



HAL
open science

Estimation non-paramétrique adaptative pour des modèles bruités

Gwennaëlle Mabon

► **To cite this version:**

Gwennaëlle Mabon. Estimation non-paramétrique adaptative pour des modèles bruités. Mathématiques générales [math.GM]. Université Sorbonne Paris Cité, 2016. Français. NNT : 2016USPCB020 . tel-01589142

HAL Id: tel-01589142

<https://theses.hal.science/tel-01589142>

Submitted on 18 Sep 2017

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.

UNIVERSITÉ PARIS DESCARTES
Laboratoire MAP5 UMR CNRS 8145
École doctorale 386 : Sciences Mathématiques de Paris Centre

THÈSE
Pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ PARIS DESCARTES

Spécialité : Mathématiques appliquées

Présentée par

Gwennaëlle MABON

Estimation non-paramétrique adaptative pour des modèles bruités

soutenue le 26 mai 2016 devant le jury composé de

Sylvain ARLOT	Université Paris Sud	Examineur
Fabienne COMTE	Université Paris Descartes	Directrice de thèse
Agathe GUILLOUX	Université Pierre et Marie Curie	Examinatrice
Jean-Michel LOUBES	Université Paul Sabatier	Rapporteur
Markus REISS	Humboldt Universität zu Berlin	Rapporteur
Vincent RIVOIRARD	Université Paris Dauphine	Examineur
Alexandre TSYBAKOV	Université Pierre et Marie Curie	Examineur

À la mémoire de mon grand-père André

“Il me suffit de m’installer dans mon fauteuil et de réfléchir. C’est ça mon instrument de travail !”

— Hercule Poirot

Dans cette thèse, nous nous intéressons à des problèmes d'estimation adaptative non-paramétrique dans des modèles bruités et tout particulièrement dans le modèle de convolution. Ce modèle correspond au cadre où l'observation est entachée d'un bruit modélisé de façon additive. Ainsi nous observons un échantillon $(Y_i)_{1 \leq i \leq n}$

$$Y_i = X_i + \varepsilon_i \quad i = 1, \dots, n,$$

où les X_i sont des variables aléatoires indépendantes et de même loi (i.i.d.), les ε_i sont aussi i.i.d. et les deux suites sont indépendantes. Le but est d'estimer la densité des X_i notée f sous diverses hypothèses sur la loi des ε_i notée f_ε .

Dans la première partie de cette thèse, nous étudions le modèle de convolution quand les variables aléatoires sont positives. Les ε_i ne s'interprètent pas nécessairement comme des erreurs de mesure mais plutôt comme un processus additionnel. Ainsi nous proposons une procédure d'estimation adaptative de la densité ainsi que de la fonction de survie dans le cas où f_ε est connue. Pour cela un angle d'attaque différent de l'approche classique en déconvolution est établi en construisant des estimateurs par projection sur une base appropriée de $\mathbb{L}^2(\mathbb{R}^+)$ appelée base de Laguerre. Ces résultats sont présentés dans le Chapitre 2. Dans le même contexte, nous étudions le problème d'estimation de fonctionnelle linéaire dans le Chapitre 3. Le but est d'estimer des fonctionnelles linéaires de f de la forme $\langle f, \psi \rangle$ avec ψ une fonction connue. Un intérêt majeur de ce développement est qu'il permet la mise en place d'une procédure d'agrégation linéaire d'estimateurs construits dans différentes bases de Laguerre. En effet la base de Laguerre dépend d'un paramètre d'échelle : se pose alors la question du choix du paramètre. Ce problème a été laissé de côté dans le Chapitre 2 en fixant le paramètre. Une procédure d'agrégation consiste à créer un dictionnaire composé de nombreux estimateurs que l'on agrège pour construire un estimateur dont le risque est au moins aussi bon que la meilleure combinaison linéaire du dictionnaire.

Dans la deuxième partie de cette thèse, nous nous intéressons au modèle de convolution dans le cas où la loi des ε_i est inconnue. Il faut alors disposer d'une autre information pour pouvoir continuer à proposer des estimateurs de f . Deux contextes sont envisagés :

- a) Le système peut être étalonné en prenant des mesures « à blanc », on dispose alors d'un échantillon d'observations du bruit seul.
- b) Plusieurs observations du même signal peuvent être obtenues, on parle alors de données répétées ou appariées, c'est le cas par exemple dans certaines applications biomédicales.

Ainsi dans le Chapitre 5, nous étudions le choix adaptatif d'un paramètre de lissage pour l'estimation de densité dans le problème de déconvolution quand la loi du bruit est inconnue. Ce travail permet dans le cas a) ou b) de proposer une méthode d'estimation adaptative de f sans contrainte de type semi-paramétrique sur la transformée de Fourier de la densité

f_ε , ni sur la taille de l'échantillon préliminaire du bruit dans le cas a). Dans le Chapitre 6, les résultats du Chapitre 5 sont étendus au cadre des modèles linéaires mixtes. Ces modèles sont proposés en pharmacologie pour tenir compte des variations des quantités mesurées non seulement temporellement mais aussi en fonction des individus. Les coefficients de ces modèles sont donc représentés comme des réalisations individuelles de variables aléatoires. Les modèles mixtes correspondent à des données longitudinales pouvant se ramener à des données répétées comme décrites dans le point b). Enfin la méthode développée dans le Chapitre 5 a encore pu être adaptée à l'estimation de la densité de somme de variables aléatoires observées avec du bruit dans le Chapitre 7.

