{E} \left[\pi_{\phi_j}(\|H'_{\phi_j}\|_F^4 \sigma^4) \right] < \infty.$$

Then, there exists a constant $C > 0$ such that

$$\mathbb{E} \left[\|\widehat{A}_j(T) - \phi_j\|_2^2 \right] \leq \frac{C}{T}. \quad (4.3.5)$$

The two last assumptions of Proposition 4.3.1 correspond to the application of Proposition 4.8.1 given in Section 4.8.1 with $g = bb^t/\sigma^2$, $p = 1$ and $p = 2$. This proposition is the key of the nonparametric estimation procedure set up in Section 4.4.

4.3.3 Key examples continued

Let us investigate the two examples given in Section 4.2.2.

Example 4.3.2. [1](continued) *In this case*

$$\frac{bb^t}{\sigma^2}(x) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & -x \\ -x & x^2 \end{pmatrix}, \quad V_j(T) = \frac{1}{\sigma^2} \begin{pmatrix} T & -\int_0^T X_j(s) ds \\ -\int_0^T X_j(s) ds & \int_0^T X_j(s)^2 ds \end{pmatrix}.$$

For simplicity we set here $\phi_j = \phi = (\phi_1, \phi_2)^t$. The limit matrix L of $V(T)/T$ is given by

$$L = \pi_\phi \left(\frac{bb^t}{\sigma^2} \right) = \frac{1}{\sigma^2} \begin{pmatrix} 1 & -\frac{\phi_1}{\phi_2} \\ -\frac{\phi_1}{\phi_2} & \frac{\sigma^2}{2\phi_2} + \frac{\phi_1^2}{\phi_2^2} \end{pmatrix}, \quad \det(L) = \frac{\sigma^2}{2\phi_2} > 0.$$

We can check that

$$H_\phi(x) = \begin{pmatrix} 1 & -\frac{1}{\phi_2\sigma^2}\left(x - \frac{\phi_1}{\phi_2}\right) \\ -\frac{1}{\phi_2\sigma^2}\left(x - \frac{\phi_1}{\phi_2}\right) & \frac{2\phi_1}{\sigma^2\phi_2^2}\left(x - \frac{\phi_1}{\phi_2}\right) + \frac{1}{2\sigma^2\phi_2}\left[\left(x - \frac{\phi_1}{\phi_2}\right)^2 - \frac{\sigma^2}{2\phi_2}\right] \end{pmatrix}.$$

Details are in Appendix. The assumptions of Proposition 4.3.1 are fulfilled if $\mathbb{E}[\phi_1^6 + \phi_2^{-10}] < \infty$.

Example 4.3.3. [2](continued) In this case,

$$\frac{bb^t}{\sigma^2}(x) = \frac{1}{\sigma^2} \begin{pmatrix} \frac{1}{x} & -1 \\ -1 & x \end{pmatrix}, \quad V_j(T) = \frac{1}{\sigma^2} \begin{pmatrix} \int_0^T \frac{1}{X_j(s)} ds & -T \\ -T & \int_0^T X_j(s) ds \end{pmatrix}.$$

For $\Phi = (\sigma^2, +\infty) \times (0, +\infty)$, $\pi_\varphi(x \mapsto 1/x^2) < \infty$ for all $\varphi \in \Phi$ and the limit matrix L is given by:

$$L = \pi_\phi \left(\frac{bb^t}{\sigma^2} \right) = \frac{1}{\sigma^2} \begin{pmatrix} \frac{2\phi_2}{2\phi_1 - \sigma^2} & -1 \\ -1 & \frac{\phi_1}{\phi_2} \end{pmatrix}, \quad \det(L) = \frac{\sigma^2}{2\phi_1 - \sigma^2} > 0.$$

One can verify (see Appendix) that

$$H_\phi(x) = \begin{pmatrix} \frac{2}{\sigma^2(2\phi_1 - \sigma^2)} \log(x) & 1 \\ 1 & \frac{1}{\phi_2\sigma^2} \left(x - \frac{\phi_1}{\phi_2}\right) \end{pmatrix}.$$

The details are relegated in Appendix. For the assumptions of Proposition 4.3.1 we must have $\pi_\phi(x \mapsto |\log(x)|)$, $\pi_\phi(x \mapsto 1/x^2)$ and $\pi_\phi(x \mapsto 1/x^4)$ finite. The last condition imposes a reduction of Φ to $\Phi = (2\sigma^2, +\infty) \times (0, +\infty)$. Moreover the other assumptions are fulfilled if $\mathbb{E}[\log^2(\phi_2) + \phi_1^5 + \phi_2^8 + \phi_2^{-4} + \psi^2(2\phi_1/\sigma^2 - 1)] < \infty$, where $\psi(x) := \Gamma'(x)/\Gamma(x)$ is the digamma function.

4.4 Nonparametric estimation

In this section we assume that $f \in \mathbb{L}^2(\mathbb{R}^2)$ and set up a nonparametric estimation procedure based on a kernel estimator. We study the obtained estimator and determinate its rate of convergence.

4.4.1 Nonparametric estimator of the density of the random effects

By Proposition 4.3.1, $\hat{A}_j(T)$ is a consistent estimator of ϕ_j when T tends to infinity. It is therefore natural to define a kernel estimator based on the $\hat{A}_j(T)$'s. Let us denote by K a kernel in $\mathcal{C}^1(\mathbb{R}^2)$ such that the partial derivatives $\frac{\partial K}{\partial u}$ and $\frac{\partial K}{\partial v}$ are in $L^2(\mathbb{R}^2)$, K is integrable, $\iint K(u, v) du dv = 1$ and $\|K\|^2 = \iint K^2(u, v) du dv < +\infty$. For all $h = (h_1, h_2)$, $h_1 > 0$, $h_2 > 0$, for all $(u, v) \in \mathbb{R}^2$, we denote

$$K_h(u, v) = \frac{1}{h_1 h_2} K\left(\frac{u}{h_1}, \frac{v}{h_2}\right).$$

For example one can consider the Gaussian kernel $K(u, v) = K_1(u) \times K_1(v)$, with $K_1(u) = (1/\sqrt{2\pi}) \exp(-u^2/2)$. We define the estimator of the density f for $x = (x_1, x_2) \in \mathbb{R}^2$ by

$$\hat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - \hat{A}_j(T)) \tag{4.4.1}$$

for a large number of observations N . Denoting $f_h(x) := K_h \star f(x) = \iint f(y_1, y_2) K_h(x_1 - y_1, x_2 - y_2) dy_1 dy_2$, the following result holds.

Proposition 4.4.1. *Under (A1)-(A5) and under the assumptions of Proposition 4.3.1,*

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + 2 \max\left(\frac{1}{h_1^3 h_2}, \frac{1}{h_1 h_2^3}\right) \left(\left\| \frac{\partial K}{\partial u} \right\|^2 + \left\| \frac{\partial K}{\partial v} \right\|^2 \right) \frac{C}{T} + \frac{\|K\|^2}{N h_1 h_2} \quad (4.4.2)$$

with C the constant which appears in Proposition 4.3.1.

This bound comes from the bias-variance decomposition:

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + 2\|\mathbb{E}[\widehat{f}_h] - f_h\|^2 + \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2].$$

The first term is a bias term due to the approximation of f by f_h , it decreases when h_1, h_2 decrease. The third term is a variance term which increases when h_1, h_2 decrease. Finally, the middle term is an error term due to the approximation of the ϕ_j 's by the $\widehat{A}_j(T)$'s also increasing when h_1, h_2 decrease. Note that the orders are consistent with the result of Comte *et al.* (2013), Proposition 1 for the multiplicative model. In fact in the case of a single random effect, multiplicative in the drift, the second term has the order $1/(Th^3)$ and the third term: $1/(Nh)$.

To choose the best h , a compromise must be done between the bias term and the variance and middle terms.

4.4.2 Rates of convergence

We consider anisotropic Nikol'ski classes of functions which are well fitted to evaluate the order of the bias term (see Goldenshluger & Lepski (2011) and Comte & Lacour (2013) for example).

Definition 4.4.2. *A function f is in the Nikol'ski class $\mathcal{N}(\beta, R)$ if $f : \mathbb{R}^2 \rightarrow \mathbb{R}$ admits partial derivatives of order $\lfloor \beta_i \rfloor$, $i \in \{1, 2\}$ such that, with $y_1 = (y, 0), y_2 = (0, y)$, for all $y \in \mathbb{R}$, for $i \in \{1, 2\}$*

$$\left(\int \left| \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}}((x_1, x_2) + y_i) - \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}}(x_1, x_2) \right|^2 dx \right)^{1/2} \leq R |y|^{\beta_i - \lfloor \beta_i \rfloor},$$

$$\left\| \frac{\partial^{\lfloor \beta_i \rfloor} f}{(\partial x_i)^{\lfloor \beta_i \rfloor}} \right\| \leq R$$

(with $\lfloor \beta \rfloor$ denotes the largest integer strictly less than β).

Recall that kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ is of order $l \in \mathbb{N}^*$ if for $j = 1, \dots, l$, $\int |x|^j |K(x)| dx < +\infty$ and $\int x^j K(x) dx = 0$ (see Tsybakov (2009)). In this context we are now able to bound the bias term $\|f - f_h\|^2$.

Proposition 4.4.3. *If $f \in \mathcal{N}(\beta, R)$ and $K(x) = K(x_1, x_2) = K_1(x_1)K_2(x_2)$ where K_1 is a kernel of order $l_1 := \lfloor \beta_1 \rfloor$, K_2 a kernel of order $l_2 := \lfloor \beta_2 \rfloor$, $\int |K_1(u)u^{\beta_1}| du < \infty$ and $\int |K_2(u)u^{\beta_2}| du < \infty$, then there exist two constants $c_1, c_2 > 0$ such that*

$$\|f - f_h\|^2 \leq c_1 h_1^{2\beta_1} + c_2 h_2^{2\beta_2}. \quad (4.4.3)$$

In order to minimize the MISE, if we only consider the first and the last term of the right hand side of (4.4.2), we have to minimize the following function of two variables:

$$(h_1, h_2) \mapsto \xi(h_1, h_2) = \frac{1}{N h_1 h_2} + h_1^{2\beta_1} + h_2^{2\beta_2}.$$

With $1/\bar{\beta} = 1/(2\beta_1) + 1/(2\beta_2)$, we get the minimizing values

$$h_1^* = O\left(N^{-\frac{1}{2\beta_1} \frac{\bar{\beta}}{\beta+1}}\right), \quad h_2^* = O\left(N^{-\frac{1}{2\beta_2} \frac{\bar{\beta}}{\beta+1}}\right) \quad (4.4.4)$$

and this implies

$$\xi(h_1^*, h_2^*) = O\left(N^{-\frac{4\beta_1\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}\right) = O\left(N^{-\frac{2\bar{\beta}}{2+2\bar{\beta}}}\right).$$

Therefore, we require (see (4.4.2))

$$\frac{1}{T} \max\left(\frac{1}{h_1^* h_2^*}, \frac{1}{h_1^* h_2^{*3}}\right) \leq \frac{1}{Th_1^* h_2^*} \left(\frac{1}{h_1^{*2}} + \frac{1}{h_2^{*2}}\right) \leq N^{-\frac{4\beta_1\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}.$$

This holds under the following condition

$$T \geq N^{1+\frac{4\beta_1}{(2\beta_1+1)(2\beta_2+1)-1}} + N^{1+\frac{4\beta_2}{(2\beta_1+1)(2\beta_2+1)-1}}. \quad (4.4.5)$$

This condition is implied by $T \geq N^2$, and if $\beta_i > 1$ by $T \geq N^{3/2}$. We are now able to state the order of the risk, for large N and T .

Corollary 4.4.4. *Under (A1)-(A5), under the assumptions of Proposition 4.4.3, under conditions (4.4.4) and (4.4.5), we have*

$$\mathbb{E}[\|\hat{f}_{h^*} - f\|^2] = O\left(N^{-\frac{2\bar{\beta}}{2\bar{\beta}+2}}\right) \text{ with } 1/\bar{\beta} = 1/(2\beta_1) + 1/(2\beta_2).$$

Notice that if $\beta_1 = \beta_2 = \beta$, thus $\bar{\beta} = \beta$ and $N^{-2\bar{\beta}/(2\bar{\beta}+2)} = N^{-2\beta/(2\beta+2)}$. The classical rate of convergence for a kernel density estimator, with one bandwidth, is $N^{-2\beta/(2\beta+1)}$. We observe here a 2 on the denominator which comes from the number of bandwidths.

In practice the regularity parameters β and R are unknown. Thus we propose a data-driven way of choosing h based on the Goldenshluger & Lepski (2011) method.

4.4.3 Data-driven bandwidth and adaptation

The idea is to replace the unknown term $\|f_h - f\|^2$ by an estimator. For this we introduce the iterated auxiliary kernel estimators

$$\hat{f}_{h,h'}(x) := K_{h'} \star \hat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_{h'} \star K_h(x - \hat{A}_j(T)).$$

Let $\mathcal{H}_{N,T}$ be a finite set of bandwidths $h = (h_1, h_2)$. Following Goldenshluger and Lepski's approach, we define the estimator of the bias term by

$$B(h) = \sup_{h' \in \mathcal{H}_{N,T}} \left(\|\hat{f}_{h,h'} - \hat{f}_{h'}\|^2 - v(h') \right)_+ \quad (4.4.6)$$

where

$$v(h) := \kappa_1 \frac{\|K\|^2 \|K\|_1^2}{Nh_1 h_2} + \kappa_2 \|K\|_1^2 \left(\frac{1}{h_1^3 h_2} + \frac{1}{h_1 h_2^3} \right) \left(\left\| \frac{\partial K}{\partial u} \right\|^2 + \left\| \frac{\partial K}{\partial v} \right\|^2 \right) \frac{C}{T}, \quad (4.4.7)$$

is a term which has the same order as the variance terms in (4.4.2), C is the constant from Proposition 4.3.1 and κ_1, κ_2 are numerical constants. The bandwidth is finally selected as follows

$$\widehat{h} = \operatorname{argmin}_{h \in \mathcal{H}_{N,T}} (B(h) + v(h)). \quad (4.4.8)$$

Then we obtain the following Theorem.

Theorem 4.4.5. *We assume that the elements of $\mathcal{H}_{N,T}$ satisfy $0 < h_i < 1, i = 1, 2, 1/(Nh_1h_2) \leq 1, 1/(h_1^3h_2) \leq T, 1/(h_1h_2^3) \leq T$, and that*

$$\forall c > 0, \exists \Sigma(c) < \infty, \quad \sum_{h \in \mathcal{H}_{N,T}} (h_1h_2)^{-1/2} e^{-c/\sqrt{h_1h_2}} \leq \Sigma(c).$$

Under (A1)-(A5), under the assumptions of Proposition 4.3.1 and if $\|K\|_{4/3} < +\infty$, there exist numerical constants κ_1, κ_2 such that

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq C_1 \inf_{h \in \mathcal{H}_{N,T}} \{\|f - f_h\|^2 + v(h)\} + \frac{C_2}{N} \quad (4.4.9)$$

with C_1 a positive constant depending on $\|K\|_1$, C_2 is a positive constant depending on $\|f\|, \|\frac{\partial K}{\partial u}\|^2 + \|\frac{\partial K}{\partial v}\|^2, \|K\|_1, \|K\|_{4/3}$ and C (the constant from Proposition 4.3.1).

This is a non-asymptotic result proving that the bias variance compromise is automatically realized by the final estimator \widehat{f}_h which is therefore adaptive. Theorem 4.4.5 is proved in Chapter 3 for a single random effect. As the proof is analogous we do not repeat it here. The key is to prove that $\mathbb{E}[B(h)] \leq \|f - f_h\|^2 + c/N$. The theoretical study gives $\kappa_1 \geq \max(40/\|K\|_1^2, 40)$ and $\kappa_2 \geq \max(10/\|K\|_1^2, 10)$. Nevertheless, the constants κ_1, κ_2 obtained in the proof are not optimal and thus it is standard to calibrate them from a simulation study (see Lacour & Massart (2016) for details).

The conditions of Theorem 4.4.5 are fulfilled for example with the set $\mathcal{H}_{N,T} = \{(1/k_1^2, 1/k_2^4), k_1 = 1 \dots, N^{1/4}, k_2 = 1 \dots, N^{1/8}, N^2 \leq T\}$.

In practice, note that in $v(h)$, the constant C is unknown, and must be replaced by a rough estimator. From the proofs of Propositions 4.3.1 and 4.8.1 it appears that $C = \mathbb{E}[S(\phi_j)]$ for an explicit function $S(\phi_j)$, thus C can be estimated by $(1/N) \sum_{j=1}^N S(\widehat{A}_j(T))$.

4.5 Discrete data

In practice, only discrete observations are available. The estimation method must take into account this fact. Let us assume that we observe synchronously $(X_j(t)), j = 1, \dots, N$ at times $t_k = k\Delta, k = 1, \dots, n$ with n tending to infinity, $\Delta = \Delta_n$ tending to zero and $T = n\Delta$ tending to infinity. One can refer to Dalalyan & Reiß (2006) for the proof of statistical equivalence between the discrete-time and continuous-time diffusion models. As it is classically done, we replace the stochastic and ordinary integrals by their discretized versions. The estimator of ϕ_j is thus defined by

$$\widehat{A}_{j,n} := A_{j,n} \mathbf{1}_{B_{j,n}}, \quad \text{with} \quad A_{j,n} = V_{j,n}^{-1} U_{j,n}, \quad (4.5.1)$$

and

$$U_{j,n} := \sum_{k=0}^{n-1} \frac{b}{\sigma^2} (X_j(k\Delta))(X_j((k+1)\Delta) - X_j(k\Delta)), \quad V_{j,n} := \sum_{k=0}^{n-1} \Delta \frac{bb^t}{\sigma^2} (X_j(k\Delta)),$$

$$B_{j,n} := \{V_{j,n} \succeq \kappa \sqrt{n\Delta} I_2\}.$$

To extend the present work to discrete data, it is enough to find conditions ensuring that $\mathbb{E}[\|\widehat{A}_{j,n} - \phi_j\|_2^2] \leq \frac{C}{n\Delta}$. The proof follows the steps of Proposition 1, with the discretized integrals instead of the continuous ones. We must study

$$\mathbb{E} \left[\left\| \frac{1}{n\Delta} (U_{j,n} - U_j(T)) \right\|_2^{2p} \right], \quad \mathbb{E} \left[\left\| \frac{1}{n\Delta} (V_{j,n} - V_j(T)) \right\|_F^{2p} \right]$$

for $p = 1, 2$ and extend Proposition 7 (see Section 7) to the discretized version of $(1/T) \int_0^T g(X_j(s)) ds$. This is relatively standard as we assume that the processes $(\phi_j, X_j(t))_{t \geq 0}$ are in stationary regime. The discretizations induce a new bias term which of order $O(\Delta)$. To make this bias negligible, a classical condition on the sampling interval is : $\Delta \lesssim 1/(n\Delta)$, *i.e.* $n\Delta^2 \lesssim 1$ (see Comte *et al.* (2007); Kessler *et al.* (2012) for example). Note that other approximations of stochastic or ordinary integrals based on discrete data are available and can improve the constraint $n\Delta^2 \lesssim 1$ (see *e.g.* Kloeden & Platen (1992) Iacus (2008)).

Moreover, due to the presence of random effects, additional moment conditions on the distribution of the initial condition (ϕ_j, γ_j) are required. To avoid technical developments, we do not give more details.

4.6 Simulation study

In this section, we illustrate on simulated data the performances of our nonparametric estimation procedure on the two examples of Sections 4.2.2 and 4.3.3. For simplicity we denote $\widehat{f} := \widehat{f}_{\widehat{h}}$ and

$$\widehat{f}_1(x_1) = \int \widehat{f}(x_1, x_2) dx_2, \quad \widehat{f}_2(x_2) = \int \widehat{f}(x_1, x_2) dx_1. \quad (4.6.1)$$

For the two examples, we simulate N independent discretized sample paths $(X_j(k\Delta), k = 1, \dots, n, j = 1, \dots, N)$. This is done by simulating independently for each j first the random variable ϕ_j , then, given $\phi_j = \varphi$, the initial variable γ_j which is simulated according to the invariant distribution π_φ corresponding to each model. Lastly, the discretized sample path is simulated. For the OU model, we use exact simulations of the discretized path. For the CIR model, we use a discretization scheme given in Alfonsi (2005).

For simplicity we simulate independent components for $\phi_j = (\phi_{j,1}, \phi_{j,2})$, thus the density f has the form $f(x_1, x_2) = f_1(x_1) \times f_2(x_2)$.

The computation of the estimators $\widehat{A}_j(T)$ given by (4.3.3) requires to choose a value for the cut-off parameter κ . To choose κ , different functions f have been investigated with different values of Δ and a large number of repetitions. We compared the MISEs of estimators as functions of the constant κ and selected the value $\kappa = 0.125$ as satisfactory. For comparison, we also computed the estimators for $\kappa = 0$ (no cut-off). For $\kappa = 0$ the estimator is denoted \widetilde{f} and its marginals \widetilde{f}_1 and \widetilde{f}_2 .

There remains the choice of the adaptive bandwidth. In Chapter 3 the kernel estimator with data-driven bandwidth selected by the Goldenshluger and Lepski criterion is implemented for one random effect. It is also implemented using the R-function *density* choosing the bandwidth by cross-validation or with the rule of thumb. The Goldenshluger and Lepski method performs better for mixture densities. Both methods are comparable for unimodal densities. This is why here we have chosen to use the R-function *kde2d* for the choice of the bandwidths with a standard bivariate Gaussian kernel.

Example 4.6.1. [1] The simulation scheme of each sample path given φ is as follows: $X^\varphi(0) = \gamma \sim \mathcal{N}\left(\frac{\varphi_1}{\varphi_2}, \frac{\sigma^2}{2\varphi_2}\right)$ and G_1, \dots, G_n i.i.d. $\mathcal{N}(0, 1)$ independent of $X^\varphi(0)$,

$$X^\varphi((k+1)\Delta) = X^\varphi(k\Delta)e^{-\Delta\varphi_2} + \frac{\varphi_1}{\varphi_2}(1 - e^{-\Delta\varphi_2}) + \sqrt{\frac{\sigma^2}{2\varphi_2}(1 - \exp(-2\Delta\varphi_2))}G_{k+1}. \quad (4.6.2)$$

The following distributions are chosen for the components of the random effects. We drop the index j :

- $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$,
- $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim 1/\Gamma(3, 0.25)$
- $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim \Gamma(2, 1)$
- $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim 1/\Gamma(3, 0.25)$

Note that, for the third case, the distribution of ϕ does not satisfy all assumptions as $\mathbb{E}[\phi_2^{-10}]$ does not exist. However the simulations show that the estimation procedure works even when theoretical assumptions do not hold.

Example 4.6.2. [2] The simulation scheme of each sample path given φ is the one given in Alfonsi (2005) which is based on the following approximation for small δ :

$$X_{(k+1)\delta}^\varphi \cong X_{k\delta}^\varphi + (\varphi_1 - \sigma^2/2 - \varphi_2 X_{(k+1)\delta}^\varphi)\delta + \sigma\sqrt{X_{(k+1)\delta}^\varphi}(W_{(k+1)\delta} - W_{k\delta}).$$

Solving for $X^\varphi((k+1)\delta)$ yields the explicit relation:

$$X^\varphi((k+1)\delta) = \left(\frac{\sqrt{\delta}\sigma G_{k+1} + \sqrt{\sigma^2\delta G_{k+1}^2 + 4(\delta(\varphi_1 - \sigma^2/2) + X^\varphi(k\delta))(1 + \varphi_2\delta)}}{2 + 2\varphi_2\delta} \right)^2 \quad (4.6.3)$$

where $G_{k+1} = (W_{(k+1)\delta} - W_{k\delta})/\sqrt{\delta}$ and $X^\varphi(0) = \gamma \sim \Gamma(2\frac{\varphi_1}{\sigma^2}, \frac{\sigma^2}{2\varphi_2})$. The values $\delta = T/20000$, $\Delta = 10 \times \delta$ and $n = T/\Delta = 2000$ are chosen.

The two random effects must be positive random variables and the first component must satisfy $\phi_1 > 2\sigma^2$. We have chosen to simulate:

- $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1 + \Gamma(1.8, 0.8)$
- $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1/\Gamma(8.1, 0.05)$.

The chosen distribution for ϕ_1 satisfies the assumption $\mathbb{E}[\psi^2(2\phi_1/\sigma^2 - 1)] < \infty$. Indeed, when $\phi_1 \sim 2\sigma^2 + \Gamma(k, \theta)$, $\mathbb{E}[\psi^2(2\phi_1/\sigma^2 - 1)] = \int_{2\sigma^2}^{+\infty} \psi^2(2x/\sigma^2 - 1)e^{-(x-2\sigma^2)/\theta}(x-2\sigma^2)^{k-1}dx$. One can use the asymptotic equivalent for the di-gamma function $\psi(x) \underset{+\infty}{\sim} \log(x)$ to see that the integral is convergent. Here, the first case does not satisfy $\mathbb{E}[\phi_2^{-4}] < \infty$. But the estimation results are satisfactory.

Figures 4.1 and 4.4 illustrate the influence of the truncation threshold on the estimators for the two examples. On top, the estimators $A_j(T)$ ($\kappa = 0$, no truncation) are plotted as function of the simulated variables ϕ_j . Top left is for the first component, top right for the second. The bottom figures represent the estimators $\hat{A}_j(T)$ ($\kappa = 0.125$) as function of the simulated ϕ_j . The random variables ϕ_j are well estimated and the cut-off seems to have the adequate effect: it sets to zero the values which are too far from the diagonal.

On Figures 4.2 (OU example) and 4.5 (CIR example), estimators of the two marginal densities, *i.e.* the densities of the two components of the random effects, are shown. The true marginals are in bold black. Three density estimators \tilde{f}_1 (left) and \tilde{f}_2 (right), built using the $A_j(T)$'s ($\kappa = 0$), are in dotted (blue) lines. Three density estimators \hat{f}_1 (left), \hat{f}_2 (right), built

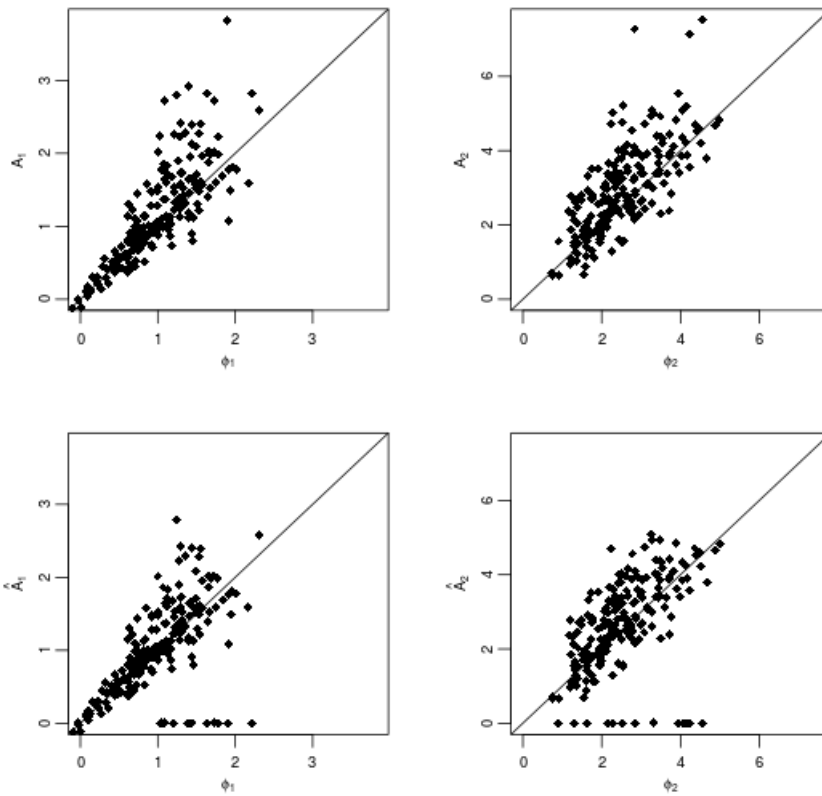


Figure 4.1 – OU example. Top left: $A_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right: $A_{j,2}(T)$'s as a function of $\phi_{j,2}$. Bottom left: $\hat{A}_{j,1}(T)$'s as a function of $\phi_{j,1}$, bottom right: $\hat{A}_{j,2}(T)$'s as a function of $\phi_{j,2}$. With $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$, $T = 10$, $N = 200$, $\sigma = 0.1$

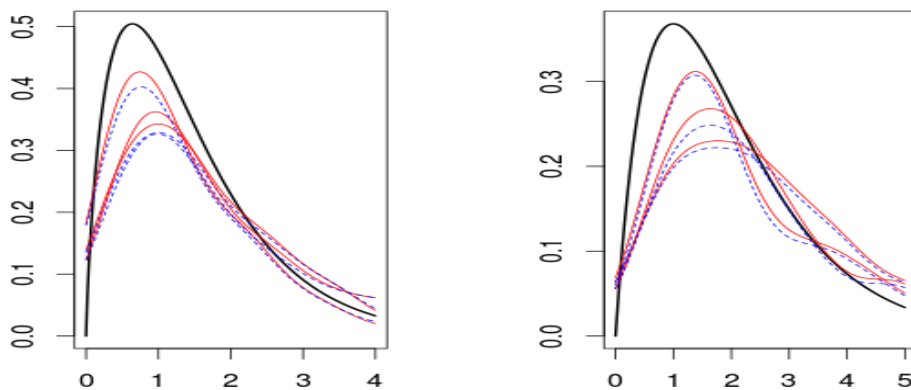


Figure 4.2 – OU example. On the left: dark bold line f_1 , 3 estimators $\tilde{f}_1(x_1)$ in dotted blue (grey) and $\hat{f}_1(x_1)$ in red (grey), on the right: dark bold line f_2 , 3 estimators $\tilde{f}_2(x_2)$ in dotted blue (grey) and $\hat{f}_2(x_2)$ in red (grey). Simulation with $\phi_1 \sim \Gamma(1.8, 0.8)$, $\phi_2 \sim \Gamma(2, 1)$, $T = 10$, $N = 200$, $\sigma = 0.1$.

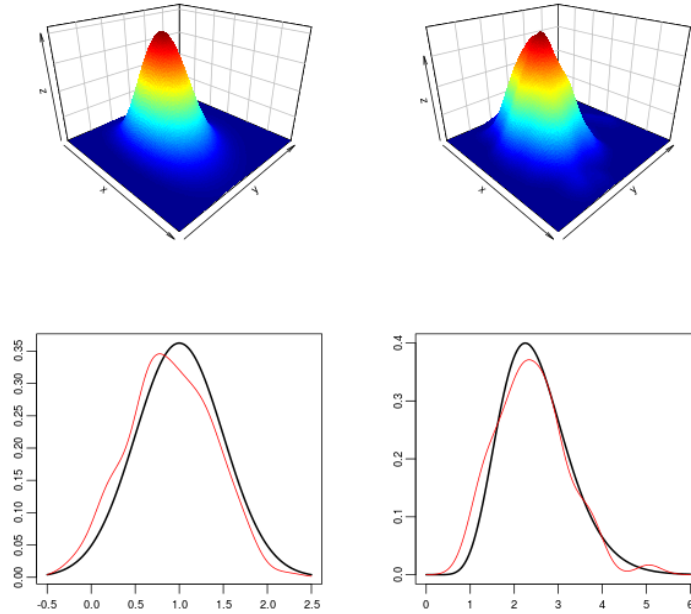


Figure 4.3 – OU example. Top left: true density f , top right: estimator \hat{f} . Bottom: sections. Bottom left: dark line for $x_1 \rightarrow f(x_1, x_2)$ versus the estimator $x_1 \rightarrow \hat{f}(x_1, x_2)$ light line (red) for x_2 fixed. Bottom right: dark line for $x_2 \rightarrow f(x_1, x_2)$ versus the estimator $x_2 \rightarrow \hat{f}(x_1, x_2)$ red line (grey), for x_1 fixed. Simulation with $\phi_1 \sim \mathcal{N}(1, 0.5)$, $\phi_2 \sim \Gamma(10.1, 0.25)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

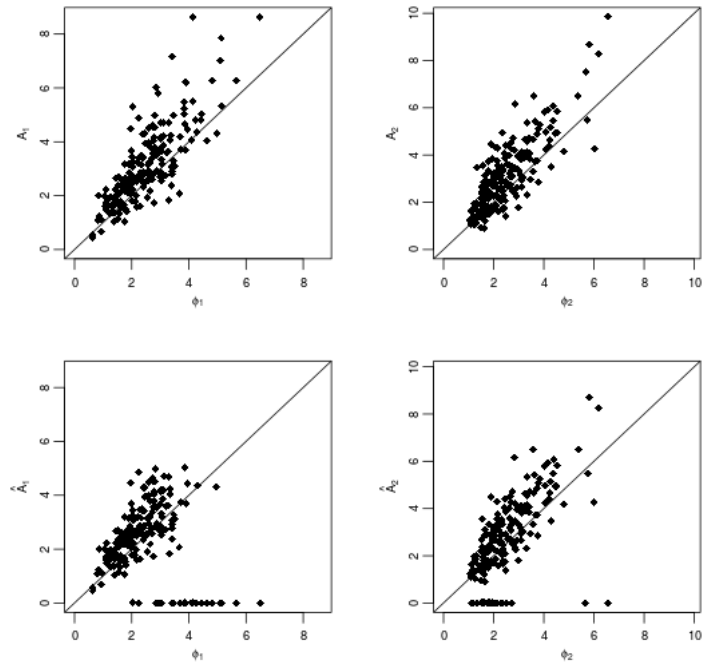


Figure 4.4 – CIR example. Top left: $A_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right $A_{j,2}(T)$'s as a function of $\phi_{j,2}$. Bottom left: $\hat{A}_{j,1}(T)$'s as a function of $\phi_{j,1}$, top right: $\hat{A}_{j,2}(T)$'s as a function of $\phi_{j,2}$. With $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1 + \Gamma(1.8, 0.8)$, $T = 10$, $N = 200$, $\sigma = 0.1$.

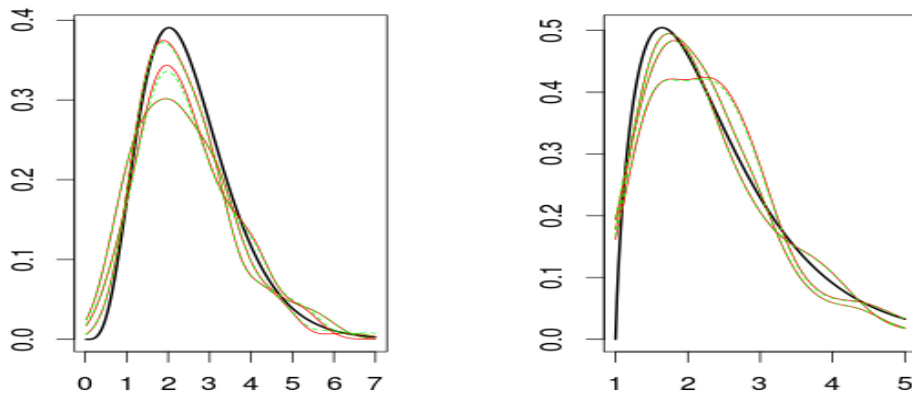


Figure 4.5 – CIR example. On the left: dark bold line f_1 , 3 estimators $\tilde{f}_1(x_1)$ in dotted grey (green) and $\hat{f}_1(x_1)$ in red (grey), on the right: dark bold line f_2 , 3 estimators $\tilde{f}_2(x_2)$ in dotted green (grey) and $\hat{f}_2(x_2)$ in red (grey). Simulation with $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim \Gamma(1.8, 0.8)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

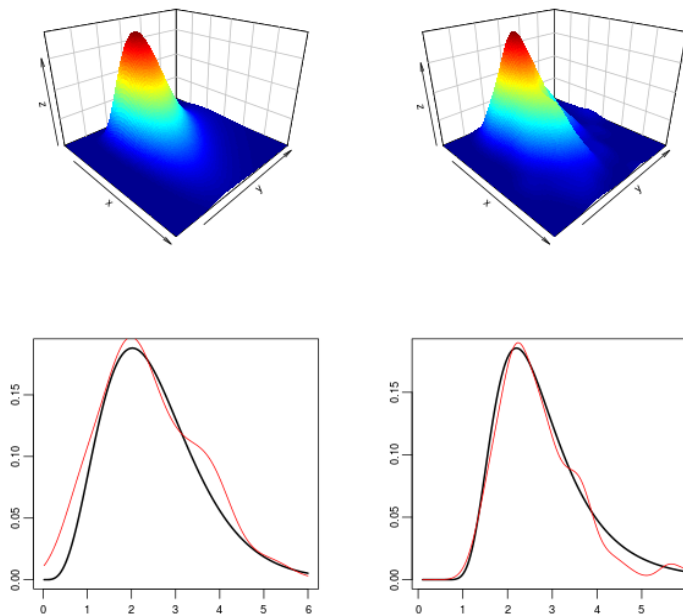


Figure 4.6 – CIR example. Top left: true density f , top right: estimator \hat{f} . Bottom: sections. Bottom left: dark line for $x_1 \rightarrow f(x_1, x_2)$ versus the estimator $x_1 \rightarrow \hat{f}(x_1, x_2)$ light line (red) for x_2 fixed. Bottom right: dark line for $x_2 \rightarrow f(x_1, x_2)$ versus the estimator $x_2 \rightarrow \hat{f}(x_1, x_2)$ red line (grey), for x_1 fixed. Simulation with $\phi_1 \sim 2\sigma^2 + \Gamma(5, 0.5)$, $\phi_2 \sim 1/\Gamma(8.1, 0.05)$, $T = 100$, $N = 200$, $\sigma = 0.1$.

Table 4.1 – OU example. MISE computed from 100 simulations

Case	σ	T	N	\tilde{f}	\hat{f}	\hat{f}_1	\hat{f}_2
$\phi_1 \sim \mathcal{N}, \phi_2 \sim \Gamma$	0.1	10	20	0.064	0.064	0.067	0.073
			200	0.039	0.038	0.038	0.047
		100	20	0.046	0.046	0.047	0.030
			200	0.010	0.010	0.009	0.006
	1	100	20	0.046	0.046	0.044	0.033
			200	0.012	0.012	0.011	0.007
$\phi_1 \sim \mathcal{N}, \phi_2 \sim 1/\Gamma$	0.1	10	20	0.077	0.075	0.113	0.081
			200	0.050	0.051	0.008	0.060
		100	20	0.068	0.068	0.083	0.040
			200	0.021	0.027	0.035	0.010
	1	100	20	0.061	0.061	0.061	0.044
			200	0.021	0.021	0.035	0.011
$\phi_1 \sim \Gamma, \phi_2 \sim \Gamma$	0.1	10	20	0.028	0.024	0.056	0.047
			200	0.015	0.013	0.030	0.028
		100	20	0.025	0.025	0.029	0.027
			200	0.009	0.009	0.007	0.011
	1	100	20	0.022	0.022	0.025	0.025
			200	0.008	0.008	0.007	0.010
$\phi_1 \sim \Gamma, \phi_2 \sim 1/\Gamma$	0.1	10	20	0.045	0.042	0.061	0.074
			200	0.031	0.030	0.030	0.076
		100	20	0.030	0.030	0.028	0.035
			200	0.008	0.008	0.007	0.009
	1	100	20	0.029	0.029	0.027	0.034
			200	0.008	0.008	0.007	0.008

using the $\hat{A}_j(T)$'s ($\kappa = 0.125$), are in continuous (red) lines. The cut-off for estimating the ϕ_j 's improves the density estimators: especially, the first marginal \hat{f}_1 fits better the true density than \tilde{f}_1 .

Figures 4.3 and 4.6 show the true bivariate density (top left) and the estimator \hat{f} (top right). The bottom graphs correspond to sections $x_1 \rightarrow f(x_1, x_2)$ (bold left) and $x_1 \rightarrow \hat{f}(x_1, x_2)$ (dotted left) for a fixed x_2 , and $x_2 \rightarrow f(x_1, x_2)$ (bold right) and $x_2 \rightarrow \hat{f}(x_1, x_2)$ (dotted right) for a fixed x_1 . The estimators are close to the true density.

In order to evaluate with more accuracy the performances of our estimator on the two examples, we have computed their empirical MISE from 100 simulated data sets, for different values of σ and T and for the chosen distributions of (ϕ_1, ϕ_2) . Results are presented in Tables 4.1 and 4.2.

On Tables 4.1 and 4.2 the influence of the cut-off is often visible when $N = 20$: the MISEs of \tilde{f} are larger than the MISEs of \hat{f} . It almost disappears on the Tables when $N = 200$. For a given value of N , when T increases from 10 to 100 the MISE decreases (twice smaller sometimes). The best results are obtained when $T = 100$ and $N = 200$. When ϕ_1 is Gaussian the estimation seems more difficult than when it is Gamma. The influence of σ may be seen when T is not very large. For example for the OU-model, the quantity σ^2/T appears is the theoretical bound of the MISE and thus it must not be too large. When T is large ($T = 100$) there is very little difference between the MISE's values for $\sigma = 0.1$ and $\sigma = 1$. Note that, the two estimators

Table 4.2 – CIR example. MISE computed from 100 simulations

Case	σ	T	N	\tilde{f}	\hat{f}	\hat{f}_1	\hat{f}_2
$\phi_1 \sim \Gamma, \phi_2 \sim \Gamma$	0.1	10	20	0.031	0.030	0.049	0.080
			200	0.019	0.018	0.050	0.051
		100	20	0.022	0.021	0.027	0.031
	1	100	200	0.005	0.004	0.006	0.006
			200	0.006	0.006	0.012	0.010
		$\phi_1 \sim \Gamma, \phi_2 \sim 1/\Gamma$	0.1	10	20	0.028	0.027
200	0.017				0.017	0.012	0.045
100	20			0.020	0.019	0.027	0.025
1	100		200	0.005	0.005	0.006	0.006
			200	0.007	0.007	0.013	0.008

Table 4.3 – CIR example. MISE computed from 100 simulations, with $N = 20$ on the left and $N = 200$ on the right, $\sigma = 0.1$, Gamma distributions for the random effects.

T	δ	Δ	T/Δ	\tilde{f}	\hat{f}	T	δ	Δ	T/Δ	\tilde{f}	\hat{f}
10	0.005	0.005×2	1000	0.038	0.037	10	0.005	0.005×2	1000	0.029	0.025
100	0.05	0.05×2	1000	0.021	0.025	100	0.05	0.05×2	1000	0.006	0.007
100	0.005	0.005×2	10000	0.019	0.020	100	0.005	0.005×2	10000	0.006	0.007

$\hat{A}_j(T)$ and $A_j(T)$ are asymptotically equivalent when T goes to infinity (in fact the proof of Proposition 4.3.1 shoes that the probability $\mathbb{P}(B_j(T)^c)$ tends to 0 when T goes to infinity), thus it could explain that the truncation effect is no more visible on the MISEs when T is large.

Table 4.3 highlights the different roles of T and Δ . When T increases (from 10 to 100) with the same Δ , the MISEs are divided by 2. When both T and Δ (from 0.005×2 to 0.05×2) increase, we still note a slight improvement. However, T/Δ has clearly no influence and this is why in the previous tables we fixed the number of observations per trajectory at $T/\Delta = 2000$ ($\Delta = 10 \times \delta$). The right table highlights again the role of N .

4.7 Concluding remarks

In this work, we first provide an estimator of a bivariate random effect in the drift of a stochastic differential equation based on the observation of one trajectory given by (4.1.1). The definition of the estimator uses a truncation threshold which allows us to study and bound its L^2 -risk. Then, using the estimators of the random effects, we build a kernel estimator of the common bivariate density of the random effects, from N *i.i.d.* observed trajectories and propose a data-driven selection of the bandwidth. The Orsntein-Uhlenbeck and the Cox-Ingersoll-Ross models both with two random effects can be studied with our estimation procedure. Illustrations on simulated data are done and show the good performances of our estimator.

As these two models have many applications in finance and in neuroscience, it would be interesting to investigate our method on a real databasis.

The method developed here for two random effects can be easily developed for d random effects. The extension to more general drift forms would require further work and other tools to accommodate the estimation procedure, especially for estimating the random effects. Models including a linear random effect in the diffusion coefficient could be investigated, for instance using the ideas developed in Delattre *et al.* (2015a).

The nonparametric estimation of the drift function in SDEs with no random effect has been largely investigated in the literature (see *e.g.* Kutoyants, 2004; Comte *et al.*, 2007; Schmisser, 2013). In the case of SDEs with random effects and a general drift $b(x, \varphi_1, \varphi_2)$, the nonparametric estimation of the function b is open and of interest.

4.8 Proofs

4.8.1 An auxiliary result and its proof

We denote \mathcal{L}_φ the generator of the process given by (4.1.1) when $\phi_j = \varphi$, given for all $F \in \mathcal{C}^2((l, r))$, by $\mathcal{L}_\varphi F(x) := (\sigma^2(x)/2)F''(x) + (b^t(x)\varphi)F'(x)$. Then when F is a matrix, the notation $\mathcal{L}_\varphi F$ only indicated that we apply the operator on each coefficient of the matrix. We also define in this sense F' as the derivative matrix coefficients by coefficients.

Proposition 4.8.1. *Consider the processes $(X_j(t))$ given by (4.1.1) under **(A1)**-**(A4)**. Let $g = (g_{i,k})_{1 \leq i,k \leq 2}$ a matrix of $S_2(\mathbb{R})$, such that $\pi_\varphi g_{i,k}^2 < +\infty$ for all $\varphi \in \Phi$. Assume that*

$$F_\varphi^g = \begin{pmatrix} F_\varphi^{g_{1,1}} & F_\varphi^{g_{1,2}} \\ F_\varphi^{g_{2,1}} & F_\varphi^{g_{2,2}} \end{pmatrix}$$

satisfies $-\mathcal{L}_\varphi F_\varphi^g = g - \pi_\varphi g$ for all $\varphi \in \Phi$. Let $J(\phi_j)$ be a non negative measurable function of ϕ_j such that for all $p \geq 1$

$$\mathbb{E} \left[J(\phi_j) \left(\frac{1}{T^p} \pi_{\phi_j} \left(\|F_{\phi_j}^g\|_F^{2p} \right) + \pi_{\phi_j} \left(\|(F_{\phi_j}^g)'\|_F^{2p} \sigma^{2p} \right) \right) \right] < \infty.$$

Then for all $p \geq 1$, there exists a constant $C_p > 0$ depending on p such that

$$\begin{aligned} \mathbb{E} \left[J(\phi_j) \left\| \sqrt{T} \left(\frac{1}{T} \int_0^T g(X_j(s)) ds - \pi_{\phi_j}(g) \right) \right\|_F^{2p} \right] &\leq C_p \left(\frac{1}{T^p} \mathbb{E} \left[J(\phi_j) \pi_{\phi_j} \left(\|F_{\phi_j}^g\|_F^{2p} \right) \right] \right. \\ &\quad \left. + \mathbb{E} \left[J(\phi_j) \pi_{\phi_j} \left(\|(F_{\phi_j}^g)'\|_F^{2p} \sigma^{2p} \right) \right] \right). \end{aligned}$$

Proof of Proposition 4.8.1.

We demonstrate the result without the function J but it can be added all along. The subscript j is omitted for simplicity. Denote $g_\varphi := g - \pi_\varphi g$. Ito's formula applied to F_φ^g term by term, and the equality $\mathcal{L}_\varphi F_\varphi^g = -g_\varphi$, lead to

$$\begin{aligned} F_\varphi^g(X^\varphi(T)) &= F_\varphi^g(X^\varphi(0)) + \int_0^T \mathcal{L}_\varphi F_\varphi^g(X^\varphi(s)) ds + \int_0^T (F_\varphi^g)'(X^\varphi(s)) \sigma(X^\varphi(s)) dW(s) \\ &= F_\varphi^g(X^\varphi(0)) - \int_0^T g_\varphi(X^\varphi(s)) ds + \int_0^T (F_\varphi^g)'(X^\varphi(s)) \sigma(X^\varphi(s)) dW(s). \end{aligned}$$

Thus,

$$\int_0^T g_\varphi(X^\varphi(s)) ds = \int_0^T (F_\varphi^g)'(X^\varphi(s)) \sigma(X^\varphi(s)) dW(s) + (F_\varphi^g(X^\varphi(0)) - F_\varphi^g(X^\varphi(T))).$$

Then with Hölder's inequality it yields

$$\mathbb{E} \left[\left\| \int_0^T g_\varphi(X^\varphi(s)) ds \right\|_F^{2p} \right] \leq 2^{2p-1} (\mathbb{T}_a + \mathbb{T}_b)$$

with

$$\mathbb{T}_a := \mathbb{E} \left[\left\| \int_0^T (F_\varphi^g)'(X^\varphi(s))\sigma(X^\varphi(s))dW(s) \right\|_F^{2p} \right], \quad \mathbb{T}_b := \mathbb{E} \left[\|F_\varphi^g(X^\varphi(0)) - F_\varphi^g(X^\varphi(T))\|_F^{2p} \right] \quad (4.8.1)$$

We study first the term \mathbb{T}_a given by (4.8.1). Again, Hölder's inequality gives

$$\begin{aligned} \mathbb{T}_a &= \mathbb{E} \left[\left(\sum_{1 \leq i, k \leq 2} \left(\int_0^T (F_\varphi^{g_{i,k}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s) \right)^2 \right)^p \right] \\ &\leq 4^{p-1} \mathbb{E} \left[\sum_{1 \leq i, k \leq 2} \left(\int_0^T (F_\varphi^{g_{i,k}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s) \right)^{2p} \right]. \end{aligned}$$

The first term of the sum can be studied with the Burkholder-Davis-Gundy (B.D.G.) inequality (see *e.g.* Revuz & Yor, 1999; Le Gall, 2010). Thus there exists a constant $c_p > 0$ such that

$$\begin{aligned} \mathbb{E} \left[\left(\int_0^T (F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s))dW(s) \right)^{2p} \right] &\leq c_p \mathbb{E} \left[\left(\int_0^T ((F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s)))^2 ds \right)^p \right] \\ &\leq c_p T^{p-1} \mathbb{E} \left[\int_0^T ((F_\varphi^{g_{1,1}})'(X^\varphi(s))\sigma(X^\varphi(s)))^{2p} ds \right] \\ &= c_p T^{p-1} T \pi_\varphi(((F_\varphi^{g_{1,1}})' \sigma)^{2p}). \end{aligned}$$

Finally there is a constant $C_p := 4^{p-1}c_p$ verifying

$$\mathbb{T}_a \leq C_p T^p \pi_\varphi \left(\left(\sum_{1 \leq i, k \leq 2} ((F_\varphi^{g_{i,k}})' \sigma)^{2p} \right) \sigma^{2p} \right) \leq C_p T^p \pi_\varphi \left(\| (F_\varphi^g)' \|_F^{2p} \sigma^{2p} \right).$$

Furthermore, we study term \mathbb{T}_b (4.8.1) and Cauchy-Schwarz's inequality leads to

$$\begin{aligned} \mathbb{T}_b &\leq \mathbb{E} \left[\left((\|F_\varphi^g(X^\varphi(0))\|_F^{2p} + \|F_\varphi^g(X^\varphi(T))\|_F^{2p})^{1/(2p)} 2^{1-1/(2p)} \right)^{2p} \right] \\ &\leq 2^{2p-1} \mathbb{E} \left[\|F_\varphi^g(X^\varphi(0))\|_F^{2p} + \|F_\varphi^g(X^\varphi(T))\|_F^{2p} \right] = 2^{2p} \pi_\varphi(\|F_\varphi^g\|_F^{2p}). \end{aligned}$$

Finally, with $C'_p := \max(2^{2p}, C_p)$ it yields

$$\mathbb{E} \left[\frac{1}{T^p} \left\| \int_0^T g_\varphi(X^\varphi(s))ds \right\|_F^{2p} \right] \leq C'_p \left(\pi_\varphi \left(\| (F_\varphi^g)' \|_F^{2p} \sigma^{2p} \right) + \frac{1}{T^p} \pi_\varphi(\|F_\varphi^g\|_F^{2p}) \right). \quad \square$$

When $g = b b^t / \sigma^2$ this result controls the distance (in Frobenius norm) between $V_j(T)/T$ and its limit matrix L_j .

4.8.2 Proof of Proposition 4.3.1

The subscript j is omitted for simplicity. First note that $\widehat{A} - \phi = (A - \phi)\mathbf{1}_{B(T)} - \phi\mathbf{1}_{B(T)^c}$. Thus:

$$T\|\widehat{A} - \phi\|_2^2 = T\|(A - \phi)\mathbf{1}_{B(T)} - \phi\mathbf{1}_{B(T)^c}\|_2^2 \leq 2(\mathbb{T}_1 + \mathbb{T}_2)$$

with

$$\mathbb{T}_1 := T\|(A - \phi)\|_2^2\mathbf{1}_{B(T)}, \quad \mathbb{T}_2 := T\|\phi\|_2^2\mathbf{1}_{B(T)^c}. \quad (4.8.2)$$

• Study of \mathbb{T}_1 . We have

$$\begin{aligned} \mathbb{T}_1 &= \mathbf{1}_{B(T)}T\|V(T)^{-1}M(T)\|_2^2 = \mathbf{1}_{B(T)}\left\|TV(T)^{-1}\frac{M(T)}{\sqrt{T}}\right\|_2^2 \\ &= \mathbf{1}_{B(T)}\left\|(TV(T)^{-1} - L^{-1} + L^{-1})\frac{M(T)}{\sqrt{T}}\right\|_2^2 \leq 2\mathbb{T}'_1 + 2\mathbb{T}_1'' \end{aligned}$$

with

$$\mathbb{T}'_1 := \mathbf{1}_{B(T)}\left\|(TV(T)^{-1} - L^{-1})\frac{M(T)}{\sqrt{T}}\right\|_2^2, \quad \mathbb{T}_1'' := \mathbf{1}_{B(T)}\left\|L^{-1}\frac{M(T)}{\sqrt{T}}\right\|_2^2 \quad (4.8.3)$$

Let us study this two terms separately. Note that

$$\mathbb{E}\left(\frac{V(T)}{T} \middle| \phi\right) = \frac{1}{T} \int_0^T \mathbb{E}\left(\frac{bb^t}{\sigma^2}(X(s)) \middle| \phi\right) ds,$$

$$\mathbb{E}\left(\frac{bb^t}{\sigma^2}(X(s)) \middle| \phi = \varphi\right) = \mathbb{E}\left(\frac{bb^t}{\sigma^2}(X^\varphi(s))\right) = \mathbb{E}\left(\frac{bb^t}{\sigma^2}(X^\varphi(0))\right) = \pi_\varphi\left(\frac{bb^t}{\sigma^2}\right).$$

Thus

$$L := \pi_\phi\left(\frac{bb^t}{\sigma^2}\right) = \mathbb{E}\left(\frac{V(T)}{T} \middle| \phi\right).$$

• Study of \mathbb{T}_1'' . As L is \mathcal{F}_0 measurable,

$$\begin{aligned} \mathbb{E}[\mathbb{T}_1''] &\leq \mathbb{E}\left[\|L^{-1}\|_F^2 \left\|\frac{M(T)}{\sqrt{T}}\right\|_2^2\right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \mathbb{E}\left[\left(\frac{M_1^2(T)}{T} + \frac{M_2^2(T)}{T}\right) \middle| \mathcal{F}_0\right]\right] \\ &= \mathbb{E}\left[\|L^{-1}\|_F^2 \left(\frac{\langle M_1 \rangle_T}{T} + \frac{\langle M_2 \rangle_T}{T}\right)\right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \text{Tr}\left(\frac{V(T)}{T}\right)\right] \end{aligned}$$

Thus we obtain:

$$\mathbb{E}[\mathbb{T}_1''] \leq \mathbb{E}\left[\|L^{-1}\|_F^2 \mathbb{E}\left[\text{Tr}\left(\frac{V(T)}{T}\right) \middle| \phi\right]\right] = \mathbb{E}\left[\|L^{-1}\|_F^2 \text{Tr}(L)\right].$$

Using Property 4.8.3 (see Appendix), we must assume :

$$\mathbb{E}\left[\frac{\|L\|_F^2}{\det(L)^2} \text{Tr}(L)\right] < \infty. \quad (4.8.4)$$

• Study of \mathbb{T}'_1 given by (4.8.3). This term will use the cut-off $B(T)$. We have:

$$\begin{aligned} TV(T)^{-1} - L^{-1} &= (TV(T)^{-1}L - I)L^{-1} \\ &= V(T)^{-1}(TL - V(T))L^{-1}. \end{aligned} \quad (4.8.5)$$

Then we obtain:

$$\begin{aligned} (TV(T)^{-1} - L^{-1}) \frac{M(T)}{\sqrt{T}} &= V(T)^{-1}(TL - V(T))L^{-1} \frac{M(T)}{\sqrt{T}} \\ &= \underbrace{TV(T)^{-1} \left(L - \frac{V(T)}{T} \right)}_{\text{matrix}} \underbrace{L^{-1} \frac{M(T)}{\sqrt{T}}}_{\text{vector}} \end{aligned}$$

and

$$\mathbb{E}[\mathbb{T}'_1] \leq \mathbb{E} \left[\|TV^{-1}(T)\|_F^2 \mathbf{1}_{B(T)} \left\| L - \frac{V(T)}{T} \right\|_F^2 \left\| L^{-1} \frac{M(T)}{\sqrt{T}} \right\|_2^2 \right].$$

But $\|V(T)^{-1}\|_F^2 = \det(V(T)^{-2})\|V(T)\|_F^2 = \lambda_1(T)^{-2} + \lambda_2(T)^{-2}$. And the definition of the set $B(T)$ implies:

$$\|TV(T)^{-1}\|_F^2 \leq \frac{2T}{\kappa^2}. \quad (4.8.6)$$

Thus the Hölder inequality yields

$$\mathbb{E}[\mathbb{T}'_1] \leq \mathbb{E} \left[\frac{2T}{\kappa^2} \left\| L - \frac{V(T)}{T} \right\|_F^2 \left\| L^{-1} \frac{M(T)}{\sqrt{T}} \right\|_2^2 \right] \leq \frac{2}{\kappa^2} (\mathbb{E}[\mathbb{T}_c] \mathbb{E}[\mathbb{T}_d])^{1/2}$$

with

$$\mathbb{T}_c := \left\| \sqrt{T} \left(L - \frac{V(T)}{T} \right) \right\|_F^4, \quad \mathbb{T}_d := \left\| L^{-1} \frac{M(T)}{\sqrt{T}} \right\|_2^4. \quad (4.8.7)$$

- Study of term \mathbb{T}_c . We apply here Proposition 4.8.1 with $p = 2$ and $g = b^t b / \sigma^2$ and get

$$\mathbb{E}[\mathbb{T}_c] \leq C \left(\frac{1}{T^2} \mathbb{E}[\|H_\phi\|_F^4] + \mathbb{E}[\|H'_\phi\|_F^4 \sigma^4] \right) < \infty. \quad (4.8.8)$$

- Study of term \mathbb{T}_d . We notice that:

$$\mathbb{E}[\mathbb{T}_d] \leq \mathbb{E} \left[\|L^{-1}\|_F^4 \left\| \frac{M(T)}{\sqrt{T}} \right\|_2^4 \right].$$

Proceeding as for \mathbb{T}_1 " and using the B.D.G. inequality implies that for some constant $c > 0$,

$$\begin{aligned} \mathbb{E} \left[\|L^{-1}\|_F^4 \left\| \frac{M(T)}{\sqrt{T}} \right\|_2^4 \right] &= \mathbb{E} \left[\|L^{-1}\|_F^4 \left(\frac{M_1^2(T) + M_2^2(T)}{T} \right)^2 \right] \\ &\leq c \mathbb{E} \left[\|L^{-1}\|_F^4 \mathbb{E} \left[\left(\text{Tr} \left(\frac{V(T)}{T} \right) \right)^2 \middle| \phi \right] \right]. \end{aligned}$$

And by the Hölder inequality,

$$\begin{aligned} \mathbb{E} \left[\left(\text{Tr} \left(\frac{V(T)}{T} \right) \right)^2 \middle| \phi = \varphi \right] &= \mathbb{E} \left[\left(\frac{1}{T} \int_0^T \left(\frac{b_1^2}{\sigma^2}(X^\varphi(s)) + \frac{b_2^2}{\sigma^2}(X^\varphi(s)) \right) ds \right)^2 \right] \\ &\leq \mathbb{E} \left[\frac{1}{T} \int_0^T \left(\frac{b_1^2}{\sigma^2}(X^\varphi(s)) + \frac{b_2^2}{\sigma^2}(X^\varphi(s)) \right)^2 ds \right] \\ &= \text{Tr} \left(\pi_\varphi \left(\left(\frac{bb^t}{\sigma^2} \right)^2 \right) \right). \end{aligned}$$

Thus,

$$\mathbb{E}[\mathbb{T}_d] \leq \mathbb{E} \left[\|L^{-1}\|_F^4 \text{Tr} \left(\pi_\phi \left(\left(\frac{bb^t}{\sigma^2} \right)^2 \right) \right) \right].$$

Finally we must assume that

$$\mathbb{E} \left[\frac{\|L\|_F^4}{\det(L)^4} \text{Tr} \left(\pi_\phi \left(\left(\frac{bb^t}{\sigma^2} \right)^2 \right) \right) \right] < \infty. \quad (4.8.9)$$

• Study of term \mathbb{T}_2 . Term \mathbb{T}_2 is given by (4.8.2). Note that

$$\mathbb{E}[\mathbb{T}_2] = T \mathbb{E}[\mathbb{E}[\|\phi\|_2^2 \mathbf{1}_{\{B(T)^c\}} | \phi]] = T \mathbb{E}[\|\phi\|_2^2 \mathbb{E}[\mathbf{1}_{\{B(T)^c\}} | \phi]].$$

We need to compute the quantity $\mathbb{P}(B(T)^c | \phi = \varphi)$ for all φ . We have:

$$B(T)^c = \left\{ \frac{V(T)}{T} \preceq \frac{\kappa}{\sqrt{T}} I_2 \right\} = \left\{ L - \frac{V(T)}{T} \succeq L - \frac{\kappa}{\sqrt{T}} I_2 \right\}.$$

If we denote $l := \pi_\varphi \left(\frac{bb^t}{\sigma^2} \right)$,

$$l - \frac{\kappa}{\sqrt{T}} I_2 \succeq \frac{l}{2} \Leftrightarrow l - \frac{l}{2} \succeq \frac{\kappa}{\sqrt{T}} I_2 \Leftrightarrow l \succeq \frac{2\kappa}{\sqrt{T}} I_2.$$

Furthermore,

$$\begin{aligned} \mathbb{P} \left(\frac{V(T)}{T} \preceq \frac{\kappa}{\sqrt{T}} I_2 \mid \phi = \varphi \right) &\leq \mathbf{1}_{l \succeq \frac{2\kappa}{\sqrt{T}} I_2} \mathbb{P} \left(\frac{V(T)}{T} \preceq \frac{\kappa}{\sqrt{T}} I_2 \mid \phi = \varphi \right) + \mathbf{1}_{l \preceq \frac{2\kappa}{\sqrt{T}} I_2} \\ &\leq \mathbb{T}_\alpha + \mathbb{T}_\beta \end{aligned}$$

with

$$\mathbb{T}_\alpha := \mathbb{P} \left(l - \frac{V(T)}{T} \succeq \frac{l}{2} \mid \phi = \varphi \right), \quad \mathbb{T}_\beta := \mathbf{1}_{l \preceq \frac{2\kappa}{\sqrt{T}} I_2}. \quad (4.8.10)$$

We study the two terms separately. First:

$$\mathbb{T}_\beta = \mathbf{1}_{l \preceq \frac{2\kappa}{\sqrt{T}} I_2} = \mathbf{1}_{l^{-1} \succeq \frac{\sqrt{T}}{2\kappa} I_2} = \mathbf{1}_{\min\{\frac{1}{\lambda_{\varphi,1}}, \frac{1}{\lambda_{\varphi,2}}\} \geq \frac{\sqrt{T}}{2\kappa}} \leq \left(\frac{4\kappa^2}{T} \right) \min \left\{ \frac{1}{\lambda_{\varphi,1}}, \frac{1}{\lambda_{\varphi,2}} \right\}$$

Secondly, for the term \mathbb{T}_α given by formula (4.8.10), we use Property 4.8.5 (Appendix)

$$l - \frac{V(T)}{T} \succeq \frac{l}{2} \Rightarrow l_{i,i} - \frac{V_{i,i}(T)}{T} \geq \frac{l_{i,i}}{2}, \quad i = 1, 2$$

then

$$\mathbb{T}_\alpha \leq \mathbb{P} \left(l_{1,1} - \frac{V_{1,1}(T)}{T} \geq \frac{l_{1,1}}{2} \mid \phi = \varphi \right) + \mathbb{P} \left(l_{2,2} - \frac{V_{2,2}(T)}{T} \geq \frac{l_{2,2}}{2} \mid \phi = \varphi \right).$$

For the two terms of the right hand side we are able now to use Markov inequality:

$$\mathbb{T}_\alpha \leq \left(\frac{2}{l_{1,1}} \right)^2 \mathbb{E} \left[\left(l_{1,1} - \frac{V_{1,1}(T)}{T} \right)^2 \mid \phi = \varphi \right] + \left(\frac{2}{l_{2,2}} \right)^2 \mathbb{E} \left[\left(l_{2,2} - \frac{V_{2,2}(T)}{T} \right)^2 \mid \phi = \varphi \right].$$

Finally we obtain

$$\mathbb{E}[T_2] \leq T\mathbb{E} \left[\|\phi\|_2^2 \left(\left(\frac{2}{L_{1,1}} \right)^2 \left(L_{1,1} - \frac{V_{1,1}(T)}{T} \right)^2 + \left(\frac{2}{L_{2,2}} \right)^2 \left(L_{2,2} - \frac{V_{2,2}(T)}{T} \right)^2 + \left(\frac{4\kappa^2}{T} \right) \min \left\{ \frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right\} \right) \right] \quad (4.8.11)$$

$$\leq \mathbb{E} \left[4\|\phi\|_2^2 \left(\frac{1}{L_{1,1}^2} + \frac{1}{L_{2,2}^2} \right) \left\| \sqrt{T} \left(L - \frac{V(T)}{T} \right) \right\|^2 \right] + \mathbb{E} \left[4\kappa^2 \|\phi\|_2^2 \min \left\{ \frac{1}{\lambda_1}, \frac{1}{\lambda_2} \right\} \right] \quad (4.8.12)$$

We are able to conclude this proof using the Proposition 4.8.1 with $J(\phi) = \|\phi\|_2^2 \left(\frac{1}{L_{1,1}^2} + \frac{1}{L_{2,2}^2} \right)$. Indeed, as $T\mathbb{E}[\|\widehat{A} - \phi\|^2] \leq 4\mathbb{E}[T_1]4\mathbb{E}[T_1] + 2\mathbb{E}[T_2]$, gathering the bounds (4.8.4), (4.8.8), (4.8.9), and (4.8.11) we obtain that there is a constant $C > 0$ such that $\mathbb{E}[\|\widehat{A}_j(T) - \phi_j\|_2^2] \leq C/T$. \square

4.8.3 Proof of proposition 4.4.1

With the Cauchy-Schwarz inequality, we obtain the following decomposition

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_h - f\|^2] &= \|f - \mathbb{E}[\widehat{f}_h]\|^2 + \mathbb{E}[\|\mathbb{E}[\widehat{f}_h] - \widehat{f}_h\|^2] \\ &\leq 2\|f - f_h\|^2 + 2\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 + \mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2]. \end{aligned}$$

Let us study the second term of the bound. For all $x = (x_1, x_2) \in \mathbb{R}^2$,

$$\mathbb{E}[\widehat{f}_h(x)] = \frac{1}{N} \sum_{j=1}^N \mathbb{E}[K_h(x - \widehat{A}_j(T))] = \mathbb{E}[K_h(x - \widehat{A}_1(T))],$$

thus $f_h \neq \mathbb{E}[\widehat{f}_h]$. The subscript j (or 1) is omitted in the following for simplicity. Note that

$$f_h(x) = f_h(x_1, x_2) = f \star K_h(x_1, x_2) = \frac{1}{h_1 h_2} \int_{\mathbb{R}^2} f(u, v) K \left(\frac{x_1 - u}{h_1}, \frac{x_2 - v}{h_2} \right) dudv = \mathbb{E}[K_h(x - \phi)].$$

Then to study $\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 = \int_{\mathbb{R}^2} \mathbb{E}[K_h(x - \widehat{A}(T)) - K_h(x - \phi)]^2 dx$ we denote $DK_h = \left(\frac{\partial K_h}{\partial u}, \frac{\partial K_h}{\partial v} \right)^t$, with

$$\frac{\partial K_h}{\partial u}(u, v) = \frac{1}{h_1^2 h_2} \frac{\partial K}{\partial u} \left(\frac{u}{h_1}, \frac{v}{h_2} \right), \quad \frac{\partial K_h}{\partial v}(u, v) = \frac{1}{h_1 h_2^2} \frac{\partial K}{\partial v} \left(\frac{u}{h_1}, \frac{v}{h_2} \right)$$

and applying Taylor's formula with $U(x, t) := (U_1(x, t), U_2(x, t)) := x - \phi + t(\phi - \widehat{A}(T))$, it yields

$$K_h(x - \widehat{A}(T)) - K_h(x - \phi) = \int_0^1 DK_h(U(x, t)) \cdot (\phi - \widehat{A}(T)) dt.$$

Thus

$$\begin{aligned} \|f_h - \mathbb{E}[\widehat{f}_h]\|^2 &\leq \int \mathbb{E} \left[(K_h(x - \phi) - K_h(x - \widehat{A}(T)))^2 \right] dx = \mathbb{E} \left[\int (K_h(x - \phi) - K_h(x - \widehat{A}(T)))^2 dx \right] \\ &\leq \mathbb{E} \left[2 \int_{\mathbb{R}^2} \left(\int_0^1 \frac{2}{h_1^2 h_2} \frac{\partial K}{\partial u} \left(\frac{U_1(x, t)}{h_1}, \frac{U_2(x, t)}{h_2} \right) dt \right)^2 dx_1 dx_2 (\phi_1 - \widehat{A}_1(T))^2 \right] \\ &\quad + \mathbb{E} \left[2 \int_{\mathbb{R}^2} \left(\int_0^1 \frac{2}{h_1 h_2^2} \frac{\partial K}{\partial v} \left(\frac{U_1(x, t)}{h_1}, \frac{U_2(x, t)}{h_2} \right) dt \right)^2 dx_1 dx_2 (\phi_2 - \widehat{A}_2(T))^2 \right]. \end{aligned}$$

The first term of the previous sum is bounded according to Proposition 4.3.1:

$$\begin{aligned} & \mathbb{E} \left[\int_{\mathbb{R}^2} \left(\int_0^1 \frac{1}{h_1^2 h_2} \frac{\partial K}{\partial u} \left(\frac{U_{1,1}}{h_1}, \frac{U_{1,2}}{h_2} \right) dt \right)^2 dx_1 dx_2 (\phi_{1,1} - \widehat{A}_{1,1}(T))^2 \right] \\ & \leq \mathbb{E} \left(\frac{1}{h_1^4 h_2^2} \int_0^1 dt \int_{\mathbb{R}^2} \left(\frac{\partial K}{\partial u}(y_1, y_2) \right)^2 dy_1 dy_2 h_1 h_2 (\phi_{1,1} - \widehat{A}_{1,1}(T))^2 \right) \\ & \leq \frac{1}{h_1^3 h_2} \int_{\mathbb{R}^2} \left(\frac{\partial K}{\partial u}(y_1, y_2) \right)^2 dy_1 dy_2 \mathbb{E}[(\phi_{1,1} - \widehat{A}_{1,1}(T))^2] \leq \frac{C}{h_1^3 h_2} T \left\| \frac{\partial K}{\partial u} \right\|^2 \end{aligned}$$

with C the constant from Equation (4.3.5). The same arguments works for the second term and finally it yields

$$\|f_h - \mathbb{E}[\widehat{f}_h]\|^2 \leq C \max \left(\frac{1}{h_1^3 h_2}, \frac{1}{h_1 h_2^3} \right) \frac{\left\| \frac{\partial K}{\partial u} \right\|^2 + \left\| \frac{\partial K}{\partial v} \right\|^2}{T}.$$

Finally, the last term is bounded by

$$\mathbb{E}[\|\widehat{f}_h - \mathbb{E}[\widehat{f}_h]\|^2] \leq \frac{\|K\|^2}{N h_1 h_2}. \quad \square$$

4.8.4 Proof of Proposition 4.4.3

For all $x = (x_1, x_2) \in \mathbb{R}^2$,

$$f_h(x) - f(x) = f \star K_h(x) - f(x) = \int_{\mathbb{R}^2} K(v)[f(x - vh) - f(x)]dv.$$

Then,

$$f(x_1 - v_1 h_1, x_2 - v_2 h_2) - f(x_1, x_2) = f(x_1 - v_1 h_1, x_2 - v_2 h_2) - f(x_1 - v_1 h_1, x_2) + f(x_1 - v_1 h_1, x_2) - f(x_1, x_2).$$

As in Tsybakov (2009) we apply Taylor's formula to the two partial functions: $t \mapsto f(x_1 - v_1 h_1, t)$ at the order l_2 and $t \mapsto f(t, x_2)$ at the order l_1 . Using the orders of the two kernels yields for the first term,

$$\begin{aligned} B_1(x) & := \int_{\mathbb{R}^2} K_1(v_1)K_2(v_2)[f(x_1 - v_1 h_1, x_2 - v_2 h_2) - f(x_1 - v_1 h_1, x_2)]dv_1 dv_2 \tag{4.8.13} \\ & = - \int_{\mathbb{R}^2} \frac{(h_2 v_2)^{l_2}}{(l_2 - 1)!} \int_0^1 (1-t)^{l_2-1} \frac{\partial^{l_2}}{\partial x_2^{l_2}} f(x_1 - v_1 h_1, x_2 - v_2 h_2 + t h_2 v_2) dt K_1(v_1)K_2(v_2)dv_1 dv_2 \\ & = - \int_{\mathbb{R}^2} \frac{(h_2 v_2)^{l_2}}{(l_2 - 1)!} \int_0^1 (1-t)^{l_2-1} \left[\frac{\partial^{l_2}}{\partial x_2^{l_2}} f(x_1 - v_1 h_1, x_2 - v_2 h_2 + t h_2 v_2) \right. \\ & \quad \left. - \frac{\partial^{l_2}}{\partial x_2^{l_2}} f(x_1 - v_1 h_1, x_2 - v_2 h_2) \right] dt K_1(v_1)K_2(v_2)dv_1 dv_2 \end{aligned}$$

and the analogue term $B_2(x)$. To evaluate $\|f - f_h\|^2 = \int (f_h(x) - f(x))^2 dx$ we remind the generalized Minkowski's inequality (see Tsybakov, 2009): for any measurable function f on \mathbb{R}^2 , we have

$$\int \left(\int f(u, x) du \right)^2 dx \leq \left[\int \left(\int f^2(u, x) dx \right)^{1/2} du \right]^2.$$

Looking at the first term (4.8.13), and applying twice the above inequality implies for $f \in \mathcal{N}(\beta, R)$:

$$\begin{aligned} \int_{\mathbb{R}^2} B_1^2(x) dx &\leq \left(\int_{\mathbb{R}^2} |K_1(v_1)K_2(v_2)| \frac{h_2|v_2|^{l_2}}{(l_2-1)!} \int_0^1 (1-t)^{l_2-1} \left[\int_{\mathbb{R}^2} \left(\frac{\partial^{l_2}}{\partial x_2^{l_2}} f(x_1 - v_1 h_1, x_2 - v_2 h_2 + t h_2 v_2) \right. \right. \right. \\ &\quad \left. \left. \left. - \frac{\partial^{l_2}}{\partial x_2^{l_2}} f(x_1 - v_1 h_1, x_2 - v_2 h_2) \right)^2 dx_1 dx_2 \right]^{1/2} dt dv_1 dv_2 \right)^2 \\ &\leq \left[\int_{\mathbb{R}^2} |K_1(v_1)K_2(v_2)| h_2^{\beta_2} \frac{|v_2|^{\beta_2}}{l_2!} R dv_1 dv_2 \right]^2 \leq C_2^2 h_2^{2\beta_2} \end{aligned}$$

with $C_2 := (R/l_2!) \int |K_2(v_2)v_2^{\beta_2}| dv_2$. Finally,

$$\begin{aligned} \int (f_h(x) - f(x))^2 dx &\leq 2 \int_{\mathbb{R}^2} B_1^2(x) dx + 2 \int_{\mathbb{R}^2} B_2^2(x) dx \\ &\leq 2C_1^2 h_1^{2\beta_1} + 2C_2^2 h_2^{2\beta_2}. \quad \square \end{aligned}$$

Appendix 1

4.8.5 Useful results of algebra

Property 4.8.2. For all $A, B \in M_2(\mathbb{R})$, $\|AB\|_F \leq \|A\|_F \|B\|_F$ and for all $x \in \mathbb{R}^2$, $\|Ax\|_2 \leq \|A\|_F \|x\|_2$ where $\|\cdot\|_2$ is the euclidean norm on \mathbb{R}^2 .

Property 4.8.3. If $A \in S_2(\mathbb{R})$ and $(\lambda_1, \lambda_2) \in \mathbb{R}^2$ are its eigenvalues, $\|A\|_F^2 = \lambda_1^2 + \lambda_2^2$. Furthermore if A is invertible, $(\lambda_1, \lambda_2) \in \mathbb{R}^{*2}$ and $\|A^{-1}\|_F^2 = \left(\frac{1}{\det(A)}\right)^2 \|A\|_F^2 = \frac{1}{\lambda_1^2} + \frac{1}{\lambda_2^2}$.

Property 4.8.4. If $A \in S_2(\mathbb{R})$ with eigenvalues λ_1, λ_2 , then $A \succeq I_2 \Leftrightarrow \min(\lambda_1, \lambda_2) \geq 1$.

Property 4.8.5. For all $A \in S_2^+(\mathbb{R})$, $a_{i,i} \geq 0$ for $i = 1, 2$. Then if $A, B \in S_2(\mathbb{R})$,

$$A \succeq B \Rightarrow a_{i,i} \geq b_{i,i} \quad i = 1, 2.$$

4.8.6 Details on examples

We refer to Genon-Catalot *et al.* (2000) and Genon-Catalot & Larédo (2016) for details and properties on the generator infinitesimal. Nevertheless we recall one of them. For $f \in \mathcal{C}^2((l, r))$,

$$\mathcal{L}_\varphi f(x) = \frac{1}{2} \sigma^2(x) f''(x) + ({}^t b(x)\varphi) f'(x) = \frac{1}{2m_\varphi(x)} \left(\frac{f'}{s_\varphi} \right)'(x). \quad (4.8.14)$$

Thus for $g \in L_{\pi_\varphi}^2$ when we are looking for the associated F_φ^g such that $\mathcal{L}_\varphi F_\varphi^g = -(g - \pi_\varphi(g))$ we can use the relation $(F_\varphi^g)'(x) = -2s_\varphi(x) \int_l^x (g(u) - \pi_\varphi(g)) m_\varphi(u) du$.

For the OU model.

We want to explicit $H_\phi = F_\phi^{bb^t/\sigma^2}$. Let us denote: $g = bb^t/\sigma^2 = (g_{i,k})_{1 \leq i,k \leq 2}$. First: $g_{1,1}(x) - \pi_\varphi g_{1,1} = 0 = -\mathcal{L}_\varphi F_\varphi^{g_{1,1}}(x)$. For example $F_\varphi^{g_{1,1}}(x) = 1$ is suitable. Then, $g_{1,2}(x) - \pi_\varphi g_{1,2} = -\frac{1}{\sigma^2} \left(x - \frac{\varphi_1}{\varphi_2} \right)$. We look for $F_\varphi^{g_{1,2}}(x) = -\frac{a}{\sigma^2} \left(x - \frac{\varphi_1}{\varphi_2} \right)$ and we obtain

$$F_\varphi^{g_{1,2}}(x) = -\frac{1}{\varphi_2 \sigma^2} \left(x - \frac{\varphi_1}{\varphi_2} \right).$$

Finally

$$g_{2,2}(x) - \pi_\varphi g_{2,2} = \frac{x^2}{\sigma^2} - \frac{1}{2\varphi_2} + \frac{\varphi_1^2}{\sigma^2 \varphi_2^2} = \frac{1}{\sigma^2} \left[\left(x - \frac{\varphi_1}{\varphi_2} \right)^2 + 2 \frac{\varphi_1}{\varphi_2} \left(x - \frac{\varphi_1}{\varphi_2} \right) \right] - \frac{1}{2\varphi_2}.$$

We look for $F_\varphi^{g_{2,2}}$ with the same functional form and we obtain

$$F_\varphi^{g_{2,2}}(x) = \frac{2\varphi_1}{\sigma^2\varphi_2^2} \left(x - \frac{\varphi_1}{\varphi_2} \right) + \frac{1}{2\sigma^2\varphi_2} \left[\left(x - \frac{\varphi_1}{\varphi_2} \right)^2 - \frac{\sigma^2}{2\varphi_2} \right].$$

For the CIR model.

We want to explicit $H_\phi = F_\phi^{b^t/\sigma^2}$. First: $g_{1,1}(x) - \pi_\varphi g_{1,1} = 1/(\sigma^2 x) - 2\varphi_2/((2\varphi_1 - \sigma^2)\sigma^2)$. Here we use formula (4.8.14). We have

$$(F_\varphi^{g_{1,1}})'(x) = -2s_\varphi(x) \int_0^x \left(\frac{1}{\sigma^2 u} - \frac{2\varphi_2}{\sigma^2(2\varphi_1 - \sigma^2)} \right) m_\varphi(u) du$$

with $s_\varphi(x) = e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2}$. This is equivalent to

$$(F_\varphi^{g_{1,1}})'(x) = -\frac{2}{\sigma^4} e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2} \left[\int_0^x e^{-2\varphi_2 u/\sigma^2} u^{2\varphi_1/\sigma^2 - 2} du - \int_0^x \frac{2\varphi_2}{2\varphi_1 - \sigma^2} e^{-2\varphi_2 u/\sigma^2} u^{2\varphi_1/\sigma^2 - 1} du \right]$$

and an integration by part in the first integral gives

$$(F_\varphi^{g_{1,1}})'(x) = -\frac{2}{\sigma^4} e^{2\varphi_2 x/\sigma^2} x^{-2\varphi_1/\sigma^2} \left[-\frac{e^{-2\varphi_2 x/\sigma^2} x^{2\varphi_1/\sigma^2 - 1}}{2\varphi_1/\sigma^2 - 1} \right] = \frac{2}{\sigma^2(2\varphi_1 - \sigma^2)x}.$$

We finally set

$$F_\varphi^{g_{1,1}}(x) = \frac{2}{\sigma^2(2\varphi_1 - \sigma^2)} \log(x).$$

Furthermore $g_{1,2}(x) - \pi_\varphi g_{1,2} = 0 = -\mathcal{L}_\varphi F_\varphi^{g_{1,2}}(x)$, and then for example $F_\varphi^{g_{1,2}}(x) = 1$. At last $g_{2,2}(x) - \pi_\varphi g_{2,2} = \frac{1}{\sigma^2} \left(x - \frac{\varphi_1}{\varphi_2} \right)$ and we obtain

$$F_\varphi^{g_{2,2}}(x) = \frac{1}{\varphi_2\sigma^2} \left(x - \frac{\varphi_1}{\varphi_2} \right).$$

Appendix 2: more about the MISE bounds

In this chapter and the previous one, we investigate the Ornstein-Uhlenbeck model with one or two random effects in the drift. In this appendix we compare the MISE bounds of the kernels estimators obtained from an estimator of the random effects.

Inference for one additive random effect

We write the model: $dX_j(t) = (\phi_j - X_j(t))dt + \sigma dW_j(t)$, $X_j(0)$. In Chapter 3 we investigate the maximum likelihood estimator of ϕ_j

$$Z_{j,T} = \phi_j + \sigma \frac{W_j(T)}{T}$$

and the kernel estimator $\hat{f}_h(x) = \frac{1}{N} \sum_{k=1}^N K_h(x - Z_{j,T})$. It yields

$$\mathbb{E}[\|\hat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{2\sigma^4\|K''\|^2}{3T^2h^5} + \frac{\|K\|^2}{Nh}.$$

If σ is not constant, we obtain the same bound as above with σ_1 instead of σ if $\sigma(x) < \sigma_1$. Moreover, if we assume that the process $(\phi_j, X_j(\cdot))$ is at the stationary regime, assuming that $\mathbb{E}[\pi_\phi(\sigma^4)] < \infty$ then we are able to bound $\mathbb{E}[(\phi_1 - Z_{1,T})^2]$ and obtain the same bound for the MISE.

Inference for one multiplicative random effect

Let us give some details about the estimation for the multiplicative model: $dX_j(t) = \phi_j X_j(t)dt + \sigma dW_j(t)$, $X_j(0)$ which is studied in Comte *et al.* (2013) nonparametrically then parametrically in Genon-Catalot & Larédo (2016). This model has not been investigate yet in this thesis and it will be in the next Chapter 5. The maximum likelihood estimator of ϕ_j is $A_{j,T} = U_j(T)/V_j(T) = \int_0^T X_j(s)dX_j(s) / \int_0^T X_j^2(s)ds$. Then a truncated estimator of the random effects is proposed in Genon-Catalot & Larédo (2016): $\widehat{A}_{j,T} = A_j \mathbb{1}_{V_j \geq \kappa \sqrt{T}}$. The authors obtain: $\forall p \geq 1$, $\mathbb{E}[(\widehat{A}_{j,T} - \phi_j)^{2p}] \leq CT^{-p}$. The first kernel estimator introduced in Comte *et al.* (2013) is $\widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - A_{j,T})$. Then it is possible to define a new kernel estimator: $\widehat{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - \widehat{A}_{j,T})$. Finally under stationary assumptions we can show the following result.

Proposition 4.8.6. *Estimator \widehat{f}_h satisfies*

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + \frac{2C\|K'\|^2}{Th^3} + \frac{\|K\|^2}{Nh}.$$

If σ is not constant, the bound remains. This is new because the previous bound obtained in Comte *et al.* (2013) has the same order for a general drift $\phi_j b(\cdot)$ and a general diffusion coefficient $\sigma(\cdot)$ for estimator \widehat{f}_h but assuming that the rate b/σ is bounded.

Inference with two random effects

Let us now recall some results of Chapter 4. The model is $dX_j(t) = (\phi_{1,j} - \phi_{2,j}X_j(t))dt + \sigma dW_j(t)$, $X_j(0)$. We have built an estimator of the density of the couple of random effects based on a truncated version of the maximum likelihood estimator of the random parameters $\widehat{A}_{j,T}$. The obtained bound of the MISE of the kernel estimator \widehat{f}_h is

$$\mathbb{E}[\|\widehat{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + 2 \max\left(\frac{1}{h_1^3 h_2}, \frac{1}{h_1 h_2^3}\right) \left(\left\| \frac{\partial K}{\partial u} \right\|^2 + \left\| \frac{\partial K}{\partial v} \right\|^2 \right) \frac{C}{T} + \frac{\|K\|^2}{Nh_1 h_2}.$$

Furthermore, another kernel estimator can be built from an estimator of the random effects proposed in Genon-Catalot & Larédo (2016). This is not a method proposed in the article but we present it here. Genon-Catalot & Larédo (2016) investigate the moment estimator of the random effects. Let us consider the new parametrisation

$$dX_j(t) = -\frac{1}{\phi_{j,2}}(X_j(t) - \phi_{j,1})dt + \sigma dW_j(t), \quad X_j(0).$$

For this model, the invariant distribution for the conditional process is $\pi_\phi = \mathcal{N}(\varphi_1, \varphi_2/2)$. Then we omit the subscript j to make the notation simpler. Thus, it comes from the Ergodic Theorem, given in Part B, that $\tilde{\phi}_1 = \frac{1}{T} \int_0^T X(s)ds$ estimates the mean ϕ_1 and $\frac{1}{T} \int_0^T X^2(s)ds - \tilde{\phi}_1^2$ estimates the variance thus $\tilde{\phi}_2 = 2 \left(\frac{1}{T} \int_0^T X^2(s)ds - \tilde{\phi}_1^2 \right)$. Let us notice that here $\mathbb{E}[\tilde{\phi}_1|\phi] = \phi_1$. We denote

$$\phi = (\phi_1, \phi_2)^t, \quad \tilde{A}_T = (\tilde{\phi}_1, \tilde{\phi}_2)^t.$$

Analogously as the Proposition 3.5 of Genon-Catalot & Larédo (2016) we have the following bounds.

Proposition 4.8.7.

$$\begin{aligned} \mathbb{E}[(\tilde{\phi}_1 - \phi_1)^2|\phi] &= \phi_2^2/T + Z_T, \quad |Z_T| \leq \phi_2^3/T^2 \\ \mathbb{E}[\tilde{\phi}_2 - \phi_2|\phi] &= -2\phi_2^2/T + 2\phi_2^3(1 - e^{-T/\phi_2})/T^2 \\ \mathbb{E}[(\tilde{\phi}_2 - \phi_2)^2|\phi] &= 2\phi_2^3/T + \psi_T, \quad |\psi_T| \leq c(\phi)/T^{3/2} \end{aligned}$$

where $c(\phi)$ is bounded if $\mathbb{E}[\phi_1^4]$ and $\mathbb{E}[\phi_2^6]$ are finite.

Based on this \tilde{A}_T we could have consider the kernel estimator: $\tilde{f}_h(x) = \frac{1}{N} \sum_{j=1}^N K_h(x - \tilde{A}_{j,T})$, $f_h(x) = K_h \star f(x) = \mathbb{E}[K_h(x - \phi)]$. We obtain the following results.

Proposition 4.8.8. *Under good stationary assumptions, and if $\mathbb{E}[\phi_1^4] < \infty$ and $\mathbb{E}[\phi_2^6] < \infty$, the kernel estimator \tilde{f}_h based of the $\tilde{A}_{j,T}$, satisfies,*

$$\mathbb{E}[\|\tilde{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + 2 \max\left(\frac{1}{h_1 h_2^3}, \frac{1}{h_1^5 h_2}, \frac{1}{h_1 h_2^5}, \frac{1}{h_1^2 h_2^2}\right) \frac{C}{T^2} + \frac{\|K\|^2}{N h_1 h_2}.$$

The point is that using these other variables $\tilde{A}_{j,T}$ instead of $\hat{A}_{j,T}$ leads to a better order of the MISE because now the middle term is of order $1/T^2$.

Proof The MISE is decomposed in

$$\mathbb{E}[\|\tilde{f}_h - f\|^2] \leq 2\|f - f_h\|^2 + 2\|\mathbb{E}[\tilde{f}_h] - f_h\|^2 + \mathbb{E}[\|\tilde{f}_h - \mathbb{E}[\tilde{f}_h]\|^2].$$

We are interested in the order of the second term:

$$I = \int \left(\mathbb{E}[K_h(x - \tilde{A}_T) - K_h(x - \phi)] \right)^2 dx$$

Let us look at the difference: $K_h(x - \tilde{A}_T) - K_h(x - \phi)$ with the Taylor formula, order 2.

$$\begin{aligned} K_h(x - \tilde{A}_T) - K_h(x - \phi) &= (\tilde{\phi}_1 - \phi_1) \frac{\partial K_h}{\partial u}(x - \phi) + (\tilde{\phi}_2 - \phi_2) \frac{\partial K_h}{\partial v}(x - \phi) \\ &\quad + \int_0^1 (1-u) D^2 K_h(U(u, x)) \cdot (\phi - \tilde{A}_T)^2 du \\ &=: T_1(x) + T_2(x) + T_3(x). \end{aligned} \tag{4.8.15}$$

with $U(x, u) := (U_1(x, u), U_2(x, u)) := x - \phi + u(\phi - \tilde{A}_T)$. Or

$$\mathbb{E}[T_1(x)] = \mathbb{E}[(\tilde{\phi}_1 - \phi_1) \frac{\partial K_h}{\partial u}(x - \phi)] = \mathbb{E} \left[\frac{\partial K_h}{\partial u}(x - \phi) \mathbb{E}[(\tilde{\phi}_1 - \phi_1) | \phi] \right] = 0$$

and

$$\mathbb{E}[T_2(x)] = \mathbb{E} \left[(\tilde{\phi}_2 - \phi_2) \frac{\partial K_h}{\partial v}(x - \phi) \right] = \mathbb{E} \left[\frac{\partial K_h}{\partial v}(x - \phi) \mathbb{E}[(\tilde{\phi}_2 - \phi_2) | \phi] \right] = \mathbb{E}[T_2'(x)].$$

Thus

$$I = \int (\mathbb{E}[T_2'(x)] + \mathbb{E}[T_3(x)])^2 dx \leq \int 2\mathbb{E}[T_2'(x)]^2 + 2\mathbb{E}[T_3(x)]^2 dx \leq 2\mathbb{E} \left[\int T_2'(x)^2 dx + \int T_3(x)^2 dx \right]$$

Study of $T_2'(x)$

$$\begin{aligned} \mathbb{E} \left[\int T_2'(x)^2 dx \right] &\leq \mathbb{E} \left[\left(-\frac{2\phi_2^2}{T} + \frac{2\phi_2^3(1 - e^{-T/\phi_2})}{T^2} \right)^2 \int \left(\frac{\partial K_h}{\partial v}(x - \phi) \right)^2 dx \right] \\ &\leq \mathbb{E} \left[\left(-\frac{2\phi_2^2}{T} + \frac{2\phi_2^3(1 - e^{-T/\phi_2})}{T^2} \right)^2 \frac{1}{h_1^2 h_2^4} \int \left(\frac{\partial K}{\partial v}(x - \phi) \right)^2 dx \right] \\ &\leq \mathbb{E} \left[\left(-\frac{2\phi_2^2}{T} + \frac{2\phi_2^3(1 - e^{-T/\phi_2})}{T^2} \right)^2 \frac{h_1 h_2}{h_1^2 h_2^4} \left\| \frac{\partial K}{\partial v} \right\|^2 \right] \\ &\leq \frac{C_1(\phi)}{T^2} \frac{1}{h_1 h_2^3} \left\| \frac{\partial K}{\partial v} \right\|^2 \end{aligned}$$

Study of $T_3(x) = \int_0^1 (1-u) D^2 K_h(U(u, x)) \cdot (\phi - \tilde{A}_T)^2 du$

$$\begin{aligned} \mathbb{E}[T_3(x)] &= \mathbb{E} \left[\mathbb{E}[(\tilde{\phi}_1 - \phi_1)^2 | \phi] \int_0^1 (1-u) \frac{\partial^2 K_h}{\partial^2 u}(U(u, x)) du \right] \\ &+ \mathbb{E} \left[\mathbb{E}[(\tilde{\phi}_2 - \phi_2)^2 | \phi] \int_0^1 (1-u) \frac{\partial^2 K_h}{\partial^2 v}(U(u, x)) du \right] \\ &+ 2\mathbb{E} \left[\mathbb{E}[(\tilde{\phi}_1 - \phi_1)(\tilde{\phi}_2 - \phi_2) | \phi] \int_0^1 (1-u) \frac{\partial K_h}{\partial u \partial v}(U(u, x)) du \right] \end{aligned}$$

thus

$$\begin{aligned} \mathbb{E} \left[\int T_3(x)^2 dx \right] &= 3\mathbb{E} \left[\left(\frac{\phi_2^2}{T} + \frac{\phi_2^3}{T^2} \right)^2 \int \left(\int_0^1 (1-u) \frac{\partial^2 K_h}{\partial^2 u}(U(u, x)) du \right)^2 dx \right] \\ &+ \left(2\frac{\phi_2^3}{T} + \frac{c(\phi)}{T^{3/2}} \right)^2 \int \left(\int_0^1 (1-u) \frac{\partial^2 K_h}{\partial^2 v}(U(u, x)) du \right)^2 dx \\ &+ 6 \left(\frac{\phi_2^2}{T} + \frac{\phi_2^3}{T^2} \right) \left(2\frac{\phi_2^3}{T} + \frac{c(\phi)}{T^{3/2}} \right) \int \left(\int_0^1 (1-u) \frac{\partial^2 K_h}{\partial u \partial v}(U(u, x)) du \right)^2 dx \right]. \end{aligned}$$

Then according to Proposition 4.8.7 and $\frac{\partial^2 K_h}{\partial^2 u}(u, v) = \frac{1}{h_1^3 h_2} \frac{\partial^2 K}{\partial^2 u}(u/h_1, v/h_2)$, it yields

$$\int \mathbb{E}[T_3(x)]^2 dx \leq C_2(\phi) \max \left(\frac{\left\| \frac{\partial^2 K}{\partial^2 u} \right\|^2}{h_1^5 h_2}, \frac{\left\| \frac{\partial^2 K}{\partial^2 v} \right\|^2}{h_1 h_2^5}, \frac{\left\| \frac{\partial^2 K}{\partial u \partial v} \right\|^2}{h_1^2 h_2^2} \right) \frac{1}{T^2}.$$

which gives the announced result. \square

Chapitre 5

Estimation dans des modèles mixtes d'équations différentielles stochastiques avec le package `R` : `mixedsde`

Nous avons mis au point le package `mixedsde` pour le langage `R`. Ce chapitre présente le package pour l'estimation de densité des effets aléatoires présents dans le coefficient de dérive de deux modèles de diffusion particuliers à partir de l'observation de M trajectoires indépendantes. Les notations diffèrent des autres chapitres de la Partie II : dorénavant M désigne le nombre de trajectoires et N le nombre d'observations discrètes par trajectoire. Ce sont les notations adoptées dans le package et dans les packages que nous utilisons.