In this thesis, we are interested in nonparametric adaptive estimation problems in measurement error models and particularly in the convolution model. This model fits standard situations in applied settings where observations are contaminated by an additive noise. Measurement error models are used in many different fields and are the subject of many theoretical studies. More precisely we assume that we observe

$$Y_i = X_i + \varepsilon_i \quad i = 1, \dots, n,$$

where the X_i 's are independent identically distributed random variables (i.i.d.), the ε_i 's are also i.i.d. and the two sequences are independent. The goal is to estimate the common density of the X_i 's denoted by f under different assumptions on the distribution of the ε_i 's denoted by f_ε .

In the first part of the thesis, we study the convolution model when random variables are nonnegative. We cannot necessarily interpret the ε_i 's as measurement errors. Thus in this work we propose an adaptive estimation procedure for the density along with the survival function. For that we describe a new angle of attack different from the classical approach in deconvolution problems by building projection estimators on an appropriate basis of $\mathbb{L}^2(\mathbb{R}^+)$ named Laguerre basis. This work is developed in Chapter 2. In the same framework, we work on linear functional adaptive estimation of density in Chapter 3. The goal is to estimate linear functionals of f such as $\langle f, \psi \rangle$ for a known function ψ . A major interest of gathering Chapters 2 and 3 is that it may allow linear aggregation of density estimators built in different Laguerre basis. Indeed the Laguerre basis can be parametrized by a scale a parameter : then arise the question of the choice of the parameter. This issue was put aside in Chapter 2 by fixing the parameter. The idea of aggregation is to create a dictionary formed with several estimators that we aggregate to build an estimator which is nearly as good as the best linear combination.

In the second part of the thesis, our aim is to study the convolution model when the common density of the ε_i 's is unknown. In this setting we need additional information to keep on proposing estimators of the density f . Two cases are considered:

- a) The system can be calibrated by taking measures without signal. Then we have only at hand an observation sample of the noise.
- b) Several observations of the same signal can be obtained. That kind of observations are called longitudinal (also panel data). They appear for instance in some biomedical applications.

In Chapter 5, we investigate the data driven choice of the cutoff parameter in density deconvolution problems with unknown error distribution. This work allows us, in the case a) or b), to propose an adaptive estimation procedure of f without semi-parametric assumptions on the Fourier transform of the density f_ε . Moreover in the case a), we do

not make any assumptions on the preliminary sample size of ε anymore. In Chapter 6, results of Chapter 5 are extended to the framework of linear mixed models. These models appear in pharmacology, for instance, to take into account fixed and random effects. Thus we can modelize individual variations. They allow for analysis of longitudinal data which can be seen as repeated measurements as described in b). Finally the method elaborated in Chapter 5 is adapted to the problem of density estimation of the sum of random variables when the latter are observed with an additive noise.

Remerciements

Mes premiers et plus sincères remerciements s'adressent à Fabienne Comte ma directrice de thèse. Je te remercie de m'avoir proposé il y a trois ans d'encadrer ma thèse. Durant cette période, tu as fait preuve d'une incroyable disponibilité et gérer mes impatiences si nombreuses. Tu n'as véritablement pas compté tes heures et tu as toujours su rester attentive à mon travail. La qualité de ce manuscrit n'est pas étrangère à tes nombreux conseils, suggestions et orientations. Je suis bien consciente de l'encadrement privilégié que tu m'as apporté. J'espère que cette collaboration fructueuse sera longue.

J'exprime toute ma gratitude à Jean-Michel Loubes et Markus Reiß pour avoir tout de suite accepté de rapporter cette thèse et l'intérêt qu'ils ont porté à mes travaux. Je remercie aussi Sylvain Arlot, Agathe Guilloux, Vincent Rivoirard et Sacha Tsybakov de me faire l'honneur de compléter ce jury.

Je remercie aussi mes différents co-auteurs : Johanna Kappus, Adeline Samson, Christophe Chesneau et Fabien Navarro.

Je suis reconnaissante envers Valentine Genon-Catalot et Céline Duval qui m'ont plusieurs fois aidée à améliorer la rédaction de certains chapitres de cette thèse, ainsi qu'à la préparation de la soutenance. Pour cette raison, entre autres, je remercie Angelina Roche.

Je remercie aussi mes prédécesseurs Claire Lacour, Sandra Plancade et Gaëlle Chagny pour leur bienveillance, de même que Ester Mariucci pour sa gentillesse et ses conseils.

Je remercie aussi le CREST qui a financé ces trois années de thèse ainsi que les différents membres du laboratoire.

La vie au MAP5 ne serait sans doute pas la même sans Marie-Hélène et sa bonne humeur ainsi que sa capacité à faire de nos démarches administratives de simples formalités. Je remercie aussi Annie Raoult, Isabelle Valéro et Marie Marduel.

Qu'aurait été cette aventure sans sa fameuse équipe de doctorants. Désolée de vous avoir poussé à cette fameuse représentation pour les 10 ans du MAP5. Mais quel succès ! Je tiens à remercier les "membres" du bureau 750 dont la capacité à débattre me semble sans fin : Alkéos, Anne-Claire, Charlotte D., Christophe, Fabien, Jean, Loïc, Maud et Thomas B. ; cette belle équipe ne saurait être complète sans Andrea, Anne-Sophie, Arthur, Charlotte L., Christèle, Julie, Léo, Pierre, Rémy et Sonia. Merci pour ces trois ans !

Je remercie aussi Anne, Étienne, Flora, Florent, Julie, Maya et Sébastien.

Je profite aussi de ce moment pour remercier mes amis qui ont suivi toutes les étapes ou péripéties de cette thèse : Rémy, Floriane, Nathalie et Thomas, Claire ainsi que Tamaki et Anissa.

Enfin merci à ma mère pour son soutien sans faille et pour m'avoir transmis cette persévérance sans fin.

Résumé		vii
Abstract		ix
Remerciements		xi
1 Introduction		1
1.1 Estimation non-paramétrique adaptative		2
1.2 Modèle de convolution		10
1.3 Perspectives		27
I Estimation adaptative dans le modèle de convolution sur \mathbb{R}^+		31
2 Adaptive deconvolution on the nonnegative real line		33
2.1 Introduction		35
2.2 Statistical model and estimation procedure		38
2.3 Bounds on the \mathbb{L}^2 risk		40
2.4 Model selection		45
2.5 Illustrations		47
2.6 Concluding remarks		50
2.7 Proofs		52
2.8 Appendix		65
3 Adaptive deconvolution of linear functionals on the nonnegative real line		67
3.1 Introduction		69
3.2 Statistical model and estimation procedure		71
3.3 Model selection and adaptive estimation		78
3.4 Particular case of pointwise estimation		80
3.5 Proofs		88
3.6 Appendix 1		97
3.7 Appendix 2		97
4 Density aggregation on the nonnegative real line in the convolution model		99
4.1 Introduction		101
4.2 Classical aggregation problem		104
4.3 Linear aggregation of Laguerre estimators		105
4.4 Illustration		110
4.5 Proofs		112

II Estimation adaptative de densité dans le modèle de convolution avec bruit inconnu	115
5 Adaptive density estimation in deconvolution problems with unknown error distribution	117
5.1 Introduction	119
5.2 Statistical model, estimation procedure and risk bounds	120
5.3 Data driven bandwidth selection and oracle bounds	124
5.4 Illustrations	126
5.5 Concluding remarks	132
5.6 Proofs	133
6 Adaptive estimation of marginal random-effects densities in linear mixed-effects model	141
6.1 Introduction	143
6.2 Statistical model and estimation procedure	144
6.3 Model selection	149
6.4 Simulation	152
6.5 Concluding remarks	154
6.6 Proofs	156
6.7 Appendix 1	169
6.8 Appendix 2: Supplementary material	170
7 Estimation of convolution in the model with noise	175
7.1 Motivations	177
7.2 Estimation with known noise density	178
7.3 Estimation with unknown noise density	182
7.4 Illustration	185
7.5 Proofs	188
7.6 Appendix	203
Bibliographie	205
Table des figures	214
Liste des tableaux	216

1

Sommaire

1.1	Estimation non-paramétrique adaptative	2
1.1.1	Estimation par projection	3
1.1.2	Étude du risque et choix non adaptatif de la dimension	4
1.1.3	Sélection de modèles et inégalités oracles	5
1.1.4	Inégalités de concentration	8
1.2	Modèle de convolution	10
1.2.1	Problème et modèle statistique	10
1.2.2	Méthodes de déconvolution à bruit connu	12
1.2.3	Méthodes de déconvolution à bruit inconnu	21
1.3	Perspectives	27

Le but de cette thèse est l'étude des méthodes d'estimation adaptative dans des modèles bruités et tout particulièrement dans le modèle avec erreur additive appelé modèle de convolution. Ainsi, ce chapitre introductif a pour but d'expliquer le point de vue statistique adopté dans cette thèse. Pour cela, nous mettons en perspective dans un premier temps les notions d'estimation non-paramétrique ou estimation fonctionnelle, ensuite nous expliquons le point de vue de la sélection de modèle. Cette partie est inspirée par [Tsybakov \(2009\)](#), [Comte \(2015\)](#) et [Massart \(2003\)](#). Certaines notions sont illustrées dans le cas de l'estimation de densité. Enfin nous décrivons le modèle de convolution et nous inscrivons les résultats de cette thèse dans ce domaine.

1.1 Estimation non-paramétrique adaptative

Dans tout le manuscrit, nous adoptons le point de vue de l'estimation *non-paramétrique* ou *estimation fonctionnelle*. Le but est d'estimer une certaine fonction s à partir d'un échantillon de variables aléatoires (X_1, \dots, X_n) . Nous appelons estimateur de s noté $\hat{s} = \hat{s}(X_1, \dots, X_n)$ une fonction mesurable des données (X_1, \dots, X_n) . Nous parlons d'estimation non-paramétrique car il est supposé que s appartient à une certaine classe de fonctions notée \mathcal{F} par exemple. Cette classe doit être assez grande pour ne pas pouvoir s'écrire sous la forme $\{f(x, \theta), \theta \in \Theta\}$ avec $\Theta \subset \mathbb{R}^k$. Ainsi \mathcal{F} peut être la classe de toutes les densités continues, classe de Lipschitz, classe des fonctions positives à support compact ...