Dans ce travail on s'intéresse à deux modèles SDMEM spécifiques : $(X_j(t), t \geq 0)$, $j = 1, \dots, M$ pour $t \in [0, T]$ décrits par :

$$\begin{cases} dX_j(t) &= (\alpha_j - \beta_j X_j(t))dt + \sigma a(X_j(t))dW_j(t) \\ X_j(0) &= x_j(0) \end{cases} \quad (5.0.1)$$

où, comme précédemment, $(W_j)_{j=1, \dots, M}$ sont M processus de Wiener indépendants. Les variables (α_j, β_j) sont deux paramètres qui peuvent être aléatoires. Enfin, $\sigma a(\cdot)$ est le coefficient de diffusion avec σ une constante (inconnue) et a une fonction connue : soit $a(x) = 1$, alors on reconnaît le modèle OU soit $a(x) = \sqrt{x}$ et c'est le modèle CIR. La condition initiale $x_j(0)$ est supposée fixée et connue et peut être différente selon les trajectoires. Pour le CIR, on suppose de plus que $x_j(0) > 0$, $\sigma > 0$, $\alpha_j > \sigma^2/2$ et $\beta_j > 0$ pour s'assurer que le processus ne croise pas 0. Le processus $(X_j(t), t \in [0, T])$ représente alors la dynamique de l'unité j . On suppose que les paramètres dépendants de j sont des variables aléatoires *i.i.d.*, on les note $\phi_j \in \mathbb{R}^d$ avec $d = 1$ ou $d = 2$:

- ($d = 1$) $\phi_j = \alpha_j$ aléatoire et pour tout $j = 1, \dots, M$, $\beta_j = \beta$ fixé
- ($d = 1$) $\phi_j = \beta_j$ aléatoire et pour tout $j = 1, \dots, M$, $\alpha_j = \alpha$ fixé
- ($d = 2$) $\phi_j = {}^t(\alpha_j, \beta_j)$ aléatoire.

Les ϕ_j sont supposés indépendants des W_j .

Le package `mixedsde` a pour but d'estimer les effets aléatoires ainsi que leur densité notée f (univariée ou bivariée), à partir d'observations discrètes répétées du processus aux temps $t_0 = 0 < t_1 < \dots < t_N = T$ (pas nécessairement équidistants).

Trois procédures sont implémentées dans le package et illustrées dans ce chapitre : un estimateur à noyau unidimensionnel et bidimensionnel vu aux Chapitres 3 et 4, un estimateur paramétrique du maximum de vraisemblance étudié dans Delattre *et al.* (2013) et un estimateur paramétrique Bayésien développé par Hermann *et al.* (2015) (tous les deux supposant les effets aléatoires gaussiens). On peut définir un cadre d'application idéal pour chaque estimateur. En effet, l'approche Baysésienne est la plus adaptée lorsque le temps observation T et le nombre d'observation M sont petits. L'estimation paramétrique fréquentiste est la plus appropriée si T et M ne sont pas trop petit et qu'une densité unimodale symétrique semble bien décrire les données. Enfin l'estimation non-paramétrique est conseillée dans tous les cas pour avoir une idée de la répartition des données. Cependant si la variance des variables estimées est trop faible et/ou si T et M sont petits elle est souvent peu fiable et donc son interprétation doit rester prudente.

Ce chapitre rappelle les trois méthodes d'estimation implémentées. Il donne une vue d'ensemble des principales fonctions et fonctionnalités du package. Enfin, deux exemples issus de simulations sont détaillés. Pour finir, le jeu de données neuronales étudiées au Chapitre 3 est repris ici de manière plus approfondie.

Remarques

- Dans ce package nous avons choisi de ne pas implémenter les méthodes de sélection de fenêtres basées sur la méthode de Goldenshluger et Lepski vues aux Chapitres 3 et 4. En effet pour l'estimateur final de la densité bidimensionnelle, trois constantes doivent être calibrées ce qu'il est compliqué de faire en pratique. Afin d'être consistant nous n'avons pas gardé la méthode de sélection de fenêtre par Goldenshluger et Lepski lorsqu'il n'y a qu'une fenêtre, qui avait été implémentée lors de l'étude du Chapitre 3. Nous avons donc utilisé une autre méthode pour l'application numérique. Dans le cas d'un seul effet aléatoire nous utilisons une méthode de *cross-validation* (en utilisant dans la fonction `density` de **R** l'argument `bw='ucv'`). La méthode de *cross-validation leave-one-out* sélectionne la fenêtre en minimisant une estimation du risque quadratique intégré. En effet,

$$\mathbb{E}[\|\hat{f}_h - f\|^2] = \mathbb{E} \left[\int \hat{f}_h^2(x) dx \right] - 2\mathbb{E} \left[\int \hat{f}_h f(x) dx \right] + \int f^2(x) dx.$$

Le dernier terme ne dépendant pas du paramètre h , on cherche à minimiser la somme des deux premiers termes. Le premier peut-être approché par $I = \int \hat{f}_h^2(x) dx$ et le second par :

$$II = \frac{1}{n(n-1)h} \sum_{i \neq j} K \left(\frac{X_i - X_j}{h} \right)$$

qui vérifie, $\mathbb{E}[II] = \mathbb{E} \left[\int \hat{f}_h f(x) dx \right]$. Alors on dispose du critère de sélection :

$$\hat{h} = \operatorname{argmin}_{h \in \mathcal{H}} \{I - 2II\}.$$

Lorsque l'on a un couple de fenêtres à sélectionner, on peut reprendre la même procédure et on sélectionne $h = (h_1, h_2)$ avec le même critère en faisant attention à ce que cette fois

$$II = \frac{1}{n(n-1)h_1 h_2} \sum_{i \neq j} K_1 \left(\frac{X_i - X_j}{h_2} \right) K_1 \left(\frac{Y_i - Y_j}{h_2} \right)$$

(si $K(u, v) = K_1(u)K_2(v)$). Nous avons implémenté cette méthode et nous l'avons comparée aux résultats obtenus par la règle du pouce (*rule of thumb*). Cette règle est donnée par Sheather (2004) pour le cas d'une seule fenêtre et par Wu (1997) dans le cadre multidimensionnel. Elle consiste en une approximation asymptotique du MISE pour un noyau gaussien et une fonction densité recherchée, gaussienne. La fenêtre sélectionnée est donnée par : $\hat{h} = 1.06 \min(\hat{\sigma}, \text{IQR}/1.34)n^{-1/5}$ où $\hat{\sigma}$ est l'écart type de l'échantillon, et IQR est l'intervalle entre les quartiles de l'échantillon. Il se trouve que la règle du pouce donne souvent de meilleurs résultats et que les temps de calculs sont très fortement réduits. Pour ces raisons nous utilisons ici cette méthode empirique.

- En pratique lorsque l'on veut estimer deux effets aléatoires, l'hypothèse de stationnarité dans laquelle nous étions au chapitre précédent (4) n'est pas indispensable. En effet, cette hypothèse, qui facilitait les calculs et nous donnait des résultats de convergence \mathbb{L}^2 des estimateurs non-paramétriques, peut dégrader en pratique l'estimation des paramètres. La partie non stationnaire des trajectoires permet une meilleure estimation de la vitesse β_j de convergence exponentielle du processus vers sa valeur moyenne α_j/β_j (voir les explications du modèle dans l'introduction 1).
- De plus, il est important de noter que lors des applications sur données réelles la difficulté de l'estimation vient de la petite variance des effets aléatoires. Ce fait est vérifié sur simulation. En effet, les estimateurs $A_j(T)$ sont dégradés lorsque le bruit, représenté par la valeur de σ , est bien supérieur à l'écart type des effets aléatoires estimés, et si cette valeur est faible alors l'estimation nécessite un plus grand nombre d'observations. De plus, l'estimation bidimensionnelle est plus difficile que l'estimation unidimensionnelle comme on l'a vu précédemment. Les vitesses sont dégradées et la sélection de deux paramètres est plus compliquée.

Ce chapitre est issu de l'article Dion *et al.* (2016) écrit en collaboration avec Adeline Samson et Simone Hermann, soumis pour publication.

Sommaire

5.1	Introduction	127
5.2	Density estimation in mixed stochastic differential models	128
5.2.1	Nonparametric estimation of the random effects density	128
5.2.2	Frequentist parametric estimation approach	130
5.2.3	Bayesian parametric approach	131
5.2.4	Predictions	132
5.3	Overview of the mixedsde functions	132
5.3.1	Data	132
5.3.2	Main function	133
5.3.3	Outputs	133
5.3.4	Validation method	135
5.3.5	Prediction methods	135
5.4	Package mixedsde through simulated examples	136
5.4.1	Frequentist nonparametric estimation	137
5.4.2	Frequentist parametric estimation	138
5.4.3	Bayesian estimation	140
5.5	Package mixedsde through a real data example	142
5.5.1	Neuronal data (<code>neuronal.data</code>)	142
Appendix 1		148
Appendix 2		148

5.1 Introduction

Continuous stochastic processes are usually observed discretely in time (with equidistant time points or not) leading to times series, although their intrinsic nature is of continuous time. More and more attention have been focused on Stochastic Differential Equations (SDE). SDE provide indeed a link with a modelling approach developed in applied sciences, parameters typically being given by an interpretation. Examples of applications where SDE have been used include dynamics of thermal systems (Bacher & Madsen, 2011), solar and wind power forecasting (Iversen *et al.*, 2014), neuronal modelling (Ditlevsen & Samson, 2014), pharmacokinetic/pharmacodynamic (PK/PD) (Hansen *et al.*, 2014), crack growth (Hermann *et al.*, 2015), breast height diameter (Rupšys, 2015).

Depending on the applications, independent repeated temporal measures might be available. For examples, drug concentration of several subjects is usually observed in PK; dynamics of several neurons is measured along time; time to crack lengths can be measured repeatedly in crack growth study. Each trajectory represents the behaviour of a unit/subject. The functional form is similar for all the trajectories. Fitting the overall data simultaneously obviously improves the quality of estimation, but one has to take into account these variabilities between experiments. This is the typical framework of mixed-effects models where individual parameters for each trajectory are introduced as random variables (called random effects). Hence the random effects represent the particularity of each subject meaning each trajectory.

In this work the model of interest is thus a mixed-effects stochastic differential equation (MSDE). The `mixededsde` package has been developed to estimate the density of the random effects from the discrete observations of M independent trajectories of a MSDE. It is available from <https://cran.r-project.org/web/packages/mixededsde/index.html>.

More precisely, we focus on MSDE with linear drift. We consider M diffusion processes $(X_j(t), t \geq 0)$, $j = 1, \dots, M$ with dynamics ruled by SDE, for $t \in [0, T]$

$$\begin{cases} dX_j(t) &= (\alpha_j - \beta_j X_j(t))dt + \sigma a(X_j(t))dW_j(t) \\ X_j(0) &= x_j \end{cases} \quad (5.1.1)$$

where $(W_j)_{1 \dots j \dots M}$ are M independent Wiener processes, (α_j, β_j) are two (random) parameters, $\sigma a(X_j(\cdot))$ is the diffusion coefficient with a a known function and σ an unknown constant. The initial condition $x_j(0)$ is assumed fixed (and known) with possibly different values for each trajectory.

In the package, we restrict the models to the two famous SDE with linear drift, namely the Ornstein-Uhlenbeck model (OU) with $a(x) = 1$ and the Cox-Ingersoll-Ross model (CIR) with $a(x) = \sqrt{x}$. For the CIR model, we assume that $x_j > 0$, $\sigma > 0$, $\alpha_j > \sigma^2/2$ and $\beta_j > 0$ to ensure that the process never crosses zero.

We assume the parameters to be random variables and denote $\phi_j \in \mathbb{R}^d$. The ϕ_j 's are assumed independent and identically distributed (*i.i.d.*) and independent of the W_j 's. Three cases are considered with either $d = 1$ or $d = 2$:

- ($d = 1$) $\phi_j = \alpha_j$ random and for all $j = 1, \dots, M$, $\beta_j = \beta$ fixed
- ($d = 1$) $\phi_j = \beta_j$ random and for all $j = 1, \dots, M$, $\alpha_j = \alpha$ fixed
- ($d = 2$) $\phi_j = (\alpha_j, \beta_j)$ random

The `mixededsde` package aims at estimating the random effects ϕ_j and their distribution whose density is denoted f , from N discrete observations of the M trajectories $(X_j(t))_j$ from equation (5.1.1) at discrete times $t_0 = 0 < t_1 < \dots < t_N = T$ (not necessarily equidistant). To the best of our knowledge, this is the first package in **R** language dedicated to the estimation of MSDE.

Estimation procedures for MSDE have been proposed in the non-parametric and the parametric frameworks, with a frequentist and a Bayesian point of view. The parametric approaches assume Gaussian random effects ϕ_j . Among other references, for parametric maximum likelihood estimation, we can cite Ditlevsen & de Gaetano (2005); Picchini *et al.* (2010) (Hermite expansion of the likelihood); Delattre *et al.* (2013) (explicit integration of the Girsanov likelihood) or Delattre *et al.* (2015b) (mixture of Gaussian distributions for the random effects); for parametric Bayesian estimation, we can cite Oravec *et al.* (2009) (restricted to Ornstein-Uhlenbeck) and Hermann *et al.* (2015) (general methodology); for non-parametric estimation, we can cite Comte *et al.* (2013), Chapter 3 and 4 (kernel estimator and deconvolution estimators).

Three estimation procedures are implemented in the `mixeddsde` package: a kernel nonparametric estimator (Chapter 4), a parametric maximum likelihood estimator (Delattre *et al.*, 2013) and a parametric Bayesian estimator (Hermann *et al.*, 2015). The parametric frequentist and Bayesian approaches assume the random effects Gaussian. The Bayesian approach seems the most appropriate method for a small time of observation T and a small number of trajectories M . The nonparametric approach can be used when no prior idea on the density is available and when T and M are both large enough. Finally, the parametric frequentist estimation can be used when data are symmetric with a large number of discrete observations.

This chapter reviews in Section 5.2 the three estimation methods. An overview of the `mixeddsde` package is given in Section 5.3 through a description of the main functions and of other related companion functions. The practical use of this package is illustrated in Section 5.4 on simulated data and in Section 5.5 a real dataset in neuronal modelling.

5.2 Density estimation in mixed stochastic differential models

We briefly recall the methodology of the three estimators implemented in the `mixeddsde` package. We start with the nonparametric approach, then the frequentist parametric Gaussian method and finally the Bayesian parametric Gaussian method.

5.2.1 Nonparametric estimation of the random effects density

The first step of the nonparametric approach is to estimate the random effects. The idea is to maximize the likelihood of the process X_j^φ solution of the stochastic differential equation with fixed φ . Assuming continuous observations of $(X_j(t), 0 \leq t \leq T)$, the likelihood function is obtained with the Girsanov formula:

$$\ell_T(\varphi) = \exp \left(\int_0^T \frac{\alpha - \beta X_j^\varphi(s)}{\sigma^2 a^2(X_j^\varphi(s))} dX_j(s) - \frac{1}{2} \int_0^T \frac{(\alpha - \beta X_j^\varphi(s))^2}{\sigma^2 a^2(X_j^\varphi(s))} ds \right).$$

Maximizing the likelihood yields to the following estimator of ϕ_j

$$A_j := V_j^{-1} U_j \tag{5.2.1}$$

where U_j and V_j are the two sufficient statistics of the model. They are explicit depending on the form of the random effects:

- α_j random and β known

$$U_j := \int_0^T \frac{1}{\sigma^2 a^2(X_j(s))} dX_j(s) + \beta \int_0^T \frac{X_j(s)}{\sigma^2 a^2(X_j(s))} ds, \quad V_j := \int_0^T \frac{1}{\sigma^2 a^2(X_j(s))} ds$$

– β_j random and α known

$$U_j := - \int_0^T \frac{X_j(s)}{\sigma^2 a^2(X_j(s))} dX_j(s) + \alpha \int_0^T \frac{X_j(s)}{\sigma^2 a^2(X_j(s))} ds, \quad V_j := \int_0^T \frac{X_j(s)^2}{\sigma^2 a^2(X_j(s))} ds$$

– (α_j, β_j) random, denote $b(x) = (1, -x)^t$, with u^t the transposition of vector u . Note that in this case U is a column vector with size 2×1 and V a 2×2 symmetric matrix:

$$U_j := \int_0^T \frac{b}{\sigma^2}(X_j(s)) dX_j(s), \quad V_j := \int_0^T \frac{bb^t}{\sigma^2}(X_j(s)) ds = \begin{pmatrix} \int_0^T \frac{b_1^2}{\sigma^2}(X_j(s)) ds & \int_0^T \frac{b_1 b_2}{\sigma^2}(X_j(s)) ds \\ \int_0^T \frac{b_1 b_2}{\sigma^2}(X_j(s)) ds & \int_0^T \frac{b_2^2}{\sigma^2}(X_j(s)) ds \end{pmatrix}$$

See Appendix 1 for details on U and V .

Truncated versions of this estimator have been introduced for theoretical reasons. In the bidimensional case $\phi_j = (\alpha_j, \beta_j)$, the Chapter 4 proposes the following estimator

$$\widehat{A}_j := A_j \mathbb{1}_{B_j}, \quad B_j := \{V_j \geq \kappa \sqrt{T} I_2\} = \{\min(\lambda_{1,j}, \lambda_{2,j}) \geq \kappa \sqrt{T}\} \quad (5.2.2)$$

with I_2 the 2×2 identity matrix and $\lambda_{i,j}, i = 1, 2$ the two eigenvalues of the symmetric non negative matrix V_j , and κ a numerical constant calibrated in practice. In the one-dimensional case $\phi_j = \beta_j$ with $\alpha = 0$, Genon-Catalot & Larédo (2016) propose

$$\widehat{A}_j := A_j \mathbb{1}_{V_j \geq \kappa \sqrt{T}} \quad (5.2.3)$$

with κ a numerical constant calibrated in practice.

Based on these estimators of the ϕ_j 's, we can proceed to the second step, the estimation of their density f . Several nonparametric estimators of f have been proposed (see Comte *et al.*, 2013, for example). In the package `mixeddsde`, we focus on the kernel estimator of f . Let us introduce the kernel function $K : \mathbb{R}^d \rightarrow \mathbb{R}$, with $d = 1, 2$ depending on the dimension of ϕ_j . We assume K to be a \mathcal{C}^2 function satisfying

$$\int K(u) du = 1, \quad \|K\|^2 = \int K^2(u) du < +\infty, \quad \int (\nabla K(u))^2 du < +\infty$$

(with ∇K is the gradient of K). A bandwidth $h \in (\mathbb{R}^+)^d$, for $d = 1, 2$, is used to define the function

$$K_h(x) = \frac{1}{h} K\left(\frac{x}{h}\right), \quad x \in \mathbb{R}^d.$$

Note that in the bidimensional case, $h = (h_1, h_2)$ and the two marginal bandwidths are different. The nonparametric estimator of the density f of ϕ_j is

$$\widehat{f}_h(x) = \frac{1}{M} \sum_{j=1}^M K_h(x - A_j). \quad (5.2.4)$$

and estimator $\widehat{f}_h(x) = \frac{1}{M} \sum_{j=1}^M K_h(x - \widehat{A}_j)$ is computed when it is different. In the `mixeddsde` package, Gaussian kernel estimators are implemented with the **R**-functions `density` (available in package `stats`) when $d = 1$ and `kde2d` (available on package `MASS`) when $d = 2$ with an automatic selection of the bandwidth h . Note that when there is only one random effect, the bandwidth is selected by cross-validation with the argument `bw="ucv"`, or as the default value given by the rule-of-thumb if the chosen bandwidth is too small. Note that the estimator is unstable for small variance of the random effects.

It is important to notice that the two random effects are not assumed independent. Furthermore, the computation of $A_j = V_j^{-1}U_j$ does not require the knowledge of σ^2 as it appears both in U_j and V_j . It requires however the evaluation of the two continuous integrals in U_j and V_j while observing the trajectories (X_j) at discrete times (t_0, t_1, \dots, t_N) . For $\Delta_k = t_{k+1} - t_k$, $k = 0, \dots, N-1$, the non-stochastic integrals $\int_0^T g(X_j(s))ds$ for any function g are approximated by

$$\int_0^T g(X_j(s))ds \approx \sum_{k=0}^{N-1} g(X_j(t_k))\Delta_k.$$

For the stochastic integrals, we use the following simple discretization

$$\int_0^T g(X_j(s))dX_j(s) \approx \sum_{k=0}^{N-1} g(X_j(t_k))(X_j(t_{k+1}) - (X_j(t_k)))\Delta_k.$$

5.2.2 Frequentist parametric estimation approach

In this section and the following one, we assume that the random parameters ϕ_j are Gaussian, and independent when $d = 2$:

- when $d = 1$, $\phi_j \sim \mathcal{N}(\mu, \omega^2)$ with $\mu \in \mathbb{R}$
- when $d = 2$, $\phi_j \sim \mathcal{N}(\mu, \Omega)$ with $\mu \in \mathbb{R}^2$ and a diagonal covariance matrix $\Omega = \text{diag}(\omega_1^2, \omega_2^2)$.

We detail the case $d = 2$. The objective is to estimate by maximum likelihood estimator the parameters $\theta := (\mu, \Omega)$. We define the likelihood function assuming first that the trajectories are continuously observed, similarly to the nonparametric approach (Section 5.2.1). Thanks to the Girsanov formula, the likelihood function of the j^{th} trajectory X_j is

$$L(X_j, \theta) = \frac{1}{\sqrt{\det(I_d + \Omega V_j)}} \exp \left[-\frac{1}{2}(\mu - V_j^{-1}U_j)'R_j^{-1}(\mu - V_j^{-1}U_j) \right] \exp \left(\frac{1}{2}U_j'V_j^{-1}U_j \right)$$

with $R_j^{-1} = (I_d + V_j\Omega)^{-1}V_j$ and I_d is the $d \times d$ identity matrix (see Delattre *et al.* (2013)). We define the maximum likelihood estimator $\hat{\theta} := (\hat{\mu}, \hat{\Omega})$ as

$$\hat{\theta} = \arg \max_{\theta} \prod_{j=1}^M L(X_j, \theta)$$

This estimator is not explicit. In the `mixedSDE` package, the function `optim` is used to maximize numerically the likelihood. Initialization of `optim` is obtained by computing the mean and the variance of the estimators A_j of the random parameters (see equation (5.2.1)). Sufficient statistics U_j and V_j are discretized as explained in Section 5.2.1.

Note that this parametric approach requires the knowledge of σ^2 to compute the sufficient statistics U_j and V_j because V_j appears alone in R_j . We plug the following estimator of σ^2

$$\widehat{\sigma^2} = \frac{1}{M} \sum_{j=1}^M \left(\frac{1}{N} \sum_{k=0}^{N-1} \frac{((X_j(t_{k+1}) - X_j(t_k)))^2}{\Delta_k a^2(X_j(t_k))} \right). \tag{5.2.5}$$

Selection of (non-nested) models can be performed with the BIC criteria, defined by $-2\ell_T(\hat{\theta}) + 2 \log(M)$ for model with one random effect and $-2\ell_T(\hat{\theta}) + 4 \log(M)$ with two random effects and the AIC criteria defined by $-2\ell_T(\hat{\theta}) + 2$ for one random effect and $-2\ell_T(\hat{\theta}) + 4$ for two random effects. These asymptotic criteria indicate the trade-off between maximizing fit and minimizing model complexity (see Delattre *et al.* (2014)).

5.2.3 Bayesian parametric approach

For the Bayesian approach we assume similarly to the frequentist parametric estimation method a Gaussian distribution for ϕ_j , with a diagonal covariance matrix $\Omega = \text{diag}(\omega_1^2, \omega_2^2)$. In this method, we estimate in the same time the diffusion coefficient σ . Let us denote $\mathbf{X}_{1:M} = (X_j(t_k))_{j=1, \dots, M, k=1, \dots, N}$ in the following. The parameters of interest are thus $\theta = (\mu, \Omega, \sigma)$ and we want to estimate their posterior distribution $p(\theta|\mathbf{X}_{1:M})$.

We now introduce prior distributions implemented in `mixeddsde` package for the parameters θ :

$$\begin{aligned} \mu &\sim \mathcal{N}(m, V), \quad V = \text{diag}(v) \\ \omega_i^2 &\sim \text{IG}(\alpha_{\omega, i}, \beta_{\omega, i}), \quad i = 1, 2 \\ \sigma^2 &\sim \text{IG}(\alpha_\sigma, \beta_\sigma), \end{aligned}$$

where IG is the inverse Gamma distribution which is conjugate to the normal likelihood and $m, V, \alpha_{\omega, i}, \beta_{\omega, i}, \alpha_\sigma, \beta_\sigma$ are hyperparameters fixed by the user. The case of only one random effect is nested by setting ω_1^2 or ω_2^2 equal to zero.

The aim is to calculate the posterior distribution $p(\theta|\mathbf{X}_{1:M})$ which is not explicit for the whole vector of parameters. Therefore, we simulate it through a Gibbs sampler (see e.g. Robert & Casella, 2004). Here, we have a true transition density of both processes that is used for the likelihood, see Iacus (2008). For a general hierarchical diffusion approach, based on the Euler approximation, see Hermann *et al.* (2015).

Analogically to the frequentist approach, there is a first step: sample from the full conditional posterior of the random effects $p(\phi_j|(X_j(t_k))_{k=1, \dots, N}, \theta), j = 1, \dots, M$. This is done by a Metropolis Hastings (MH) algorithm. The second step is the estimation of the hierarchical parameters μ and Ω . Full conditional posteriors $p(\mu|\phi_1, \dots, \phi_M, \Omega)$ (resp. $p(\Omega|\phi_1, \dots, \phi_M, \mu)$) are Gaussian (resp. inverse Gamma) and can, for example, be found in Hermann *et al.* (2015).

The last step of the Gibbs sampler is sampling from the full conditional posterior of σ^2 . For the CIR model, this is also conducted by a MH step. For the OU model, the inverse Gamma distribution is conjugate to the normal likelihood. The full conditional posterior distribution is given by

$$\sigma^2|\mathbf{X}_{1:M}, \phi_1, \dots, \phi_M \sim \text{IG}\left(\alpha_\sigma + \frac{MN}{2}, \beta_\sigma + \frac{1}{2} \sum_{j=1}^M \sum_{k=1}^N \frac{\beta_j}{1 - e^{-2\beta_j \Delta_k}} \left(X_j(t_k) - \frac{\alpha_j}{\beta_j} - \left(X_j(t_{k-1}) - \frac{\alpha_j}{\beta_j}\right) e^{-\beta_j \Delta_k}\right)^2\right).$$

In the case of one random effect, there is one additional Gibbs sampler step for the fixed effect, that is also conducted through a MH algorithm.

In the package, the starting values for the Gibbs sampler are set equal to the mean of the prior distributions. In all the MH algorithms, one each has to choose a proposal density. In the package `mixeddsde`, we use a normal density for all location parameters with mean equal to the last chain iteration and a proposal variance that has to be chosen. For the CIR model, the proposal distribution for σ^2 is chosen by $\sqrt{\sigma^2} \sim \mathcal{N}(\sqrt{\sigma_{\text{prev}}^2}, \text{variance})$ where σ_{prev}^2 is the previous value of σ^2 . The remaining question is how to choose the suitable proposal variance. This variance controls the chain dependence and the acceptance rate. If the variance is small, the acceptance rate is large and the chains gets very dependent. If the proposal variance is large, only few candidates are accepted with the advantage of weakly dependent chains. This problem is solved in the package with an adaptive Metropolis-within Gibbs algorithm (Rosenthal, 2011)

using the proposal distribution $\mathcal{N}(0, e^{2l})$ where l the logarithm of the standard deviation of the increment. This parameter is chosen so that the acceptance rate is approximately 0.44 which is proposed to be optimal in the Metropolis-within Gibbs sampler. It is proposed to add/subtract an adoption amount $\delta(n) = \min(0.1, n^{-1/2})$ to/from t after every 50th iteration and adapt the proposal variance if the acceptance rate is smaller than 0.3 or larger than 0.6.

5.2.4 Predictions

In many cases, one is not only interested in parameter estimation but also in the prediction for future observations. The first step is the prediction of a future random effect ϕ_{pred} . The simulation of a new random effect is direct for the frequentist parametric approach sampling from $\mathcal{N}(\hat{\mu}, \hat{\omega}^2)$. For the nonparametric approach, first note that \hat{f}_h is an estimator given on a discrete grid $\{x_1, \dots, x_n\}$, i.e. a vector of corresponding $\{p_1, \dots, p_n\}$ after normalisation. Simulating from the estimator \hat{f}_h can therefore be performed simulating a discrete variable from vector $\{x_1, \dots, x_n\}$ with (normalized) probabilities $\{p_1, \dots, p_n\}$.

For the Bayesian approach, a new ϕ_{pred} is sampled from the predictive distribution $p(\phi_{\text{pred}}|\mathbf{X}_{1:M}) = \int p(\phi_{\text{pred}}|\mu, \Omega)p(\mu, \Omega|\mathbf{X}_{1:M}) d(\mu, \Omega)$ where the posterior of μ and Ω is approximated by the results of the Gibbs sampler. This distribution is not explicit, we propose to sample over a grid through inversion method, equal to the nonparametric case.

Given a new random effect ϕ_{pred} , we are able to simulate predictive trajectories. This is performed using the transition density $p(X(t_k)|X(t_{k-1}), \phi_{\text{pred}}, \sigma^2)$ for the frequentist approach. And the starting points of the process x_j are the observed ones. For the Bayesian approach, we implemented two prediction settings. Firstly, analogously to the frequentist approach a new trajectory is simulated using the transition density $p(X(t_k)|X(t_{k-1}), \phi_{\text{pred}}, \sigma^2)$ where ϕ_{pred} is sampled from the MCMC posterior distribution $p(\phi|\mathbf{X}_{1:M})$. Secondly, we can calculate the predictive distribution

$$p(X(t_i)|\mathbf{X}_{1:M}) = \int p(X(t_i)|\phi_{\text{pred}}, \sigma^2)p(\phi_{\text{pred}}, \sigma^2|\mathbf{X}_{1:M}) d(\phi_{\text{pred}}, \sigma^2)$$

in each time point. We can then calculate only the quantiles for a prediction interval or to draw directly samples from the predictive distribution. For this predictive distribution, we take the starting point $x_j = x_0$ to be the same for all series. If the starting points would vary, this is an additional random effect whose density has to be estimated. This is not implemented in the estimation procedure and will, therefore, left out for the prediction.

It is then interesting to compare the new trajectories to the real ones and if the number of new trajectories is large enough, to compute an empirical confidence interval.

5.3 Overview of the mixedside functions

This Section presents an overview of the functions implemented in the package, illustrations of the code are given in Section 5.4.

5.3.1 Data

Data is a matrix \mathbf{X} of size $M \times N$ for M trajectories with N time points, not necessarily equidistant but similar for the M trajectories. The vector of times is a vector `times` of length N . Real datasets are available on the package, and detailed on Section 5.5.

To lead a simulation study, the function `mixedside.sim` allows to generate a list with a $M \times N$ matrix \mathbf{X} of M trajectories on the interval $[0, T]$ with N equidistant points (default value

Table 5.1 – Summary of the different methods for the two S4-classes `Freq.fit` and `Bayes.fit` resulting of the function `mixededsde`

class	Freq.fit	Bayes.fit
method	out	out
method	plot	plot
method	–	plot.compare
method	print	print
method	summary	summary
method	pred	pred
method	valid	valid

100) and a vector `times` with the equidistant times. This function leans on function `sde.sim` available via package `sde` from the web address <http://cran.r-project.org/package=sde>, to simulate SDE. One has to choose: `model` either OU or CIR; `random` that fixes the position and the number of random effects: `random = 1` for α_j random, `random = 2` for β_j random or `random = c(1,2)` for α_j and β_j random; σ the diffusion coefficient; `invariant`, default value 0 means that X_0 is 0 (default) or fixed by the user, value 1 means that X_0 is generated from the invariant distribution (see details in the package documentation); `density.phi` to choose the distribution of the random effect (see package documentations).

5.3.2 Main function

Main function is `mixededsde.fit` producing estimation of the random effects and their common density. Inputs of `mixededsde.fit` are

- `X` a $M \times N$ matrix containing the trajectories by rows
- `times` the vector of observations times
- `model` the chosen model either OU or CIR
- `random` that fixes the position and the number of random effects: `random = 1` for α_j random, `random = 2` for β_j random or `random = c(1,2)` for α_j and β_j random
- `fixed` the value of the fixed effect β (resp. α) when `random = 1` (resp. `random = 2`), default 0 (only for the frequentist approaches)
- `estim.method` the estimation method: `nonparam` (see Section 5.2.1), `paramML` (see Section 5.2.2) or `paramBayes` (see Section 5.2.3)
- `gridf` the x-axis grid on which the random effect distribution is computed: we recommend a fine grid with at least 200 points, default value is a sequence of length 500 starting in $0.8 \times \min_j \hat{\phi}_j$ and ending in $1.2 \times \max_j \hat{\phi}_j$ (only for the frequentist approaches)
- `prior` the list of prior parameters `m`, `v`, `alpha.omega`, `beta.omega`, `alpha.sigma`, `beta.sigma` for `paramBayes` method: default values are calculated based on the estimations $(A_j)_j$ for the first $\min(3, \lceil M \cdot 0.1 \rceil)$ series and main estimation is only made with the remaining $\lfloor M \cdot 0.9 \rfloor$ (only for the Bayesian approach)
- `nMCMC` length of the Markov chain for `paramBayes` method (only for the Bayesian approach)

In the following we describe the related methods, proposed in the package, they are summarised in Table 5.1.

5.3.3 Outputs

Output of `mixededsde.fit` is a S4 class called `Freq.fit` for the frequentist approaches and `Bayes.fit` for the Bayesian approach. Results of the estimation procedure are available as

a list applying function `out` to the `Freq.fit` (resp. `Bayes.fit`) object.

Elements of `Freq.fit` are:

- `sigma2` estimation $\widehat{\sigma^2}$ of the diffusion coefficient
- `estimphi` estimation $(A_j)_j$ of the random effects
- `estimphi.trunc` truncated estimation $(\widehat{A}_j)_j$ of the random effects
- `gridf` the x-axis grid on which the random effect distribution is computed
- `estimf` estimation of the density of the random effects (for both `paramML` and `nonparam` methods).
- `cutoff` binary M -vector of binary values indicated the truncated trajectories, default `FALSE` when no truncation
- `estimf.trunc` truncated estimation of the density of the random effects
- `mu` estimation of Gaussian mean of the random effects (only for `paramML` method)
- `omega` estimation of Gaussian variance matrix of the random effects (only for `paramML` method)
- `aic` and `bic` AIC and BIC criteria (only for `paramML` method)
- `index` index of trajectories used for the estimation, excluded are trajectories with $V_j = 0$ or $V_j = +\infty$ (one random effect) or $\det V_j = +\infty$ (two random effects), trajectories containing negative values for CIR model,

Elements of `Bayes.fit` are:

- `sigma2` trace of the Markov chain simulated from the posterior of σ^2
- `mu` trace of the Markov chain simulated from the posterior of μ
- `omega` trace of the Markov chain simulated from the posterior of ω^2
- `alpha` trace of the Markov chain simulated from the posterior of α_j , $\mathbf{nMCMC} \times M$ matrix if α is random effect, $\mathbf{nMCMC} \times 1$ otherwise
- `beta` trace of the Markov chain simulated from the posterior of β_j , $\mathbf{nMCMC} \times M$ matrix if β is random effect, $\mathbf{nMCMC} \times 1$ otherwise
- `burnIn` a proposal for the burn-in phase
- `thinning` a proposal for the thin rate
- `ind.4.prior` the indices used for the prior parameter calculation, $M+1$ if prior parameters were specified.

Outputs `burnIn` and `thinning` are only proposals for a burn-in phase and a thin rate. The proposed `burnIn` is calculated by dividing the Markov chains into 10 blocks and calculate the 95% credibility intervals and the respective mean. Starting in the first one, the block is taken as burn-in as long as the mean of the current block is not in the credibility interval of the following block or vice versa. The thinning rate is proposed by the first lag which leads to a chain autocorrelation of less than 80%. It is not easy to automate these choices, so it is highly recommended by the authors to verify the chains manually.

Command `plot` applied to a `Freq.fit` object produces a frequencies histogram of $(A_j(T))_j$ (one or two according to the number of random effects) with the estimated density (red curve) and the truncated estimator if available (dotted red curve) and a quantile-quantile graph with the quantiles of the A_j 's versus the quantiles of a normal sample of the same length, with the same empirical mean and standard deviation. This illustrates the normality of the sample. Applying this function to the nonparametric results indicates if the Gaussian assumption of the parametric approach is appropriate. When `plot` is applied to a `Bayes.fit` object, one can choose four different options, named `style`. The default value is `chains`, it plots the Markov chains for the different parameter values. `acf` leads to the corresponding autocorrelation functions, `density` to the approximated densities for each parameter and `cred.int` leads to the credibility intervals of the random parameters with the input parameter `level` with default 0.05. For all options,

with the input parameter `reduced = TRUE`, the burn-in period is excluded and a thinning rate is taken, default is `FALSE`. There is also a possibility to include the prior means in the plots by lines with `plot.priorMean = TRUE`, default is `FALSE`.

In the Bayesian estimation the influence of prior parameters is interesting, thus for the `Bayes.fit` object, there is a second plot method, named `plot.compare` where three estimation objects can be compared. For reasons of clarity, only the densities are compared, with the default `reduced = TRUE`. Here, there is also a possibility to include `true.values`, a list of the true parameters for the comparison in a simulation example.

Command `summary` applied to a `Freq.fit` object computes the kurtosis and the skewness of the distribution, $\widehat{\sigma}^2$, the empirical mean and standard deviation computed from the estimators $(A_j)_j$, $\hat{\mu}$, $\hat{\Omega}$, AIC, BIC criteria for the frequentist MLE method. When applied to a `Bayes.fit` object, it computes means and credibility interval (default level 95%) for each parameter $(\mu, \Omega, \sigma, \alpha, \beta)$. Here, there is also a possibility to choose the burn-in and the thinning rate manually by the input parameters `burnIn` and `thinning`.

Command `print` applied to a `Freq.fit` object returns the use or not of the cut-off and the vector of excluded trajectories. When applied to a `Bayes.fit` object, it returns the acceptance rates of the MCMC procedure.

5.3.4 Validation method

Validation of a mixed model, obtained with function `valid`, is an individual validation. Indeed, the validation of trajectory j is obtained comparing it to M new trajectories simulated with parameters (α, β) fixed to the estimator A_j (or \hat{A}_j) in the frequentist approaches and to the posterior means in the Bayesian approach. Inputs of the function are

- `Freq.fit` object
- `Xtrue` matrix of the true trajectories (only `Freq.fit` class)
- `times` vector of observation times (only `Freq.fit` class)
- `plot.valid` 1 to generate a figure
- `numj` a specific individual trajectory to validate (default all trajectories)
- `Mrep` number of simulated trajectories (default value 100)

Outputs are two plots. Left: plot of the `Mrep` new trajectories (black) and the trajectory `numj` (in grey/red). Right: quantile-quantile plot of the quantiles of a uniform distribution and the N quantiles obtained comparing each observation $X_{\text{numj}}(t_k)$ with the `Mrep` simulated values $(X_{\text{numj}}^1(t_k), \dots, X_{\text{numj}}^{M_{\text{rep}}}(t_k))$, for $k = 1, \dots, N$. If `numj` is not given, the validation is done for all the trajectories and if `plot.valid=1`, a number of trajectory is chosen randomly for the graph.

This is an empiric method. The recent work Kuelbs & Zinn (2015) on depth and quantile regions for stochastic processes (see for example Zuo & Serfling (2000) for depth functions definitions) should provide the theoretical context for a more extensive study. This could be done in further works.

5.3.5 Prediction methods

Prediction (see Section 5.2.4) is implemented in function `pred`. Applied to a `Freq.fit` object it returns a list with predicted random effects `phipred`, predicted trajectories `Xpred` and indexes of the corresponding true trajectories `indexpred` (see Section 5.2.4 for details of simulation). If `plot.pred = 1` (default) three plots are produced. Left predicted random effects versus estimated random effects. Middle: true trajectories. Right predicted trajectories and their empirical 95% prediction intervals (default value `level=0.05`).

Function `pred` applied to a `Bayes.fit` object returns a S4 class object `Bayes.pred`. The first element of this class is `Xpred`, which depends on the input parameters. In the case of `trajectories = TRUE`, matrix `Xpred` contains the M drawn trajectories by rows (see first method described for the Bayesian approach in Section 5.2.4). If `plot.pred=TRUE`, the observed trajectories are compared with the M drawn predictive trajectories.

Default is `trajectories = FALSE` which leads to the calculation of the predictive distribution explained in Section 5.2.4. If `only.interval = TRUE` is specified (default), only the quantiles for the $1-\alpha$ (α is input parameter `level`) prediction interval are calculated, stored in `qu.l` and `qu.u`, and no `Xpred` is given out. If `only.interval = FALSE`, additionally `Xpred` contains `sample.length` (default 500) samples from the predictive distribution in each time point of the observations (except the first). In both cases, with `plot.pred=TRUE`, two figures are produced. On the left side, the data trajectories are compared with the prediction intervals and on the right side, the coverage rate is depicted which is stored in entry `coverage.rate`, namely the amount of series covered by the prediction intervals for each time point. The last class entry `estim` stores the results from the `Bayes.fit` object in a list. Other input parameters are `burnIn` and `thinning` which allow for the choice of other burn-in phase and thinning rate than proposed in the `Bayes.fit` object.

For the `Bayes.pred` class object, two plot methods are available. `plot()` repeats the figures that are created with the `plot.pred=TRUE` command in the `pred` method. `plot.compare()` compares up to three `Bayes.pred` objects, where in a first figure the prediction intervals are presented in colors black, red and green and the observed data series in grey and in a second figure the corresponding coverage rates are compared. With the input parameter `names` a vector of characters to be written in a legend can be indicated.

5.4 Package `mixedside` through simulated examples

In this part two simulated examples are given to illustrate the strengths of each proposed method. Two datasets are simulated according to:

1. CIR model with one non-Gaussian random effect $\beta_j \sim \Gamma(1.8, 0.8)$, $fixed = 1$, $T = 50$, $M = 200$, $N = 1000$:

```
model1 <- "CIR"; random <- 2; fixed1 <- 1; sigma1 <- 0.1 ; M1 <- 200;
  T1 <- 50; N1 <- 1000; X01 <- 1; density.phi1 <- "gamma";
  param1 <- c(1.8, 0.8); t01 <- 0
simu1 <- mixedside.sim(M1, T1, N1, model1, random1, fixed1, density.phi1,
  param1, sigma1, t01, X01, op.plot=0)
X1<- simu1$X; phi1 <- simu1$phi; times1 <-simu1$times
```

2. OU model with one Gaussian random effect $\alpha_j \sim \mathcal{N}(3, 0.5^2)$, $\beta_j = 5$, $T = 1$, $M = 50$, $N = 500$:

```
model2 <- "OU"; random2 <- 1; sigma2 <- 0.1; fixed2 <- 5; M2 <- 50;
  T2 <- 1; N2 <- 500;X02 <- 0; density.phi2 <- "normal";
  param2 <- c(3, 0.5); t02<-0;
simu2 <- mixedside.sim(M2, T2, N2, model2, random2, fixed2, density.phi2,
  param2, sigma2, t02, X02, op.plot=0)
X2 <- simu2$X; phi2 <- simu2$phi; times2 <- simu2$times
```

Example 1 has non Gaussian random effect, the nonparametric method is the most appropriate approach. Example 2 has T small and Gaussian random effect, nonparametric method is therefore not the most appropriate approach. Parametric methods should performed well with a preference to the Bayesian as the number of trajectories $M2 = 50$ is not large.

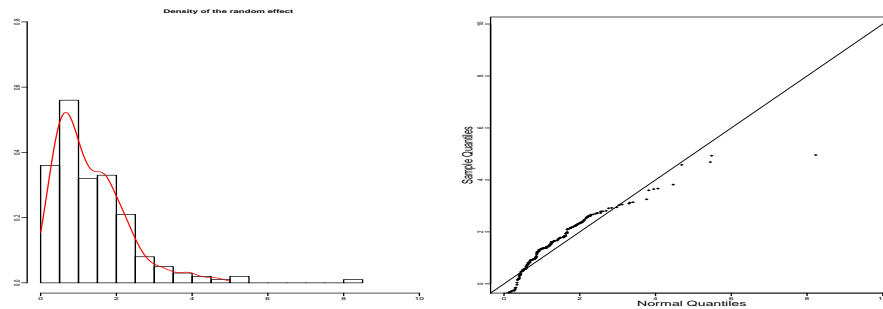


Figure 5.1 – Simulated example 1 (CIR with one Gamma random effect), nonparametric estimation. Left: histogram of estimated random effects (A_j) and nonparametric estimation of f . Right: qqplot of (A_j) versus a normal sample (true distribution is Gamma).