Pour mesurer la qualité de l'estimateur ou contrôler l'erreur induite par l'estimation, il faut se donner une fonction de perte, aussi nommée risque, de la forme $\mathbb{E}[d(s, \hat{s})]$ où d est une distance ou une semi-distance sur l'ensemble des fonctions. Dans le cadre qui nous concerne les distances les plus usitées sont la distance ponctuelle ou les normes \mathbb{L}^p , c'est-à-dire que pour deux fonctions s et t nous définissons

$$\begin{aligned} d(s, t) &= |s(x_0) - t(x_0)|^p \\ d(s, t) &= \|s - t\|_p^p = \int |s - t|^p. \end{aligned}$$

Ainsi on appelle risque quadratique ponctuel l'erreur $\mathbb{E}[|s(x_0) - \hat{s}(x_0)|^2]$ pour $x_0 \in \mathbb{R}^d$ avec $d \geq 1$ associée à la distance ponctuelle et risque quadratique intégré $\mathbb{E}[\|s - \hat{s}\|_2^2]$ l'erreur associée à la norme \mathbb{L}^2 . L'erreur ponctuelle est aussi appelée MSE (mean squared error) et l'erreur quadratique intégrée MISE (mean integrated squared error). Dans la suite de ce travail nous n'examinerons que ces deux types de risque. En effet, nous ne considérerons que des estimateurs par projection ce qui induit naturellement des normes \mathbb{L}^2 .

De plus, pour apprécier la qualité d'un estimateur, nous nous intéressons au comportement du risque en fonction de la taille des données n . En effet nous nous attendons à ce que le risque décroisse quand n augmente. Quand n est grand, nous recueillons plus d'information sur s et devons ainsi améliorer la précision de l'estimation. Nous adoptons alors le cadre de vitesse optimale au sens *minimax*.

Premièrement, nous cherchons à établir le risque maximum de l'estimateur \hat{s} pour une certaine classe de fonctions \mathcal{F} et une certaine semi-distance d en étudiant

$$\sup_{\hat{s} \in \mathcal{F}} \mathbb{E}[d^2(s, \hat{s})] \leq C\psi_n^2 \quad \text{avec } C > 0 \quad \text{et } \psi_n \xrightarrow{n \rightarrow +\infty} 0. \quad (1.1)$$

Deuxièmement, pour montrer qu'il s'agit de la meilleure vitesse de convergence, c'est-à-dire qu'elle ne peut être améliorée, il faut établir la minoration suivante

$$\inf_{\hat{s}} \sup_{s \in \mathcal{F}} \mathbb{E}[d^2(s, \hat{s})] \geq c\psi_n^2 \quad \text{avec } c > 0 \quad (1.2)$$

l'infimum est pris sur tous les estimateurs. Si les Équations (1.1) et (1.2) sont vérifiées, la vitesse est dite *minimax optimale*.

Dans la suite nous nous ne considérerons que les risques quadratiques ponctuel ou intégré.

1.1.1 Estimation par projection

Nous allons maintenant brièvement présenter le principe de l'estimation par projection. Supposons que la fonction cible s appartienne à un certain espace $\mathbb{L}^2(A)$, avec $A \subseteq \mathbb{R}^d$ muni du produit scalaire usuel $\langle \cdot, \cdot \rangle$ et de la norme en découlant $\| \cdot \|$. Soit $(\varphi_j)_{j \geq 1}$ une base hilbertienne de $\mathbb{L}^2(A)$, la fonction s peut s'écrire sous la forme

$$s = \sum_{j=1}^{\infty} a_j \varphi_j \quad \text{avec} \quad a_j = \langle s, \varphi_j \rangle.$$

Ne pouvant estimer une infinité de coefficients, le but de l'estimation par projection est de se restreindre à un nombre fini (et raisonnable) de sous-espaces vectoriels $\mathcal{S}_m \subset \mathbb{L}^2(A)$ de dimension finie D_m . Le choix d'une dimension pertinente sera abordé dans la section suivante. Considérons maintenant les espaces d'approximation, $\mathcal{S}_m = \text{Vec}(\varphi_1, \dots, \varphi_{D_m})$, aussi appelé modèle, la projection de s sur \mathcal{S}_m est

$$s_m = \sum_{j=1}^{D_m} a_j \varphi_j.$$

Un estimateur de s_m s'écrit alors sous la forme suivante

$$\hat{s}_m = \sum_{j=1}^{D_m} \hat{a}_j \varphi_j,$$

où les $(\hat{a}_j)_{j \in [1, D_m]}$ estiment les $(a_j)_{j \in [1, D_m]}$. Afin d'estimer les coefficients de la fonction s sur l'espace \mathcal{S}_m , nous adoptons le principe de minimisation de contraste. En effet, rappelons que la projection s sur \mathcal{S}_m est définie par $s_m = \arg \min_{t \in \mathcal{S}_m} \|s - t\|^2$. Posons $\gamma(t) = \|s - t\|^2$ et remarquons que pour toute fonction t , $\gamma(t) \geq 0$ et $\gamma(s) = 0$. Une première idée serait de chercher à minimiser cette fonction γ . Cependant celle-ci n'est pas calculable car la fonction s est inconnue. Ainsi la méthode d'estimation consiste à définir un équivalent empirique du contraste γ que nous notons γ_n , que l'on cherche à minimiser. De manière générale le contraste empirique est de la forme

$$\gamma_n(t) = \frac{1}{n} \sum_{i=1}^n \mu(X_i, t)$$

et vérifie $\mathbb{E}[\gamma_n(t)] = \gamma(t) + c_0(s)$.