5.4.1 Frequentist nonparametric estimation

We illustrate nonparametric estimation on Example 1. Code for the nonparametric estimation is

```
estim.method <- 'nonparam'
estim_nonparam <- mixedside.fit(times1, X1, model1, random1, fixed1, estim.method)
outputsNP <- out(estim_nonparam) # stores the results in a list
```

Summary function provides:

```
summary(estim_nonparam, newwindow=FALSE)
      [,1]      [,2]
[1,] "kurtosis" "11.0008360676508"
[2,] "skewness" "2.16789214842904"
      [,1]      [,2]
[1,] "sigma"    "0.100164891397244"
[2,] "empiric mean" "1.38241196209623"
[3,] "empiric sd"  "1.10890963312551"
```

As expected kurtosis is larger than 3 and skewness is positive which means that the distribution is right-tail. Figure 5.1 is provided by

```
plot(estim_nonparam)
```

Nonparametric estimation fits well the histogram of (A_j) (left plot) and we see that the random effects are non-Gaussian (right plot). Because we are working on simulated data, we can compare the estimations with the true random effects and the true f :

```
# comparison of the true f and its estimation
gridf1 <- outputsNP$gridf
f1 <- dgamma(gridf1, param1[1], param1[2]) # true density function
fhat <- outputsNP$estimf # nonparametric estimated density function
plot(gridf1, f1, type='l', lwd=2, xlab='', ylab='') # plot of the true density
lines(gridf1, fhat, col='red') # plot of the nonparametric estimated density
# comparison of the true random effects and their estimations
phi1 <- outputsNP$estimphi # estimated random effects
plot(phi1, phi1, type = 'p', pch = 18, xlab='', ylab='') # true vs estimated
abline(0, 1)
```

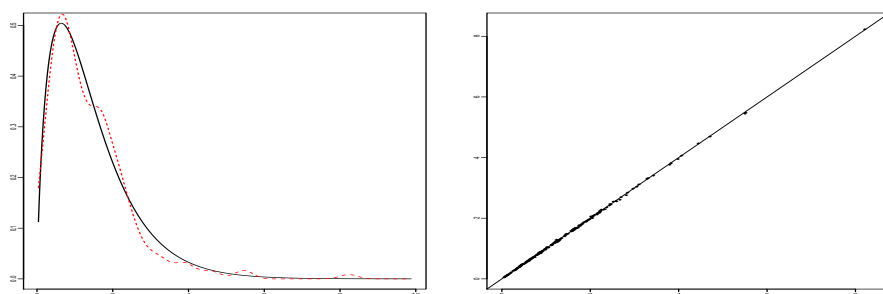


Figure 5.2 – Simulated example 1 (CIR with one Gamma random effect), nonparametric estimation, comparison to the truth. Left: estimation \hat{f} (dotted line) and true density f (plain line). Right: Estimated random effects A_j versus true random effects ϕ_j

It yields to Figure 5.2. On the left plot, the estimated density (dotted curve) is very close to the true density f (plain line). The right plot shows that A_j is a good estimation of ϕ_j . This confirms that the nonparametric approach performs well for this settings. Validation of the MSDE is produced by function `valid`. The two graphs on the right of Figure 5.5 are obtained by

```
choicel <- floor(runif(1,1,M))
validationCIR <- valid(estim_nonparam, X1 = X1, times1 = times1, numj = choicel)
```

Prediction are obtained with `pred` and similar Figure 5.6 (not shown) can be obtained with

```
predNPCIR <- pred(estim_nonparam, X1 = X1, times1 = times1)
```

5.4.2 Frequentist parametric estimation

We present the parametric estimation on Example 2. The code is

```
#- parametric estimation
estim.method<-'paramML';
estim_param <- mixedsde.fit(times2, X2 = X2, model2 = model2, random2 = random2,
    fixed = fixed, estim.method = 'paramML' )
outputsP <- out(estim_param)      # stores the results in a list
```

Summary function provides:

```
summary(estim_param)
  [,1] [,2]
[1,] "BIC" "-3727.5433017198"
[2,] "AIC" "-3733.36734773066"
  [,1]      [,2]
[1,] "kurtosis" "2.64251059224136"
[2,] "skewness" "0.207314458035785"
  [,1]      [,2]
[1,] "sigma"      "0.107968951754119"
[2,] "empiric mean" "2.95224502520867"
[3,] "MLE mean"    "2.95250185124178"
[4,] "empiric sd"  "0.585479181167053"
[5,] "MLE sd"     "0.570077670106673"
```

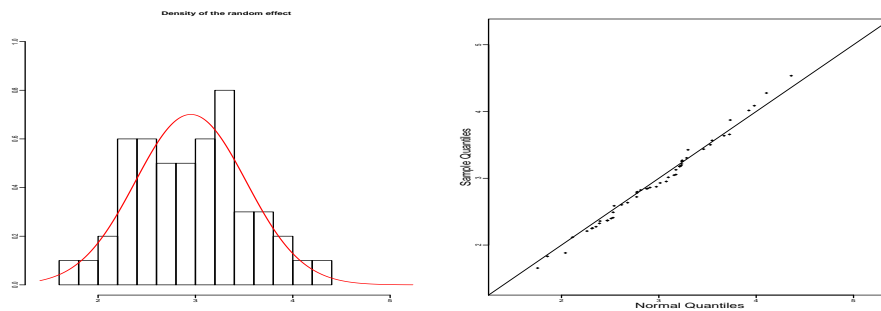


Figure 5.3 – Simulated example 2 (OU with one Gaussian random effect) frequentist parametric estimation. Left: histogram of the (A_j) and Gaussian parametric estimation of f . Right parametric qqplot of (A_j) versus a normal sample.

Kurtosis is, as expected, close to 3 and skewness close to 0. The diffusion parameter σ is well estimated (true value 0.1). Empirical mean and standard deviations are very close to MLE (estimator of the mean is the same in that case) and close to the real ones (3, 0.5). Then, Figure 5.3 (left and right) is provided by

```
plot(estim_param)
```

The small number of observations makes the frequentist estimation harder, nevertheless here, the histogram seems pretty well fitted by the parametrically estimated density. Because we are working on simulated data, we can compare the estimations with the true random effects and the true f :

```
# comparison of the true f and its estimation
gridf2 <- outputsP$gridf
f2 <- dnorm(gridf, param[1], param[2]) # true density
fhat_param <- outputsP$estimf # parametric estimated density
plot(gridf2, f2, type = 'l', lwd = 2, xlab = '', ylab = '')
# plot of the true density
lines(gridf2, fhat_param, col='red', lty = 2, lwd = 2)
# plot of the estimated density
# comparison of the true random effects and their estimations
phihat2 <- outputsP$estimphi # estimated random effects
plot(phi2, phihat2, type="p", pch=18, xlab='', ylab='') # true vs estimated
abline(0, 1)
```

It yields to Figure 5.4. It shows that estimation of the density is satisfactory (left graph) and estimation of the random effects is very good (right graph).

Validation of the MSDE is produced by function `valid`. For example the individual validation of the first trajectory is plotted Figure 5.5, the first two graphs on the left, using

```
choice2 <- floor(runif(1,1,M))
validationOU <- valid(estim_param, X2 = X2, times2 = times2, numj =choice2)
```

This illustrates the good estimation of the random effects: a beam of trajectories with the true one in the middle and the lining up of the quantiles.

Finally, we can predict some trajectories using `pred`. Predictions are shown on Figure 5.6, as a result of

```
predPOU <- pred(estim_param, X2 = X2, times2 = times2)
```

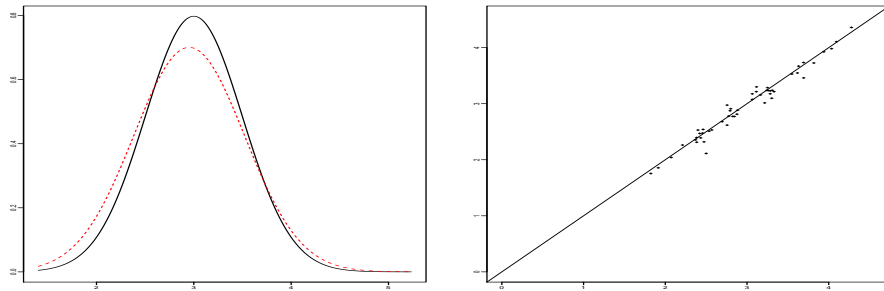


Figure 5.4 – Simulated example 2 (OU with one Gaussian random effect) frequentist parametric estimation, comparison to the truth. Left: parametric estimation $\mathcal{N}(\hat{\mu}, \hat{\omega}^2)$ (dotted line) and true f (plain line). Right: true ϕ_j versus estimated random effects A_j .

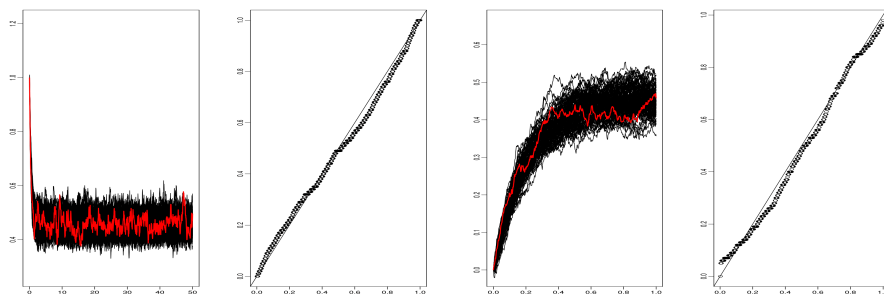


Figure 5.5 – Simulated examples frequentist approaches, outputs of valid method. Two left plots: frequentist nonparametric estimation on example 1 (CIR process). Two right plots: frequentist parametric estimation on example 2 (OU process).

Beam of predicted trajectories (right) is close to the true one (middle). The lining up of the predicted random effects versus the estimated random effects (left) shows the goodness of the prediction from the estimated density, thus of the estimation of the density.

5.4.3 Bayesian estimation

Bayesian method is applied to Example 2.

```
prior2 <- list( m = c(param[1], fixed2), v = c(param[1], fixed), alpha.omega = 11,
               beta.omega = param[2]^2*10, alpha.sigma=10, beta.sigma = sigma^2*9)
estim.method <- 'paramBayes'
estim_bayes <- mixedsde.fit(times2, X2, model = 'OU', random = 2, estim.method,
                           prior = prior, nMCMC = 10000)
outputsBayes <- out(estim_bayes)
```

Figure 5.7 is produced by

```
plot(estim)
```

Traces of the Markov chains of μ_1 , β , ω_1^2 and σ are plotted, showing that all chains converge and have the correct location. Command `print()` yields acceptance rates of the MH algorithm:

```
print(estim_bayes)
[1] "acceptance rates for random effect:"
Min. 1st Qu. Median Mean 3rd Qu. Max.
0.5607 0.5675 0.5694 0.5705 0.5730 0.5828
[1] "acceptance rate for fixed effect:" "0.3591"
```

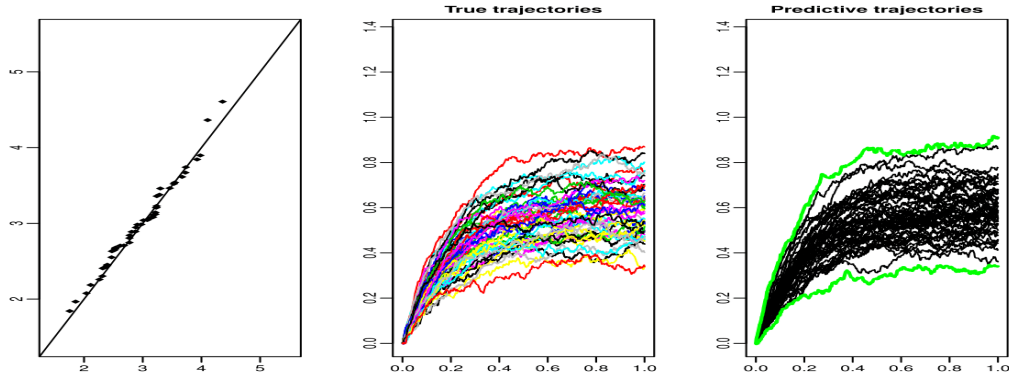


Figure 5.6 – Simulated example 2 (OU with one Gaussian random effect), frequentist parametric estimation. Left: predicted random effects versus estimated random effects. Middle: true trajectories. Right: predicted trajectories in black and 95% prediction interval in green (grey)

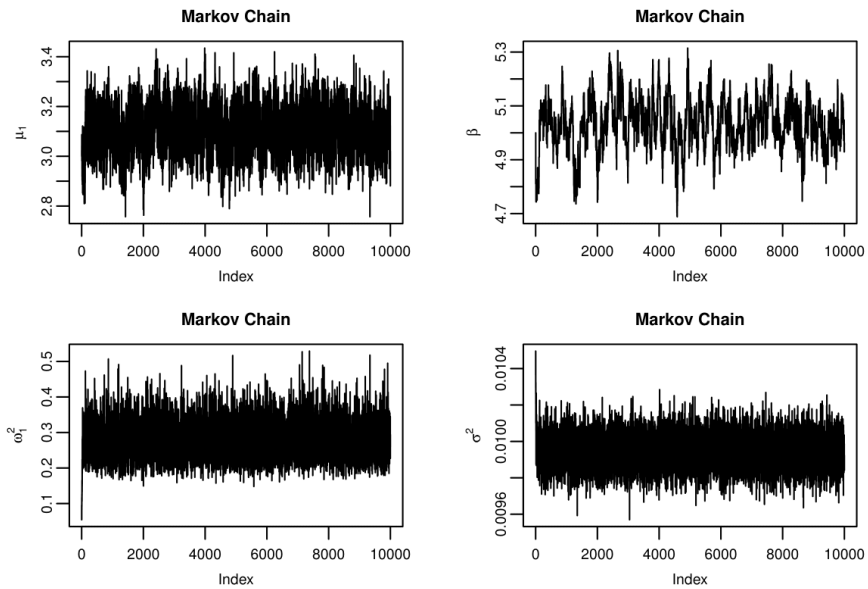


Figure 5.7 – Simulated example 2 (OU with one Gaussian random effect) Bayesian estimation. Markov chains of μ_1 , β , ω_1^2 and σ^2

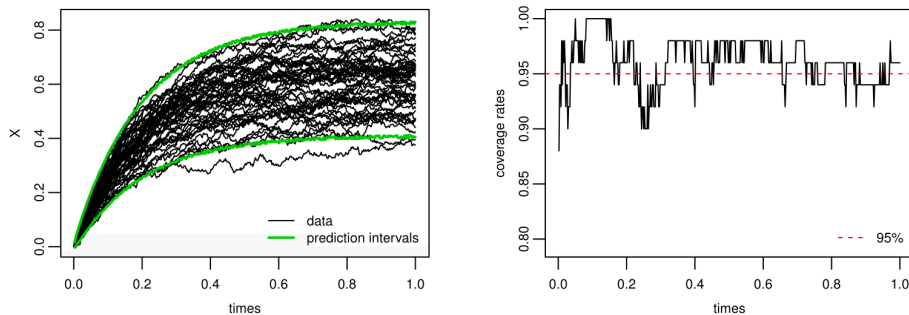


Figure 5.8 – Simulated example 2 (OU with one Gaussian random effect) Bayesian estimation. Left: predicted trajectories in black and 95% prediction interval in green (grey). Right: coverage rates: amount of observed values covered by the prediction intervals

The fixed effect β has a small acceptance rate, explaining the dependent chain (Figure 5.7 top right). This is due to a very sharp likelihood because of the large amount of observations ($N \cdot M$) in comparison to the random effect (N).

Predictions in the Bayesian framework and the corresponding Figure 5.8 is obtained by

```
pred.result <- pred(estim_bayes)
```

Figure 5.8 shows the beam of simulated data trajectories together with the 95% prediction interval. Coverage rates are shown on the right plot and we see that the intervals hold the level.

5.5 Package mixedside through a real data example

A real dataset is available (`neuronal .data.rda`) through lists of a matrix `X` and a vector `times`. We detail below the analysis of the data, following the next steps: run the two random effects model with both the parametric and nonparametric procedure; choose the number of random effects depending on the variability of the estimators ($A_{j,1}, A_{j,2}$), on the shape of \hat{f}_h and the variance $\hat{\Omega}$.

These data are available thanks to Rune Berg and Jufang He. Details on data acquisition can be found in Lansky *et al.* (2006).

5.5.1 Neuronal data (`neuronal.data`)

Neurons are the basement of nervous system and each neuron is connected with around 1000 other neurons. They are communicating through emission of electrical signal. We focus on the dynamic of the neuron membrane potential between two spikes emission measured in volts as the difference of ions concentration between the exterior and the interior of the cell. Data are obtained from one single neuron of a pig. Data are composed of $M = 240$ membrane potential trajectories with $N = 2000$ equidistant observation times. Time step is $\delta = 0.00015$ [s] and observation time is $T = 0.3$ [s]. Data are uploaded using `data(neuronal.data)`. They are presented on Figure 5.9. These data have been previously analysed with a Ornstein-Uhlenbeck model with one additive random effect (α_j): Picchini *et al.* (2008) and Picchini *et al.* (2010) use parametric methods assuming the normality of the random effect, and Chapter 3 with a nonparametric method. Here α_j represents the local average input that the neuron receives after the j^{th} spike. The initial voltage (the value following a spike) is assumed to be equal to

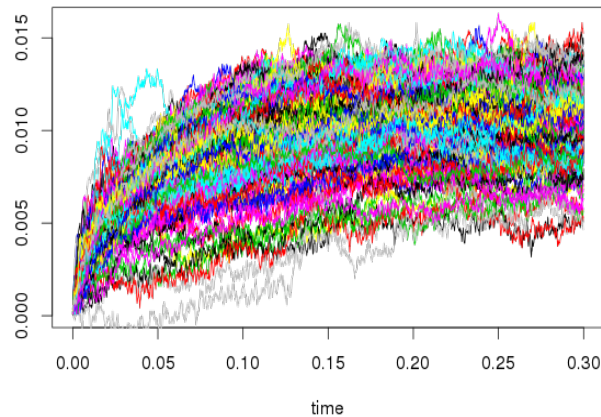


Figure 5.9 – Neuronal data

the resting potential and set to zero: $x_j = 0$. Parameter β_j (non negative) is the time constant of the neuron. It was fixed in Picchini *et al.* (2008) and Picchini *et al.* (2010).

In this new analysis, we assume that both α_j and β_j may change from one trajectory to another because of other neurons or environment influence, for example. Indeed, the form of each trajectory lead us to think that this is the good model: each one has its mean-revert value and its own speed to reach this value. There is no reason to assume that the speed is always the same, but looking at the trajectories the stationary state seem reached nearly at the same time, thus the second random effect should have a small variance if the model is the good one.

Our goal is also to compare the two models OU and CIR, both with two random effects, and two approaches: the nonparametric density estimation and the parametric density estimation. Let us remark that for the CIR model the algorithm removes two trajectories: 168 and 224, because they contain negatives values. For two random effects the commands have the form

```
estim <- mixedsde.fit(times, X, model= model, random = random,
  estim.method = estim.method)
```

and they can be found in the help data file (command `?neuronal.data`). We first apply the two frequentist approaches on models with two random effects. Kurtosis and skewness of the distribution of the estimation A_j of the random effects given in Table 5.2 are not closed to a symmetric distribution. The bidimensional density of (α_j, β_j) is estimated for both models with the parametric and nonparametric methods running function `mixedsde.fit`. Figure 5.10 gives the 4 estimated marginals. The blue (black) is for the OU model and the green (grey) for the CIR model. The dotted lines are the estimations from the parametric method, the plain lines for the nonparametric estimation. Parametric and nonparametric estimators are close, except for the second random effect with the OU model. Indeed, parametric estimation produces a very small variance for the second random effect, suggesting it could be fixed. Would this assumption be valid, it explains the difference with the nonparametric estimator which is not stable if the variance is too small. Estimation of σ is $\hat{\sigma} = 0.0136$ for the OU model and $\hat{\sigma} = 0.163$ for the CIR model.

To compare with previous literature and from knowledge on the biological phenomenon, we focus on the OU model. To select the number and the position of the random effect, we run the code with one random effect, additive or multiplicative: `random=1` or `random = 2`, for both

Table 5.2 – Neuronal data. Kurtosis and skewness estimations for samples $(A_{j,1})$'s and $(A_{j,2})$'s, for OU and CIR models

		OU	CIR
$(A_{j,1})$	Kurtosis	6.16	11.69
	Skewness	0.95	2.32
$(A_{j,2})$	Kurtosis	6.67	7.07
	Skewness	1.63	1.86

models fixing the other parameter to the estimate obtained with two random effects. Estimators of the means μ_1, μ_2 and standard deviations ω_1, ω_2 are given in Table 5.3 and compared to values obtained in Picchini *et al.* (2008) and Picchini *et al.* (2010). Criteria AIC and BIC are also given in Table 5.3. The preferred model is the one minimizing both criteria. Thus, the OU model with one additive random effect $\phi_j = \alpha_j$ and $\beta = 37$ seems to be the best model to describe these data. The summary method gives for the kurtosis: 4.55 and for the skewness -0.95. Also $\sigma = 0.0136$. Estimated densities obtained for this model with β to 37 are given in Figure 5.11. The dotted line is the result of the parametric estimation and the plain line of the nonparametric estimation, plotted on the histogram of the $A_j(T)$'s. On the graph we see the nonparametric estimation detects a left tail that is not detected by the parametric one. Otherwise both estimators are very close.

The OU model with `random= 1` is then validated with `valid` function. Figure 5.12 illustrates the result for a random trajectory (number 232): 100 simulated trajectories (black) and true trajectory (X_{232} , red) (left plot) and quantiles of the true measurement among the 100 simulated points at each time points versus uniform quantiles. The qq-plot is satisfactory (compared to the graph obtained on simulated data Figure 5.5). Finally some prediction plot are done (not shown) with the `pred` method and they confirm that model OU with `random = c(1,2)` with the parameters obtain from the parametric estimation, and the OU model with `random = 1` and $\beta = 37$ produce very close trajectories and could be both validated.

We then apply the Bayesian procedure. As already mentioned, for the Bayesian procedure, large data sets are a problem because of the very long running time. Therefore, we thin the data set by 10. That means, every 10th data point of the series is used for the estimation and also for the prediction. Even with this thinning, one estimation with 20000 samples takes half an hour and the corresponding prediction, only the prediction intervals, takes about 15 minutes.

Based on the best model selected by the frequentist approach, the OU model with one random effect $\phi_j = \alpha_j$ is fitted. No prior knowledge is available, we therefore leave this information out and let the algorithm take the first 10%, i.e. 24, series for the calculation of the prior parameter, as described in Section 5.3.2. Figure 5.13 plots the Markov chains estimated from the remaining $M - 24 = 216$ trajectories and show good convergence of the chains. Bayesian point estimations, i.e. posterior means, are $\hat{\mu}_1 = 0.34$, $\hat{\omega}_1 = \sqrt{\hat{\omega}_1^2} = 0.06$, $\hat{\beta} = 33$ and $\hat{\sigma} = \sqrt{\hat{\sigma}^2} = 0.01$. Compared to frequentist estimation (Table 5.3), we notice that these results are a compromise between Picchini *et al.* (2010) and frequentist estimation.

The estimated values themselves are not very meaningful in the case of real data because we do not know the truth. Therefore, prediction is made, first for the random effect α_f , and then for the process himself, see description in Section 5.2.3. The results can be seen in Figure 5.14, on the left the prediction intervals and on the right the appropriate coverage rates. They all hold the level. Therefore, we can say, even with different estimated values to the frequentist approach, the Bayesian procedure works.

Table 5.3 – Neuronal data. MLE and BIC AIC criteria for OU model, depending on the number of random effects

	μ_1	ω_1	μ_2	ω_2	BIC	AIC
Picchini <i>et al.</i> (2008)						
random=1	0.28	0.04	25.64	0	-2969.7	-2978.71
Picchini <i>et al.</i> (2010)						
random=1	0.49	0.07	47	0	-3043.89	-3052.85
random=1	0.37	0.06	37.00	0	-3240.45	-3249.41
random=2	0.38	0	38.82	7.76	-3093.12	-3102.08
random=c(1,2)	0.38	0.06	37.70	1.10	-3229.67	-3247.5

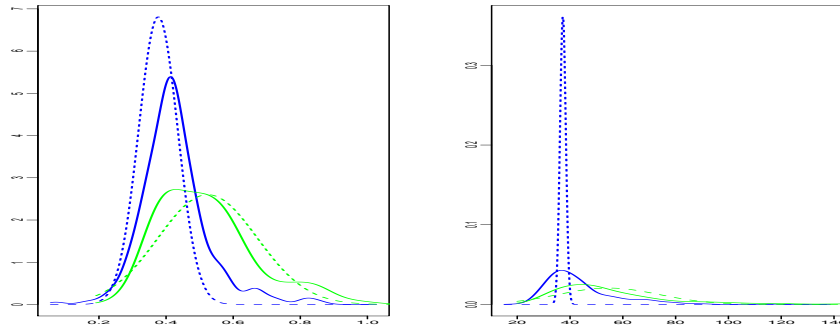


Figure 5.10 – Neuronal data. Frequentist estimated marginals of the bidimensional density of the random effects obtained from 4 estimators. Left: α_j 's density, right: β_j 's density. CIR model in green (grey), OU in blue (black). Nonparametric in plain line, parametric in dotted line

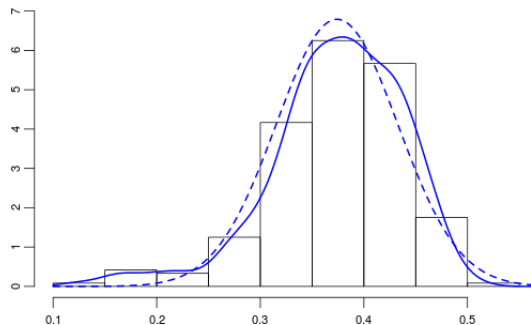


Figure 5.11 – Neuronal data, OU model, α random, β fixed, frequentist approaches. Estimation of the density f : parametric estimation in blue (black) plain line, non-parametric estimation blue (black) dotted line and histogram of the A_j 's

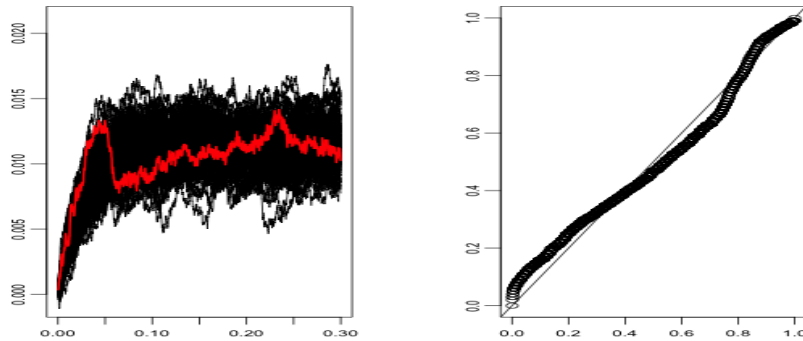


Figure 5.12 – Neuronal data, OU model, α random, β fixed, validation of the frequentist approaches. Individual validation of trajectory 232. Left: 100 simulated trajectories in black and true trajectory (X_j) in grey (red). Right: quantiles of the true measurement among the 100 simulated points at each time points versus uniform quantiles

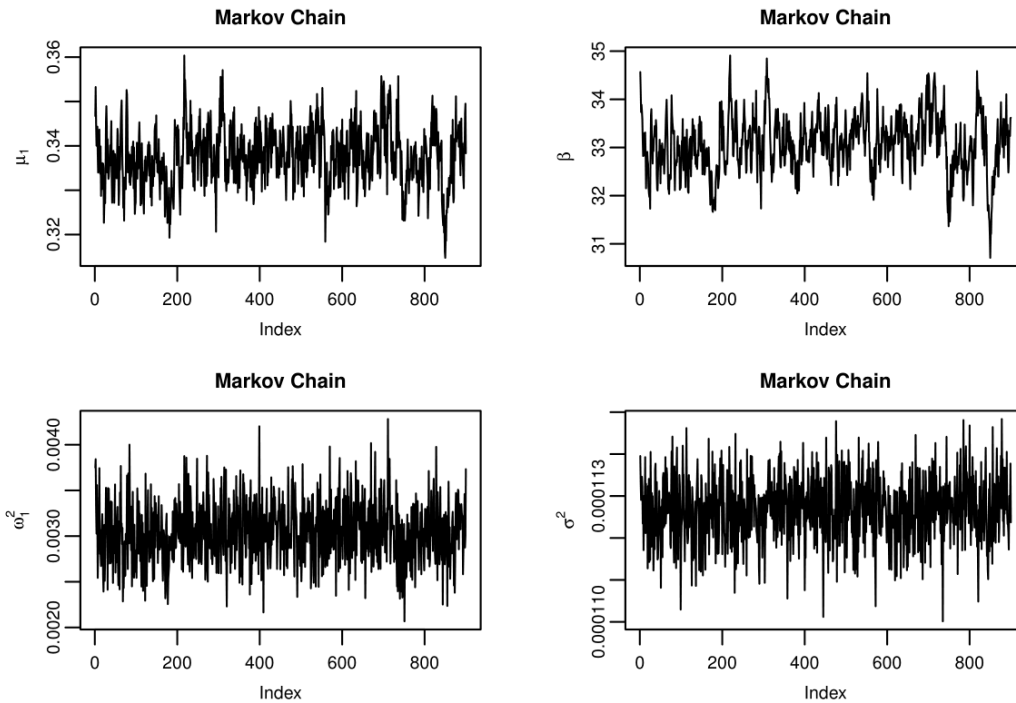


Figure 5.13 – Neuronal data, OU model, α random, β fixed, Bayesian estimation. Reduced Markov chains (less the burn-in phase and the thinning rate).

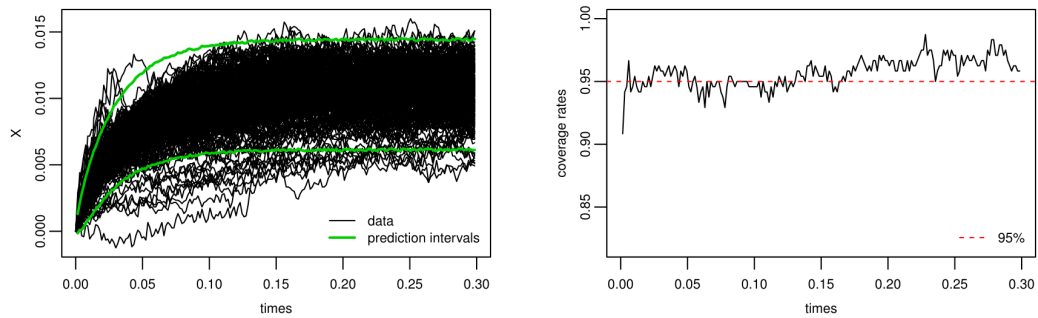


Figure 5.14 – Prediction results for the ISI data using the O-U model with $\text{random} = 1$, left: pointwise 95% prediction intervals for the 230 data series (the first 10 are used to generated prior knowledge), right: coverage rates: amount of observed values covered by the prediction intervals.

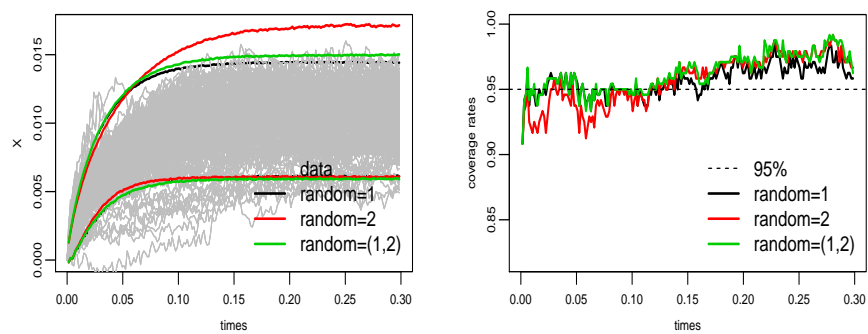


Figure 5.15 – Neuronal data. Bayesian prediction results using the OU model, left: pointwise 95% prediction intervals and data series, right: coverage rates: amount of observed values covered by the prediction intervals

Appendix 1

In this appendix we explicit the computation of U and V for each model.

Random = 1

OU

$$U_j(T) := \frac{1}{\sigma^2} \left(X_j(T) - X_j(0) + \beta \int_0^T X_j(s) ds \right), \quad V_j(T) := \frac{T}{\sigma^2}.$$

CIR

$$U_j(T) := \frac{1}{\sigma^2} \left(\int_0^T \frac{dX_j(s)}{X_j(s)} + \beta T \right), \quad V_j(T) := \frac{1}{\sigma^2} \int_0^T \frac{ds}{X_j(s)}.$$

Random = 2

OU

$$U_j(T) := \frac{1}{\sigma^2} \left(- \int_0^T X_j(s) dX_j(s) + \alpha \int_0^T X_j(s) ds \right), \quad V_j(T) := \frac{1}{\sigma^2} \int_0^T X_j(s)^2 ds.$$

$$U := -\frac{1}{\sigma^2} \int_0^T X_j(s) dX_j(s), \quad V := \frac{1}{\sigma^2} \int_0^T X_j(s)^2 ds$$

CIR

$$U_j(T) := \frac{1}{\sigma^2} (-X_j(T) + X_j(0) + \alpha T), \quad V_j(T) := \frac{1}{\sigma^2} \int_0^T X_j(s) ds.$$

Random = c(1,2)

OU

$$U_j(T) := \frac{1}{\sigma^2} \left(X_j(T) - X_j(0), - \int_0^T X_j(s) dX_j(s) \right)^t; \quad V_j(T) = \frac{1}{\sigma^2} \begin{pmatrix} T & - \int_0^T X_j(s) ds \\ - \int_0^T X_j(s) ds & \int_0^T X_j(s)^2 ds \end{pmatrix}.$$

CIR

$$U_j(T) := \frac{1}{\sigma^2} \left(\int_0^T \frac{1}{X_j(s)} dX_j(s), -(X_j(T) - X_j(0)) \right)^t; \quad V_j(T) = \frac{1}{\sigma^2} \begin{pmatrix} \int_0^T \frac{1}{X_j(s)} ds & -T \\ -T & \int_0^T X_j(s) ds \end{pmatrix}$$

Appendix 2: another application on a real dataset

We have another neuronal dataset available, and we thank Rune Berg for that. They are measurements of a turtle motoneuron which is responsible for a scratch reflex when the turtle receives a mechanical stimulation. The membrane potential is measured during 25s, 16 times with 5 minutes break between each measurement. More details on data acquisition are available in Berg *et al.* (2007). A trajectory of the potential begins (after the stimulation phase) by a quiescence phase, during 5s, represented of Figure 5.16. We focus on this phase here and the data have been translated to be positive. A previous study of these data exists in Jahn *et al.* (2011). The validated model to describe these trajectories is an Ornstein-Uhlenbeck model. At

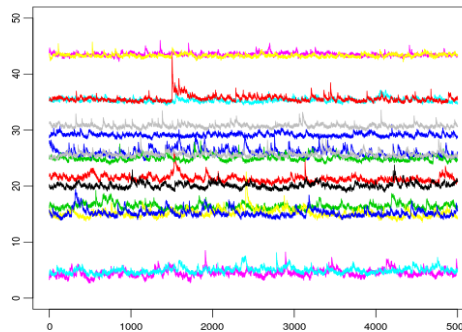


Figure 5.16 – Neuronal data 2: quiescence trajectories

Table 5.4 – Neuronal data 2. MLE and AIC and BIC criteria for OU model, depending on the number of random effects

	μ_1	ω_1	μ_2	ω_2	BIC	AIC
random=c(1,2)	8.97	8.32	0.31	0.16	-11785.7	-11792.8
random=1	7.35	3.41	0.31	0	-9209.8	-9222.1

first it seems natural to add one additive random effect $\phi_j = \alpha_j$ in the drift function, to model the mean-revert value, different for each trajectory, and to fix β_j .

Thus we investigate the data with both frequentist methods with the package for the model OU, with `random=c(1,2)` at first then with `random=1`. Indeed, the first step will give us an initial value for β_j if we want it fixed, and some information about both variables. From the parametric method, the obtained MLE and AIC, BIC criteria are given in Table 5.4. The estimated σ is 0.347. The criteria indicate to choose the two random effects model. The validation method `valid` applied on the outputs of the estimation with `random=c(1,2)`, thus based on the $A_{j,1}, A_{j,2}$, gives Figure 5.17. We see that this model seems very appropriate to describe the data.

Both variables have a kurtosis close to 3 but a skewness close to 1 which reflects a right tail on the distributions. Thus the nonparametric strategy seems more appropriate. Figure 5.18 shows the estimated marginals from the nonparametric estimator in green (grey) and the parametric ones in dotted blue (black). The dotted green (grey) lines are the marginal of the nonparametric kernel estimator $\hat{f}_h(x) = \frac{1}{M} \sum_{j=1}^M K_h(x - \hat{A}_j)$ based on the truncated A_j (studied theoretically in Chapter 4).

Finally we simulate new trajectories from the estimated densities with \hat{f}_h and \hat{f}_h . The first estimator gives the best prediction. Figure 5.19 shows the results obtained with function `pred`.

We conclude from this study that the best model to describe the data is $dX_j(t) = (\phi_{j,1} - \phi_{j,2}X_j(t))dt + \sigma dW_j(t)$ and the bidimensional random effect have the density \hat{f}_h .

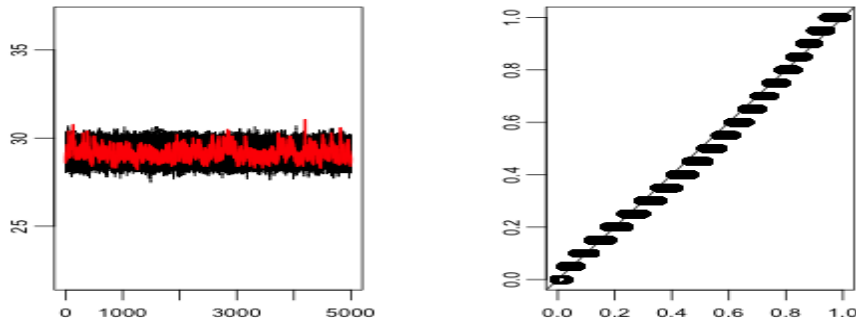


Figure 5.17 – Neuronal data 2. Validation of OU model with $\text{random}=c(1,2)$ for the first trajectory, with method `valid` of the package

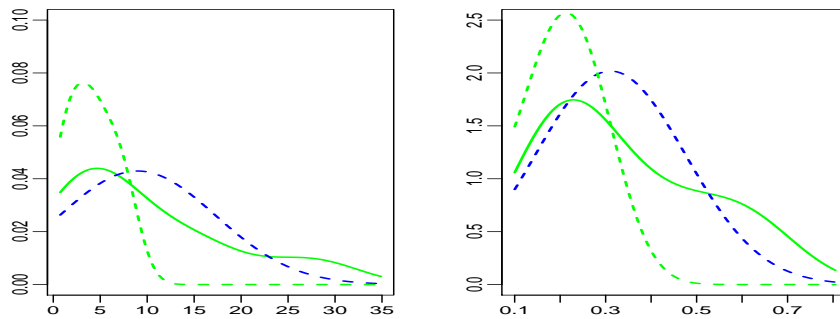


Figure 5.18 – Neuronal data 2. Frequentist estimated marginals: in dotted blue (black) the parametric, in green (grey) the nonparametric with in plain \hat{f}_h and dotted $\widehat{\widehat{f}}_h$.

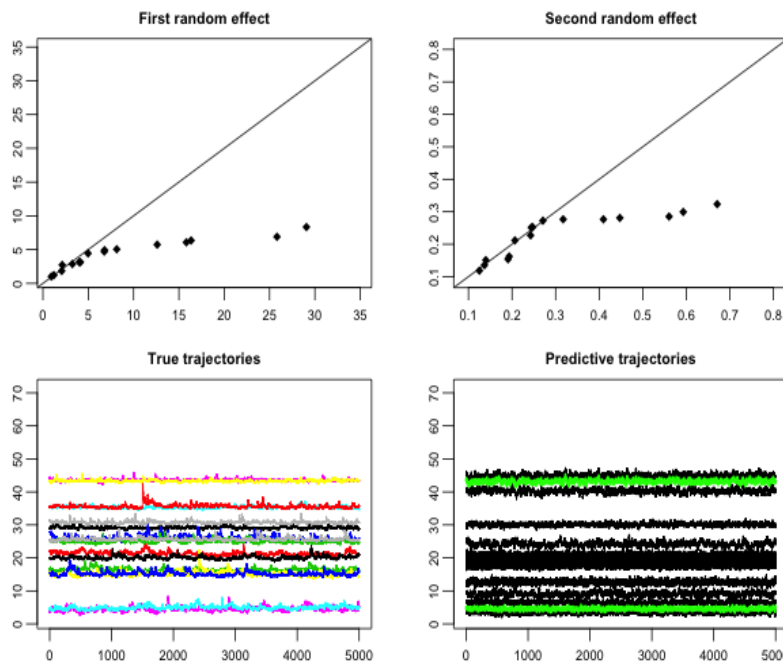


Figure 5.19 – Neuronal data 2. Prediction from the nonparametric estimation $\widehat{\widehat{f}}_h$ with model OU and $\text{random}=c(1,2)$ with method `pred` of the package

Troisième partie

Modèle de bruit multiplicatif

Chapitre 6

Estimation non-paramétrique dans un modèle de censure multiplicative avec un bruit symétrique

Dans ce travail nous considérons le modèle : $Y_i = X_i U_i$, $i = 1, \dots, n$, où les X_i , U_i et donc les Y_i sont des variables indépendantes et identiquement distribuées. Les X_i ont pour densité f et sont les variables d'intérêt, les U_i sont des variables de bruit multiplicatif de densité uniforme sur l'intervalle $[1 - a, 1 + a]$, avec $0 < a < 1$, et ces deux séquences sont indépendantes. Cependant, seulement les Y_i sont observés.

Ce modèle peut par exemple être utilisé pour la protection de l'information. L'enjeu est de masquer assez les données à l'aide du bruit multiplicatif pour transformer leur structure, tout en sachant retrouver l'information importante contenue dans les données d'origine. Notre but est d'estimer de manière non-paramétrique la densité f des variables X_i puis la fonction de survie associée \bar{F} , lorsque a est connu. Remarquons que les variables observées Y_i sont *i.i.d.* de densité commune f_Y sur \mathbb{R}^+ donnée par la formule suivante

$$f_Y(y) = \frac{1}{2a} \int_{\frac{y}{1+a}}^{\frac{y}{1-a}} \frac{f(x)}{x} dx, \quad y \in]0, +\infty[.$$

L'inversion de cette formule pour atteindre f n'est pas évidente. Nous proposons une méthode en deux étapes. Tout d'abord, nous estimons une fonction auxiliaire g par projection. La fonction f est approchée par une série de fonctions dépendant de g , qui converge vers f en norme \mathbb{L}^2 . Alors il se déduit de l'estimateur de g un estimateur naturel de f . Cette même procédure est utilisée pour estimer \bar{F} la fonction de survie. Dans chaque cas une borne du risque quadratique intégré de l'estimateur est établie. Nous proposons une méthode de sélection de modèle par pénalisation. Enfin nous prouvons une inégalité de type oracle pour chaque estimateur final. Nous atteignons les vitesses minimax prouvées dans Brunel *et al.* (2015) pour le bruit uniforme sur $[0, 1]$. Enfin nous donnons des pistes pour l'estimation du paramètre a que l'on suppose connu dans cette étude. Une étude sur simulation puis sur données réelles illustre notre propos et l'intérêt de ce modèle. En particulier nous comparons notre méthode à une procédure classique par déconvolution après l'application de la fonction logarithme au modèle. Nous observons que cette nouvelle procédure donne de meilleurs résultats théoriques et numériques.

Ce chapitre est issu de l'article Comte & Dion (2016) écrit avec Fabienne Comte, soumis pour publication. La section Appendix a été rajoutée pour figurer dans cette thèse.

Sommaire

6.1	Introduction	157
6.2	Multiplicative denoising of density and survival function	158
6.2.1	Notations	158
6.2.2	Estimation strategy	159
6.2.3	Risk bound for density estimator	160
6.2.4	Model selection for density estimation	161
6.2.5	Survival function estimation	162
6.2.6	Case of unknown a	164
6.3	Model transformation and deconvolution approach	164
6.4	Numerical study	166
6.4.1	Simulated data	166
6.4.2	Application	170
6.5	Proofs	171
6.5.1	Proof of Lemma 6.2.1	171
6.5.2	Useful properties of the Laguerre basis	171
6.5.3	Proof of Proposition 6.2.2	174
6.5.4	Proof of Theorem 6.2.4	176
6.5.5	Proof of Lemma 6.5.4	177
6.5.6	Proof of Lemma 6.2.5	180
6.5.7	Proof of Proposition 6.2.6	181
6.5.8	Proof of Theorem 6.2.7	181
6.5.9	Proof of Theorem 6.3.2	182
6.5.10	Proof of the Lemma 6.5.6	183
Appendix : observations and outcomes		183

6.1 Introduction

We consider the following model

$$Y_i = X_i U_i, \quad i = 1, \dots, n, \quad U_i \sim \mathcal{U}_{[1-a, 1+a]}, \quad 0 < a < 1 \quad (6.1.1)$$

where $(X_i)_{\{i=1, \dots, n\}}$ and $(U_i)_{\{i=1, \dots, n\}}$ are two independent samples. The U_i 's are independent and identically distributed (*i.i.d.*) random variables from uniform density on an interval $[1 - a, 1 + a]$ of \mathbb{R}^+ with $0 < 1 - a < 1 + a$ and a is assumed to be known. The X_i 's are *i.i.d.* from an unknown density f on \mathbb{R}^+ . Both sequences are unobserved. Only the Y_i 's are observed. The model implies that they are *i.i.d.* and we denote by f_Y their density on \mathbb{R}^+ . Our goal is to estimate nonparametrically the density f of the X_i 's from the observations $Y_i, i = 1, \dots, n$.

Equation (6.1.1) can be obtained as follows. Classical models involving measurement errors are often additive and state that the variable of interest X_i is not directly observed because an additive noise hides it: only samples of $X_i + \xi_i$ are available, where ξ_i is an *i.i.d.* centred sequence. Then, in many contexts, this noise depends on the level of the signal and the simplest strategy is to consider that it is proportional to the signal. Thus, the model for the observations becomes $X_i + \alpha X_i \xi_i, \alpha \in \mathbb{R}$. Rewriting this $X_i(1 + \alpha \xi_i)$, we obtain a multiplicative noise model with noise $1 + \alpha \xi_i$ with mean 1. This corresponds to model (6.1.1) where we specified the final distribution of the noise as uniform, and symmetric around one¹.

In any case, Equation (6.1.1) models an approximate transmission of the information: the recorded values Y_i correspond to the value of interest X_i , up to an error of order of $\pm 100\%$. This represents rather standard situations, when people have to give their height or the amount of money they devote to some specific expenses, *i.e.* quantities they may not know exactly with no intention to change them (for instance, weight or income may be intentionally biased). However, very few studies of this model have been conducted in the literature. We mainly found it in Sinha *et al.* (2011), who study "noise multiplied magnitude microdata" as a form of data masking in contexts where one needs to protect the privacy of survey respondents. The authors mainly study quantile estimation.

Nevertheless, multiplicative noise models can be found with other distributions for the noise U . The case of U following a uniform distribution on $[0, 1]$ ($\mathcal{U}([0, 1])$) has been introduced by Vardi (1989) who called it a "multiplicative censoring" model. This model was studied by Vardi & Zhang (1992), Asgharian *et al.* (2012), Abbaszadeh *et al.* (2013), Brunel *et al.* (2015), and is mostly applied in survival analysis, see van Es *et al.* (2000). In these papers, nonparametric estimators of the density f or of the survival function $\bar{F} = 1 - F, F(x) = \int_0^x f(u)du$, of the unobserved random variable X are built and studied, but in different contexts. For instance, Asgharian *et al.* (2012) assume that part of the observations are directly observed and the proposed method is no longer valid if this proportion is null as in our model. In Brunel *et al.* (2015), kernel estimators are studied, while Abbaszadeh *et al.* (2013) build wavelet estimators of the density and its derivatives. The case of Gaussian U , for variables on \mathbb{R} , has also been considered in financial context and studied from statistical point of view by *e.g.* Es *et al.* (2005).