Exemple : estimation de densité. Afin d'illustrer la méthode de minimisation de contraste, supposons dans ce paragraphe que nous observons un n -uplet (X_1, \dots, X_n) de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de densité f . Nous cherchons alors à minimiser $\|f - t\|^2$ sur \mathcal{S}_m . Notons que

$$\|f - t\|^2 = \|t\|^2 - 2\langle f, t \rangle + \|f\|^2,$$

ce qui implique

$$\arg \min_{t \in \mathcal{S}_m} \|f - t\|^2 = \arg \min_{t \in \mathcal{S}_m} \|t\|^2 - 2\langle f, t \rangle.$$

Enfin il suffit de remarquer que

$$\langle f, t \rangle = \int_A f(u)t(u) \, du = \mathbb{E}[t(X_1)].$$

Ainsi pour

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(X_i),$$

nous obtenons d'après la loi des grands nombres $\gamma_n(t) \xrightarrow[n \rightarrow +\infty]{\text{ps}} \mathbb{E}[\gamma_n(t)] = \|f - t\|^2 - \|f\|^2$. Finalement les estimateurs s'expriment comme $\hat{f}_m = \sum_{j=1}^n \hat{a}_j \varphi_j$ avec $\hat{a}_j = (1/n) \sum_{i=1}^n \varphi_j(X_i)$.

Ainsi pour estimer la projection s_m , il suffit de chercher à minimiser le contraste empirique sur les sous-espaces \mathcal{S}_m

$$\hat{s}_m = \arg \min_{t \in \mathcal{S}_m} \gamma_n(t).$$

La question se pose désormais de choisir de manière pertinente un espace \mathcal{S}_m pour estimer la fonction s . Pour cela nous allons étudier le risque des estimateurs s_m pour un ensemble donné de m dépendant de n , appelé collection de modèle et noté \mathcal{M}_n .

1.1.2 Étude du risque et choix non adaptatif de la dimension

D'après le théorème de Pythagore, le risque quadratique peut se décomposer sous la forme biais variance suivante

$$\mathbb{E}\|s - \hat{s}_m\|^2 = \|s - s_m\|^2 + \mathbb{E}\|s_m - \hat{s}_m\|^2. \quad (1.3)$$

Le premier terme $\|s - s_m\|^2$ appelé biais est l'erreur d'approximation liée à l'aspect fini dimensionnel. Le deuxième terme $\mathbb{E}\|s_m - \hat{s}_m\|^2$, appelé variance ou erreur stochastique (car $s_m = \mathbb{E}[\hat{s}_m]$) est engendrée par le caractère aléatoire des observations. Nous pouvons remarquer que si la dimension D_m croît alors l'erreur d'approximation diminue alors que l'erreur stochastique augmente puisque le nombre de coefficients à estimer augmente. Donc pour minimiser le risque quadratique intégré, il faut déterminer la dimension optimale D_m^* définie par $D_m^* = \arg \min_{D_m} \mathbb{E}\|s - \hat{s}_m\|^2$. L'estimateur résultant \hat{s}_{m^*} est appelé oracle. C'est l'estimateur qui minimise le risque quadratique intégré.

Pour déterminer D_m^* , il faut résoudre le compromis biais-variance de l'Équation (1.3). Pour cela, il est usuel de supposer que la fonction s appartient à un certain espace de régularité indexé par β (Hölder, Sobolev, Nikolskii, ...). En général une base est associée à un espace de régularité (voir Barron et al. (1999)), c'est pourquoi nous n'entrerons pas dans les détails ici. Pour une régularité β , le biais au carré est de l'ordre de $D_m^{-2\beta}$. Le terme de variance correspondant à l'erreur stochastique est une quantité aléatoire. Son espérance est souvent majorée par une quantité déterministe. La borne de la variance est alors de l'ordre d'une fonction croissante de la dimension D_m divisée par le nombre d'observations n . Il suffit ensuite de minimiser la fonction de D_m obtenue en additionnant biais et variance pour obtenir la dimension optimale. Cependant, ce choix de dimension est dit *non adaptatif* puisqu'il dépend de paramètres inconnus du problème, c'est-à-dire la régularité de la fonction à estimer.

Exemple : estimation de densité. Soit \hat{f}_m l'estimateur par projection dans la base orthonormée $(\varphi_j)_{[1, D_m]}$. Supposons qu'il existe une constante strictement positive Φ_0 telle que $\|\sum_{j=1}^{D_m} \varphi_j^2\|_\infty \leq \Phi_0 D_m$, alors le risque quadratique est majoré comme suit

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{\Phi_0 D_m}{n}.$$

De plus supposons que dans la base considérée f ait une régularité β de telle sorte que

$$\|f - f_m\|^2 \leq L D_m^{-2\beta} \quad \text{avec } L > 0.$$

Le risque quadratique est alors majoré par

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq L D_m^{-2\beta} + \frac{\Phi_0 D_m}{n} := \psi(D_m).$$

Il suffit alors de minimiser la fonction ψ pour obtenir un choix pertinent de D_m . En dérivant on trouve

$$\psi'(D_m) = -2\beta L D_m^{-2\beta-1} + \frac{\Phi_0}{n},$$

la solution est alors $D_m^* = (2L/\Phi_0)^{1/(2\beta+1)} n^{1/(2\beta+1)}$ et $\psi''(D_m) > 0$. Donc le minimum est bien atteint en D_m^* . Donc D_m^* dépend de L et β dépendant eux-même de f qui est inconnue.

De plus, on déduit que la vitesse de convergence de l'estimateur est de l'ordre de $n^{-\frac{2\beta}{2\beta+1}}$, i.e.

$$\sup_{\hat{f}_m} \mathbb{E}\|f - \hat{f}_m\|^2 \leq O\left(n^{-\frac{2\beta}{2\beta+1}}\right).$$

Il peut aussi être démontré qu'il s'agit de la vitesse optimale au sens minimax, voir [Tsybakov \(2009\)](#).

Par la suite, nous allons présenter une méthode permettant de sélectionner une dimension adéquate à partir des données, validée par une borne du risque de l'estimateur final. Cette méthode permet de construire un estimateur dont le risque quadratique est proche de celui de l'oracle. Cette méthode est uniquement conduite par les données. Ces méthodes sont dites *data driven*. De tels estimateurs sont appelés estimateurs adaptatifs car sans connaître la régularité associée à la fonction s , l'estimateur adaptatif atteint la même vitesse que si la régularité était connue. L'estimateur s'adapte donc à la régularité de la fonction à estimer.

1.1.3 Sélection de modèles et inégalités oracles

Dans cette thèse nous utiliserons deux méthodes différentes pour sélectionner une dimension \hat{m} afin que l'estimateur $\hat{s}_{\hat{m}}$ soit optimal ou adaptatif au sens de l'oracle, c'est-à-dire que cet estimateur vérifie une inégalité oracle

$$\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 \leq \mathfrak{C} \inf_{m \in \mathcal{M}_n} \mathbb{E}\|s - \hat{s}_m\|^2 + \Delta_n$$

avec Δ_n un terme négligeable devant $\inf_{m \in \mathcal{M}_n} \mathbb{E}\|s - \hat{s}_m\|^2$ et \mathfrak{C} une constante strictement positive qui dans l'idéal devrait être proche de 1 et \mathcal{M}_n une collection de modèle.

De plus, nous nous plaçons dans un cadre *non-asymptotique*, c'est-à-dire que les résultats que nous établissons sont valides à distance finie quand le nombre d'observations n est fixe.

Avant de présenter les méthodes adaptatives, il nous faut formuler quelques hypothèses sur la nature des modèles \mathcal{S}_m . Ils doivent au moins vérifier trois des quatre hypothèses suivantes :

\mathcal{H}_1 Connexion de normes : $\exists \Phi_0 > 0, \forall t \in \mathcal{S}_m, \|t\|_\infty \leq \Phi_0 \sqrt{D_m} \|t\|$.

\mathcal{H}_2 La dimension de \mathcal{S}_m, D_m , est majorée par n le nombre d'observations.

\mathcal{H}_3 Modèles emboîtés : $\forall m' \leq m, \mathcal{S}_{m'} \subset \mathcal{S}_m$.

\mathcal{H}'_3 Espace englobant : $\exists m_n \in \mathcal{M}_n, \forall m \in \mathcal{M}_n, \mathcal{S}_m \subset \mathcal{S}_{m_n}$

L'hypothèse \mathcal{H}_1 de connexion de normes est la plus importante des trois hypothèses. Tout au long des chapitres, nous utilisons des bases vérifiant cette hypothèse. L'hypothèse \mathcal{H}_2 est peu contraignante, dans l'exemple d'estimation de densité, le terme de variance est de l'ordre de D_m/n donc il est raisonnable que la dimension reste plus petite que n pour avoir une erreur d'estimation qui tende vers 0. L'hypothèse \mathcal{H}_3 est relativement forte. Si la base ne la vérifie pas, nous pouvons nous satisfaire de l'hypothèse \mathcal{H}'_3 . Cette dernière hypothèse assure que pour tous modèles $\mathcal{S}_{m'}$ et \mathcal{S}_m avec m' et m dans \mathcal{M}_n alors $\mathcal{S}_{m'} + \mathcal{S}_m \subset \mathcal{S}_{m_n}$.

La première méthode que nous présentons est la sélection de modèle par pénalisation inspirée des travaux de [Barron et al. \(1999\)](#), [Birgé and Massart \(1997\)](#) et [Massart \(2003\)](#). La deuxième est une méthode dite de Lepski plus précisément issue des travaux de [Goldenshluger and Lepski \(2011\)](#).