In this work, we build estimators of the density f and of the survival function $\bar{F} = 1 - F$. The operator linking the density of the observations and the density of interest is given by

$$f_Y(y) = \frac{1}{2a} \int_{\frac{y}{1+a}}^{\frac{y}{1-a}} \frac{f(x)}{x} dx, \quad y \in]0, +\infty[, \quad (6.1.2)$$

and the inversion of formula (6.1.2) is not obvious. This is why our strategy relies on two steps. Let us give here a sketch of the procedure. First, we approach an auxiliary function g expressed

1. Extension to general $\mathcal{U}([a, b])$ distribution for the U_i 's is possible.

as a function of f and a . We prove that for an explicit transformation $t \in \mathbb{L}^2(\mathbb{R}^+) \mapsto \psi_t$ and this function g in $\mathbb{L}^2(\mathbb{R}^+)$, we have

$$\mathbb{E}[\psi_t(Y_1)] = \langle t, g \rangle, \tag{6.1.3}$$

where $\langle s, t \rangle = \int_{\mathbb{R}^+} s(x)t(x)dx$ denotes the scalar product of two functions of $\mathbb{L}^2(\mathbb{R}^+)$. The two functions g and ψ_t are given in Section 6.2. Relation (6.1.3) is used to build projection estimators of g . Indeed, considering the collection of spaces

$$S_m = \text{Vect}\{\varphi_0, \varphi_1, \dots, \varphi_{m-1}\}$$

where $(\varphi_j)_{j \geq 0}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$, the orthogonal projection g_m of g on S_m is given by $g_m = \sum_{j=0}^{m-1} a_j \varphi_j$, with $a_j = \langle g, \varphi_j \rangle$. From relation (6.1.3), we notice that $a_j = \mathbb{E}[\psi_{\varphi_j}(Y_1)]$ and replacing the expectation by its empirical counterpart \hat{a}_j , we obtain the estimator $\hat{g}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j$. With similar ideas, we also define $\check{G}_m = \sum_{j=0}^{m-1} \check{b}_j \varphi_j$ where \check{b}_j are also computed from the observations Y_1, \dots, Y_n . Then we deduce, by inverting the relation between f and g (see Section 6.2.2) and between \bar{F} and \bar{G} (see Section 6.2.5), collections of estimators of f and \bar{F} , for which risk bounds are provided, in term of mean integrated squared error (MISE) on \mathbb{R}^+ . Model selection criterion are proposed to automatically select m in both cases, and they are proven to make the adequate tradeoff between bias (m must be large enough for the projection bias to be small) and variance (estimating too many coefficients increases the estimation error), see Theorems 6.2.4 and 6.2.7.

Finally we illustrate our method on simulated and real data. Our purpose is to propose a new method of privacy protection by the mean of our multiplicative censoring model. On the data given in Sinha *et al.* (2011), knowing the level of noise a , we show how to recover the hidden information about the original data from the noisy observations.

The plan of the Chapter is the following. In Section 6.2, we describe our estimation method and the model selection procedure, for the density in Sections 6.2.2 to 6.2.4 and for the survival function in Section 6.2.5. We compute bounds on the integrated quadratic risk associated to the estimators and deduce rates of convergence. The strategy and the results are detailed for density estimation and then extended to the case of survival function estimation. In Section 6.3, we describe a deconvolution strategy based on the additive model obtained by taking the logarithm of (6.1.1): we compare our method to this one from theoretical point of view here and in practice in Section 6.4. Finally Section 6.4 illustrates the theoretical results, on simulated data (Section 6.4.1) and on real data (Section 6.4.2). Simulation experiments show the good performances of our method, and estimation on real data is presented through an example of application. Lastly, most proofs are gathered in Section 6.5.

6.2 Multiplicative denoising of density and survival function

6.2.1 Notations

The space $\mathbb{L}^2(\mathbb{R}^+)$ is the space of square integrable functions on the positive real line. The associated \mathbb{L}^2 -norm is denoted $\|t\|^2 = \int_{\mathbb{R}^+} |t(x)|^2 dx$. The Fourier transform of $t \in \mathbb{L}^1$, for $x \in \mathbb{R}$ is: $t^*(x) = \int t(u)e^{iux} du$. Finally, the supremum norm of a bounded function t is denoted by $\|t\|_\infty = \sup_{x \in \mathbb{R}^+} |t(x)|$. The Laguerre basis is defined by:

$$\varphi_0(x) = \sqrt{2}e^{-x}, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \text{ for } k \geq 1, \quad x \geq 0, \tag{6.2.1}$$

with L_k the Laguerre polynomials

$$L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}. \quad (6.2.2)$$

It satisfies the orthonormality property $\langle \varphi_j, \varphi_k \rangle = \delta_{j,k}$ where $\delta_{j,k}$ is the Kronecker symbol equal to 1 if $j = k$ and to zero otherwise; and the following relations on the norms (see Abramowitz & Stegun (1966)):

$$\forall j \geq 0, \|\varphi_j\|_\infty \leq \sqrt{2}, \text{ and } \|\varphi'_j\|_\infty \leq 2\sqrt{2}(j+1), \quad (6.2.3)$$

where φ'_j is the derivative of φ_j . Any function of $\mathbb{L}^2(\mathbb{R}^+)$ can be decomposed on this basis. Lastly, we state a useful lemma, proven in Section 6.5, relying on the fact that the density f_Y is given by (6.1.2).

Lemma 6.2.1. *The density f_Y defined in (6.1.2) satisfies $\lim_{y \rightarrow 0} y f_Y(y) = 0$ and $\lim_{y \rightarrow +\infty} y f_Y(y) = 0$.*

Lemma 6.2.1 is a useful property to justify the construction of the estimator.

6.2.2 Estimation strategy

Recall that f_Y is given by (6.1.2). Now let g be given by

$$g(x) := \frac{1}{2a} \left[f\left(\frac{x}{1+a}\right) - f\left(\frac{x}{1-a}\right) \right], \quad (6.2.4)$$

and consider a bounded function t , derivable and with derivative function t' in $\mathbb{L}^2(\mathbb{R}^+)$. Then, an integration by part and the Lemma 6.2.1 imply

$$\begin{aligned} \mathbb{E}[t(Y_1) + Y_1 t'(Y_1)] &= \frac{1}{2a} \int_0^{+\infty} t(y) \left[f\left(\frac{y}{1+a}\right) - f\left(\frac{y}{1-a}\right) \right] dy \\ &= \langle t, g \rangle. \end{aligned} \quad (6.2.5)$$

In other words

$$\mathbb{E}[\psi_t(Y_1)] = \langle t, g \rangle \quad \text{with} \quad \psi_t(y) := t(y) + y t'(y).$$

Our strategy is to use equation (6.2.5) to build a projection estimator of g , and then to look for an inversion of formula (6.2.4) to recover f . Precisely, it follows from (6.2.4) that

$$f(x) - f\left(\left(\frac{1+a}{1-a}\right)x\right) = 2a g((1+a)x)$$

and iterating the relation (by changing x into $(1+a)x/(1-a)$, $x > 0$), it yields

$$f(x) - f\left(\left(\frac{1+a}{1-a}\right)^N x\right) = 2a \sum_{k=0}^{N-1} g\left(\left(\frac{1+a}{1-a}\right)^k (1+a)x\right)$$

Thus a sequence of approximations of f , for $x > 0$, is

$$f_N(x) = 2a \sum_{k=0}^{N-1} g\left(\left(\frac{1+a}{1-a}\right)^k (1+a)x\right). \quad (6.2.6)$$

Besides, using that $f(x) - f_N(x) = f(((1+a)/(1-a))^N x)$, it is easy to check for $f \in \mathbb{L}^2(\mathbb{R}^+)$ that $\|f - f_N\|$ tends to 0 when N tends to infinity. Now, if f is square-integrable, so is g and therefore we can write its decomposition on the Laguerre basis:

$$g(x) = \sum_{j=0}^{\infty} a_j(g) \varphi_j(x), \quad \text{with } a_j(g) = \langle \varphi_j, g \rangle.$$

Recall that $g_m := \sum_{j=0}^{m-1} a_j(g) \varphi_j$ is the orthogonal projection of g on \mathcal{S}_m . According to (6.2.5), we have $a_j(g) = \mathbb{E}[\varphi_j(Y_1) + Y_1 \varphi_j'(Y_1)] = \langle \varphi_j, g \rangle$. Then the projection g_m of g on \mathcal{S}_m is estimated by

$$\hat{g}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{1}{n} \sum_{i=1}^n [Y_i \varphi_j'(Y_i) + \varphi_j(Y_i)] = n^{-1} \sum_{i=1}^n \psi_{\varphi_j}(Y_i), \quad (6.2.7)$$

with m in a finite collection $\mathcal{M}_n \subset \mathbb{N}$ that will be given later. Finally, plugging estimator (6.2.7) into (6.2.6), gives the collection of estimators of f , for $m \in \mathcal{M}_n$,

$$\hat{f}_{N,m}(x) = 2a \sum_{k=0}^{N-1} \hat{g}_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right). \quad (6.2.8)$$

6.2.3 Risk bound for density estimator

We first state a bound on the mean integrated squared error (MISE) of $\hat{f}_{N,m}$ as an estimator of f .

Proposition 6.2.2. *Assume that $f \in \mathbb{L}^2(\mathbb{R}^+)$ and $\mathbb{E}[X_1^2] < +\infty$.*

(i) *The estimator \hat{g}_m of g defined by (6.2.7) satisfies*

$$\mathbb{E}[\|\hat{g}_m - g\|^2] \leq \|g - g_m\|^2 + c_1 \frac{m}{n} + c_2 \frac{m^3}{n}, \quad c_1 = 4, \quad c_2 = 16\mathbb{E}[Y_1^2]. \quad (6.2.9)$$

(ii) *The estimator $\hat{f}_{N,m}$ of f defined by (6.2.8) satisfies*

$$\mathbb{E}[\|\hat{f}_{N,m} - f\|^2] \leq \frac{8a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} \left(\|g - g_m\|^2 + c_1 \frac{m}{n} + c_2 \frac{m^3}{n} \right) + 2 \left(\frac{1-a}{1+a} \right)^N \|f\|^2. \quad (6.2.10)$$

Both risk bounds involve a bias term (proportional to $\|g - g_m\|^2$) which decreases when m increases, and a variance term with main order m^3/n , which increases with m . The last term of (6.2.10) is clearly exponentially decreasing with N . As the value of N is chosen by the statistician, the choice

$$N \geq N_{\min} = \log(n) / |\log((1-a)/(1+a))| \quad (6.2.11)$$

makes this term negligible (numerically if $a \in (0.1; 0.9)$ and $n \in [100; 10000]$ then $N_{\min} \in (15; 45)$).

Rates of convergence of estimators can be computed more precisely. To evaluate the order of $\|g - g_m\|^2$, the regularity of the function g has to be specified. Let us assume, in this paragraph, that g belongs to a Sobolev-Laguerre space (see Bongioanni & Torrea (2009)), defined by

$$W^s(\mathbb{R}^+, L) := \{f : \mathbb{R}^+ \rightarrow \mathbb{R}, f \in \mathbb{L}^2(\mathbb{R}^+), \sum_{j \geq 0} j^s \langle f, \varphi_j \rangle^2 \leq L < +\infty\}, \quad (6.2.12)$$

with $s > 0$ (see Comte & Genon-Catalot (2015) for equivalent definitions in case s is an integer). Then we get the following order for the squared bias term:

$$\|g_m - g\|^2 = \sum_{j=m}^{\infty} a_j^2(g) = \sum_{j=m}^{\infty} a_j^2(g) j^s j^{-s} \leq Lm^{-s}.$$

Therefore we look for the choice $m = m_{\text{opt}}$ which minimizes $Lm^{-s} + c_2 m^3/n$. We obtain $m_{\text{opt}} = Cn^{1/(s+3)}$ with $C := (3c_2/(sL))^{-1/(s+3)}$, which implies $\mathbb{E}[\|\widehat{g}_{m_{\text{opt}}} - g\|^2] = O(n^{-s/(s+3)})$. This rate is the classical one in the multiplicative censoring model, and it is minimax optimal in case $U \sim \mathcal{U}([0, 1])$, see Belomestny *et al.* (2016) with additional log terms.

Remark 6.2.3. *Almost parametric rates for Gamma distributions.*

More specifically, the Laguerre basis yields better speeds of convergence than the previous ones, when f is a Gamma density (see for example Comte & Genon-Catalot (2015) or Mabon (2015a) for similar conclusions in the additive noise model). If X follows Gamma(p, θ) distribution, a consequence of Section 3.3, formula (18) in Mabon (2015a), is that

$$\|g - g_m\|^2 \leq C(a, p, \theta) m^{2(p-1)} \left[\left(\frac{\theta/(1+a) - 1}{\theta/(1+a) + 1} \right)^{2m} + \left(\frac{\theta/(1-a) - 1}{\theta/(1-a) + 1} \right)^{2m} \right]. \quad (6.2.13)$$

Indeed the coefficients $a_j(g)$ can be seen as a linear combination of coefficients of a Gamma($p, \theta/(1+a)$) and a Gamma($p, \theta/(1-a)$). The rate of decrease of the bias in (6.2.13) is thus exponential and choosing $m_{\text{opt}} = \lambda \log(n)$ with $\lambda = 2 \max \left(\left| \frac{\theta/(1-a)-1}{\theta/(1-a)+1} \right|, \left| \frac{\theta/(1+a)-1}{\theta/(1+a)+1} \right| \right)$ gives a negligible bias and a variance-governed rate for the risk of \widehat{g}_m , of order $O(\log^3(n)/n)$.

6.2.4 Model selection for density estimation

As the regularity s of g is unknown, the choice $m = m_{\text{opt}}$ cannot be performed in practice. Therefore, a selection method must be set up to choose automatically the best m among the discrete collection $\mathcal{M}_n = \{m \in \llbracket 1, n \rrbracket, m^3 \leq n\}$, realizing the bias-variance trade-off. We want to choose m minimizing the MISE of $\widehat{f}_{N,m}$. Considering bound (6.2.10), the theoretical value is

$$m_{th} := \operatorname{argmin}_{m \in \mathcal{M}_n} \left\{ \|g - g_m\|^2 + c_1 \frac{m}{n} + c_2 \frac{m^3}{n} \right\} = \operatorname{argmin}_{m \in \mathcal{M}_n} \left\{ -\|g_m\|^2 + c_1 \frac{m}{n} + c_2 \frac{m^3}{n} \right\}$$

as $\|g - g_m\|^2 = \|g\|^2 - \|g_m\|^2$ and $\|g\|^2$ does not depend on m . But functions g_m are unknown, thus we replace them by estimators. Therefore, we may select m as the minimizer of the sum $-\|\widehat{g}_m\|^2 + \text{pen}(m)$ with

$$\text{pen}(m) := \kappa_1 \frac{m}{n} + \kappa_2 \mathbb{E}[Y_1^2] \frac{m^3}{n} =: \text{pen}_1(m) + \text{pen}_2(m). \quad (6.2.14)$$

The penalty terms have the order of the variance term in (6.2.9). Note that the definition of \mathcal{M}_n ensures that it is bounded. As $\mathbb{E}[Y_1^2]$ is unknown, we finally propose to replace it by its empirical counterpart and we get:

$$\widehat{m} = \operatorname{argmin}_{m \in \mathcal{M}_n} \{-\|\widehat{g}_m\|^2 + \widehat{\text{pen}}(m)\}, \quad (6.2.15)$$

where

$$\widehat{\text{pen}}(m) = 2\kappa_1 \frac{m}{n} + 2\kappa_2 \widehat{C}_2 \frac{m^3}{n} := 2\text{pen}_1(m) + 2\widehat{\text{pen}}_2(m), \quad \widehat{C}_2 = \frac{1}{n} \sum_{k=1}^n Y_k^2. \quad (6.2.16)$$

The constants κ_1 and κ_2 are numerical constants which are calibrated in the simulations. Note that $\|\widehat{g}_m\|^2 = \sum_{j=0}^{m-1} \widehat{a}_j^2$ with \widehat{a}_j given in (6.2.7) is easy to compute. Our final estimator is

$$\widehat{f}_{N,\widehat{m}}(x) = 2a \sum_{k=0}^{N-1} \widehat{g}_{\widehat{m}} \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right). \quad (6.2.17)$$

We can prove the following result.

Theorem 6.2.4. *Assume that $f \in \mathbb{L}^2(\mathbb{R}^+)$, that f is bounded and that $\mathbb{E}[X_1^8] < +\infty$. For the final estimator $\widehat{f}_{N,\widehat{m}}$ defined by (6.2.7), (6.2.15) and (6.2.17), there exists κ_0 such that for $\kappa_1, \kappa_2 \geq \kappa_0$,*

$$\mathbb{E}[\|\widehat{f}_{N,\widehat{m}} - f\|^2] \leq \frac{16a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} \left(6 \inf_{m \in \mathcal{M}} \{\|g - g_m\|^2 + \text{pen}(m)\} + \frac{C_a}{n} \right) + \left(\frac{1-a}{1+a} \right)^N \|f\|^2,$$

where pen is given by (6.2.14), and C_a is a positive constant depending on a and $\|f\|_\infty$.

The theoretical study gives the bounds: $\kappa_1 \geq 32$ and $\kappa_2 \geq 288$. But it is well known that these theoretical constants are too large in practice: this is why the calibration step for choosing the values of the constants is done through simulations. Theorem 6.2.4 is a non-asymptotic bound for the MISE of the adaptive estimator $\widehat{f}_{N,\widehat{m}}$. It shows that the selection method leads to an estimator with smallest possible risk among all the estimators in the collection. Note that as previously, the choice $N = \log(n)/\log((1+a)/(1-a))$ is suitable for the last term to be negligible.

6.2.5 Survival function estimation

In this section, we extend the previous procedure to provide an estimator of the survival function of X , defined on \mathbb{R}^+ by

$$\overline{F}(x) = 1 - F(x) = \int_x^{+\infty} f(u)du. \quad (6.2.18)$$

We denote by \overline{F}_Y the survival function of Y , defined accordingly. We also define a similar function \overline{G} associated with g (which is not a density). We can prove the following Lemma.

Lemma 6.2.5. *For all x in \mathbb{R}^+ ,*

$$\overline{G}(x) := \int_x^\infty g(u)du = \frac{1}{2a} \left[(1+a)\overline{F}\left(\frac{x}{1+a}\right) - (1-a)\overline{F}\left(\frac{x}{1-a}\right) \right] = x f_Y(x) + \overline{F}_Y(x). \quad (6.2.19)$$

By integrating relation (6.2.6), we also get a relation between \overline{F} and \overline{G} : for $x > 0$, let

$$\overline{F}_N(x) := \frac{2a}{1+a} \sum_{k=0}^{N-1} \left(\frac{1-a}{1+a} \right)^k \overline{G} \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right), \quad (6.2.20)$$

then

$$\overline{F}(x) - \overline{F}_N(x) = \left(\frac{1-a}{1+a} \right)^N \overline{F} \left(\left(\frac{1+a}{1-a} \right)^N x \right).$$

Note that $\bar{G}(0) = 1$ and thus $\lim_{N \rightarrow \infty} \bar{F}_N(0) = 1$, which is coherent with $\bar{F}(0) = 1$. Moreover, if $\mathbb{E}[X_1^2] < +\infty$ the function \bar{F} , and thus \bar{G} , is square integrable on \mathbb{R}^+ . Denoting by \bar{G}_m the orthogonal projection of \bar{G} on \mathcal{S}_m , we have

$$\bar{G}_m = \sum_{j=0}^{m-1} b_j(\bar{G}) \varphi_j, \quad \text{with } b_j(\bar{G}) := \langle \bar{G}, \varphi_j \rangle.$$

According to relation (6.2.19), the coefficients $b_j(\bar{G})$ can also be written as follows: $b_j(\bar{G}) = \mathbb{E}[Y \varphi_j(Y)] + \langle \bar{F}_Y, \varphi_j \rangle$. Thus we estimate the projection \bar{G}_m of \bar{G} on \mathcal{S}_m by

$$\check{\bar{G}}_m = \sum_{j=0}^{m-1} \check{b}_j \varphi_j, \quad \check{b}_j = \frac{1}{n} \sum_{i=1}^n \left[\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_i \geq x} dx + Y_i \varphi_j(Y_i) \right]. \quad (6.2.21)$$

Finally, plugging (6.2.21) into (6.2.20), an estimator of \bar{F} is given by

$$\check{\bar{F}}_{N,m} = \frac{2a}{1+a} \sum_{k=0}^{N-1} \left(\frac{1-a}{1+a} \right)^k \check{\bar{G}}_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right). \quad (6.2.22)$$

We can prove the following bound.

Proposition 6.2.6. *Assume that $\mathbb{E}[X_1^2] < +\infty$. Then, \bar{F} is square integrable and the estimator $\check{\bar{F}}_{N,m}$ of \bar{F} given by (6.2.22) satisfies*

$$\mathbb{E}[\|\check{\bar{F}}_{N,m} - \bar{F}\|^2] \leq \mathfrak{C}(a) \left(\|\bar{G} - \bar{G}_m\|^2 + 4\mathbb{E}[Y_1^2] \frac{m}{n} + \frac{2\mathbb{E}[Y_1]}{n} \right) + \left(\frac{1-a}{1+a} \right)^{3N} \|\bar{F}\|^2, \quad (6.2.23)$$

where $\mathfrak{C}(a) = 8a^2 / ((1+a)^{3/2} - (1-a)^{3/2})^2$.

Inequality (6.2.23) provides a squared-bias/variance decomposition with bias proportional to $\|\bar{G} - \bar{G}_m\|^2$ and variance proportional to $\mathbb{E}[Y_1^2]m/n$. The term of order $\mathbb{E}[Y_1]/n$ is negligible, as well as the last one, for $N \geq \log(n) / [3 \log((1+a)/(1-a))]$ (if $a = 0.5$, and $n = 1000$, the condition is $N \geq 3$). If \bar{G} belongs to $W^s(\mathbb{R}^+, L)$ defined by (6.2.12), then choosing m_{opt}^* proportional to $n^{1/(s+1)}$ yields $\mathbb{E}[\|\check{\bar{F}}_{N,m_{\text{opt}}^*} - \bar{F}\|^2] = O(n^{s/(s+1)})$. The rate is better than the one obtained for density estimation. However, it remains a nonparametric rate while cumulative distribution functions are estimated with parametric rates in direct problems.

Then we proceed as in the density case for selecting m and set:

$$\check{m} = \operatorname{argmin}_{m \in \mathcal{M}_n} \{-\|\check{\bar{G}}_m\|^2 + \text{pen}(m)\}, \quad \text{pen}(m) = 2\check{\kappa} \hat{C}_2 \frac{m}{n} \quad (6.2.24)$$

where \hat{C}_2 is given by (6.2.16). The constant $\check{\kappa}$ is calibrated in the simulation part. We can prove the following oracle-type inequality of the final estimator $\check{\bar{F}}_{N,\check{m}}$.

Theorem 6.2.7. *If $\bar{F} \in \mathbb{L}^2(\mathbb{R}^+)$ and $\mathbb{E}[X_1^4] < \infty$, the final estimator $\check{\bar{F}}_{N,\check{m}}$ defined by (6.2.22) and (6.2.24) satisfies*

$$\mathbb{E}[\|\check{\bar{F}}_{N,\check{m}} - \bar{F}\|^2] \leq \mathfrak{C}(a) \left(6 \inf_{m \in \mathcal{M}_n} \{\|\bar{G} - \bar{G}_m\|^2 + \text{pen}(m)\} + \frac{D_a}{n} \right) + \left(\frac{1-a}{1+a} \right)^{3N} \|\bar{F}\|^2 \quad (6.2.25)$$

where $\mathfrak{C}(a)$ is defined in Proposition 6.2.6 and D_a is a constant depending on a .

Only a sketch of proof of the Theorem is given in Section 6.5.8, and we find $\check{\kappa} \geq 192$.

6.2.6 Case of unknown a

Parameter a is not identifiable, unless additional information is available. Two cases can be considered. First, if an additional K -sample is available, where the signal is a deterministic known constant, then we have a set of observations of U , say $U_1^{(1)}, \dots, U_K^{(1)}$. In this case, we can use the maximum likelihood estimator $\max_{1 \leq i \leq K} (|U_i^{(1)} - 1|)$ as an estimator of a with rate of convergence K (i.e. the mean square risk is of order $1/K^2$). Secondly, we can consider the model of repeated observations, where the variable X_i can be observed repeatedly, with independent errors:

$$Y_{i,k} = X_i U_{i,k}, \quad k \in \{1, 2\}, \quad i = 1, \dots, n,$$

where $(U_{i,1})_i$ and $(U_{i,2})_i$ are independent i.i.d. samples with distribution $\mathcal{U}([1 - a, 1 + a])$. Then we have

$$\mathbb{E} \begin{bmatrix} Y_{i,1}^2 \\ Y_{i,2}^2 \end{bmatrix} = \mathbb{E}[U_{i,1}^2] \mathbb{E} \begin{bmatrix} 1 \\ U_{i,2}^2 \end{bmatrix}, \quad E[U_{i,1}^2] = \frac{a^2}{3} + 1, \quad \mathbb{E} \begin{bmatrix} 1 \\ U_{i,2}^2 \end{bmatrix} = \frac{1}{(1 - a)(1 + a)},$$

which yields $\mathbb{E}[Y_{i,1}^2/Y_{i,2}^2] = (1 + a^2/3)/(1 - a^2)$. Therefore, we make the proposal

$$\hat{a}_n = \sqrt{\frac{\bar{W}_n - 1}{\bar{W}_n + 1/3}}, \quad \text{with } \bar{W}_n = \frac{1}{n} \sum_{i=1}^n W_i, \quad W_i := \frac{Y_{i,1}^2}{Y_{i,2}^2}. \quad (6.2.26)$$

Clearly, \hat{a}_n is a consistent estimator of a and by the limit central Theorem and the delta-method, we obtain the convergence in distribution

$$\sqrt{n}(\hat{a}_n - a) \xrightarrow{\mathcal{L}} Z, \quad Z \sim \mathcal{N}(0, \sigma^2(a)), \quad \sigma^2(a) = \frac{1 - a^2}{40}(15 + 8a^2 + a^4) \in (0, 0.375).$$

This estimator can be plugged into the previous estimation procedure.

6.3 Model transformation and deconvolution approach

We present now another estimation strategy, to which ours may be compared. The idea is to rewrite the model under an additive form by taking logarithm of (6.1.1) (see Es *et al.* (2005)). We obtain

$$Z_j := \log(Y_j) = \log(X_j) + \log(U_j) =: T_j + \varepsilon_j, \quad j = 1 \dots, n. \quad (6.3.1)$$

Estimating the density of T_1 in model (6.3.1) is a classical deconvolution problem on \mathbb{R} (see for example Comte *et al.* (2006)). Each sample $(Z_j)_j, (T_j)_j, (\varepsilon_j)_j$ is *i.i.d.* from density f_Z, f_T, f_ε respectively, and $(T_j)_j, (\varepsilon_j)_j$ are independent. They satisfy $f_Z = f_T \star f_\varepsilon$ where \star denotes the convolution product. Taking the Fourier transform of the equality implies $f_Z^* = f_T^* f_\varepsilon^*$. Then using the Fourier inversion formula, we get the following closed form for the density f_T ,

$$f_T(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-iux} \frac{f_Z^*(u)}{f_\varepsilon^*(u)} du, \quad x \in \mathbb{R}. \quad (6.3.2)$$

An estimator of f_T is obtained by replacing f_Z^* by its empirical counterpart, $\hat{f}_Z^*(u) = (1/n) \sum_{j=1}^n e^{iuZ_j}$. However, although formula (6.3.2) is well defined, the ratio $\hat{f}_Z^*/f_\varepsilon^*$ is not integrable on the whole real line, since f_ε^* tends to zero near infinity. Therefore, we do not only plug \hat{f}_Z^* in equation

(6.3.2) but we also introduce a cut-off which avoids integrability problems. Finally the estimator is defined by:

$$\tilde{f}_{T,\ell}(x) = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-iux} \frac{\widehat{f}_Z^*(u)}{f_\varepsilon^*(u)} du = \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-iux} \frac{1}{n} \sum_{j=1}^n \frac{e^{iuZ_j}}{f_\varepsilon^*(u)} du. \quad (6.3.3)$$

Clearly $\mathbb{E}[\tilde{f}_{T,\ell}(x)] = f_{T,\ell}(x)$ with

$$f_{T,\ell}(x) := \frac{1}{2\pi} \int_{-\pi\ell}^{\pi\ell} e^{-iux} f_T^*(u) du.$$

We can remark that, by Plancherel-Parseval formula, $\|f_{T,\ell} - f_T\|^2 = (2\pi)^{-1} \int_{|u| \geq \pi\ell} |f_T^*(u)|^2 du$. Then, with an additional bound on the variance, we recall the following result.

Proposition 6.3.1. *If $f_\varepsilon(u) \neq 0$, for all $u \in \mathbb{R}$, the estimator $\tilde{f}_{T,\ell}$ defined by (6.3.3), satisfies*

$$\mathbb{E}[\|\tilde{f}_{T,\ell} - f_T\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq \pi\ell} |f_T^*(u)|^2 du + \frac{1}{2\pi n} \int_{-\pi\ell}^{\pi\ell} \frac{du}{|f_\varepsilon^*(u)|^2}.$$

Several proofs of this bound can be found in the literature, see for example Comte & Lacour (2011), Chapter 2. Using $\varepsilon_j = \log(U_j)$, we have

$$f_\varepsilon^*(u) = \frac{1}{2a} \int e^{iu \log(t)} \mathbb{1}_{[1-a, 1+a]}(t) dt = \frac{(1+a)e^{\log(1+a)iu} - (a-1)e^{\log(1-a)iu}}{2a(1+iu)},$$

and

$$|f_\varepsilon^*(u)|^2 = \frac{1+a^2 - (1-a^2)\cos(u \log((1+a)/(1-a)))}{2a^2(1+u^2)} \quad (6.3.4)$$

which never reaches zero, as $0 < a < 1$. Besides, $1/|f_\varepsilon^*(u)|^2 \leq 2a^2(1+u^2)/(2a^2) = 1+u^2$, for $u \in \mathbb{R}$. Therefore, Proposition 6.3.1 writes in the present case

$$\mathbb{E}[\|\tilde{f}_{T,\ell} - f_T\|^2] \leq \frac{1}{2\pi} \int_{|u| \geq \pi\ell} |f_T^*(u)|^2 du + \frac{\ell}{n} + \frac{\pi^2 \ell^3}{3n}. \quad (6.3.5)$$

We can see that here ℓ plays the role of m previously, and we have to choose it in order to make a compromise between the squared bias term $(2\pi)^{-1} \int_{|u| \geq \pi\ell} |f_T^*(u)|^2 du$ which decreases when ℓ increases and the variance term (with main term $\pi^2 \ell^3 / (3n)$) which increases when ℓ increases. Thus as previously, writing that $(2\pi)^{-1} \int_{|u| \geq \pi\ell} |f_T^*(u)|^2 du = \|f_T\|^2 - \|f_{T,\ell}\|^2$, we omit the constant term $\|f_T\|^2$ and estimate the second term by $-\|f_{T,\ell}\|^2$; then we replace the variance by its upper bound, up to a multiplicative constant. Finally, we set

$$\tilde{\ell} = \operatorname{argmin}_{\ell \in \mathcal{M}_n} \{-\|\tilde{f}_{T,\ell}\|^2 + \widehat{\text{pen}}(\ell)\}, \quad \text{with } \widehat{\text{pen}}(\ell) := \tilde{\kappa} \left(\frac{\ell}{n} + \frac{\pi^2 \ell^3}{3n} \right), \quad (6.3.6)$$

where $\tilde{\kappa}$ is a numerical constant calibrated in Section 6.4.1. We can prove for the estimator $\tilde{f}_{T,\tilde{\ell}}$ a non-asymptotic oracle-type inequality.

Theorem 6.3.2. *The estimator $\tilde{f}_{T,\tilde{\ell}}$ defined by (6.3.3) and (6.3.6) satisfies*

$$\mathbb{E}[\|\tilde{f}_{T,\tilde{\ell}} - f_T\|^2] \leq 4 \inf_{\ell \in \mathcal{M}_n} \{\|f_{T,\ell} - f_T\|^2 + \widehat{\text{pen}}(\ell)\} + \frac{K}{n}$$

with K a numerical constant and $\tilde{\kappa} \geq 4$.

Finally to estimate f (the density of X) we have to apply the following relations:

$$f(v) = f_T(\log(v))/v, \quad f_T(v) = f_{\log(X)}(v) = f(e^v)e^v.$$

We define the estimator of f by

$$\tilde{f}_{\tilde{\ell}}(x) := \tilde{f}_{T, \tilde{\ell}}(\log(x))/x. \tag{6.3.7}$$

We can see on this definition that the estimator is not defined near zero, thus we have to consider the truncated integral

$$\int_{\alpha}^{+\infty} \left(\tilde{f}_{\tilde{\ell}}(x) - f(x) \right)^2 dx \leq \frac{1}{\alpha} \|\tilde{f}_{T, \tilde{\ell}} - f_T\|^2$$

to obtain a bound on the risk: for any $\alpha > 0$

$$\mathbb{E} \left[\int_{\alpha}^{+\infty} (\tilde{f}_{\tilde{\ell}}(x) - f(x))^2 dx \right] \leq \frac{4}{\alpha} \inf_{\ell \in \mathcal{M}_n} \{ \|f_{T, \ell} - f_T\|^2 + \widetilde{\text{pen}}(\ell) \} + \frac{K}{\alpha n}.$$

We can see on these bounds that, the smaller α , the larger the bound. This is clearly confirmed by the simulations hereafter.

6.4 Numerical study

6.4.1 Simulated data

In this Section we evaluate our estimators of the density and the survival function on simulated data. We compute three estimators: the estimators of f , $\hat{f}_{N, \hat{m}}$ given by (6.2.17) and $\tilde{f}_{\tilde{\ell}}$ given by (6.3.7) and the estimator of \bar{F} , $\tilde{F}_{N, \tilde{m}}$, given by (6.2.22). For each estimator, there is a preliminary step before estimating the target function. Indeed, we first compute the collection of projection estimators of function g : \hat{g}_m . Then we implement the selection procedure for the dimension parameter m . We obtain the final estimator of g : $\hat{g}_{\hat{m}}$. Finally, applying formula (6.2.17) with $N = 30$ (which satisfies constraint (6.2.11)), we obtain our final estimator $\hat{f}_{30, \hat{m}}$ of f . The estimation procedure is implemented similarly for the survival function \bar{F} .

For the deconvolution density estimator we first estimate the density $f_T = f_{\log(X)}$ with the collection $\tilde{f}_{T, \ell}$ as given by (6.3.3). The integrals are computed using Riemann approximations with thin discretisations. We select the best cut-off parameter ℓ among the collection, according to the criterion given in Section 6.3. Finally we use formula (6.3.7) to obtain $\tilde{f}_{\tilde{\ell}}$.

Each selection procedure depends on a parameter which has to be calibrated, namely κ_1, κ_2 in (6.2.16), $\tilde{\kappa}$ in (6.2.24), $\tilde{\kappa}$ in (6.3.6). They are chosen from preliminary simulation experiments. Different cases of density f have been investigated with different parameter values, and a large number of repetitions. Comparing the MISE obtained as functions of the constants of interest, yields to select values making a good compromise over all experiences. We choose: $\kappa_1 = 0.5$, $\kappa_2 = 0.01$, $\tilde{\kappa} = 0.3$, $\tilde{\kappa} = 4$. In the following we investigate 3 densities for X :

- $\Gamma(4, 0.5)$.
- $\mathcal{E}(1)$
- $0.5\Gamma(2, 0.4) + 0.5\Gamma(11, 0.4)$

The first one is uni-modal and 0 in 0, the second one is decreasing and is 1 in 0: we are indeed interested in the behaviour near 0 of the estimators. The last one is bi-modal. For each density, we could start the estimation procedure near $x = 0$ for our estimator $\hat{f}_{N, \hat{m}}$. But in order to compare our estimator with the deconvolution estimator $\tilde{f}_{\tilde{\ell}}$ which is not defined in 0, we start in $x = 0.1$ for all the grids of density estimation. Figure 6.1 illustrates the kind of data generated

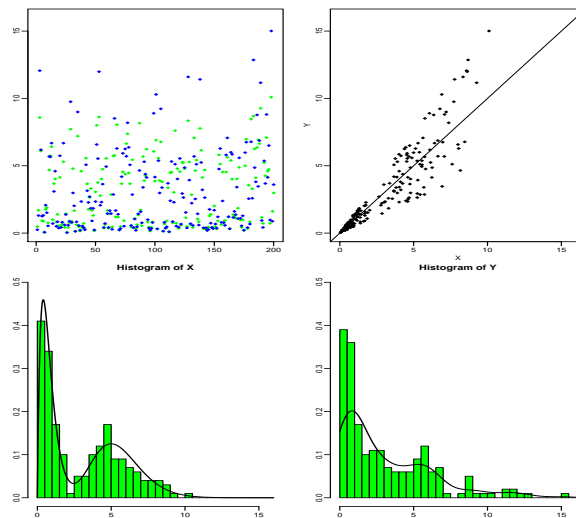


Figure 6.1 – Example of database when $X \sim 0.5\Gamma(2, 0.4) + 0.5\Gamma(11, 0.5)$, $a = 0.5$, $n = 200$. Top left: plot of X green (grey) and Y blue (black). Top right: Y as a function of X. Bottom left: histogram of X with the true density f , bottom right: histogram of Y with a projection Laguerre estimator of f_Y applied on the (Y_i) 's.

by the model and the effect of the censoring variable. It is a real issue to successfully reconstruct the density of X from the censored data Y .

Let us first comment the density estimation procedure. For the projection estimator $\hat{f}_{N,\hat{m}}$ we choose $m_{max} = 10$ or 15 because the selected m are small most of the time. For the deconvolution estimator: $\ell_{max} = 10$ and the selected ℓ are often small $(1,2,3)$.

Figure 6.2 illustrates the good performances of our estimation procedure by projection. We represent 20 estimators $\hat{f}_{N,\hat{m}}$ of f (for 20 simulated samples) in the exponential case and the mixed-gamma case, and the beam of estimators are very close and close to the true density. On Figure 6.3, we can see both estimators $\hat{f}_{N,\hat{m}}$, \tilde{f}_{ℓ} and the true density. We also plot on this graph the projection estimator of density f_Y from the observations $(Y_i)_i$. It is defined for observations $(Z_i)_i$,

$$\hat{f}_{Z,m} = \sum_{j=0}^{m-1} \hat{b}_j \varphi_j \quad \text{with} \quad \hat{b}_j = \frac{1}{n} \sum_{i=1}^n \varphi_j(Z_i), \tag{6.4.1}$$

and $\hat{\hat{m}} = \operatorname{argmin}_{m \in \mathcal{M}_n} \{-\|\hat{f}_{Z,m}\|^2 + m/n\}$ (the calibration constant has been chosen equal to 1 here).

We can notice from the graph that estimator \tilde{f}_{ℓ} is closer to $\hat{f}_{Y,\hat{\hat{m}}}$ (6.4.1) and f_Y than to f , the target function. However, estimator $\hat{f}_{N,\hat{m}}$ catches the difference between f and f_Y which is the aim here, and fits well the true density f of sample $(X_i)_i$.

Then we compute approximation of the MISE from 100 or 200 Monte-Carlo simulations. The number of repetitions has been checked to be large enough to insure the stability of the MISEs. They are multiplied by 100 and summed up in Table 6.1 for different values of parameter a and of the number of observations n , to complete the illustration. When a goes from 0.25 to 0.5 the estimation is more difficult, and this increases the value of the errors. Likewise, when n increases from 200 to 10000 the estimation is easier and the MISEs are smaller. We can see again that the results are specifically good for exponential densities. For the mixed-gamma case function g is hard to estimate because it has 2 modes, thus the estimation of f is also difficult and requires more observations, see the third line of Table 6.1. Still according to Table 6.1, the

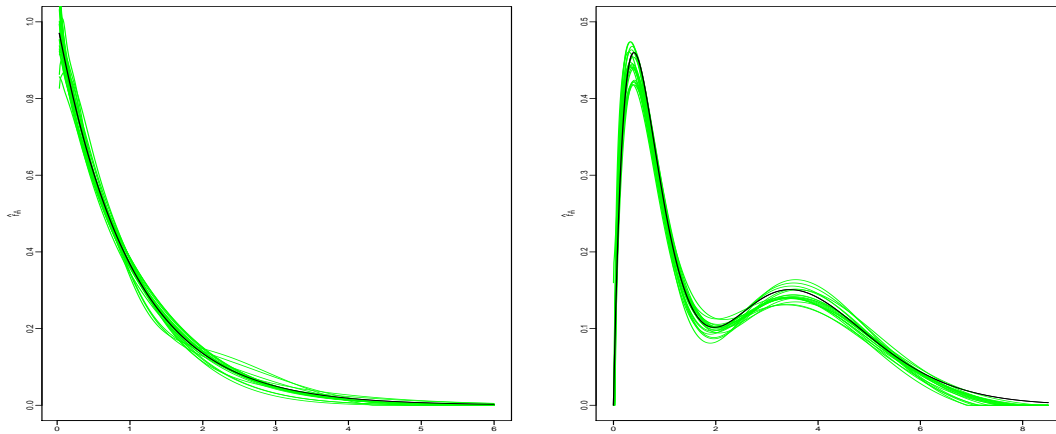


Figure 6.2 – 20 estimator $\hat{f}_{N,\hat{m}}$ of f in plain green line (grey) versus the true density f in black bold plain line: on the left when $X \sim \mathcal{E}(1)$ with $a = 0.5$, $n = 1000$; on the right when $X \sim 0.5\Gamma(4, 0.25) + 0.5\Gamma(20, 0.5)$, with $a = 0.25$, $n = 1000$.

projection estimator performs better than the deconvolution method. All along, it has been seen that the projection method is computationally faster than the deconvolution strategy. Besides, the deconvolution estimator is very unstable around zero.

Remark. Estimation at $x = 0$. Note that by definition of function g (6.2.4) $g(0) = 0$, then $f_N(0) = 0$ and the projection estimator $\hat{f}_{N,\hat{m}}$ may have to be corrected in point zero if the true density is non-zero in zero. But, we can see that, if the function f is continuous in 0^+ , then $\lim_{y \rightarrow 0} f_Y(y) = f(0) \log((1+a)/(1-a))/(2a)$. This implies that estimating f_Y in zero by a direct the projection estimator of f_Y relying on the Laguerre basis ($\hat{f}_{Y,\hat{m}}$) and applying the multiplicative correction factor $2a/\log((1+a)/(1-a))$ should be an adequate approximation of f near zero. On Figure 2 the grid begins in 0.03 for the exponential density and in 0 for the mixed-gamma density. If the statistician wants to start the estimation in 0, the plugging of corrected $\hat{f}_{Y,\hat{m}}(0)$ for the first value of estimator $\hat{f}_{N,\hat{m}}$ is a good strategy.

For the estimation of the survival function the grid of estimation begins in 0. We choose for the maximal dimension $m_{max} = 10, 15, 20$ ($n = 200, 1000, 10000$) and the selected m are small most of the time. The left graph of Figure 6.4 illustrates the good estimation of the survival function of X when it has an exponential distribution with parameter 1 from observations $(Y_i)_i$. On the right, the second graph shows the mixed-gamma case: our estimator $\check{\bar{F}}_{N,\check{m}}$ (plain grey line) detects well the bimodal character of the density (true \bar{F} in plain black line). We also represent the empirical distribution function $\bar{F}_{Y,n}$ in dotted grey line, given for a sample $(Z_i)_i$ by:

$$\bar{F}_{Z,n}(t) = 1 - \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{Z_i \leq t}. \tag{6.4.2}$$

We can see that this function is not a good approximation of \bar{F} when $a = 0.5$. To confirm this fact, Table 6.2 provides the MISEs (times 100) for estimator $\check{\bar{F}}_{N,\check{m}}$ of \bar{F} . They can be compared to the MISEs of estimators of \bar{F} : $\bar{F}_{Y,n}$ (available in practice) and $\bar{F}_{X,n}$ (not available in practice). This table highlights the quality of our estimator when $a = 0.5$ (results in bold black). When $a = 0.25$ as expected $\bar{F}_{Y,n}$ can be considered as a satisfying approximation of the survival function of X , except for the exponential case, where our estimator is the best. Again, when a increases the MISEs are higher and with n increases the MISEs are smaller.

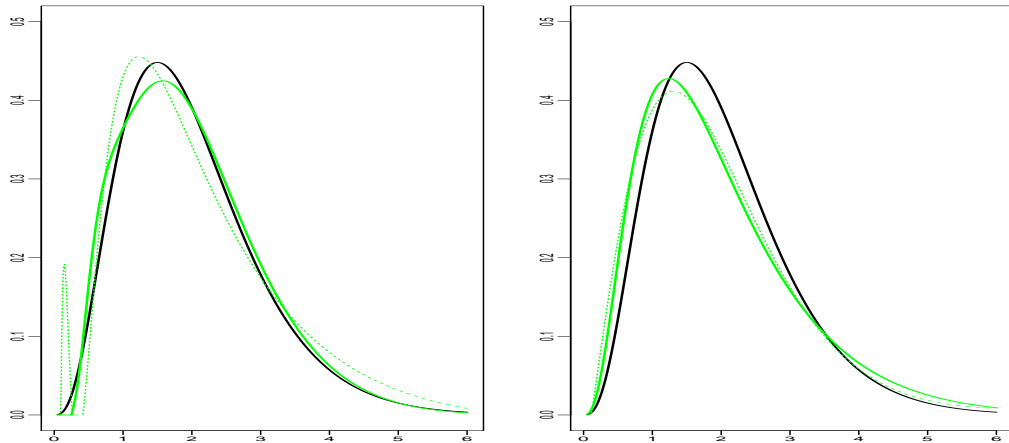


Figure 6.3 – Gamma case: $X \sim \Gamma(4, 0.5)$, $a = 0.5$, $n = 1000$. Left graph: f in bold black line, estimator $\hat{f}_{N,\hat{m}}$ of f in plain bold green (grey) line, estimator $\tilde{f}_{\hat{\ell}}$ of f in thin dotted green line (grey). Right graph: f in bold black line, f_Y in bold green line (grey), estimator of f_Y by projection $\hat{f}_{Y,\hat{m}}$ in dotted green line (grey).

Table 6.1 – MISE for the estimators of f : $\hat{f}_{N,\hat{m}}$ and $\tilde{f}_{\hat{\ell}}$, times 100, with 200 repetitions for $n = 200, 1000$ and 100 for $n = 10000$.

Distribution of f	a	Estimator $\hat{f}_{N,\hat{m}}$			Estimator $\tilde{f}_{\hat{\ell}}$		
		$n = 200$	$n = 1000$	$n = 10000$	$n = 200$	$n = 1000$	$n = 10000$
Exponential	0.25	0.386	0.075	0.006	0.703	0.153	0.024
	0.5	0.470	0.095	0.009	0.964	0.231	0.030
Gamma	0.25	0.538	0.110	0.014	1.122	0.987	0.017
	0.5	0.972	0.394	0.154	1.589	1.851	0.217
Mixed-gamma	0.25	1.070	0.146	0.015	1.603	0.346	0.048
	0.5	1.441	0.563	0.208	2.703	2.833	0.337

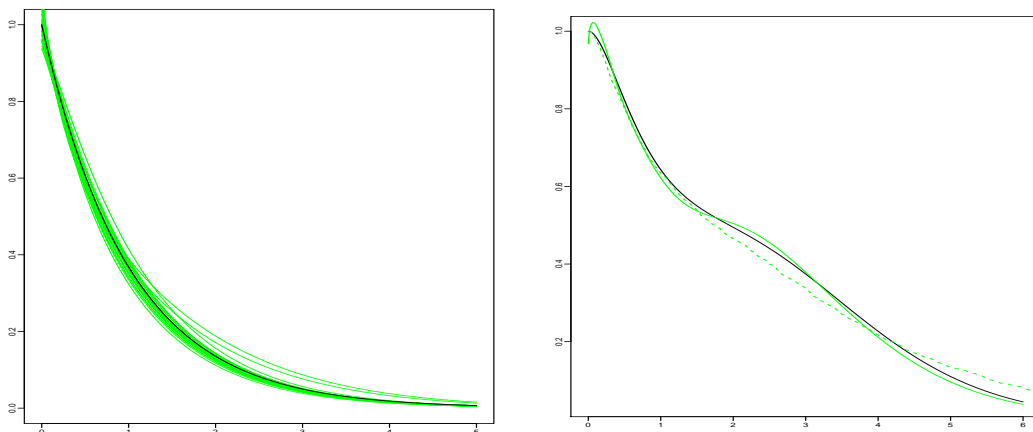


Figure 6.4 – Left: 20 estimators $\check{\check{F}}_{N,\check{m}}$ of $\check{\check{F}}$ in plain green line (grey) versus the true function $\check{\check{F}}$ in black bold plain line: when $X \sim \mathcal{E}(1)$, $a = 0.25$, $n = 200$. Right: estimator $\check{\check{F}}_{N,\check{m}}$ in bold plain green line (grey) and empirical distribution $\check{\check{F}}_n$ of Y in dotted bold green, line (grey), when $X \sim 0.5\Gamma(4, 0.25) + 0.5\Gamma(20, 0.5)$ ($\check{\check{F}}$ in black plain bold line), $a = 0.5$, $n = 1000$.

Table 6.2 – MISE for the estimators of \bar{F} : $\check{\bar{F}}_{N,\check{m}}$, $\bar{F}_{Y,n}$, $\bar{F}_{X,n}$, times 100, with 200 repetitions for $n = 200, 1000$ and 100 for $n = 10000$.

Distribution f	a	$\check{\bar{F}}_{N,\check{m}}$			$\bar{F}_{Y,n}$		$\bar{F}_{X,n}$	
		$n = 200$	$n = 1000$	$n = 10000$	$n = 200$	$n = 1000$	$n = 200$	$n = 1000$
Exponential	0.25	0.194	0.043	0.026	0.253	0.055	0.248	0.054
	0.5	0.269	0.072	0.027	0.277	0.106	0.234	0.054
Gamma	0.25	0.269	0.151	0.121	0.260	0.081	0.245	0.054
	0.5	0.500	0.133	0.121	0.756	0.497	0.281	0.057
Mixed-gamma	0.25	0.677	0.175	0.098	0.610	0.126	0.557	0.097
	0.5	0.888	0.225	0.126	0.855	0.430	0.517	0.102

6.4.2 Application

Klein *et al.* (2013) detail the problem of confidential protection of data. The issue is how to alter the data before releasing it to the public in order to minimize the risk of disclosure and at the same time to remain able to find the main characteristics of the original dataset when the level of noise is known.

The multiplicative noise perturbation can be proposed in this context. Sinha *et al.* (2011) investigate this method on $n = 51$ data (percentages of people under poverty level for each state of United States), different noise distributions, among which a uniform density $\mathcal{U}_{[1-a, 1+a]}$ for $a = 0.1$ (the data set is publicly available from the American Community Survey (ACS) via <http://factfinder.census.gov>). The question is: how can the moments, the quantiles, the minimal value, maximal value of the sample X be estimated from the observations Y_i . They propose a strategy which delivers good results. But, looking at the noisy data Y_i one can see that they are very close from the true ones and thus in that case the privacy may be not insured. We illustrate this fact on Figure 6.5: it represents the multiplicative noise scenario, with $a = 0.1$ on the left and $a = 0.5$ on the right, for the original data $(X_i)_{i=1, \dots, n=51}$ from Sinha *et al.* (2011). The three graphs are: top left the histogram of the $(X_i)_i$ the real data, top right an histogram of $(Y_i = X_i U_i)_i$ and on the bottom a plot of Y versus X .

Thus here we choose to illustrate the second choice: $a = 0.5$. What is the estimated density of X from these observations $(Y_i = X_i U_i)_i$? Are we capable of giving predictions of the data from this estimated density? What are the mean, the min, the max, the main quantiles of our new sample?

Figure 6.6 shows the estimator $\hat{f}_{30, \hat{m}}$ of f from the $(Y_i)_i$, the projection estimator of f on the sample $(X_i)_i$: $\hat{f}_{X, \hat{m}}$ (a benchmark, not available in practice) and $\hat{f}_{Y, \hat{m}}$ the projection estimator of f_Y on the $(Y_i)_i$. It seems that the two densities are very different. The quality of the method is asserted by the fact that $\hat{f}_{X, \hat{m}}$ and $\hat{f}_{30, \hat{m}}$ are very close. Then, from the estimator $\hat{f}_{30, \hat{m}}$ we simulate a new sample $(X_{\text{pred}_i})_i$ of length $n = 51$. To do so, we generate a *discrete variable* because we have a discrete version of the estimator of the density function f . The graph of the sorted new sample versus the sorted original sample is presented of Figure 6.7. The lining up of the values confirms the goodness of our estimator $\hat{f}_{N, \hat{m}}$ from the noisy observations $(Y_i)_i$. Finally, from Monte-Carlo simulations of 1000 new samples $(X_{\text{pred}_i})_i$ we can compare the quantities of interest of $(X_i)_i$ (not available), of $(Y_i)_i$ (noisy sample) and the empiric quantities obtained from the new samples $(X_{\text{pred}_i})_i$. They are given in Table 6.3. Except for the third quantile Q3 at (75 %), the information we get from our new sample is very close from the information from X .

The proposed procedure allows to correctly mask the data and to recover the main informa-

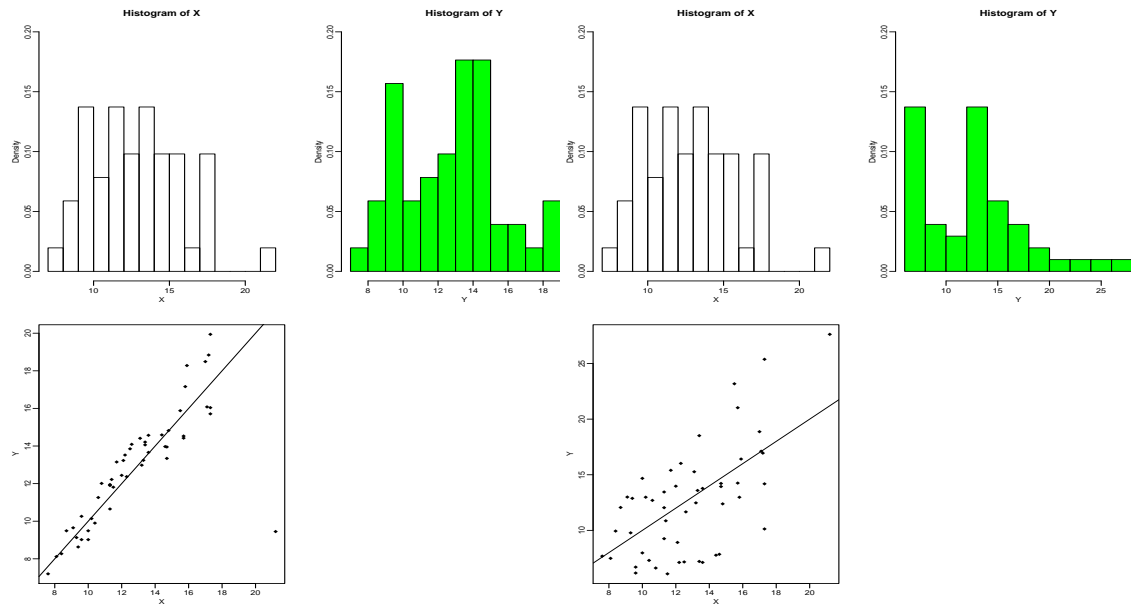


Figure 6.5 – Illustration of uniform noise multiplication on real data. Three graphs for $a = 0.1$ on the left and $a = 0.5$ on the right. Top left histogram of $(X_i)_i$, top right histogram of $(Y_i)_i$, bottom plot of Y_i versus X_i .

	Mean	Standard deviation	Minimum	Maximum	Q1	Median	Q3
X	12.82	2.98	7.6	21.2	10.5	12.5	14.75
Y	12.59	4.93	6.10	27.6	7.91	12.70	14.46
X_{pred}	12.78	3.09	7.18	19.31	10.42	12.77	14.54

Table 6.3 – Comparison of characteristic quantities from samples $(X_i)_i, (Y_i)_i, (X_{\text{pred}})_i$

tion from the original sample, as soon as the level of noise (given by a) is known. The method is easy to use in practice and insures the privacy protection of the data.

6.5 Proofs

6.5.1 Proof of Lemma 6.2.1

Denote F the cumulative distribution function of X_1 , it comes the bounds

$$\frac{1-a}{2a} \int_{\frac{y}{1+a}}^{\frac{y}{1-a}} f(x)dx \leq yf_Y(y) \leq \frac{1+a}{2a} \int_{\frac{y}{1+a}}^{\frac{y}{1-a}} f(x)dx$$

$$\frac{1-a}{2a} \left[F\left(\frac{y}{1-a}\right) - F\left(\frac{y}{1+a}\right) \right] \leq yf_Y(y) \leq \frac{1+a}{2a} \left[F\left(\frac{y}{1-a}\right) - F\left(\frac{y}{1+a}\right) \right] \quad (6.5.1)$$

Equation (6.5.1) shows that $yf_Y(y) \xrightarrow{y \rightarrow 0} 0$ and $yf_Y(y) \xrightarrow{y \rightarrow +\infty} 0$. \square

6.5.2 Useful properties of the Laguerre basis

Property 6.5.1. *If $t \in \mathcal{S}_m$, (1) $\|t\|_\infty \leq \sqrt{2m}\|t\|$, (2) $\|t'\|_\infty \leq 2\sqrt{2m}^{3/2}\|t\|$ and (3) If $\|t\| = 1$, $\|t'\| \leq 1 + \sqrt{2m(m-1)}$.*

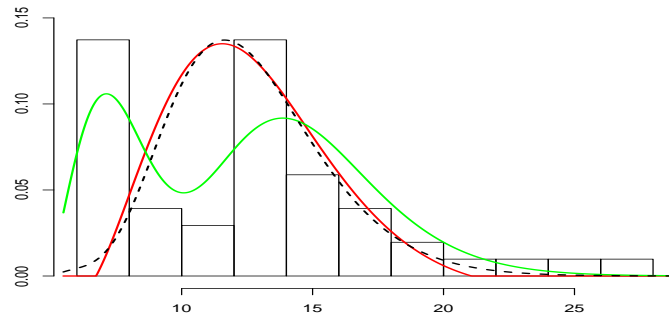


Figure 6.6 – Histogram of the real data X_i 's with full multiplicative noise, with $a = 0.5$, $Y_i = X_i U_i$. Dotted black line estimator $\hat{f}_{X, \hat{m}}$ of f on the $(X_i)_i$, plain red line (black) $\hat{f}_{N, \hat{m}}$ estimator of f on the $(Y_i)_i$, plain green line (grey) line estimator $\hat{f}_{Y, \hat{m}}$ of f_Y on the $(Y_i)_i$.

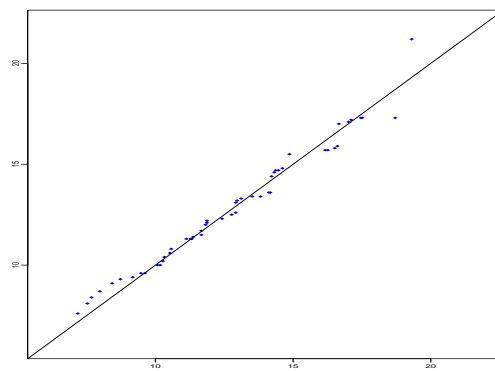


Figure 6.7 – Plot of the new predictive sample $(X_{\text{pred}_i})_i$ versus original data $(X_i)_i$.

The two first points are direct consequences of (6.2.3). The last point comes from the following Lemma.

Lemma 6.5.2. *For all $j \in \mathbb{N}$, the Laguerre basis function $(\varphi_j)_j$ satisfies:*

$$\varphi'_0(x) = -\varphi_0(x), \quad \varphi'_j(x) = -\varphi_j(x) - 2 \sum_{k=0}^{j-1} \varphi_k(x), \quad j \geq 1. \quad (6.5.2)$$

Considering $t \in S_m$, such that $\|t\| = 1$, $t(x) = \sum_{j=0}^{m-1} a_j \varphi_j(x)$, then

$$\begin{aligned} t'(x) &= \sum_{j=0}^{m-1} a_j \varphi'_j(x) = \sum_{j=1}^{m-1} a_j \left(-\varphi_j(x) - 2 \sum_{k=0}^{j-1} \varphi_k(x) \right) - a_0 \varphi_0(x) \\ &= - \sum_{j=0}^{m-1} a_j \varphi_j(x) - 2 \sum_{j=1}^{m-1} a_j \left(\sum_{k=0}^{j-1} \varphi_k(x) \right). \end{aligned}$$

Then, $\|t'\| \leq \|t\| + 2 \left\| \sum_{j=1}^{m-1} a_j \left(\sum_{k=0}^{j-1} \varphi_k \right) \right\|$

$$\left(\sum_{j=1}^{m-1} a_j \left(\sum_{k=0}^{j-1} \varphi_k(x) \right) \right)^2 \leq \sum_{j=1}^{m-1} a_j^2 \left(\sum_{j=1}^{m-1} (\varphi_0 + \varphi_1 + \dots + \varphi_{j-1}) \right)^2 = \|t\|^2 \sum_{j=1}^{m-1} (\varphi_0 + \varphi_1 + \dots + \varphi_{j-1})^2$$

thus using that $\|t\| = 1$ and integrating, as the (φ_j) 's form a b.o.n, it yields

$$\left\| \sum_{j=1}^{m-1} a_j \left(\sum_{k=0}^{j-1} \varphi_k \right) \right\|^2 \leq \sum_{j=1}^{m-1} (\|\varphi_0\|^2 + \|\varphi_1\|^2 + \dots + \|\varphi_{j-1}\|^2) \leq \sum_{j=1}^{m-1} j = m(m-1)/2.$$

Finally $\|t'\| \leq 1 + \sqrt{2m(m-1)}$. \square

Proof of Lemma 6.5.2.

The following equality holds $\varphi'_j(x) = -\varphi_j(x) + 2\sqrt{2}e^{-x}L'_j(2x)$ which is a polynomial function of degree j multiplied by e^{-x} . Thus, it could be decomposed as $\varphi'_j(x) = \sum_{k=0}^j a_k^{(j)} \varphi_k(x)$ with

$$\begin{aligned} a_k^{(j)} &= \langle \varphi'_j, \varphi_k \rangle = \int_0^{+\infty} \varphi'_j(x) \varphi_k(x) dx = [\varphi_j(x) \varphi_k(x)]_0^{+\infty} - \int_0^{+\infty} \varphi_j(x) \varphi'_k(x) dx \\ &= -\varphi_j(0) \varphi_k(0) - \int_0^{+\infty} \varphi_j(x) \varphi'_k(x) dx = -2 - 2 \langle \varphi_j, \varphi'_k \rangle = -2 - 2a_j^{(k)} \end{aligned}$$

Notice that this formula is also true when $k = j$: $\langle \varphi'_j, \varphi_j \rangle = \int_0^{+\infty} \varphi'_j(x) \varphi_j(x) dx = -(1/2) \varphi_j^2(0) = -2/2 = -1$. Thus we obtain:

$$\begin{aligned} \varphi'_j(x) &= \sum_{k=0}^j (-2 - \langle \varphi'_j, \varphi_k \rangle) \varphi_k(x) = -2 \sum_{k=0}^j \varphi_k(x) - \sum_{k=0}^j \langle \varphi_j, \varphi'_k \rangle \varphi_k(x) \\ &= -\varphi_j(x) - 2 \sum_{k=0}^{j-1} \varphi_k(x) - \sum_{k=0}^{j-1} \langle \varphi_j, \varphi'_k \rangle \varphi_k(x) \end{aligned}$$

Or the $\langle \varphi_j, \varphi'_k \rangle$ are zero for $k \leq j-1$. Thus we obtain (6.5.2). \square

6.5.3 Proof of Proposition 6.2.2

Proof of (i). To compute $\mathbb{E}[\|g_m - \widehat{g}_m\|^2]$ we start by noting that $\|g_m - \widehat{g}_m\|^2 = \sum_{j=0}^{m-1} (\widehat{a}_j - a_j(g))^2$.

This implies

$$\mathbb{E}[\|g_m - \widehat{g}_m\|^2] = \sum_{j=0}^{m-1} \text{Var}(\widehat{a}_j) = \frac{1}{n} \sum_{j=0}^{m-1} \text{Var}(Y_1 \varphi'_j(Y_1) + \varphi_j(Y_1)) \leq \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E}[(Y_1 \varphi'_j(Y_1) + \varphi_j(Y_1))^2].$$

Now, Equation (6.2.5) applied with $t = \varphi_j^2$ and (6.2.3) lead to

$$\begin{aligned} \mathbb{E}[(Y_1 \varphi'_j(Y_1) + \varphi_j(Y_1))^2] &\leq \mathbb{E}[2Y_1^2 \varphi_j'^2(Y_1) + 2\varphi_j(Y_1)^2] \leq 2\|\varphi_j'\|_\infty^2 \mathbb{E}[Y_1^2] + 2\|\varphi_j\|_\infty^2 \\ &\leq 16(j+1)^2 \mathbb{E}[Y_1^2] + 4. \end{aligned}$$

As $8 \sum_{j=0}^{m-1} (j+1)^2 = 8 \sum_{j=1}^m j^2 \leq 8m^3$, it yields

$$\mathbb{E}[\|g_m - \widehat{g}_m\|^2] \leq 16\mathbb{E}[Y_1^2] \frac{m^3}{n} + 4 \frac{m}{n}, \quad (6.5.3)$$

which is the result (i). \square

Proof of (ii). Let us study the mean of the estimator of f :

$$\begin{aligned} \mathbb{E}[\widehat{f}_{N,m}(x)] &= 2a \sum_{k=0}^{N-1} \mathbb{E} \left[\widehat{g}_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right) \right] = 2a \sum_{k=0}^{N-1} \sum_{j=1}^m a_j \varphi_j \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right) \\ &= 2a \sum_{k=0}^{N-1} g_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right) := f_{N,m}(x). \end{aligned}$$

we denote

$$f_{N,m}(x) := 2a \sum_{k=0}^{N-1} g_m \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right). \quad (6.5.4)$$

Thus the estimator $\widehat{f}_{N,m}$ is an unbiased estimator of $f_{N,m}$ and

$$\mathbb{E}[\|\widehat{f}_{N,m} - f\|^2] = \|f - f_{N,m}\|^2 + \mathbb{E}[\|f_{N,m} - \widehat{f}_{N,m}\|^2]. \quad (6.5.5)$$

In the following we denote by h_k the composition of h with the function

$$x \mapsto \left(\frac{1+a}{1-a} \right)^k (1+a)x.$$

We note that for any function $h \in \mathbb{L}^2(\mathbb{R}^+)$,

$$\|h_k\|^2 = \int h^2 \left(\left(\frac{1+a}{1-a} \right)^k (1+a)x \right) dx = \frac{1}{1+a} \left(\frac{1-a}{1+a} \right)^k \int h^2(y) dy = \frac{1}{1+a} \left(\frac{1-a}{1+a} \right)^k \|h\|^2. \quad (6.5.6)$$

Let us study the bias term $\|f - f_{N,m}\|^2$:

$$\begin{aligned} (f - f_{N,m})(x) &= 2a \sum_{k=0}^{N-1} g_k(x) + f\left(\left(\frac{1+a}{1-a}\right)^N x\right) - 2a \sum_{k=0}^{N-1} g_{m,k}(x) \\ &= 2a \sum_{k=0}^{N-1} (g_k - g_{m,k})(x) + f\left(\left(\frac{1+a}{1-a}\right)^N x\right). \end{aligned}$$

The triangular inequality gives

$$\|f - f_{N,m}\| \leq 2a \sum_{k=0}^{N-1} \|g_k - g_{m,k}\| + \left\| f\left(\left(\frac{1+a}{1-a}\right)^N \cdot\right) \right\|. \quad (6.5.7)$$

As a consequence, using (6.5.6), we get

$$\sum_{k=N}^{\infty} \|g_k\| = \|g\| \sum_{k=N}^{\infty} \frac{1}{\sqrt{1+a}} \left(\frac{1-a}{1+a}\right)^{k/2} = \frac{\left(\frac{1-a}{1+a}\right)^{N/2}}{\sqrt{1+a} - \sqrt{1-a}} \|g\|. \quad (6.5.8)$$

Furthermore

$$\begin{aligned} \sum_{k=0}^{N-1} \|g_k - g_{m,k}\| &= \sum_{k=0}^{N-1} \frac{1}{\sqrt{1+a}} \left(\frac{1-a}{1+a}\right)^{k/2} \|g - g_m\| = \frac{1 - \left(\frac{1-a}{1+a}\right)^{N/2}}{\sqrt{1+a} - \sqrt{1-a}} \|g - g_m\| \\ &\leq \frac{\|g - g_m\|}{\sqrt{1+a} - \sqrt{1-a}}. \end{aligned} \quad (6.5.9)$$

Furthermore, $\|f\left(\left(\frac{1+a}{1-a}\right)^N \cdot\right)\| = \left(\frac{1-a}{1+a}\right)^{N/2} \|f\|$, and plugging this and (6.5.9) in (6.5.7), we obtain

$$\|f - f_{N,m}\|^2 \leq \frac{8a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} \|g - g_m\|^2 + 2 \left(\frac{1-a}{1+a}\right)^N \|f\|^2. \quad (6.5.10)$$

For the variance term, we study $\|f_{N,m} - \widehat{f}_{N,m}\|^2$. We easily obtain

$$\|f_{N,m} - \widehat{f}_{N,m}\| = 2a \left\| \sum_{k=0}^{N-1} (\widehat{g}_{m,k} - g_{m,k}) \right\| \leq 2a \sum_{k=0}^{N-1} \|\widehat{g}_m - g_m\| \frac{1}{\sqrt{1+a}} \left(\frac{1-a}{1+a}\right)^{k/2}$$

and finally

$$\|f_{N,m} - \widehat{f}_{N,m}\| \leq 2a \frac{1 - \left(\frac{1-a}{1+a}\right)^{N/2}}{\sqrt{1+a} - \sqrt{1-a}} \|\widehat{g}_m - g_m\| \quad (6.5.11)$$

and $\mathbb{E}[\|\widehat{g}_m - g_m\|]$ has been evaluated in (6.5.3). Gathering (6.5.5), (6.5.10) and (6.5.11) implies (ii). \square

6.5.4 Proof of Theorem 6.2.4

First, by the Cauchy-Schwarz inequality, we have

$$\mathbb{E}[\|\widehat{f}_{N,\widehat{m}} - f\|^2] \leq 2\mathbb{E}[\|f - f_{\widehat{m},N}\|^2] + 2\mathbb{E}[\|f_{N,\widehat{m}} - \widehat{f}_{N,\widehat{m}}\|^2]$$

Then we apply (6.5.10) and (6.5.11):

$$\begin{aligned} \mathbb{E}[\|\widehat{f}_{N,\widehat{m}} - f\|^2] &\leq 4 \left(\frac{1-a}{1+a} \right)^N \|f\|^2 + \frac{16a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} (\mathbb{E}[\|g - g_{\widehat{m}}\|^2] + \mathbb{E}[\|\widehat{g}_{\widehat{m}} - g_{\widehat{m}}\|^2]) \\ &= 4 \left(\frac{1-a}{1+a} \right)^N \|f\|^2 + \frac{16a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} (\mathbb{E}[\|g - \widehat{g}_{\widehat{m}}\|^2]). \end{aligned} \quad (6.5.12)$$

The last term is the MISE of the estimator $\widehat{g}_{\widehat{m}}$ which follows from the following Lemma.

Lemma 6.5.3. *Under the assumptions of Theorem 6.2.4, the estimator $\widehat{g}_{\widehat{m}}$ defined by (6.2.7) and (6.2.15), satisfies*

$$\mathbb{E}[\|\widehat{g}_{\widehat{m}} - g\|^2] \leq 6 \inf_{m \in \mathcal{M}} \{\|g - g_m\|^2 + \text{pen}(m)\} + \frac{C'}{n}$$

with C' a positive constant depending on a and $\|f_Y\|_\infty$.

Gathering Lemma 6.5.3 and Inequality (6.5.12) ends the proof of Theorem 6.2.4. \square

Proof of Lemma 6.5.3. Let us define the contrast

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n [t(Y_i) + Y_i t'(Y_i)]. \quad (6.5.13)$$

It is easy to check that $\widehat{g}_m = \underset{t \in \mathcal{S}_m}{\text{argmin}} \gamma_n(t)$, *i.e.* the estimator \widehat{g}_m is also a minimum contrast estimator, and to compute that $\gamma_n(\widehat{g}_m) = -\|\widehat{g}_m\|^2$. We notice that

$$\gamma_n(t) - \gamma_n(s) = \|t - g\|^2 - \|s - g\|^2 - 2\nu_n(t - s) \quad (6.5.14)$$

with

$$\begin{aligned} \nu_n(t) &= \frac{1}{n} \sum_{i=1}^n t(Y_i) + Y_i t'(Y_i) - \langle t, g \rangle \\ &= \frac{1}{n} \sum_{i=1}^n t(Y_i) + Y_i t'(Y_i) - \mathbb{E}[t(Y_i) + Y_i t'(Y_i)] = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t) \end{aligned}$$

where $\nu_{n,1}(t) := (1/n) \sum_{i=1}^n t(Y_i) - \mathbb{E}[t(Y_i)]$ and

$$\begin{aligned} \nu_{n,2}(t) &:= \frac{1}{n} \sum_{i=1}^n Y_i t'(Y_i) \mathbf{1}_{Y_i \leq c_n} - \mathbb{E}[Y_i t'(Y_i) \mathbf{1}_{Y_i \leq c_n}] \\ \nu_{n,3}(t) &:= \frac{1}{n} \sum_{i=1}^n Y_i t'(Y_i) \mathbf{1}_{Y_i > c_n} - \mathbb{E}[Y_i t'(Y_i) \mathbf{1}_{Y_i > c_n}] \end{aligned}$$

with

$$c_n := C_3 \mathbb{E}[Y_1^2] \sqrt{n} / (\log(n)). \quad (6.5.15)$$

By definition of $\widehat{g}_{\widehat{m}}$, for all $m \in \mathcal{M}_n$, we have

$$\gamma_n(\widehat{g}_{\widehat{m}}) + \widehat{\text{pen}}(\widehat{m}) \leq \gamma_n(g_m) + \widehat{\text{pen}}(m).$$

Denoting $m \vee m' = m^*$,

$$\mathcal{B}_{m,m'} = \{t \in \mathcal{S}_{m \vee m'}, \|t\| = 1\}, \quad (6.5.16)$$

and using (6.5.14) we get

$$\begin{aligned} \|\widehat{g}_{\widehat{m}} - g\|^2 &\leq \|g - g_m\|^2 + \|\widehat{g}_{\widehat{m}} - g\|^2 - \|g - g_m\|^2 \\ &\leq \|g - g_m\|^2 + \widehat{\text{pen}}(m) + 2\nu_n(\widehat{g}_{\widehat{m}} - g_m) - \widehat{\text{pen}}(\widehat{m}) \\ &\leq \|g - g_m\|^2 + \frac{1}{4}\|\widehat{g}_{\widehat{m}} - g_m\|^2 + 4 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\widehat{m}) \\ &\leq \|g - g_m\|^2 + \frac{1}{2}\|\widehat{g}_{\widehat{m}} - g\|^2 + \frac{1}{2}\|g_m - g\|^2 + 4 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) + \widehat{\text{pen}}(m) - \widehat{\text{pen}}(\widehat{m}) \end{aligned}$$

Therefore we get

$$\begin{aligned} \|\widehat{g}_{\widehat{m}} - g\|^2 &\leq 3\|g - g_m\|^2 + 8 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_n^2(t) + 2\widehat{\text{pen}}(m) - 2\widehat{\text{pen}}(\widehat{m}) \\ &\leq 3\|g - g_m\|^2 + 24 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,1}^2(t) + 24 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,2}^2(t) + 24 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,3}^2(t) \\ &\quad + 2\widehat{\text{pen}}(m) - 2\widehat{\text{pen}}(\widehat{m}) + 2\text{pen}(\widehat{m}) - 2\text{pen}(\widehat{m}) + 2\text{pen}(m) - 2\text{pen}(m) \\ &\leq 3\|g - g_m\|^2 + 24 \left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,1}^2(t) - p_1(m, \widehat{m}) \right)_+ + 24 \left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,2}^2(t) - p_2(m, \widehat{m}) \right)_+ \\ &\quad + 24 \sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,3}^2(t) + 2\widehat{\text{pen}}(m) - 2\widehat{\text{pen}}(\widehat{m}) + 2\text{pen}(\widehat{m}) + 2\text{pen}(m) \end{aligned} \quad (6.5.17)$$

with $p_1(m, m') = 6m^*/n$ satisfying $12p_1(m, m') \leq \text{pen}_1(m) + \text{pen}_1(m')$ for $\kappa_1 \geq 72$ and

$$p_2(m, m') = 24\mathbb{E}[Y_1^2]m^{3^*}/n$$

and $12p_2(m, m') \leq \text{pen}_2(m) + \text{pen}_2(m')$ for $\kappa_2 \geq 288$. Let us state intermediate results.

Lemma 6.5.4. *Under the assumption of Theorem 6.2.4,*

- (i) $\mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,1}^2(t) - p_1(m, \widehat{m}) \right)_+ \right] \leq K_1/n,$
- (ii) $\mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,2}^2(t) - p_2(m, \widehat{m}) \right)_+ \right] \leq K_2/n,$
- (iii) $\mathbb{E} \left[\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,3}^2(t) \right] \leq K_3/n,$

where K_1, K_2, K_3 are constants which do not depend on n .

- (iv) *There exists a positive constant K_4 depending on a such that,*

$$\mathbb{E}[\{\text{pen}(\widehat{m}) - \widehat{\text{pen}}(\widehat{m})\}_+] \leq \frac{K_4}{n}.$$

Taking expectation of (6.5.17), using $\mathbb{E}[\widehat{\text{pen}}(m)] = 2\text{pen}(m)$, and plugging the results of Lemmas 6.5.4 implies Lemma 6.5.3. \square

6.5.5 Proof of Lemma 6.5.4

First notice that, for $i = 1, 2$,

$$\mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,\widehat{m}}} \nu_{n,i}^2(t) - p_i(m, \widehat{m}) \right)_+ \right] \leq \sum_{m' \in \mathcal{M}_n} \left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,i}^2(t) - p_i(m, m') \right)_+.$$

In the following we apply Talagrand's inequality to the two above terms. For that purpose, we compute the terms denoted by H^2 , v and M in Theorem B.0.1 (Annexe B).

Proof of (i). We bound $\mathbb{E}[\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,1}^2(t)]$. For $t \in \mathcal{B}_{m,m'}$, using that $\sum_{j=0}^{m^*-1} \langle t, \varphi_j \rangle^2 = 1$, we get

$$\nu_{n,1}^2(t) = \left(\nu_{n,1} \left(\sum_{j=0}^{m^*-1} \langle t, \varphi_j \rangle \varphi_j \right) \right)^2 = \left(\sum_{j=0}^{m^*-1} \langle t, \varphi_j \rangle \nu_{n,1}(\varphi_j) \right)^2 \leq \sum_{j=0}^{m^*-1} \nu_{n,1}^2(\varphi_j)$$

$$\mathbb{E}[\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,1}^2(t)] \leq \sum_{j=0}^{m^*-1} \mathbb{E}[\nu_{n,1}^2(\varphi_j)] = \sum_{j=0}^{m^*-1} \frac{1}{n} \text{Var}(\varphi_j(Y_1)) \leq \frac{2m^*}{n} =: H^2,$$

as $\varphi_j^2(x) \leq 2, \forall j, \forall x$. Now, $\|f\|_\infty = \sup_{x \in \mathbb{R}^+} |f(x)| < \infty$ implies that $|f_Y(y)| \leq (\|f\|_\infty/2a) \log((1+a)/(1-a))$ and $\|f_Y\|_\infty < \infty$. Thus

$$\text{Var}(t(Y_1)) \leq \mathbb{E}[t(Y_1)^2] \leq \|f_Y\|_\infty \|t\|^2 = \|f_Y\|_\infty =: v$$

Finally, point (1) of Property 6.5.1 gives

$$\sup_{t \in \mathcal{B}_{m,m'}} \|t\|_\infty = \sqrt{2m^*} \sup_{t \in \mathcal{B}_{m,m'}} \|t\| = \sqrt{2m^*} =: M.$$

We obtain (for $\alpha = 1/2$ in Theorem B.0.1):

$$\mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,1}^2(t) - 6 \frac{m^*}{n} \right)_+ \right] \leq 36 \left(\frac{\|f_Y\|_\infty}{n} e^{-m^*/(6\|f_Y\|_\infty)} + C_1 \frac{m^*}{n^2} e^{-C_2 \sqrt{n}} \right)$$

with $C_1 = \frac{588}{5/2 - \sqrt{6}}, C_2 = \frac{\sqrt{3/2} - 1}{42}$. Consequently,

$$\begin{aligned} \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,1}^2(t) - 6 \frac{m'}{n} \right)_+ \right] &\leq \sum_{m' \in \mathcal{M}_n} 36 \left(\frac{\|f_Y\|_\infty}{n} e^{-m'/(6\|f_Y\|_\infty)} + C_1 \frac{m'}{n^2} e^{-C_2 \sqrt{n}} \right) \\ &\leq \sum_{m' \in \mathcal{M}_n} 36 \left(\frac{\|f_Y\|_\infty}{n} e^{-m'/(6\|f_Y\|_\infty)} + C_1 \frac{m'}{n^2} e^{-C_2 \sqrt{m'}} \right) \leq \frac{K_1}{n} \end{aligned}$$

with C a positive constant depending on $\|f_Y\|_\infty$. This explains the choice $p_1(m, m') = 6m^*/n$, and the constraint $\kappa_1 \geq 12 \times 6 = 72$.

Proof of (ii). As before

$$\begin{aligned} \mathbb{E}[\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,2}^2(t)] &\leq \sum_{j=0}^{m^*-1} \mathbb{E}[\nu_{n,2}^2(\varphi_j)] = \sum_{j=0}^{m^*-1} \frac{1}{n} \text{Var}(Y_1 \varphi_j'(Y_1) \mathbb{1}_{Y_1 \leq c_n}) \leq \sum_{j=0}^{m^*-1} \frac{1}{n} \mathbb{E}[Y_1^2 (\varphi_j')^2(Y_1)] \\ &\leq \sum_{j=0}^{m^*-1} \frac{1}{n} \|\varphi_j'\|_\infty^2 \mathbb{E}[Y_1^2] \leq 8 \mathbb{E}[Y_1^2] \frac{m^*3}{n} =: H^2 \end{aligned}$$

We introduce the following result

Lemma 6.5.5. $\mathbb{E}[Y_i^2 \psi^2(Y_i)] \leq \mathbb{E}[X_i^2 \psi^2(X_i U_i)] \leq (1+a)^2 \|\psi\|^2 \mathbb{E}[X_i]$.

Using also Lemma 6.5.2, we obtain

$$\text{Var}(Y_1 t'(Y_1) \mathbf{1}_{Y_1 \leq c_n}) \leq \mathbb{E}[Y_1^2 t'^2(Y_1)] \leq \|t'\|^2 \mathbb{E}[X] \leq 3(1+a)^2 m^{*2} \mathbb{E}[X] =: v$$

Finally: $\sup_{t \in \mathcal{B}_{m,m'}} (\sup_x |xt'(x) \mathbf{1}_{x \leq c_n}| \leq \sup_{t \in \mathcal{B}_{m,m'}} c_n \|t'\|_\infty \leq c_n 2\sqrt{2} m^{*3/2} =: M$. We obtain, applying Theorem B.0.1 with $\alpha = 1/2$ again:

$$\begin{aligned} \sum_{m' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,2}^2(t) - 24\mathbb{E}[Y_1^2] \frac{m'}{n} \right)_+ \right] &\leq \sum_{m' \in \mathcal{M}_n} 24 \left(\frac{(3(1+a)^2 m'^2 \mathbb{E}[X]) e^{-\frac{C_1 \mathbb{E}[Y_1^2] m'}{(3(1+a)^2 \mathbb{E}[X])}}}{n} \right. \\ &\quad \left. + C_2 \frac{c_n^2 m'^3}{n^2} e^{-C_3 \mathbb{E}[Y_1^2] \sqrt{n}/c_n} \right) \\ &\leq \frac{K_2}{n} \end{aligned}$$

with $C_1 = 8$, $C_2 = 2352/(5/2 - \sqrt{6})$, $C_3 = (\sqrt{3/2} - 1)/42$ with c_n given by (6.5.15) and C' a constant depending on $\mathbb{E}[X_1]$ and $\mathbb{E}[Y_1^2]$. We choose

$$p_2(m, m') = 24\mathbb{E}[Y_1^2] \frac{m^3}{n}, \quad \text{pen}_2(m) = \kappa_2 \mathbb{E}[Y_1^2] \frac{m^{*3}}{n}, \quad \kappa_2 \geq 288$$

and obtain (ii). \square

Proof of Lemma 6.5.5

We have, as $U \leq (1+a)$ a.s.,

$$\begin{aligned} \mathbb{E}[Y^2 \psi^2(Y)] &\leq \mathbb{E}[(1+a)^2 X^2 \psi^2(XU)] = (1+a)^2 \int_0^{+\infty} \int_{(1-a)}^{(1+a)} x^2 \psi^2(xu) f(x) du dx \\ &\leq (1+a)^2 \int_0^{+\infty} \psi^2(v) dv \int_0^{+\infty} x f(x) dx \\ &= (1+a)^2 \|\psi\|^2 \mathbb{E}[X]. \quad \square \end{aligned}$$

Proof of (iii). We use that $m^{*3} \leq n$, it yields

$$\begin{aligned} \mathbb{E} \left[\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,3}^2(t) \right] &\leq \frac{1}{n} \sum_{j=0}^{m^*-1} \text{Var}(Y_1 \varphi'_j(Y_1) \mathbf{1}_{Y_1 > c_n}) \leq \frac{1}{n} \sum_{j=0}^{m^*-1} \|\varphi'_j\|_\infty^2 \mathbb{E}[Y_1^2 \mathbf{1}_{Y_1 > c_n}] \\ &\leq 8 \frac{m^{*3}}{n} \mathbb{E}[Y_1^2 c_n^p \mathbf{1}_{Y_1 > c_n}] c_n^{-1} \leq 8 \frac{\mathbb{E}[Y_1^{2+p}]}{c_n^p} \end{aligned}$$

with the choice of c_n (6.5.15) we obtain

$$\mathbb{E} \left[\sup_{t \in \mathcal{B}_{m,m'}} \nu_{n,3}^2(t) \right] \leq 8 \frac{\mathbb{E}[Y_1^{2+p}] \text{Card}(\mathcal{M}_n)}{C_3^p \mathbb{E}[Y_1^2]^p \log(n)^p n^{p/2}} \leq \frac{K_3}{n}$$

for $p = 4$, using that $\text{card}(\mathcal{M}_n) \leq n^{1/3}$ and that the function $\log(n)^4/n^{2/3}$ is bounded with C'' a positive constant depending on $\mathbb{E}[Y_1^4]$. \square

Proof of (iv). Let us study the difference

$$\mathbb{E}[\{\text{pen}(\widehat{m}) - \widehat{\text{pen}}(\widehat{m})\}_+] = \mathbb{E} \left[2\kappa_2 \left\{ \frac{\mathbb{E}[Y_1^2]}{2} - \widehat{C}_2 \right\}_+ \frac{\widehat{m}^3}{n} \right].$$

Denote $\Omega = \{|\mathbb{E}[Y_1^2] - \widehat{C}_2| \leq \mathbb{E}[Y_1^2]/2\}$. Then $\mathbb{E}[Y_1^2]/2 - \widehat{C}_2 \leq 0$ on Ω , thus

$$\mathbb{E}[\{\text{pen}(\widehat{m}) - \widehat{\text{pen}}(\widehat{m})\}_+] = \mathbb{E} \left[2\kappa_2 \left(\frac{\mathbb{E}[Y_1^2]}{2} - \widehat{C}_2 \right) \frac{\widehat{m}^3}{n} \mathbf{1}_{\Omega^c} \right] \leq \mathbb{E} \left[2\kappa_2 \left(\mathbb{E}[Y_1^2] - \widehat{C}_2 \right) \frac{\widehat{m}^3}{n} \mathbf{1}_{\Omega^c} \right].$$

By Cauchy-Schwarz we have

$$\mathbb{E} \left[\left| \mathbb{E}[Y_1^2] - \widehat{C}_2 \right| \mathbf{1}_{\Omega^c} \right] \leq \mathbb{E}[|\mathbb{E}[Y_1^2] - \widehat{C}_2|^2]^{1/2} \mathbb{P}(\Omega^c)^{1/2}.$$

First, Markov's inequality implies

$$\mathbb{P}(\Omega^c) = \mathbb{P} \left(\left| \mathbb{E}[Y_1^2] - \widehat{C}_2 \right| \geq \frac{\mathbb{E}[Y_1^2]}{2} \right) \leq \frac{2^4}{\mathbb{E}[Y_1^2]^4} \mathbb{E}[|\mathbb{E}[Y_1^2] - \widehat{C}_2|^4].$$

Then an inequality of Rosenthal (given in Annexe (B.0.2)) implies that there exists a constant C , such that

$$\mathbb{E}[|\mathbb{E}[Y_1^2] - \widehat{C}_2|^4] \leq Cn^{-2} \mathbb{E} \left[(Y_1^2 - \mathbb{E}[Y_1^2])^4 \right].$$

Gathering the results we obtain:

$$\mathbb{E} \left[\left| \mathbb{E}[Y_1^2] - \widehat{C}_2 \right| \mathbf{1}_{\Omega^c} \right] \leq \text{Var}(\widehat{C}_2)^{1/2} \frac{4}{\mathbb{E}[Y_1^2]^2} \left(\frac{c_1}{n^3} + \frac{c_2}{n^2} \right)^{1/2} (\mathbb{E}[(Y_1^2 - \mathbb{E}[Y_1^2])^4])^{1/2}.$$

Thus, as $\mathbb{E}[Y_1^k] = \mathbb{E}[X_1^k] \mathbb{E}[U_1^k]$ and the moments of $\mathbb{E}[U_1^k]$ are finite depending on a , if $\mathbb{E}[X_1^8] < \infty$ the quantities $m_4 := \mathbb{E}[(Y_1^2 - \mathbb{E}[Y_1^2])^4]$, $\text{Var}(\widehat{C}_2)$ are bounded, we have the announced result. \square

6.5.6 Proof of Lemma 6.2.5

The survival function of Y satisfies

$$\begin{aligned} 2a\overline{F}_Y(y) &= \int_y^{+\infty} f_Y(z) dz = \int_y^{+\infty} \left(\int_{\frac{z}{1+a}}^{\frac{z}{1-a}} \frac{f(x)}{x} dx \right) dz \\ &= \int_{\frac{y}{1+a}}^{+\infty} \left(\int_{y \vee (1-a)x}^{x(1+a)} dz \right) \frac{f(x)}{x} dx \\ &= \int_{\frac{y}{1+a}}^{+\infty} \frac{f(x)}{x} [x(1+a) - y \vee (1-a)x] dx \\ &= (1+a) \int_{\frac{y}{1+a}}^{+\infty} f(x) dx - (1-a) \int_{\frac{y}{1+a}}^{+\infty} f(x) \mathbf{1}_{y < (1-a)x}(x) dx - y \int_{\frac{y}{1+a}}^{+\infty} \frac{f(x)}{x} \mathbf{1}_{y > (1-a)x}(x) dx. \end{aligned}$$

with $x \vee y = \max(x, y)$. Finally it yields:

$$\overline{F}_Y(y) = \frac{1}{2a} \left[(1+a)\overline{F} \left(\frac{y}{1+a} \right) - (1-a)\overline{F} \left(\frac{y}{1-a} \right) \right] - y f_Y(y). \quad (6.5.18)$$

But, looking at the definition of g given in (6.2.4), we define analogously the function

$$\overline{G}(x) := \int_x^\infty g(y) dy = \frac{1}{2a} \left[(1+a)\overline{F} \left(\frac{x}{1+a} \right) - (1-a)\overline{F} \left(\frac{x}{1-a} \right) \right].$$

Thus relation (6.5.18) becomes:

$$\overline{G}(x) = x f_Y(x) + \overline{F}_Y(x). \quad \square$$

6.5.7 Proof of Proposition 6.2.6

First note that $\mathbb{E}[X_1^2] < +\infty$ implies that \bar{F} integrable. Indeed

$$\int_0^{+\infty} \bar{F}^2(x)dx \leq \int_0^{+\infty} \bar{F}(x)dx = \mathbb{E}[X_1] \leq \mathbb{E}^{1/2}(X_1^2).$$

The result follows if we prove that

$$\mathbb{E}[\|\bar{G}_m - \check{\check{G}}_m\|^2] \leq \frac{2\mathbb{E}[Y_1]}{n} + 4\mathbb{E}[Y_1^2] \frac{m}{n}.$$

The MISE of estimator $\check{\check{G}}_m$ is:

$$\mathbb{E}[\|\check{\check{G}}_m - \bar{G}\|^2] = \|\mathbb{E}[\check{\check{G}}_m] - \bar{G}\|^2 + \mathbb{E}[\|\mathbb{E}[\check{\check{G}}_m] - \check{\check{G}}_m\|^2].$$

First, $\mathbb{E}[\bar{G}_m] = \sum_{j=0}^{m-1} \mathbb{E}[\check{b}_j] \varphi_j = \sum_{j=0}^{m-1} b_j(\bar{G}) \varphi_j = \bar{G}_m$. Then to compute the variance term $\mathbb{E}[\|\bar{G}_m - \check{\check{G}}_m\|^2]$ we start with the relation: $\|\bar{G}_m - \check{\check{G}}_m\|^2 = \sum_{j=0}^{m-1} (\check{b}_j - b_j)^2$ and then

$$\begin{aligned} \mathbb{E}[\|\bar{G}_m - \check{\check{G}}_m\|^2] &= \sum_{j=0}^{m-1} \text{Var}(\check{b}_j) \leq \frac{1}{n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left(Y_1 \varphi_j(Y_1) + \int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_1 \geq x}(x) dx \right)^2 \right] \\ &\leq \frac{2}{n} \sum_{j=0}^{m-1} \mathbb{E}[Y_1^2 \varphi_j(Y_1)^2] + \frac{2}{n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left(\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_1 \geq x}(x) dx \right)^2 \right] \\ &\leq \frac{2m}{n} \|\varphi_j\|_\infty^2 \mathbb{E}[Y_1^2] + \frac{2}{n} \sum_{j=0}^{m-1} \mathbb{E} \left[\left(\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_1 \geq x}(x) dx \right)^2 \right]. \end{aligned}$$

The last term:

$$\frac{1}{n} \mathbb{E} \left[\sum_{j=0}^{m-1} \left(\int_{\mathbb{R}^+} \varphi_j(x) \mathbf{1}_{Y_1 \geq x}(x) dx \right)^2 \right] = \frac{1}{n} \mathbb{E} \left[\sum_{j=0}^{m-1} \langle \varphi_j \mathbf{1}_{Y_1 \geq \cdot} \rangle^2 \right] \leq \frac{1}{n} \mathbb{E} [\|\mathbf{1}_{Y_1 \geq \cdot}\|^2] = \frac{\mathbb{E}[Y_1]}{n}.$$

Thus it comes

$$\mathbb{E}[\|\bar{G}_m - \check{\check{G}}_m\|^2] \leq \frac{2\mathbb{E}[Y_1]}{n} + 4\mathbb{E}[Y_1^2] \frac{m}{n}. \quad \square$$

6.5.8 Proof of Theorem 6.2.7

This proof follows the same line as the proof of Theorem 6.2.4. We define the contrast

$$\gamma_n^{(2)}(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \left[\int_{\mathbb{R}^+} t(x) \mathbf{1}_{Y_i \geq x} dx + Y_i t(Y_i) \right]. \quad (6.5.19)$$

It is such that $\gamma_n^{(2)}(\check{\check{G}}_m) = -\|\check{\check{G}}_m\|^2$ and $\check{\check{G}}_m = \underset{t \in \mathcal{S}_m}{\text{argmin}} \gamma_n^{(2)}(t)$. Then, let

$$\nu_n^{(2)}(t) = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^+} t(x) \mathbf{1}_{Y_i \geq x} dx + Y_i t(Y_i) - \mathbb{E} \left[\int_{\mathbb{R}^+} t(x) \mathbf{1}_{Y_i \geq x} dx + Y_i t(Y_i) \right] = \nu_{n,1}(t) + \nu_{n,2}(t) + \nu_{n,3}(t)$$

with

$$\begin{aligned}\nu_{n,1}^{(2)}(t) &:= \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^+} t(x) \mathbb{1}_{Y_i \geq x} dx - \mathbb{E} \left[\int_{\mathbb{R}^+} t(x) \mathbb{1}_{Y_i \geq x} dx \right] \\ \nu_{n,2}^{(2)}(t) &:= \frac{1}{n} \sum_{i=1}^n Y_i t(Y_i) \mathbb{1}_{Y_i \leq c_n} - \mathbb{E}[Y_i t(Y_i) \mathbb{1}_{Y_i \leq c_n}] \\ \nu_{n,3}^{(2)}(t) &:= \frac{1}{n} \sum_{i=1}^n Y_i t(Y_i) \mathbb{1}_{Y_i > c_n} - \mathbb{E}[Y_i t(Y_i) \mathbb{1}_{Y_i > c_n}]\end{aligned}$$

with c_n a numerical constant depending on n . Following the steps which lead to Equation (6.5.17),

6.5.9 Proof of Theorem 6.3.2

Denote:

$$\phi_t(x) = \frac{1}{2\pi} \int t^*(-u) \frac{e^{iux}}{f_\varepsilon^*(u)} du$$

and $\gamma(t) := \|t\|^2 - \frac{2}{n} \sum_{j=1}^n \phi_t(Z_j) = \|t\|^2 - 2\langle t, \tilde{f}_{T,\ell} \rangle$. Let us define

$$\nu(t) := \frac{1}{2\pi} \langle t^*, \tilde{f}_{T,\ell}^* - f_{T,\ell}^* \rangle = \frac{1}{n} \sum_{j=1}^n (\phi_t(Z_j) - \mathbb{E}[\phi_t(Z_j)]).$$

The two functions $\gamma(t)$ and $\nu(t)$ satisfy the following relation, for $t, s \in S_\ell$:

$$\begin{aligned}S_\ell &= \{t \in \mathbb{L}^1(\mathbb{R} \cap \mathbb{L}^2(\mathbb{R}), \text{support}(t^*) \subset [-\pi\ell, \pi\ell])\}, \\ \gamma(t) - \|t - f\|^2 - (\gamma(s) - \|s - f\|^2) &= -2\nu(t - s).\end{aligned}\tag{6.5.20}$$

Thus writing this relation with $\tilde{f}_{T,\tilde{\ell}}$ and $f_{T,\ell}$ and as, by definition, $\gamma(\tilde{f}_{T,\tilde{\ell}}) + \widetilde{\text{pen}}(\tilde{\ell}) \leq \gamma(f_{T,\ell}) + \widetilde{\text{pen}}(\ell)$, it yields

$$\begin{aligned}\|\tilde{f}_{T,\tilde{\ell}} - f_T\|^2 &= \|f_{T,\ell} - f_T\|^2 + \|\tilde{f}_{T,\tilde{\ell}} - f_T\|^2 - \|f_{T,\ell} - f_T\|^2 \\ &\leq \|f_{T,\ell} - f_T\|^2 + 2\nu(\tilde{f}_{T,\tilde{\ell}} - f_{T,\ell}) + \widetilde{\text{pen}}(\ell) - \widetilde{\text{pen}}(\tilde{\ell}).\end{aligned}$$

Let us remark that $\nu(\tilde{f}_{T,\tilde{\ell}} - f_{T,\ell}) = \|\tilde{f}_{T,\tilde{\ell}} - f_{T,\ell}\| \nu \left(\frac{\tilde{f}_{T,\tilde{\ell}} - f_{T,\ell}}{\|\tilde{f}_{T,\tilde{\ell}} - f_{T,\ell}\|} \right)$. This leads, as in the previous proofs, to

$$\|\tilde{f}_{T,\tilde{\ell}} - f_T\|^2 \leq 3\|f_{T,\ell} - f_T\|^2 + 4\widetilde{\text{pen}}(\ell) + 8 \sum_{\ell' \in \mathcal{M}_n} \left(\sup_{t \in B_{\ell,\ell'}} \nu^2(t) - p(\ell, \ell') \right)_+$$

with a function p such that $\forall \ell, \ell', 4p(\ell, \ell') \leq \text{pen}(\ell) + \text{pen}(\ell')$.