Sélection de modèles par pénalisation

Idéalement, il faudrait trouver la dimension m qui minimise le risque quadratique $\mathbb{E}\|s - \hat{s}_m\|^2$. Cette quantité ne peut évidemment pas être minimisée directement pour les raisons déjà évoquées précédemment. Une première idée serait alors de minimiser en m le risque empirique $\gamma_n(\hat{s}_m)$. Néanmoins si l'on considère que les espaces \mathcal{S}_m sont emboîtés, cela implique que pour $\mathcal{S}_{m'} \subset \mathcal{S}_m, \hat{s}_{m'} \in \mathcal{S}_m$ et donc $\gamma_n(\hat{s}_m) \leq \gamma_n(\hat{s}_{m'})$. Ainsi le risque empirique décroît avec la taille du modèle contrairement au véritable risque. Cela conduit à une sous-estimation systématique de l'erreur quadratique. Il faut alors contrebalancer le risque empirique en pénalisant cette quantité en fonction de la dimension de chaque modèle et de la complexité associée à la collection. Ainsi la dimension est déterminée comme

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{ \gamma_n(\hat{s}_m) + \text{pen}(m) \}.$$

Dans le cas particulier de l'estimation de densité, qui nous intéresse particulièrement, ce choix peut encore se voir d'une autre manière. Nous cherchons m tel que $\|s - \hat{s}_m\|^2$ soit minimale. En remarquant que pour s_m la projection de s sur \mathcal{S}_m

$$\|s - \hat{s}_m\|^2 = \|s\|^2 - \|s_m\|^2 + \|s_m - \hat{s}_m\|^2,$$

il est alors équivalent de chercher à minimiser $-\|s_m\|^2 + \|s_m - \hat{s}_m\|^2$. La quantité $-\|s_m\|^2$ est alors remplacée par son pendant empirique $-\|\hat{s}_m\|^2 = \gamma_n(\hat{s}_m)$. Le terme de variance est quant à lui approché par son ordre de grandeur moyen. Si nous nous plaçons dans le cas classique d'estimation de densité, un majorant du terme de variance serait $\Phi_0 D_m/n$. La pénalité serait alors posée comme $\text{pen}(m) = \kappa \Phi_0 D_m/n$ avec κ une constante numérique strictement positive. Dans ce cas particulier, les résultats théoriques indiquent que $\kappa \geq 4$ convient. Cependant ce résultat est obtenu après moult majorations et ne correspond pas aux majorations les plus fines possibles. Ainsi en pratique κ est pris plus petit.

La question de la calibration de la constante κ est le centre de nombreux questionnements depuis les travaux de [Birgé and Massart \(2007\)](#). En pratique, il existe des heuristiques de pente qui permettent de déterminer une valeur de κ en fonction des données comme le package CAPUSHE développé en Matlab[®] par [Baudry et al. \(2012\)](#) et plus récemment en R. Dans la suite, nous utiliserons une méthode de calibration basée sur des répétitions préliminaires intensives afin de déterminer la meilleure constante possible. Une fois la constante déterminée, elle est fixée une fois pour toute pendant l'évaluation numérique de la procédure adaptative.

Enfin si la pénalité est bien choisie, nous obtenons une inégalité dite de type oracle

$$\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 \leq C \inf_{m \in \mathcal{M}_n} \left\{ \|s - s_m\|^2 + \text{pen}(m) \right\} + \Delta_n. \quad (1.4)$$

avec C une constante strictement positive et Δ_n un terme négligeable qui dans la suite de la thèse sera de l'ordre de $1/n$.

Méthodes de Goldenshluger et Lepski

La méthode à la Lepski est une méthode visant, à l'origine, à choisir la fenêtre d'un estimateur à noyau de telle manière que l'estimateur vérifie une inégalité de type oracle. La méthode présentée par la suite est la méthode développée dans [Goldenshluger and Lepski \(2011\)](#) pour l'estimation de densité qui permet d'obtenir des résultats non-asymptotiques comparables à l'Équation (1.4). La méthode présentée dans [Goldenshluger and Lepski \(2011\)](#) est un raffinement de techniques antérieures se trouvant dans les travaux de [Kerkycharian et al. \(2001\)](#) où les premières inégalités oracles sont obtenues, dans [Goldenshluger and Lepski \(2008, 2009\)](#) pour une application au modèle de bruit blanc. Cette méthode est aussi reprise dans [Goldenshluger and Lepski \(2013\)](#) pour des estimateurs linéaires et [Goldenshluger and Lepski \(2014\)](#) pour une estimation minimax adaptative de densité sur \mathbb{R}^d . [Birgé \(1999\)](#) fait le lien entre la sélection de modèle et la méthode de Lepski introduite à l'origine dans [Lepskii \(1991\)](#).

La différence avec la sélection modèle présentée précédemment réside dans l'estimation du biais. En effet, dans les deux cas nous remplaçons le terme de variance de l'Équation (1.3) par son ordre de grandeur moyen. Précédemment le biais était estimé à l'aide du contraste empirique γ_n . Désormais nous n'utilisons plus le paradigme de minimisation de contraste. L'estimation du biais repose sur une comparaison par paire des estimateurs comme suit

$$A(m) = \sup_{m' \in \mathcal{M}_n} \left\{ \|\hat{s}_{m'} - \hat{s}_{m' \wedge m}\|^2 - V(m') \right\}_+$$

où V est un majorant de la variance dans la décomposition biais-variance de l'Équation (1.3).