Lemma 6.5.6. *There exists a constant $C > 0$ such that*

$$\sum_{\ell' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in B_{\ell,\ell'}} \nu^2(t) - p(\ell, \ell') \right)_+ \right] \leq \frac{C}{N}.$$

We conclude that there exist two numerical constants $C_1, C_2 > 0$ such that

$$\mathbb{E}[\|\tilde{f}_{\tilde{\ell}} - f_T\|^2] \leq C_1 \inf_{\ell \in \mathcal{M}_n} \{\|f_{T,\ell} - f_T\|^2 + \widetilde{\text{pen}}(\ell)\} + \frac{C_2}{N}. \quad \square$$

6.5.10 Proof of the Lemma 6.5.6

For $\ell \in \mathcal{M}_n$, we consider $t \in S_\ell$. We use Talagrand's inequality. We denote $B_{\ell, \ell'} = \{t \in S_{\ell \vee \ell'}, \|t\| = 1\}$ and $\ell^* = \ell \vee \ell'$. Using Proposition 6.3.1, we obtain

$$\mathbb{E} \left[\sup_{t \in B_{\ell, \ell'}} \nu^2(t) \right] = \mathbb{E} \|\tilde{f}_{T, \ell^*} - f_{T, \ell}^*\|^2 \leq \frac{\ell^*}{n} + \frac{\pi^2 (\ell^*)^3}{3n} \leq \frac{\ell^*}{n} + \frac{\pi^2 (\ell^*)^3}{n} := H^2.$$

Then, by the Plancherel-Parseval inequality, it yields

$$\sup_{t \in B_{\ell, \ell'}} \|\phi_t\|_\infty \leq \frac{1}{\sqrt{2\pi}} \sqrt{2\pi\ell^* + \frac{2\pi^3\ell^{*3}}{3}} \leq \sqrt{\ell^* + \pi^2\ell^{*3}} := M.$$

Finally, for $t \in B_{\ell, \ell'}$, as we know the characteristic function f_ε^* , we have

$$\begin{aligned} \text{Var}(\phi_t(Z_1)) &\leq \mathbb{E}[|\phi_t(Z_1)|^2] = \frac{1}{4\pi^2} \mathbb{E} \left[\left| \int_{-\pi\ell^*}^{\pi\ell^*} t^*(-u) \frac{e^{iuZ_1}}{f_\varepsilon^*(u)} du \right|^2 \right] \\ &\leq \frac{\|f_Z\|_\infty}{4\pi^2} \int \left| \int_{-\pi\ell^*}^{\pi\ell^*} \frac{t^*(-u)e^{iuz}}{f_\varepsilon^*(u)} du \right|^2 dz = \frac{\|f_Z\|_\infty}{2\pi} \int_{-\pi\ell^*}^{\pi\ell^*} \frac{|t^*(-u)|^2}{|f_\varepsilon^*(u)|^2} du \\ &\leq \frac{\|f_Z\|_\infty}{2\pi} (1 + (\pi\ell^*)^2) \int_{-\pi\ell^*}^{\pi\ell^*} |t^*(-u)|^2 du \leq \frac{1+a}{2a} (1 + (\pi\ell^*)^2) := v. \end{aligned}$$

Indeed $\|f_Z\|_\infty \leq (1+a)/(2a) < +\infty$ since $\|f_Z\|_\infty = \|f_T \star f_\varepsilon\|_\infty \leq \|f_\varepsilon\|_\infty = (1+a)/(2a)$, with $f_\varepsilon(x) = e^x/(2a)\mathbf{1}_{[\log(1-a), \log(1+a)]}(x)$. According to Talagrand's inequality, for $\alpha = 1/2$, we obtain

$$\begin{aligned} \mathbb{E} \left[\left(\sup_{t \in B_{\ell, \ell'}} \nu^2(t) - 4H^2 \right)_+ \right] &\leq 24 \left(\frac{1 + \pi^2\ell^{*2}}{n} \exp(-\ell^*/12) \right. \\ &\quad \left. + \frac{294}{(\frac{3}{2} - 1)^2} \left[\frac{\ell^*}{n^2} + \frac{\pi^2\ell^{*3}}{n^2} \right] \exp \left(-\frac{(\sqrt{3/2} - 1)}{42} \sqrt{\ell^* + \pi^2\ell^{*3}} \right) \right) \end{aligned}$$

Then we use that $1 \leq \ell^{*3} \leq n$, thus for $\kappa \geq 4$ we obtain that there exist four numerical constants A_1, A_2, A_3, A_4 and a constant $C > 0$ such that

$$\begin{aligned} \sum_{\ell' \in \mathcal{M}_n} \mathbb{E} \left[\left(\sup_{t \in B_{\ell, \ell'}} \nu^2(t) - p(\ell, \ell') \right)_+ \right] &\leq \sum_{\ell' \in \mathcal{M}_n} A_1 \frac{1 + \pi^2\ell'^2}{n} \exp(-A_2\ell') + \frac{A_3}{n} \exp(-A_4\ell'^{3/2}) \\ &\leq \frac{C}{n} \end{aligned}$$

with $p(\ell, \ell') = \frac{4\ell^*}{n} + \frac{4\pi^2\ell^{*3}}{3n}$. \square

Appendix: observations and outcomes

On the application

For the real data investigated in this chapter, as the number of observations n is small ($n = 51$), we tried to apply a resampling method (coined in Efron (1979) and explained in details in

Ruppert (2011)). From the *i.i.d.* sample of observations that we have, we build 1000 new samples drawn with replacement (we use the R-function `sample` with argument `replace=TRUE`). For each new sample of length $n = 51$ we estimate the density of X . Then we average all the 1000 estimators and obtain a new one. This estimator is very close to the one obtained on one single sample.

On the validation of the model for some repeated data

It is possible to check easily if some data are well fitted by this multiplicative uniform noise model if repeated observations are available. We denote

$$Y_{i,j} = X_i U_{i,j}, \quad U_{i,j} \sim \mathcal{U}_{[1-a,1+a]}, \quad j \in \{1, 2\}, \quad i \in \{1, \dots, N\}. \quad (6.5.21)$$

In this case, we saw that we are able to estimate the noise parameter a with the estimator \hat{a}_n given by equation (6.2.26). Thus, we can compare the histogram of $(Y_{i,1}/Y_{i,2})_i$ and the histogram of $(U_{i,1}^{\text{th}}/U_{i,2}^{\text{th}})_i$ where the U_j^{th} are two simulated samples of length n from the theoretical density $\mathcal{U}_{[1-\hat{a}_n,1+\hat{a}_n]}$. If n is large enough, one can say that the data can be analysed by our model if both histograms are visually similar.

Moreover, the density of the ratio of two independent uniform random variables on interval $[1 - a, 1 + a]$ is computable. We find the density h_a

$$h_a(u) = \frac{1}{4a^2} \left(\frac{1}{2} \left(\frac{1+a}{u} \right)^2 - \frac{1}{2} (1-a)^2 \right) \mathbb{1}_{\left[1, \frac{1+a}{1-a}\right]}(u) + \frac{1}{4a^2} \left(\frac{1}{2} \left(\frac{1}{2} (1+a)^2 - \frac{1-a}{u} \right)^2 \right) \mathbb{1}_{\left[\frac{1-a}{1+a}, 1\right]}(u).$$

We can also compare the theoretical density $h_{\hat{a}_n}$ and a kernel estimator obtained on the $Y_{i,1}/Y_{i,2}$'s.

Let us highlight our point on a very real application. A web site realized a survey and asked to users their range of age. The results on men are presented on Table 6.4 and available via this link <http://www.eos-numerique.com/forums/f11/sondage-quelle-est-votre-tranche-dage-93974/>.

Table 6.4 – Survey for $n = 2129$ men

From 15 to 20 years old	94
From 21 to 30 years old	652
From 31 to 40 years old	683
From 41 to 50 years old	359
From 51 to 60 years old	240
From 61 to 70 years old	101

The results of this study can typically be the outputs of our multiplicative uniform noise model. Indeed, people responding to this question can give an age a little bit higher or smaller than their real age.

We are able to generate two samples Y_1 and Y_2 corresponding to the Table 6.4: for the first category we simulate 94 realizations of a uniform random variable on $\{15, 16, 17, 18, 19, 20\}$ etc. We assume now that Y_1, Y_2 are two samples corresponding to the repeated data (6.5.21). The estimator of a gives $\hat{a}_n = 0.12$. Then, we compare the histogram of the $Y_{i,1}/Y_{i,2}$'s and the one of $(U_{i,1}^{\text{th}}/U_{i,2}^{\text{th}})_i$ where the $U_{i,j}^{\text{th}}$ are two simulated samples from $\mathcal{U}_{[1-\hat{a}_n,1+\hat{a}_n]}$. The results are presented on Figure 6.8, on the left for the simulated U and on the right for the real data Y_1, Y_2 . Also we

compare the density $h_{\hat{a}_n}$ and a kernel estimator. This Figure confirms our model for these data. Indeed, the density $h_{\hat{a}_n}$ fits very well the right histogram of the real data.

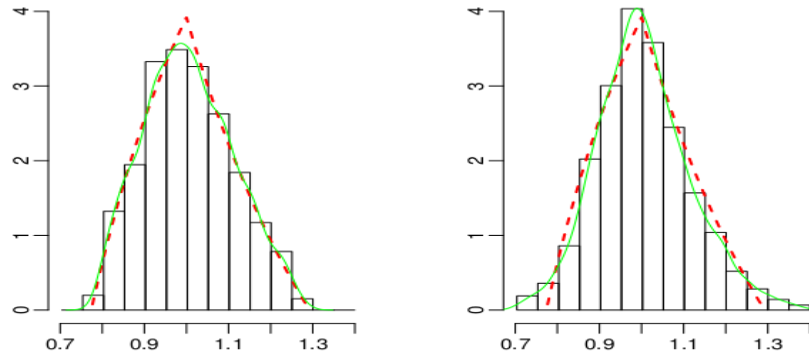


Figure 6.8 – Left graph: histogram of $(U_{i,1}^{\text{th}}/U_{i,2}^{\text{th}})_i$, the continuous green (grey) line is the kernel estimator and the dotted red (black) line is the theoretical density $h_{\hat{a}_n}$. On the right: histogram of the data $Y_{i,1}/Y_{i,2}$'s, the continuous green (grey) line is the kernel estimator and the dotted red (black) line is the theoretical density $h_{\hat{a}_n}$.

On the choice of projection estimator

Instead of the projection strategy, we could have used a kernel estimator, as it has been done in Brunel *et al.* (2015) for the noise $U \sim \mathcal{U}_{[0,1]}$. To define a kernel estimator of the function g we have to take a kernel function satisfying: $K : \mathbb{R}^+ \rightarrow \mathbb{R}$, derivable such that $\int_0^\infty K'^2(u)du < \infty$, and such that $\int_0^\infty K(u)du = 1$, $\int_0^\infty K^2(u)du < \infty$, $\int_0^\infty K'^2(u)du < \infty$ (for example a mixed of gamma densities). In this case we would have defined:

$$\hat{g}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{Y_i - x}{h}\right) + \frac{Y_i}{h} K'\left(\frac{Y_i - x}{h}\right) \tag{6.5.22}$$

and then

$$\mathbb{E}[\hat{g}_h(x)] = \int_0^\infty K\left(\frac{z - x}{h}\right) g(z) dz := g_h(x). \tag{6.5.23}$$

Besides, we estimate f with:

$$\hat{f}_{N,h}(x) = 2a \sum_{k=0}^{N-1} \hat{g}_h\left(\left(\frac{1+a}{1-a}\right)^k (1+a)x\right). \tag{6.5.24}$$

It follows from 6.5.23 that $\mathbb{E}[\hat{f}_{N,h}(x)] = 2a \sum_{k=0}^{N-1} g_h\left(\left(\frac{1+a}{1-a}\right)^k (1+a)x\right) =: f_{N,h}(x)$. Then, $\hat{f}_{N,h}$ is an unbiased estimator of $f_{N,h}$ and we obtain the following result.

Proposition 6.5.7. *If $f \in \mathbb{L}^2(\mathbb{R}^+)$ and $\mathbb{E}[X_1^2] < \infty$, the estimator \hat{g}_h of g given by 6.5.22, satisfies*

$$\mathbb{E}[\|\hat{g}_h - g\|^2] \leq \|g - g_h\|^2 + \frac{\mathbb{E}[Y_1^2] \|K'\|^2}{nh^3} + \frac{\|K\|^2}{nh},$$

and the estimator $\widehat{f}_{N,h}$ given by (6.5.24) satisfies,

$$\mathbb{E}[\|\widehat{f}_{N,h} - f\|^2] \leq \frac{8a^2}{(\sqrt{1+a} - \sqrt{1-a})^2} \left(\|g - g_h\|^2 + \frac{\mathbb{E}[Y_1^2]\|K'\|^2}{nh^3} + \frac{\|K\|^2}{nh} \right) + 2 \left(\frac{1-a}{1+a} \right)^N \|f\|^2.$$

Proof

The MISE is $\mathbb{E}[\|\widehat{g}_h - g\|^2] \leq \|g - g_h\|^2 + \mathbb{E}[\|\widehat{g}_h - g_h\|^2]$ with

$$\begin{aligned} \int_0^\infty \text{Var}(\widehat{g}_h(x))dx &\leq \frac{1}{nh^4} \mathbb{E} \left[Y_1^2 \int_0^\infty K'^2 \left(\frac{Y_1 - x}{h} \right) dx \right] + \frac{1}{nh^2} \int_0^\infty \int_0^\infty K^2 \left(\frac{z-x}{h} \right) g(z) dz dx \\ &\leq \frac{\mathbb{E}[Y_1^2]\|K'\|^2}{nh^3} + \frac{\|K\|^2}{nh}. \end{aligned}$$

We deduce the result for $\widehat{f}_{N,h}$ with the same steps of the proof 6.5.3 (ii). \square

It is interesting to note that for any estimator of the function g we can deduce an estimator of f . And especially, from a bound of the MSE (mean squared error) of the estimator of g we deduce a bound for the MSE of the estimator of f .

Proposition 6.5.8. *If $\mathbb{E}[X_1] < \infty$, the point-wise risk of the estimator $\widehat{g}_h(x_0)$ of $g(x_0)$ given by 6.5.22, satisfies:*

(i) for $x_0 > 0$

$$\begin{aligned} \mathbb{E}[(\widehat{g}_h(x_0) - g(x_0))^2] &\leq (g(x_0) - g_h(x_0))^2 + \frac{(1+a)^2 \mathbb{E}[Y_1]\|K'\|^2}{nh^3} \\ &\quad + \frac{\min(h\|g\|_\infty\|K\|^2, \|K\|_\infty^2)}{nh^2} \end{aligned} \tag{6.5.25}$$

(ii) for $x_0 = 0$

$$\mathbb{E}[(\widehat{g}_h(0) - g(0))^2] \leq (g(0) - g_h(0))^2 + \frac{\min(h\|g\|_\infty\|K\|^2, \|K\|_\infty^2) + \int zK'^2(z)dz}{nh^2} \tag{6.5.26}$$

Proof We first study the mean squared error (MSE):

$$x_0 \in \mathbb{R}^+, \quad \mathbb{E}[(\widehat{g}_h(x_0) - g(x_0))^2] = (g(x_0) - g_h(x_0))^2 + \text{Var}(\widehat{g}_h(x_0)).$$

$$\begin{aligned} \text{Var}(\widehat{g}_h(x_0)) &\leq \frac{1}{nh^2} \mathbb{E} \left[\left(K \left(\frac{Y_1 - x_0}{h} \right) + \frac{Y_1}{h} K' \left(\frac{Y_1 - x_0}{h} \right) \right)^2 \right] \\ &= \frac{1}{nh^2} \mathbb{E} \left[\frac{Y_1^2}{h^2} K'^2 \left(\frac{Y_1 - x_0}{h} \right) \right] + \frac{1}{nh^2} \mathbb{E} \left[K^2 \left(\frac{Y_1 - x_0}{h} \right) \right] \\ &\quad + 2 \frac{Y_1}{h} K \left(\frac{Y_1 - x_0}{h} \right) K' \left(\frac{Y_1 - x_0}{h} \right) \end{aligned}$$

For the first term:

$$\begin{aligned} \mathbb{E} \left[\frac{Y_1^2}{h^2} K'^2 \left(\frac{Y_1 - x_0}{h} \right) \right] &\leq (1+a)^2 \int_0^\infty \int_{1-a}^{1+a} \frac{x^2}{h^2} K'^2 \left(\frac{xu - x_0}{h} \right) f(x) du dx \\ &\leq (1+a)^2 \int_0^\infty \frac{x}{h} f(x) dx \int_0^\infty K'^2(v) dv \leq \frac{(1+a)^2}{h} \mathbb{E}[X] \|K'\|^2 \end{aligned}$$

For the second term we use the relation (6.2.5) and it yields

$$\begin{aligned} \mathbb{E} \left[K^2 \left(\frac{Y_1 - x_0}{h} \right) + 2 \frac{Y_1}{h} K \left(\frac{Y_1 - x_0}{h} \right) K' \left(\frac{Y_1 - x_0}{h} \right) \right] &= \int_0^\infty K^2 \left(\frac{z - x_0}{h} \right) g(z) dz \\ &\leq \min(h \|g\|_\infty \|K\|^2, \|K\|_\infty^2). \end{aligned}$$

which is the bound (6.5.25).

Then notice that for all $z \in \mathbb{R}^{+*}$, $z f_Y(z) \leq (1+a)/(2a)$, thus

$$\begin{aligned} \mathbb{E} \left[\frac{Y_1^2}{h^2} K'^2 \left(\frac{Y_1 - x_0}{h} \right) \right] &= \frac{1}{h^2} \int_0^\infty (x_0 + hz)^2 K'^2(z) f_Y(x_0 + hz) h dz \\ &\leq \frac{1}{h^2} \int_0^\infty (x_0 + hz) K'^2(z) h dz \end{aligned}$$

and for $x_0 = 0$ it yields

$$\mathbb{E} \left[\frac{Y_1^2}{h^2} K'^2 \left(\frac{Y_1 - x_0}{h} \right) \right] \leq \int_0^\infty z K'^2(z) dz.$$

This gives the bound (6.5.26). \square

Here we see that for $x_0 > 0$ $\widehat{g}_h(x_0)$ reaches the optimal speed of convergence in minimax sense. Indeed, if we assume that g is in a Hölder ball $\Sigma_I(\beta, L)$ (see definitions in Chapter 1), for I a neighbourhood of x_0 , and K is a kernel of order $\ell = \lfloor \beta \rfloor$ with $\int |u|^\beta |K(u)| du < \infty$ then $(g(x_0) - g_h(x_0))^2 \leq Ch^{2\beta}$. Thus, if h is of order $n^{-1/(2\beta+3)}$ the estimator of $g(x_0)$ has rate $n^{-2\beta/(2\beta+3)}$. Following the steps of Brunel *et al.* (2015) we can obtain the lower bound proving the minimax optimality.

Then, we want to compute the MSE of estimator $\widehat{f}_{N,h}$ from the MSE of \widehat{g}_h . By Cauchy-Schwarz inequality we get:

$$\mathbb{E} [(f - f_{N,h})^2(x_0)] \leq 8a^2 N \sum_{k=0}^{N-1} \mathbb{E}[(\widehat{g}_{h,k} - g_k)^2(x_0)] + 2f^2 \left(\left(\frac{1+a}{1-a} \right)^N x_0 \right).$$

If we assume that g in a Hölder class $\Sigma_{\mathbb{R}^+}(\beta, L)$ (defined in Chapter 1 Section 1.2.4), the squared bias term satisfies for all x : $|g_h(x) - g(x)|^2 \leq Ch^{2\beta}$ (with C a constant depending on K , see Tsybakov (2009)). Thus we get the following bound

$$\begin{aligned} \mathbb{E} [(f - f_{N,h})^2(x_0)] &\leq 8a^2 N^2 \left(Ch^{2\beta} + \frac{(1+a)^2 \mathbb{E}[Y_1] \|K'\|^2}{nh^3} + \frac{\min(h \|g\|_\infty \|K\|^2, \|K\|_\infty^2)}{nh^2} \right) \\ &\quad + 2f^2 \left(\left(\frac{1+a}{1-a} \right)^N x_0 \right) \end{aligned}$$

For example if f as the form $x \mapsto 1/(1+x^2)$ then taking $N = \log(n)/(4 \log((1+a)/(1-a)))$ the last term of the sum is a rest of order $1/n$. Thus the bound is deteriorate to a coefficient $\log(n)$.

On the estimation of the survival function by deconvolution

In this chapter we have investigated the estimation of the density of X by a deconvolution method, considering the transformed model by the logarithm function

$$Z_j = \log(X_j) + \log(U_j) = T_j + \varepsilon_j.$$

We could estimate also the survival function of T_j then of X_j , with this strategy. It is a difficult problem because the random variables T_j are now on \mathbb{R} , and their survival function is not integrable.

One can notice that

$$F_Z = f_\varepsilon \star F_T, \quad \bar{F}_Z = f_\varepsilon \star \bar{F}_T,$$

this is the starting point of the estimation of \bar{F}_T through a sieve given in Schwarz *et al.* (2011) for a Gaussian noise. Dattner *et al.* (2011) build a nonparametric estimator of $F_T(t_0)$ for this additive model. It is based on the definition formula:

$$F_T(t_0) = \mathbb{P}(T \leq t_0) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty w^{-1} \text{Im}(e^{-it_0 w} f_T^*(w)) dw, \quad t_0 \in \mathbb{R}.$$

We are able to use the same equation to build an estimator of \bar{F}_T . Denote $\widehat{f}_Z^*(x) = (1/n) \sum_{j=1}^n e^{iZ_j x}$, the collection of estimators deriving is, for $t_0 \in \mathbb{R}$,

$$\begin{aligned} \widetilde{\bar{F}}_{T,m}(t_0) &= \frac{1}{2} + \frac{1}{\pi} \int_0^{\pi m} w^{-1} \text{Im} \left(e^{-it_0 w} \frac{\widehat{f}_Z^*(w)}{f_\varepsilon^*(w)} \right) dw \\ &= \frac{1}{2} + \frac{1}{\pi n} \sum_{j=1}^n \int_0^{\pi m} w^{-1} \left(\frac{\sin(w(Z_j - t_0))}{f_\varepsilon^*(w)} \right) dw \end{aligned}$$

Then we have to apply the transformation $\widetilde{\bar{F}}_m(x) := \widetilde{\bar{F}}_{T,m}(\log(x))$ to define an estimator of \bar{F}_X .

It is possible to compute a point-wise risk for this estimator. But is it more difficult to compute the MISE of estimator $\widetilde{\bar{F}}_T$. Comte *et al.* (2015b) propose an estimation method to get around this difficulty for positive random variable, which is no more the case here. This is why we do not investigate more this estimator in the article.

Quatrième partie

Perspectives

Chapitre 7

Perspectives

Sommaire

7.1	Sur les modèles linéaires mixtes	192
7.2	Sur la modélisation neuronale	193
7.2.1	Étude dans le cas des processus à sauts	194
7.2.2	Modèle non-homogène en temps	195
7.3	Sur le modèle de bruit multiplicatif	198

7.1 Sur les modèles linéaires mixtes

Dans cette section nous donnons quelques perspectives qui font suite au Chapitre 2 sur le modèle linéaire mixte (2.1.1).

Estimer le couple (f_α, f_β)

Nous construisons dans le Chapitre 2 deux estimateurs de la densité f_β des effets aléatoires β_j présents dans les observations $(Y_{j,k})_{j=1,\dots,N; k=0,\dots,\mathcal{K}}$ du modèle linéaire mixte (2.1.1). Les vitesses obtenues sont celles que l'on obtient pour des estimateurs non-paramétriques de la densité d'un échantillon observé. C'est à dire qu'en supposant $f_\beta \in \mathcal{A}_b(L)$ la vitesse de l'estimateur est en $N^{-2b/(2b+1)}$ (vitesse optimale obtenue dans Delaigle *et al.* (2008)) alors que la vitesse de convergence de l'estimateur de f_α est une vitesse typique de déconvolution, donc à priori moins bonne.

De cette remarque vient une question naturelle : serait-il possible d'estimer la loi du couple (α, β) (rappelons que les effets aléatoires ne sont pas supposés indépendants) en profitant de la stratégie présentée ici pour l'estimation de f_β , pour améliorer la vitesse de convergence ?

Partons des variables $Y_{j,0} = \alpha_j + \varepsilon_{j,0}$ et $Z_{j,m} = \beta_j + (\varepsilon_{j,m} - \varepsilon_{j,0})/\Delta m$. Il faut remarquer alors que pour $(u, v) \in \mathbb{R}^2$,

$$\mathbb{E} \left[e^{i(uY_{j,0} + vZ_{j,m})} \right] = \mathbb{E} \left[e^{i(u\alpha_j + v\beta_j)} \right] \mathbb{E} \left[e^{iu\varepsilon_{j,0}} \right] \mathbb{E} \left[e^{iv \left(\frac{\varepsilon_{j,m} - \varepsilon_{j,0}}{\Delta m} \right)} \right].$$

C'est à dire :

$$f_{\alpha,\beta}^*(u, v) = \frac{\mathbb{E} \left[e^{i(uY_{j,0} + vZ_{j,m})} \right]}{f_\varepsilon^*(u) \left| f_\varepsilon^* \left(\frac{v}{m\Delta} \right) \right|^2}.$$

Puis après inversion de Fourier l'estimateur qui vient s'écrit :

$$\widehat{f}_{m,\ell}(u, v) = \frac{1}{2\pi} \int_{-m\Delta}^{m\Delta} \int_{-\pi\ell}^{\pi\ell} \frac{1}{N} \sum_{j=1}^N \frac{e^{i(uY_{j,0} + vZ_{j,m})}}{f_\varepsilon^*(u) \left| f_\varepsilon^* \left(\frac{v}{m\Delta} \right) \right|^2} dudv.$$

Il semble possible maintenant de choisir le couple de paramètres de réglage (m, ℓ) avec une méthode de Goldenshluger et Lepski, comme mise en place dans le Chapitre 4 dans le cadre d'un estimateur à noyau bidimensionnel. Ceci est une direction que j'aimerais suivre.

Sélection adaptative dans le cas f_ε inconnue

Nous proposons en Section 2.6.7 du Chapitre 2, deux estimateurs de la densité f_β lorsque la loi du bruit f_ε est inconnue, issus directement du cas f_ε connue. En effet, à partir des deux estimateurs $\widehat{f}_{\beta,m}$ et $\widehat{f}_{\beta,m}^\mathcal{K}$ proposés, nous remplaçons le module $\|f_\varepsilon^*\|^2$ par un estimateur. Nous obtenons alors une borne de risque pour chaque nouvel estimateur. Pour cela, nous n'avons pas eu besoin de supposer que le bruit est symétrique. Les majorations des MISEs laissent penser qu'il est possible de sélectionner le meilleur m et d'obtenir une vitesse de convergence similaire à celle obtenue pour f_ε connue, grâce par exemple à une inégalité de Rosenthal pour contrôler la pénalité aléatoire (méthode appliquée dans le Chapitre 6). En effet, par exemple pour l'estimateur $\widehat{f}_{\beta,m}$ (2.6.14) on sélectionne $\widehat{m} = \operatorname{argmin}_{m \in \{3, \dots, N \wedge \mathcal{K}\}} \left\{ -\|\widehat{f}_{\beta,m}\|^2 + \widehat{\text{pen}}(m) \right\}$ avec

$\widehat{\text{pen}}(m) = \widehat{C}_\varepsilon m\Delta / (\pi N)$ où C_ε est donnée dans la Proposition 2.6.5. L'article Mabon (2015b) s'appuyant sur un estimateur de Comte & Samson (2012) fournit un résultat adaptatif dans

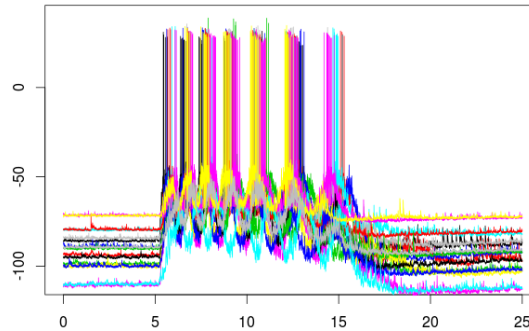


Figure 7.1 – Données réelles : 16 potentiels de membrane d’un neurone de tortue enregistrés pendant 25s, 5 minutes s’écoulent entre chaque mesure, le pas de temps est $\Delta = 0.0001$ s

ce cadre, pour un bruit symétrique. Puis Comte & Kappus (2015) ont proposé une nouvelle stratégie en déconvolution pour un bruit non symétrique. Il sera intéressant d’approfondir cette piste.

7.2 Sur la modélisation neuronale

Nous avons présenté à la Partie II trois chapitres qui traitent l’estimation de la loi des effets aléatoires présents dans la dérive d’un processus de diffusion, en supposant connues les fonctions de dérive et de diffusion. Mais que sait-on sur l’inférence de ces fonctions pour une trajectoire donnée ?

Cette question est venue naturellement lors de l’étude de potentiels neuronaux. En effet, la trajectoire complète du potentiel d’un neurone responsable d’un phénomène de contraction chez la tortue soumise à une stimulation mécanique humaine, peut être décomposée en plusieurs phases, comme le montre la Figure 7.1 qui représente des trajectoires de potentiels. Il y a tout d’abord une phase de repos appelée quiescence (de 0 à 5s) puis une phase qui comporte des cycles qui correspond au déclenchement du réflexe de la tortue. Ils peuvent être eux-même décomposés en *on-cycles* et *off-cycles* (qui indiquent respectivement le haut des cycles comportant les *spikes* et le bas des cycles), puis une nouvelle phase de repos (à partir de 20s environ). La phase de quiescence a été étudiée dans l’Appendice 2 du Chapitre 5 par un modèle OU à deux effets aléatoires.

Pour décrire ces données le modèle doit pouvoir générer des sauts à partir d’un certain seuil, et comporter une partie non-homogène en temps.

Pour ces deux types de modèles, peut-on mener une inférence statistique sur les fonctions qui les caractérisent : coefficient de dérive, de diffusion et de sauts ? Et que sait-on faire si on ajoute en plus des effets aléatoires pour décrire l’aléa présent entre ces trajectoires qui ont une forme commune mais avec une grande variabilité ?

Dans la suite de cette partie nous présentons quelques méthodes existantes ainsi que quelques axes de recherche pour la modélisation qui peuvent conduire à des nouveautés en méthodologie.

Traitement des données réelles

Pour étudier les données réelles représentées Figure 7.1, on les découpe selon les phases détaillées ci dessus. Nous extrayons d'abord les phases de quiescence, représentées Figure 5.16 au Chapitre 5 après renormalisation. Ensuite pour extraire les cycles il faut tronquer les *spikes*. Pour cela on fixe le seuil -30mV : dès que la trajectoire dépasse ce seuil on coupe 4ms avant et 4ms après (soit 1000 points avant et après chaque *spikes*). Enfin après avoir calculé la valeur moyenne de chaque trajectoire, on considère que les phases *on-cycles* sont constituées des points 12% au dessus et les *off-cycles* sont les points 12% au dessous.

7.2.1 Étude dans le cas des processus à sauts

Pour un modèle de diffusion (1.3.2) : $dX(t) = b(X(t))dt + \sigma(X(t))dW(t)$, $X(0) = x_0$ il existe une vaste littérature pour l'estimation non-paramétrique de la fonction de dérive b et de diffusion σ . On se place dans le cas où $n + 2$ observations discrètes $(X(0), X(\Delta), \dots, X((n + 1)\Delta))$ sont disponibles, avec $\Delta = \Delta_n \rightarrow 0$ et $n\Delta_n \rightarrow +\infty$ lorsque n temps vers l'infini. Par exemple Höpfner (2007) propose deux estimateurs à noyau basés sur les limites suivantes :

$$b(a) = \lim_{s \rightarrow 0} \mathbb{E} \left(\frac{X(t+s) - X(t)}{\sqrt{s}} \middle| X(t) = a \right), \quad \sigma^2(a) = \lim_{s \rightarrow 0} \mathbb{E} \left(\left(\frac{X(t+s) - X(t)}{\sqrt{s}} \right)^2 \middle| X(t) = a \right).$$

Puis Comte *et al.* (2007) mettent en place une procédure par projection adaptative inspirée des méthodes de régression, basée sur les incréments

$$Y_{k\Delta} = \frac{X((k+1)\Delta) - X(k\Delta)}{\Delta}, \quad U_{k\Delta} = \frac{(X((k+1)\Delta) - X(k\Delta))^2}{\Delta}$$

qui vérifient $Y_{k\Delta} = b(X(k\Delta)) + 2$ termes, et $U_{k\Delta} = \sigma^2(X(k\Delta)) + 2$ termes avec les termes supplémentaires qui tendent vers 0 lorsque l'on est dans les conditions asymptotiques. Enfin Robert *et al.* (2012) proposent une méthode en transformant la procédure non-paramétrique du maximum de vraisemblance grâce à un temps local.

Pour étudier les données représentées en Figure 7.1 nous avons implémenté les deux premières méthodes. Nous avons estimé les coefficients b et σ de façon non-paramétrique sur les phase de quiescence et sur les phases *off-cycle*. Nous rejoignons les conclusions de Jahn *et al.* (2011) en estimant deux fonctions affines pour b et σ^2 pour les phases *off-cycles* donc un modèle CIR, puis b affine et σ constant pour les phases de quiescence, c'est à dire un modèle OU.

La littérature sur les processus à sauts est très vaste. Les processus qui nous intéressent ici sont les processus de Lévy (voir par exemple à Sato (1999), Miermont (2012)). Le modèle de diffusion avec sauts s'écrit

$$dX(t) = b(X(t))dt + \sigma(X(t))dW(t) + \xi(X(t-))dL(t) \quad (7.2.1)$$

où $L(t)$ est un processus de Levy à sauts purs, centré, et $\sigma_0^2 < \sigma^2(x) < \sigma_1^2$, $0 < \xi(x) < \xi_1$. Le processus de Lévy le plus simple est un processus de Poisson composé. Alors il vérifie $\int_0^t \xi(X(s-))dL(s) = \sum_{k=0}^{N_t} \xi(t_{k-})J_k$ où N_t est un processus de Poisson avec t_1, \dots, t_k les temps de sauts, et ξ_k la taille des sauts.

Pour estimer b , une procédure adaptative est développée dans Schmisser (2014a), et pour estimer σ^2 et $\sigma^2 + \xi^2$ dans Schmisser (2014b). Comme en pratique l'estimation des fonctions se fait sur un compact, on se place dans l'espace $\mathbb{L}^2([0, 1])$. La méthode consiste à tronquer les

incréments précédents lorsqu'il y a un saut :

$$\begin{aligned}\bar{Y}_{k\Delta} &= Y_{k\Delta} \mathbf{1}_{\Omega_{X,k}}, \quad \Omega_{X,k} = \{\omega, |X((k+1)\Delta) - X(k\Delta)| \leq C_{\Delta}^Y\} \\ \bar{U}_{k\Delta} &= U_{k\Delta} \mathbf{1}_{\Omega_{X,k}}, \quad \Omega_{X,k} = \{\omega, |X((k+1)\Delta) - X(k\Delta)| \leq C_{\Delta}^U\}\end{aligned}$$

avec C_{Δ}^Y qui dépend de $b, \sigma_1, \xi_1, n, \Delta$, et C_{Δ}^U dépend de $\sigma_1, \xi_1, n, \Delta$. Ces variables s'écrivent encore $\bar{Y}_{k\Delta} = b(X(k\Delta)) +$ des termes, $\bar{U}_{k\Delta} = \sigma^2(X(k\Delta)) +$ d'autres termes et Schmisser (2014a) et Schmisser (2014b) traitent les termes supplémentaires. Les estimateurs adaptatifs non-paramétriques des fonctions b et σ^2 sont définis par projection et la dimension est sélectionnée par critère pénalisé.

Puis Schmisser (2014b) propose une méthode pour estimer la somme $\sigma^2 + \xi^2$. Cependant on ne peut pas retrouver ξ à partir de ces deux estimations en pratique car elles ne peuvent pas être menées dans le même cadre. En effet, pour estimer le coefficient σ^2 il faut que n soit très grand et Δ petit. Par contre pour l'estimation de $\sigma^2 + \xi^2$ il faut n grand et Δ pas trop petit (ce phénomène est visible et expliqué sur l'étude numérique menée dans Schmisser (2014b)).

Les bases utilisées dans ces articles sont des bases de splines. Nous avons implémenté les estimateurs de b et σ^2 pour la base de Legendre (qui dépend de deux paramètres).

Nous étudions les trajectoires *on-cycles* avec les estimateurs construits dans Schmisser (2014b) sur les variables tronquées \bar{Y} et \bar{U} . En pratique les constantes C_{Δ} sont trop grandes et sont fixées à une valeur arbitraire. Les estimateurs de b et σ^2 sont encore des fonctions affines.

Il serait intéressant d'étudier l'inférence d'un paramètre θ dans la fonction de *drift* b du modèle (7.2.1), puis d'un paramètre aléatoire ϕ_j , profitant des données répétées disponibles.

Mais comme l'avaient fait remarquer Picchini *et al.* (2008) peut-être que ce n'est pas un second effet aléatoire qu'il faut intégrer dans la dérive du processus pour décrire les données neuronales mais plutôt une fonction qui dépend du temps (et non de l'individu). Cette idée est reprise et développée dans Jahn *et al.* (2011) pour étudier les données de la Figure 7.1. Ce qui amène la section suivante.

7.2.2 Modèle non-homogène en temps

Le modèle a la forme suivante

$$dX(t) = b(t, X(t))dt + \sigma(X(t))dW(t), \quad X(0).$$

Le coefficient de dérive est alors trop compliqué à estimer comme fonction de deux variables de nature différente. Alors on suppose que c'est la somme d'une fonction dépendant du temps et une fonction dépendant du processus, ce qui rend le problème identifiable. De plus, à la vue des trajectoires nous nous intéressons à une fonction S périodique. Il existe dans ce cas un travail de Diether (2012) qui traite le cas d'un processus non-homogène en temps dont la dynamique est décrite par l'équation :

$$dX(t) = (S(t) - b(X(t)))dt + \sigma(X(t))dW(t), \quad X_0 = x_0 \in \mathbb{R}, \quad (7.2.2)$$

avec $\sigma^2(x) \leq \sigma_1^2$, et S une fonction T_{per} -périodique, déterministe. La fonction S est inconnue et sa période T_{per} est connue. La fonction b est connue. Le processus est observé pendant n_{per} périodes. La méthode est mise en place pour des bases d'ondelettes, nous la transposons ici pour la base trigonométrique : $\varphi_1(x) = 1$, $\varphi_{2j}(x) = \sqrt{2} \cos(2\pi jx)$, $\varphi_{2j+1}(x) = \sqrt{2} \sin(2\pi jx)$, avec $\mathcal{S}_m = \text{Vect}\{\varphi_1, \varphi_{2j}, \varphi_{2j+1}, j = 1, \dots, m\}$. L'estimateur de S est donné par $\hat{S}_{n_{\text{per}}, m}(x) =$

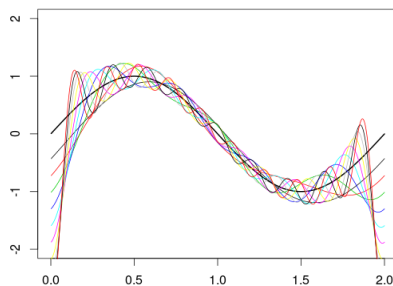


Figure 7.2 – Simulation. Collection d’estimateurs de $S : \widehat{S}_{n_{\text{per}},m}$ en couleur (gris fin), avec $S(t) = \sin(\pi t)$ en noir épais, $T_{\text{per}} = 2$, $\Delta = 0.002$ et $n_{\text{per}} = 1000$, $\sigma(x) = 0.1$, $b(x) = x$.

$\sum_{k=1}^{D_m} \widehat{\alpha}_k \varphi_k(x)$ avec φ_k les vecteurs de la base trigonométrique sur l’intervalle $[0, T_{\text{per}}]$, $D_m = 2m + 1$

et $\widehat{\alpha}_{k,n_{\text{per}}}$ sont les estimateurs de $\alpha_k = \int_0^{T_{\text{per}}} S(t) \varphi_k(t) dt$. L’auteur propose le choix

$$\widehat{\alpha}_{k,n_{\text{per}}} = \frac{1}{n_{\text{per}}} \sum_{i=1}^{n_{\text{per}}} \int_0^T \varphi_k(t) (d\mathbb{X}_i(t) + b(\mathbb{X}_i(t)) dt)$$

avec $\mathbb{X}_i := (X((i-1)T_{\text{per}} + s))_{s \in [0, T_{\text{per}}]}$, $\mathbb{W}_i := (W((i-1)T_{\text{per}} + s) - W((i-1)T_{\text{per}}))_{s \in [0, T_{\text{per}}]}$.

Dans ce cas on a $\widehat{\alpha}_{k,n_{\text{per}}} - \alpha_k = \frac{1}{n_{\text{per}}} \sum_{i=1}^{n_{\text{per}}} \int_0^T \varphi_k(t) \sigma(\mathbb{X}_i(t)) d\mathbb{W}_i(t)$ ce qui nous permet de calculer le MISE de l’estimateur. Nous obtenons pour cette base simple la borne suivante

$$\mathbb{E}[\|\widehat{S}_{n_{\text{per}},m} - S\|^2] \leq \|S - S_m\|^2 + \frac{2D_m}{n_{\text{per}}} T_{\text{per}} \sigma_1^2.$$

La Figure 7.2 montre la collection d’estimateurs $\widehat{S}_{n_{\text{per}},m}$ de la fonction $S(t) = \sin(\pi t)$ sur une période, à partir de l’observation du processus sur 1000 périodes. Il serait intéressant de mener ensuite la sélection du paramètre m pour choisir parmi la collection d’estimateurs et obtenir une procédure adaptative et une inégalité de type oracle.

Sur données simulées l’estimateur est très sensible aux valeurs de Δ , (qui régit le nombre de points par période lorsque l’on discrétise) et à n_{per} qui doit être très grand pour une estimation satisfaisante. Par exemple $n_{\text{per}} = 100$ ne suffit pas, or c’est ce que l’on peut espérer au maximum sur les données réelles. Pour cette raison nous n’avons pas traité les données avec cet estimateur.

De plus, pour l’étude sur données réelles il faut une estimation de la période. On peut obtenir une valeur approchée empiriquement. Mais les travaux de Höpfner & Kutoyants (2011) et Höpfner & Kutoyants (2012) fournissent des résultats théoriques pour des estimateurs du paramètre de fréquence θ (notamment le maximum de vraisemblance dans ce cas) lorsque la fonction S s’écrit $S(t) = A(\theta t)$ où A est une fonction périodique de période 1, connue. Une procédure regroupant ces deux idées pourrait conduire à une méthode complète.

Il me semble possible d’autre part de supposer que la fréquence θ est un paramètre dépendant de chaque trajectoire, aléatoire, et d’étudier les données de potentiel neuronal comme un groupe avec le modèle : $dX_j(t) = (S(t, \theta_j) - b(X_j)) dt + \sigma(X_j(t)) dW_t$. On pourra partir du maximum de vraisemblance grâce aux travaux de Höpfner & Kutoyants (2012) pour construire un estimateur, comme on l’a fait au Chapitre 4.

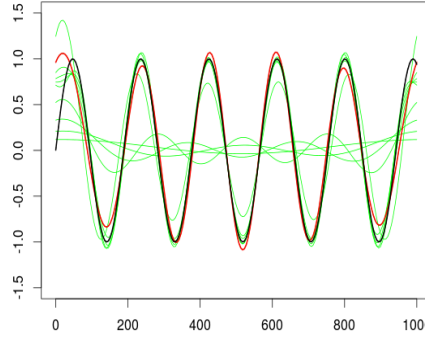


Figure 7.3 – Simulation. Collection des estimateur \widehat{S}_m de S en vert (gris) avec la vraie fonction $S(t) = \sin(t/30)$ en noir épais, et l'estimateur final $\widehat{S}_{\widehat{m}}$ en rouge (gris) épais, avec $n = 10000$, $\Delta = 0.1$.

Idée simple d'une méthode de régression

Reprenons le modèle précédent (7.2.2). Toujours en supposant que b est une fonction connue, les variables aléatoires suivantes sont accessibles :

$$G_{k\Delta} := \frac{X((k+1)\Delta) - X(k\Delta) + \Delta b(X(k\Delta))}{\Delta} = S(k\Delta) + \frac{1}{\sqrt{\Delta}}\sigma(X(k\Delta))\varepsilon$$

avec $\varepsilon \sim \mathcal{N}(0, 1)$. Il est alors possible d'estimer S par une méthode de régression linéaire. Nous considérons l'estimateur du minimum de contraste $\gamma_n^G(t) = \frac{1}{n} \sum_{k=1}^n (G_{k\Delta} - t(k\Delta))^2$, qui est un estimateur par projection. Notons cet estimateur $\widehat{S}_m = \operatorname{argmin}_{m \in \mathcal{S}_m} \gamma_N^G(t)$ avec \mathcal{S}_m le sous espace vectoriel de $\mathbb{L}^2([0, 1])$ de dimension $D_m = 2m + 1$, engendré par les vecteurs de la base trigonométrique. L'estimateur final est défini comme $\widehat{m} = \operatorname{argmin}_{m \in \mathcal{M}_n} \gamma_N^G(\widehat{S}_m) + \operatorname{pen}(m)$ avec $\operatorname{pen}(m) = \kappa D_m/n$. La constante de calibration κ est choisie selon la méthode d'heuristique des pentes donnée dans Baudry *et al.* (2012). Nous avons implémenté et testé cette méthode sur simulations. La Figure 7.3 représente les estimateurs \widehat{S}_m de la vraie fonction $S(t) = \sin(t/30)$ en noir épais et l'estimateur final $\widehat{S}_{\widehat{m}}$ en rouge (gris) épais.

Enfin, pour étudier la phase contenant les cycles entiers des données réelles, nous supposons que nos données suivent un modèle non-homogène en temps de type (7.2.2) et nous estimons la fonction S avec l'estimateur $\widehat{S}_{\widehat{m}}$ que nous venons de construire. Ceci est une approximation car les cycles contiennent les sauts des phases *on-cycles* puis l'on a tronqué les *spikes* lors du traitement des données. Le résultat pour les cycles de la trace 14 est donné Figure 7.4 avec $b(x) = 1/\tau(a - x)$ et $\tau = 13.1$ (valeur moyenne obtenue dans Jahn *et al.* (2011)) et a la moyenne de la trace. Il semble que ce modèle pourrait convenir, et la fonction S serait responsable du caractère sinusoïdal de la trajectoire.

En fait pour modéliser les potentiels neuronaux, un modèle générant des sauts et non-homogène en temps est nécessaire. Ce modèle combinerait les deux modèles (7.2.1) et (7.2.2). De plus, l'étude des données de quiescences étudiées dans l'Appendice 2 du Chapitre 5 montre la présence de deux effets aléatoires linéaires dans le coefficient de dérive. Pour décrire les données de la Figure 7.1 un modèle possible aurait la forme

$$dX_j(t) = (S(t, \phi_j) - \beta_j X_j(t))dt + \sigma \sqrt{X_j(t)} dW_j(t) + \xi(X_j(t-)) dL_j(t)$$

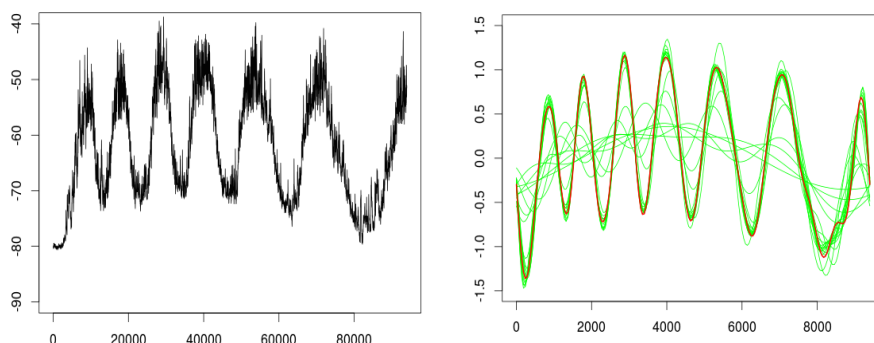


Figure 7.4 – Données réelles. Gauche : cycle de la trace 14. Droite : collection d’estimateur \widehat{S}_m de S en vert (gris) et estimateur final $\widehat{S}_{\widehat{m}}$ en rouge (noir), avec $n = 94000$, $\Delta = 0.1$

où l’indice j est le numéro de la trace, avec les bonnes hypothèses impliquant existence et unicité d’une solution.

7.3 Sur le modèle de bruit multiplicatif

Dans cette section nous donnons quelques perspectives qui font suite au Chapitre 6 sur le modèle multiplicatif (6.1.1).

Dériver la fonction de survie pour estimer la densité ?

On reprend ici les notations du Chapitre 6. Supposons que l’on dispose d’observations directes de la variable positive $X : X_1, \dots, X_n$. Si $\mathbb{E}[X_1] < +\infty$ alors la fonction de survie \overline{F} est de carré intégrable et on peut l’estimer par projection sur $\mathbb{L}^2(\mathbb{R}^+)$. On note \overline{F}_m le projeté orthogonal de

\overline{F} sur \mathcal{S}_m , $\overline{F}_m = \sum_{j=0}^{m-1} b_j \varphi_j$, avec $b_j := \langle \overline{F}, \varphi_j \rangle$. Puis on définit l’estimateur :

$$\widehat{\overline{F}}_m = \sum_{j=0}^{m-1} \widehat{b}_j \varphi_j, \quad \widehat{b}_j = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}^+} \varphi_j(x) \mathbb{1}_{Y_i \geq x} dx. \quad (7.3.1)$$

Mais alors est-ce que la stratégie qui consiste à dériver $\widehat{\overline{F}}$ pour estimer f est envisageable ? ($\overline{F}' = -f$). La dérivée de $\widehat{\overline{F}}$ se calcule grâce au Lemme 6.5.2 Chapitre 6 : $\varphi'_j = -\varphi_j - 2 \sum_{k=0}^{j-1} \varphi_k(x)$ et $\varphi'_0 = -\varphi_0$:

$$\widehat{\overline{F}}'_m(x) = \sum_{j=0}^{m-1} \widehat{b}_j \varphi'_j(x) = \sum_{j=0}^{m-1} \widehat{b}_j \left(-\varphi_j(x) - 2 \sum_{k=0}^{j-1} \varphi_k(x) \right) - \widehat{b}_0 \varphi_0(x).$$

On réécrit cette dérivée pour faire apparaître les coefficients devant chaque vecteur de la base, on obtient :

$$\widehat{\overline{F}}'_m(x) = - \sum_{j=0}^{m-2} \left(\widehat{b}_j + 2(\widehat{b}_{j+1} + \dots + \widehat{b}_{m-1}) \right) \varphi_j(x) - \widehat{b}_{m-1} \varphi_{m-1}(x). \quad (7.3.2)$$

On définit alors $\widehat{f}_m = -\widehat{\overline{F}}'_m$. Cet estimateur correspond-il à celui que l’on obtient en construisant un estimateur par projection de f directement à partir des observations X_i ?

Définissons l'estimateur par projection de $f = \sum_{j \in \mathbb{N}} a_j \varphi_j$, avec $a_j := \langle f, \varphi_j \rangle$,

$$\tilde{f}_m = \sum_{j=0}^{m-1} \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{1}{n} \sum_{i=1}^n \varphi_j(X_i). \quad (7.3.3)$$

On se demande alors si l'égalité suivante est vraie :

$$\begin{cases} \hat{a}_j &= \hat{b}_j + 2\hat{b}_{j+1} + \dots + 2\hat{b}_{m-1}, \quad j = 0, \dots, m-2 \\ \hat{a}_{m-1} &= \hat{b}_{m-1}. \end{cases} \quad (7.3.4)$$

En calculant les coefficients \hat{a}_j on obtient la relation suivante en fonction des \hat{b}_j :

$$\hat{a}_j = \sqrt{2} - \hat{b}_j - 2 \sum_{k=0}^{j-1} \hat{b}_k, \quad m-1 \geq j \geq 0 \quad (7.3.5)$$

Il faut une condition supplémentaire pour avoir une égalité entre les formules (7.3.4) et (7.3.5). Toute fonction de survie vérifie $\overline{F}(0) = 1$. Une solution est d'imposer $\widehat{F}_m(0) = 1$ c'est à dire $\sqrt{2} \sum_{j=0}^{m-1} \hat{b}_j = 1$. Ainsi, les deux formules coïncident exactement. Il semble à priori maladroit d'utiliser $-(\widehat{F}_m)'$ comme estimateur de f puisque l'on a un estimateur direct. Pour cette raison, la procédure qui consiste à estimer f directement est plus simple.

Définir un estimateur contraint de la survie ?

Néanmoins le paragraphe précédent ouvre une nouvelle idée de recherche. En effet, en construisant un estimateur de \overline{F} directement sous la contrainte $\widehat{F}_m(0) = 1$ qu'obtient-on comme estimateur ? La vitesse de convergence est-elle la même ? D'autre part qu'est-ce que l'on obtient dans notre cadre de bruit multiplicatif ? J'aimerais par la suite étudier l'estimateur de \overline{F} sous contrainte (à l'aide d'un multiplicateur de Lagrange) et obtenir les vitesses de convergence dans ce cadre de bruit multiplicatif.

Lorsque a est inconnu

Lorsque la valeur du paramètre a est inconnue en pratique, si nous sommes dans l'un des deux cas décrits dans le Chapitre 6 : lorsque un échantillon du bruit seul est disponible ou lorsque l'on a des données répétées, on peut l'estimer. On dispose de l'estimateur \widehat{a}_n donné en (6.2.26). Mais peut-on évaluer l'erreur commise en remplaçant a par \widehat{a}_n ? Notons $\widehat{f}_{N, \widehat{m}}$ le nouvel estimateur de f . Il faut alors étudier la différence $\widehat{f}_{\widehat{m}, N} - \widehat{f}_{\widehat{m}, N}$, en norme \mathbb{L}^2 . Il vient

$$\begin{aligned} \|\widehat{f}_{\widehat{m}, N} - \widehat{f}_{m, N}\| &\leq 2\widehat{a}_n \sum_{k=0}^{N-1} \|\widehat{g}_{m, \widehat{k}} - \widehat{g}_{m, k}\| + 2|\widehat{a}_n - a| \|\widehat{g}_m\| \sum_{k=0}^{N-1} \left(\frac{1+a}{1-a}\right)^{k/2} \frac{1}{\sqrt{1+a}} \\ &\leq 2\widehat{a}_n \sum_{k=0}^{N-1} \|\widehat{g}_{m, \widehat{k}} - \widehat{g}_{m, k}\| + \frac{2|\widehat{a}_n - a| \|\widehat{g}_m\|}{\sqrt{1+a} - \sqrt{1-a}}, \end{aligned}$$

en notant $\widehat{g}_{m, k}$ la composition de \widehat{g}_m avec la fonction $x \mapsto \left(\frac{1+a}{1-a}\right)^k (1+a)x$ et $\widehat{g}_{m, \widehat{k}}$ la composition avec la fonction $x \mapsto \left(\frac{1+\widehat{a}_n}{1-\widehat{a}_n}\right)^k (1+\widehat{a}_n)x$. Dans cette majoration on voit que le second

terme peut être contrôlé. En effet $\|\widehat{g}_m\| \leq m^{3/2}$ et $|\widehat{a}_n - a| = O_{\mathbb{P}}(n^{-1/2})$. Pour le premier terme il y a plus de travail. En jouant sur le fait que $\widehat{a}_n = \widehat{a}_n - a + a$ et avec une décomposition de $\varphi_j(x) - \varphi_j(y)$ (pour x et y dépendant de a et \widehat{a}_n) il est peut-être possible de majorer l'erreur commise. Cette piste sera étudiée dans des recherches à venir.

Extension du modèle

Dans le cadre d'un bruit multiplicatif plusieurs lois ont été proposées pour U_i , pour lesquelles on sait estimer f et \overline{F} à partir de l'observation des données bruitées $Y_i = X_i U_i$. Notamment : le bruit $U_i \sim \mathcal{U}_{[0,1]}$ pour Brunel *et al.* (2015) où la procédure repose sur la formule $\mathbb{E}[Yt'(Y) + t(Y)] = \mathbb{E}[t(X)]$ si t est une fonction bornée dérivable de dérivée de carré intégrable. La procédure se généralise sans peine au cas $U_i \sim \beta(b, 1)$ et alors on se base sur la formule analogue $\mathbb{E}[Yt'(Y) + bt(Y)] = b\mathbb{E}[t(X)]$. En particulier la loi $\beta(1, 1)$ est la loi uniforme sur $[0, 1]$. Puis nous avons traité le cas $U_i \sim \mathcal{U}_{[1-a, 1+a]}$ dans le Chapitre 6. La méthode repose sur la formule plus compliquée $\mathbb{E}[Yt'(Y) + t(Y)] = \mathbb{E}[\psi_t(Y)] = \langle t, g \rangle$. Enfin le cas $U_i \sim \beta(1, k)$ est étudié dans Belomestny *et al.* (2016) pour le fonction k -monotones. Une question ouverte est alors : comment traiter le cas $U \sim \beta(b, k)$?

Application à l'imagerie

La communauté des mathématiciens travaillant sur les images bruitées, a besoin d'informations concernant la distribution de la norme du gradient de l'image observée. Elle est généralement modélisée par une exponentielle décroissante.

A partir de l'image X de Lena par exemple, on obtient un matrice de pixels 512×512 on peut regarder le norme du gradient de cet objet : $\widetilde{X} : (i, j) \mapsto \|\nabla X(i, j)\|^2$. Cela donne un vecteur de taille 262144, que nous pouvons supposer être des réalisations i.i.d. d'une variable aléatoire. Ensuite on rajoute un bruit multiplicatif uniforme U sur l'intervalle $[1 - a, 1 + a]$: $\widetilde{Y} := \widetilde{X}U$. Est-on capable alors de donner une estimation de la distribution de \widetilde{X} ? La Figure 7.5 prouve que oui. On voit sur l'histogramme de \widetilde{X} , que l'estimateur par déconvolution $\widehat{f}_{\widehat{m}}$ (ligne rouge) semble mieux s'adapter aux données que l'estimateur par projection $\widehat{f}_{N, \widehat{m}}$ (bleu). Pour calculer l'estimateur $\widehat{f}_{N, \widehat{m}}$ on est obligé de prendre un sous-échantillon des 262144 valeurs, de taille $n = 10000$ car les calculs de la base de Laguerre prennent trop de mémoire sinon. Le code de l'estimation pourrait sans doute être optimisé pour contourner cette difficulté et améliorer le résultat.

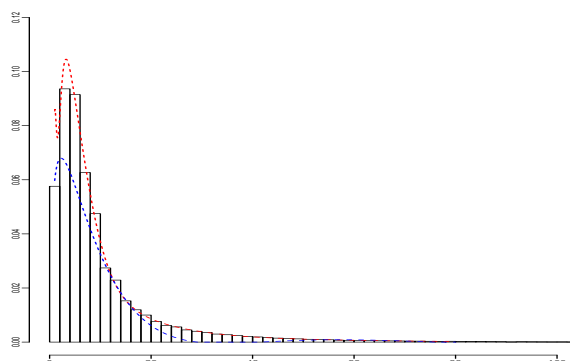


Figure 7.5 – Estimateur \widehat{f} ligne rouge (grise) pointillée et \widehat{f} ligne bleue (noire) pointillée sur l'histogramme des \widetilde{X}

En pratique, pour que cette procédure soit utile il faudrait que l'on ait connaissance d'une

transformation sur l'image de départ X qui se traduit par un bruit multiplicatif uniforme sur le gradient. Alors, on saurait retrouver le gradient d'origine. Pour le moment nous n'avons pas connaissance de tels phénomènes, c'est pour cela que nous n'avons pas poussé l'étude plus loin. Pour approfondir cette piste, un travail récent Comte *et al.* (2015a) propose une application à l'imagerie avec une procédure pouvant rejoindre celles vues pour le modèle de bruit multiplicatif.

Annexe A

Liste de publications

- ▶ New adaptive strategies for nonparametric estimation in linear mixed models
Journal of Statistical Planning and Inference
DOI : 10.1016/j.jspi.2014.03.006 (2014)
Chapitre 2

- ▶ Bidimensional random effect estimation in mixed stochastic differential model, jointly with V. Genon-Catalot
Stochastic Inference for Stochastic Processes
DOI : 10.1007/s11203-015-9122-0 (2015)
Chapitre 4

- ▶ Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model
Metrika
DOI : 10.1007/s00184-016-0583-y (2016)
Chapitre 3

Soumis pour publication

- ▶ Nonparametric Estimation in a multiplicative censoring model with symmetric noise, jointly with F. Comte
Chapitre 6

- ▶ `MixedSDE` : a R package to fit mixed stochastic differential equations, jointly with S. Hermann and A. Samson
Chapitre 5

Annexe B

Résultats d'analyse

La version du Théorème de Talagrand donnée si dessous est une conséquence du Théorème de Talagrand (1996) donné dans Birgé & Massart (1997).

Proposition B.0.1 (Talagrand). *Soit $N \in \mathbb{N}^*$, et \mathcal{F} une classe au plus dénombrable de fonctions et $(X_i)_{i \in \{1, \dots, N\}}$ une famille de variable aléatoire réelles indépendantes. On définit pour tout $f \in \mathcal{F}$,*

$$\nu_N(f) = \frac{1}{N} \sum_{i=1}^N (f(X_i) - \mathbb{E}[f(X_i)]).$$

En supposant qu'il existe trois constantes positives M , H et v telles que $\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M$,

$\mathbb{E}[\sup_{f \in \mathcal{F}} |\nu_N f|] \leq H$, et $\sup_{f \in \mathcal{F}} (1/N) \sum_{i=1}^N \text{Var}(f(X_i)) \leq v$, alors pour tout $\alpha > 0$,

$$\mathbb{E} \left[\left(\sup_{f \in \mathcal{F}} |\nu_N(f)|^2 - 2(1 + 2\alpha)H^2 \right)_+ \right] \leq \frac{4}{a} \left(\frac{v}{N} \exp \left(-a\alpha \frac{NH^2}{v} \right) + \frac{49M^2}{aC^2(\alpha)N^2} \exp \left(-\frac{\sqrt{2}aC(\alpha)\sqrt{\alpha} NH}{7} \right) \right)$$

avec $C(\alpha) = (\sqrt{1 + \alpha} - 1) \wedge 1$, et $a = \frac{1}{6}$.

L'inégalité de Young qui suit est donnée par exemple dans Briane & Pagès (2006).

Théorème B.0.1 (Young). *Soit f une fonction de $\mathbb{L}^p(\mathbb{R})$ et g dans $\mathbb{L}^q(\mathbb{R})$, soit p, q, r des réels de $[1, +\infty]$ tels que*

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1.$$

Alors,

$$\|f \star g\|_r \leq \|f\|_p \|g\|_q.$$

Nous donnons maintenant l'inégalité de Burkholder-Davis-Gundy, donnée par exemple dans Revuz & Yor (1999).

Théorème B.0.2 (Burkholder-Davis-Gundy). *Pour tout $p > 0$ il existe deux constantes $c_p, C_p \in (0, +\infty)$, telles que pour toute martingale locale continue M avec $M_0 = 0$,*

$$c_p \mathbb{E}[\langle M \rangle_\infty^{p/2}] \leq \mathbb{E}[\sup_{t \geq 0} |M_t|^p] \leq C_p \mathbb{E}[\langle M \rangle_\infty^{p/2}].$$

Si T est un temps d'arrêt, alors en remplaçant M par sa martingale arrêtée M^T on obtient la même inégalité avec T au lieu de ∞ .

Rappelons enfin d'inégalité de Marcinkiewicz et Zygmund (1937), donnée par exemple dans Shiryayev (1996) p.498.

Lemme B.0.2. *Soit n un entier naturel et $p \geq 2$, soit ξ_1, \dots, ξ_n des variables indépendantes centrées telles que $\sup_{i=1, \dots, n} \mathbb{E}[|\xi_i|^p] < \infty$. Alors,*

$$\mathbb{E} \left[\left| \sum_{i=1}^n \xi_i \right|^p \right] \leq C(p) \left(\sum_{i=1}^n \mathbb{E}[\xi_i^2] \right)^{p/2}$$

avec $C(p) = (18p)^p (p/((p-1))^{p/2})$.

Principaux résultats sur les processus stochastiques utilisés dans cette thèse

Dans cette partie nous décrivons certains outils classiques qui sont utiles pour le Chapitre 4.

Généralités sur la loi invariante

Il est intéressant d'obtenir la propriété de Markov et de récurrence positive pour la solution d'équation

$$dX^\varphi(t) = b(X^\varphi(t), \phi^\varphi)dt + \sigma(X^\varphi(t))dW_j(t), \quad X^\varphi(0) = x, \quad j = 1, \dots, N, \quad t \in [0, T] \quad (\text{B.0.1})$$

Pour une diffusion X^φ solution de (B.0.1), notons $(\ell, r) \subset \mathbb{R}$ tel que pour tout $(x, y) \in (\ell, r)$ $\mathbb{P}_x(T_y < \infty) > 0$ avec $T_a = \inf\{t \geq 0 : X^\varphi(t) = a\}$ (et la convention $\inf(\emptyset) = 0$). Le processus est dit récurrent sur (ℓ, r) si $\mathbb{P}_x(T_y < \infty) = 1$ pour tout $(x, y) \in (\ell, r)$, si $\mathbb{E}_x[T_y] < \infty$, le processus est dit récurrent positif sur (ℓ, r) (voir Kutoyants, 2004, page 40).

D'après Kutoyants (2004) (ou Kessler *et al.*, 2012, chapitre 1) nous avons la propriété suivante :

Propriété B.0.3. *Si la dérivée de la fonction d'échelle vérifie, pour tout $x, x^* \in (\ell, r)$*

$$s_\varphi(x) = \exp \left(-2 \int_{x^*}^x \frac{b(u, \varphi)}{\sigma^2(u)} du \right) \quad \text{est telle que} \quad \int_{x^*}^r s_\varphi(x) dx = \int_\ell^{x^*} s_\varphi(x) dx = \infty$$

et si la fonction de vitesse

$$m_\varphi(x) = \frac{1}{\sigma^2(x)s(x)} \quad \text{est intégrable} \quad \int_\ell^r m_\varphi(x) dx = M(\varphi) < \infty$$

alors il existe une unique densité stationnaire associée à l'équation différentielle donnée par

$$\pi_\varphi(x) = \mathbb{1}_{(\ell, r)}(x) \frac{m_\varphi(x)}{M(\varphi)}.$$

Le processus X^φ démarré de $X^\varphi(0) \sim \pi_\varphi$ est strictement stationnaire et ergodique de densité marginale $\pi_\varphi(x)dx$.

La première condition de la propriété B.0.3 garantit la récurrence du processus et la seconde la récurrence positive. Dans ce cas, alors on a le théorème ergodique suivant, donné par exemple dans Hasminskii (1980).

Théorème B.0.3 (Théorème ergodique). *Soit g une fonction mesurable et π_φ -intégrable alors,*

$$\mathbb{P} \left(\frac{1}{T} \int_0^T g(X^\varphi(v)) dv \xrightarrow{T \rightarrow \infty} \int_{\mathbb{R}} g(x) \pi_\varphi(x) dx \right) = 1.$$

Dans le Chapitre 4 de cette thèse, on utilise quelques propriétés du générateur, on se réfère à Karlin & Taylor (1981).

La fonction de vraisemblance

Dans le cadre des diffusions, Iacus (2008) (page 41) ou Kessler *et al.* (2012) (page 159) considèrent les deux diffusions :

$$dX^\varphi(t) = b(X^\varphi(t), \varphi) dt + \sigma(X^\varphi(t)) dW^\varphi(t), X^\varphi(0), 0 \leq t \leq T$$

$$dX_0(t) = \sigma(X_0(t)) dW(t), X_0(0), 0 \leq t \leq T$$

avec φ un paramètre fixe et soient, P_φ^T, P_0^T deux probabilités induites par les deux solutions de ces équations sur $(\mathcal{C}[0, T], \mathcal{B}(\mathcal{C}[0, T])) = \sigma(X(s), s \leq T)$, $X(t)(\omega) = \omega(t), w \in \mathcal{C}[0, T]$. Sous les bonnes conditions, si $X_0(0), X^\varphi(0)$ sont soit deux constantes égales soit deux variables aléatoires avec des densités de même support, f_φ et f , et si de plus

$$\mathbb{P}_\varphi^T \left(\int_0^T \frac{b^2(X(s), \varphi)}{\sigma^2(X(s))} ds < \infty \right) = 1, \quad \mathbb{P}_0^T \left(\int_0^T \frac{b^2(X(s), \varphi)}{\sigma^2(X(s))} ds < \infty \right) = 1,$$

(voir Lipster & Shiryaev, 2001, page 257), alors P_0^T et P_φ^T sont équivalentes et la dérivée de Radon-Nikodym est

$$\frac{dP_\varphi^T}{dP_0^T}(X) = \frac{f_\varphi(X(0))}{f(X(0))} \exp \left\{ \int_0^T \frac{b(X(s), \varphi)}{\sigma^2(X(s))} dX(s) - \frac{1}{2} \int_0^T \frac{b^2(X(s), \varphi)}{\sigma^2(X(s))} ds \right\}.$$

Cette formule est issue d'une application du Théorème de Girsanov (voir Kutoyants, 2004, page 37). Ce ratio est utilisé comme ratio de vraisemblance.

Alors, si $X^\varphi(0) = x = X_0(0)$, et $b(x, \varphi) = \varphi b(x)$, l'estimateur du maximum de vraisemblance de φ est :

$$\hat{\varphi}_T = \frac{\int_0^T \frac{b}{\sigma^2}(X^\varphi(s)) dX^\varphi(s)}{\int_0^T \frac{b^2}{\sigma^2}(X^\varphi(s)) ds}.$$

Bibliographie

- Abbaszadeh, M., Chesneau, C. & Doosti, H. (2013). Multiplicative censoring : Estimation of a density and its derivatives under the l-p-risk. *Revstat Statistical Journal* **11**, 255–276.
- Abramowitz, M. & Stegun, I. A. (1966). Handbook of mathematical functions. *Applied Mathematics Series* **55**.
- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Second International Symposium on Information Theory* **11**, 267–281.
- Alarcon, F., Bonaiti-Pellié, C. & Harari-Kermadec, H. (2009). A nonparametric method for penetrance function estimation. *Genetic epidemiology* **33**, 38–44.
- Alfonsi, A. (2005). On the discretization schemes for the cir (and bessel squared) processes. *Monte Carlo Methods and Applications* **11**, 355–384.
- Asgharian, M., Carone, M. & Fakoor, V. (2012). Large-sample study of the kernel density estimators under multiplicative censoring. *Ann. Statist.* **40**, 159–187.
- Bacher, P. & Madsen, H. (2011). Identifying suitable models for the heat dynamics of buildings. *Energy and Buildings* **43**, 1511–1522.
- Barron, A., Birgé, L. & Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113**, 301–413.
- Baudry, J.-P., Maugis, C. & Michel, B. (2012). Slope heuristics : overview and implementation. *Stat. Comput.* **22**, 455–470.
- Belomestny, D., Comte, F. & Genon-Catalot, V. (2016). Laguerre estimation for k-monotone densities observed with noise. *hal-01122847* .
- Berg, R., Alaburda, W. & Hounsgaard, J. (2007). Balanced inhibition and excitation drive spike activity in spinal half centers. *Science* **315**, 390–393.
- Birgé, L. & Massart, P. (1997). *From model selection to adaptive estimation*. Springer New York.
- Birgé, L. & Massart, P. (1998). Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli* **4**, 329–375.
- Birgé, L. & Massart, P. (2006). Minimal penalties for gaussian model selection. *Probability Theory and Related Fields* **138**, 33–73.
- Bissantz, N., Dümbgen, L., Holzmann, H. & Munk, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *Journal of the Royal Statistical Society : Series B (Statistical Methodology)* **69**, 483–506.

-
- Bongioanni, B. & Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.* **192**, 147–172.
- Briane, M. & Pagès, G. (2006). *Théorie de l'intégration*. Vuibert, Paris.
- Brunel, E., Comte, F. & Genon-Catalot, V. (2015). Nonparametric density and survival function estimation in the multiplicative censoring model. *hal-01122847 to appear in Test* .
- Burkitt, A. (2006). A review of the integrate-and-fire neuron model : I. homogeneous synaptic input. *Biol Cybern* **95**, 1–19.
- Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *The Canadian Journal of Statistics* **32**, 181–192.
- Butucea, C. & Tsybakov, A. (2007). Sharp optimality in density deconvolution with dominating bias. ii. *Teor. Veroyatnost. i Primenen.* **52**, 336–349.
- Carroll, R. & Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* **83**, 1184–1186. ISSN 01621459.
- Celisse, A. (2008). Model selection via cross-validation in density estimation, regression, and change-points detection. Theses, Université Paris Sud - Paris XI.
- Celisse, A. (2014). Optimal cross-validation in density estimation with the L2-loss. *The Annals of Statistics* **42**, 1879–1910.
- Chagny, G. (2013). Warped bases for conditional density estimation. *Mathematical Methods of Statistics* **22**, 253–282.
- Comte, . & Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *Journal of Multivariate Analysis* **140**, 31–46.
- Comte, F., Cuénod, C.-A., Pensky, M. & Rozenholc, Y. (2015a). Laplace deconvolution and its application to dynamic contrast enhanced imaging. *preprint arXiv :1207.2231 to appear* .
- Comte, F. & Dion, C. (2016). Nonparametric estimation in a multiplicative censoring model with symmetric noise. *hal-01252780* .
- Comte, F. & Genon-Catalot, V. (2015). Adaptive laguerre density estimation for mixed poisson models. *Electronnic Journal of Statistics* **9**, 1113–1149.
- Comte, F., Genon-Catalot, V. & Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13**, 514–543.
- Comte, F., Genon-Catalot, V. & Samson, A. (2013). Nonparametric estimation for stochastic differential equation with random effects. *Stochastic Processes and their Applications* **7**, 2522–2551.
- Comte, F. & Johannes, J. (2012). Adaptive functional linear regression. *The Annals of Statistics* **40**, 2765–2797.
- Comte, F. & Lacour, C. (2010). Pointwise deconvolution with unknown error distribution. *C. R. Math. Acad. Sci. Paris* **348**, 323–326.
- Comte, F. & Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **73**, 601–627.

-
- Comte, F. & Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Annales de l'Institut Henri Poincaré (B) Probabilités et Statistiques* **49**, 569–609.
- Comte, F., Rozenholc, Y. & Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.* **34**, 431–452.
- Comte, F. & Samson, A. (2012). Nonparametric estimation of random-effects densities in linear mixed-effects model. *J. Nonparametr. Statist.* **24**, 951–975.
- Comte, F., Samson, A. & Stirnemann, J. (2015b). Censored data and measurement error. *MAP5 2015-18, hal-01150296* .
- Cox, J., Ingersoll, J. & Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica* **53**, 385–407.
- Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Annals of Statistics* **33**, 2507–2528.
- Dalalyan, A. & Reiß, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134**, 248–282.
- Dattner, I., Goldenshluger, A. & Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.* **39**, 2477–2501.
- Davidian, M. & Giltinan, D. (1995). *Nonlinear models for repeated measurement data*. Chapman and Hall.
- Delaigle, A. & Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* **56**, 19–47.
- Delaigle, A., Hall, P. & Meister, M. (2008). On deconvolution with repeated measurements. *The Annals of Statistics* **36**, 665–685.
- Delattre, M., Genon-Catalot, V. & Samson, A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics* **40**, 322–543.
- Delattre, M., Genon-Catalot, V. & Samson, A. (2015a). Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. *ESAIM : Probability and Statistics* **19**, 671–688.
- Delattre, M., Genon-Catalot, V. & Samson, A. (2015b). Mixtures of stochastic differential equations with random effects : application to data clustering. *Journal of Statistical Planning and Inference, to appear* .
- Delattre, M. & Lavielle, M. (2013). Coupling the SAEM algorithm and the extended Kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Statistics and Its Interface* **6**, 519–532.
- Delattre, M., Lavielle, M. & Poursat, M.-A. (2014). A note on BIC in mixed-effects models. *Electron. J. Statist.* **8**, 456–475.
- DeVore, R. & Lorentz, G. (1993). *Constructive approximation*, vol. 303. Springer-Verlag, Berlin.
- Diether, M. (2012). Wavelet estimation in diffusions with periodicity. *Statistical Inference for Stochastic Processes* **15**, 257–284.

-
- Diggle, P., Heagerty, P., Liang, K. & Zeger, S. (2002). *Analysis of longitudinal data*. Oxford University Press.
- Dion, C. (2014). New adaptive strategies for nonparametric estimation in linear mixed models. *Journal of Statistical Planning and Inference* **150**, 30 – 48.
- Dion, C. (2016). Nonparametric estimation in a mixed-effect Ornstein-Uhlenbeck model. *Metrika* pp. 1–33.
- Dion, C. & Genon-Catalot, V. (2015). Bidimensional random effect estimation in mixed stochastic differential model. *Stochastic Inference for Stochastic Processes* **18**, 1–28.
- Dion, C., Hermann, S. & Samson, A. (2016). MixedSDE : a R package to fit mixed stochastic differential equations. *hal-01305574* .
- Ditlevsen, S. & de Gaetano, A. (2005). Mixed effects in stochastic differential equation. *REVSTAT - Statistical Journal* **3**, 137–153.
- Ditlevsen, S. & Lansky, P. (2005). Estimation of the input parameters in the Ornstein-Uhlenbeck neuronal model. *Physical review E* **71**, 011907.
- Ditlevsen, S. & Samson, A. (2014). Estimation in the partially observed stochastic morris–lecar neuronal model with particle filter and stochastic approximation methods. *The Annals of Applied Statistics* **8**, 674–702.
- Donnet, S., Foulley, J. & Samson, A. (2010). Bayesian analysis of growth curves using mixed models defined by stochastic differential equations. *Biometrics* **66**, 733–741.
- Donnet, S. & Samson, A. (2008). Parametric inference for mixed models defined by stochastic differential equations. *ESAIM P&S* **12**, 196–218.
- Donnet, S. & Samson, A. (2013). A review on estimation of stochastic differential equations for pharmacokinetic - pharmacodynamic models. *Advanced Drug Delivery Reviews* **65**, 929–939.
- Donnet, S. & Samson, A. (2014). Using PMCMC in EM algorithm for stochastic mixed models : theoretical and practical issues. *Journal de la Société Française de Statistique* **155**, 49–72.
- Donoho, D. L., Johnstone, I. M., Kerkycharian, G. & Picard, D. (1996). Density estimation by wavelet thresholding. *Ann. Statist* **24**, 183–218.
- Efron, B. (1979). Bootstrap methods : Another look at the jackknife. *Ann. Statist.* **7**, 1–26.
- Es, B. V., Spreij, P. & Zanten, H. V. (2005). Nonparametric volatility density estimation for discrete time models. *Journal of Nonparametric Statistics* **17**, 237–249.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19**, 1257–1272.
- Genon-Catalot, V. & Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques* **29**, 119–151.
- Genon-Catalot, V., Jeantheau, T. & Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli* **6**, 1051–1079.

-
- Genon-Catalot, V. & Larédo, C. (2016). Estimation for stochastic differential equations with mixed effects. *Statistics* URL <http://dx.doi.org/10.1080/02331888.2016.1141910>.
- Gikhman, I. & Skorokhod, A. (1969). *Introduction to the Theory of Random Processes*. W.B. Saunders Philadelphia.
- Goldenshluger, A. & Lepski, O. (2011). Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. *The Annals of Statistics* **39**, 1608–1632.
- Gurland, J. (1948). Inversion formulae for the distribution of ratios. *Ann. Math. Statist.* **19**, 228–237.
- Hansen, A., Duun-Henriksen, A., Juhl, R., Schmidt, S., Norgaard, K., Jorgensen, J. & Madsen, H. (2014). Predicting plasma glucose from interstitial glucose observations using bayesian methods. *Journal of diabetes science and technology* **8**, 321–330.
- Hasminskii, R. (1980). Stochastic stability of differential equations. *Sijthoff and Noordhoff* .
- Hermann, S., Ickstadt, K. & Müller, C. (2015). Bayesian prediction of crack growth based on a hierarchical. *Submitted* .
- Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications* **79**, 135 – 163.
- Höpfner, R. (2007). On a set of data for the membrane potential in a neuron. *Mathematical Biosciences* **207**, 275 – 301.
- Höpfner, R. & Brodda, K. (2006). A stochastic model and a functional limit theorem for information processing in large systems of neurons. *Journal of Mathematical Biology* **52**, 439 – 457.
- Höpfner, R. & Kutoyants, Y. (2011). Estimating a periodicity parameter in the drift of a time inhomogeneous diffusion. *Mathematical Methods of Statistics* **20**, 58–74.
- Höpfner, R. & Kutoyants, Y. A. (2012). On frequency estimation for a periodic ergodic diffusion process. *Problems of Information Transmission* **48**, 127–141.
- Iacus, S. (2008). *Simulation and Inference for stochastic differential equation*. Springer Series in Statistics.
- Iversen, E., Morales, J., Moller, J. & Madsen, H. (2014). Probabilistic forecasts of solar irradiance using stochastic differential equations. *Environmetrics* **25**, 152–164.
- Izhikevich, E. (2003). Simple model of spiking neurons. *Transactions on neural networks* **14**, 1569–1572.
- Jahn, P., Berg, R., Hounsgaard, J. & Ditlevsen, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *J. Comput. Neurosci.* **31**, 563–579.
- Kappus, J. & Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution (work in progress). *Electron. J. Statist.* **8**, 2879–2904.
- Karatzas, I. & Shreve, S. (1988). *Brownian motion and stochastic calculus*, vol. 133. Graduate Texts in Mathematics. Springer-Verlag, New York.
- Karlin, S. & Taylor, H. (1981). *A Second Course in Stochastic Processes*. Academic Press.

-
- Kerkycharian, G., Lepski, O. & Picard, D. (2007). Nonlinear estimation in anisotropic multiindex denoising. *Theory Probab. Appl.* **52**, 150–171.
- Kessler, M., Lindner, A. & Sorensen, M. (2012). *Statistical methods for stochastic differential equations*, vol. 124. CRC Press.
- Klein, M., Mathew, T. & Sinha, B. (2013). A comparison of statistical disclosure control methods : Multiple imputation versus noise multiplication. *Research report series* .
- Kloeden, P. & Platen, E. (1992). *Numerical solution of stochastic differential equations*. Springer, Berlin.
- Komárek, A. & Lesaffre, E. (2008). Generalized linear mixed model with a penalized gaussian mixture as a random effects distribution. *Comput. Statist. Data Anal* **52**, 3441–3458.
- Kuelbs, J. & Zinn, J. (2015). Limit theorems for quantile and depth regions for stochastic processes. *High dimensional probability VII-Progress in Probability, to appear* .
- Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, London.
- Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Math. Acad. Sci. Paris* **342**, 877–882.
- Lacour, C. & Massart, P. (2016). Minimal penalty for Goldenshluger-Lepski method. *Stochastic Processes and their Applications* URL <http://www.sciencedirect.com/science/article/pii/S0304414916300370>.
- Lamberton, D. & Lapeyre, B. (2012). *Introduction au Calcul Stochastique Appliqué à la Finance*. Broché.
- Lansky, P., Sanda, P. & He, J. (2006). The parameters of the stochastic leaky integrate-and-fire neuronal model. *Journal of Computational Neuroscience* **21**, 211–223.
- Lapicque, L. (1907). Recherches quantitatives sur l’excitation électrique des nerfs traitée comme une polarisation. *J. Physiol. Pathol. Gen* pp. 620–635.
- Le Gall, J. (2010). Calcul stochastique et processus de markov. *Notes de cours* .
- Leborgne, G. (2013). Introduction à la théorie des distributions. *Notes de cours* URL <http://www.isima.fr/leborgne>.
- Lerner, N. (2008). Lecture notes on real analysis. *Université Pierre et Marie Curie (Paris 6)* .
- Lipster, R. & Shiryaev, A. (2001). *Statistics for random processes I*. Springer, Berlin.
- Mabon, G. (2015a). Adaptive deconvolution of linear functionals on the nonnegative real line. *MAP5 2015-24 hal-01195711* .
- Mabon, G. (2015b). Adaptive estimation of marginal random-effects densities in linear mixed-effects model. *Mathematical Methods of Statistics* **24**, 81–102.
- Mallows, C. (1973). Comments on C_p . *Technometrics* **15**, 661–675.
- Massart, P. (2007). Concentration inequalities and model selection. In *Lectures on Probability Theory and Statistics, Ecole d’Eté de Probabilité de St-Flour 2003*. Springer-Verlag, Berlin.

-
- Meister, A. & Neumann, M. H. (2010). Deconvolution from non-standard error densities under replicated measurements. *Statist. Sinica* **20**, 1609–1636.
- Miermont, G. (2012). Théorème limites et processus de Poisson. *Notes de cours. Orsay* .
- Neumann, M. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* **7**, 307–330.
- Nikol'skii, S. (1975). Approximation of functions of several variables and imbedding theorems. *Springer-Verlag, New York* .
- Oravec, Z., Tuerlinckx, F. & Vandekerckhove, J. (2009). A hierarchical ornstein-uhlenbeck model for continuous repeated measurement data. *Psychometrika* **74**, 395–418.
- Papageorgiou, G. & Hinde, J. (2012). Multivariate generalized linear mixed models with semi-nonparametric and smooth nonparametric random effects densities. *Stat. Comput.* **22**, 79–92.
- Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* **33**, 1065–1076.
- Pensky, M. & Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27**, 2033–2053.
- Picchini, U., De Gaetano, A. & Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scandinavian Journal of statistics* **37**, 67–90.
- Picchini, U. & Ditlevsen, S. (2011). Practical estimation of high dimensional stochastic differential mixed-effects models. *Computational Statistics & Data Analysis* **55**, 1426–1444.
- Picchini, U., Ditlevsen, S., De Gaetano, A. & Lansky, P. (2008). Parameters of the diffusion leaky integrate-and-fire neuronal model for a slowly fluctuating signal. *Neural Computation* **20**, 2696–2714.
- Pinheiro, J. & Bates, D. (2000). *Mixed-effect models in S and Splus*. Springer-Verlag, New York.
- Revuz, D. & Yor, M. (1999). *Continuous martingales and Brownian motion*, vol. 293. Springer-Verlag, Berlin.
- Robert, C., Papaspiliopoulos, O. & Stuart, A. (2012). Nonparametric estimation of diffusions : a differential equation approach. *Biometrika* pp. 1–21.
- Robert, C. P. & Casella, G. (2004). *Monte Carlo Statistical Methods*. Springer-Verlag, New York.
- Rosenthal, J. (2011). Optimal proposal distributions and adaptive mcmc. *Handbook of Markov Chain Monte Carlo* pp. 93–112.
- Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. Springer.
- Rupšys, P. (2015). Generalized fixed-effects and mixed-effects parameters height–diameter models with diffusion processes. *International Journal of Biomathematics* **8**, 1550060.
- Sato, K. (1999). *Lévy processes and infinitely divisible distributions*. Cambridge university press.
- Schmisser, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47**, 61–84.

-
- Schmitter, E. (2014a). Nonparametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications* pp. 883–914.
- Schmitter, E. (2014b). Nonparametric estimation of coefficients of a diffusion with jumps. *Hal*.
- Schwarz, M., Van Bellegem, S. & Florens, J.-P. (2011). Nonparametric frontier estimation from noisy data. In *Exploring Research Frontiers in Contemporary Statistics and Econometrics*, pp. 45–64. Springer.
- Sheather, S. (2004). Density estimation. *Statistical Science* **19**, 588–597.
- Shiryayev, A. N. (1996). *Probability*, vol. 95 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edn. Translated from the first (1980) Russian edition by R. P. Boas.
- Sinha, B., Nayak, T. K. & Zayatz, L. (2011). Privacy protection and quantile estimation from noise multiplied data. *Sankhya B* **73**, 297–315. ISSN 0976-8386.
- Stone, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *Ann. Statist.* **8**, 1348–1360.
- Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.* **126**, 505–563.
- Tsybakov, A. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- Tuckwell, H. & Richter, W. (1978). Neuronal interspike time distributions and the estimation of neurophysiological and neuroanatomical parameters. *Journal of theoretical Biology* **71**, 167–183.
- van Es, B., Klaassen, C. & Oudshoorn, K. (2000). Survival analysis under cross-sectional sampling : length bias and multiplicative censoring. *J. Statist. Plann. Inference* **91**, 295–312. Prague Workshop on Perspectives in Modern Statistical Inference : Parametrics, Semi-parametrics, Non-parametrics (1998).
- Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density : Nonparametric estimation. *Biometrika* **76**, 751–761.
- Vardi, Y. & Zhang, C.-H. (1992). Large sample study of empirical distributions in a random-multiplicative censoring model. *Ann. Statist.* **20**, 1022–1039.
- Vasicek, O. & Gifford Fong, H. (1982). Term structure modeling using exponential splines. *The Journal of Finance* **37**, 339–348.
- Wu, C. (1997). A cross-validation bandwidth choice for kernel density estimates with selection biased data. *Journal of Multivariate Analysis* **61**, 38–60.
- Wu, P. & Zhu, L. (2010). An orthogonality-based estimation of moments for linear mixed models. *Ann. Statist.* **37**, 253–263.
- Yu, Y., Xiong, Y., Chan, Y. & He, J. (2004). Corticofugal gating of auditory information in the thalamus : an in vivo intracellular recording study. *The Journal of Neuroscience* **24**, 3060–3069.

Zuo, Y. & Serfling, R. (2000). General notions of statistical depth function. *Annals of statistics* **28**, 461–482.