Pour comprendre cette estimation particulière du biais, nous proposons l'heuristique suivante. Nous voulons estimer le terme $\|s - s_m\|^2$ avec s_m la projection de s sur l'espace \mathcal{S}_m avec $s_m = \Pi_{\mathcal{S}_m}(s)$, d'où $\|s - s_m\|^2 = \|s - \Pi_{\mathcal{S}_m}(s)\|^2$. La fonction s étant inconnue, elle est alors remplacée par une quantité empirique $s_{m'} = \Pi_{\mathcal{S}_{m'}}(s)$ avec $m' \in \mathcal{M}_n$ ce qui donne $\|\Pi_{\mathcal{S}_{m'}}(s) - \Pi_{\mathcal{S}_m}(\Pi_{\mathcal{S}_{m'}}(s))\|^2$. Sous l'hypothèse \mathcal{H}_3 , les modèles considérés sont emboîtés, donc $\Pi_{\mathcal{S}_m}(\Pi_{\mathcal{S}_{m'}}(s)) = \Pi_{\mathcal{S}_{m \wedge m'}}(s)$ et $\|\Pi_{\mathcal{S}_{m'}}(s) - \Pi_{\mathcal{S}_{m \wedge m'}}(s)\|^2$. Il faut encore retrancher le terme de variance $V(m)$ car cet estimateur du biais introduit de la variabilité absente de la quantité de départ. Ensuite il faut calculer cette quantité pour tout m' dans la collection \mathcal{M}_n . Enfin la dimension est choisie comme

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{A(m) + V(m)\}.$$

Finalement, l'inégalité oracle suivante est obtenue

$$\mathbb{E}\|s - \hat{s}_{\hat{m}}\|^2 \leq \mathfrak{C} \inf_{m \in \mathcal{M}} \left\{ \|s - s_m\|^2 + V(m) \right\} + \Delta_n$$

où \mathfrak{C} et Δ_n sont de même nature que dans l'inégalité (1.4). Dans le cas de l'estimation de densité $V(m) = \kappa' \Phi_0 D_m / n$ avec κ' une constante strictement positive. Ainsi comme dans les méthodes par pénalisation, le terme de variance dépend d'une constante κ' qu'il faut calibrer. La méthode de Goldenshluger et Lepski est évidemment toujours valide si le risque quadratique intégré est considéré à la place du risque ponctuel.

Arrêtons-nous sur le problème de calibration de la constante dans les procédures de Goldenshluger et Lepski. De récents travaux de [Lacour and Massart \(2015\)](#) proposent une méthode pour calibrer la constante dans le cas d'estimation de densité avec des noyaux. Nous illustrons cette méthode dans le Chapitre 3 pour des estimateurs par projection dans le cas du modèle de convolution.

Si V est assez grand, la méthode atteint ainsi de bons résultats. [Lacour and Massart \(2015\)](#) mettent en évidence, d'un point de vue théorique, un phénomène d'explosion. Si V est en-dessous d'un certain seuil noté V_0 alors la procédure échoue et le risque augmente de façon spectaculaire. Tandis que pour $V \geq V_0$ le risque est quasi optimal. Ce phénomène est illustré par des simulations. Les auteurs trouvent que $\kappa = 1$ est la valeur seuil pour l'estimation de densité à noyau. Ils proposent aussi de légèrement modifier la procédure de la manière suivante

$$A(m) = \sup_{m' \in \mathcal{M}_n} \left\{ \|\hat{s}_{m'} - \hat{s}_{m' \wedge m}\|^2 - \kappa_1 V(m') \right\}_+$$

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{A(m) + \kappa_2 V(m)\},$$

avec $\kappa_2 \neq \kappa_1$. Une distinction entre les constantes permettrait une meilleure calibration. Des simulations préliminaires montrent que dans le contexte de [Lacour and Massart \(2015\)](#) $\kappa_1 = 1$ et $\kappa_2 = 2$ seraient des choix pertinents. Les auteurs préconisent une procédure en deux étapes pour calibrer les constantes : premièrement, implémenter la méthode avec $\kappa_2 = \kappa_1$ et déterminer κ_1 tel qu'il y ait un saut du risque ; deuxièmement prendre κ_2 tel que $\kappa_2 = 2\kappa_1$.

1.1.4 Inégalités de concentration

L'obtention d'inégalités oracles de même que la construction de la pénalité reposent en grande partie sur des inégalités de concentration de supremum de processus empiriques autour de leur espérance.

En effet, l'approche adoptée peut être résumée de la manière suivante. Posons $\bar{\gamma}_n$ le processus empirique recentré

$$\bar{\gamma}_n(t) = \gamma_n(t) - \mathbb{E}[\gamma_n(t)].$$

L'estimateur adaptatif $\hat{s}_{\hat{m}}$ par définition vérifie les inégalités suivantes $\forall m \in \mathcal{M}_n$

$$\gamma_n(\hat{s}_{\hat{m}}) + \text{pen}(\hat{m}) \leq \gamma_n(\hat{s}_m) + \text{pen}(m) \leq \gamma_n(s_m) + \text{pen}(m),$$

ce qui implique

$$\bar{\gamma}_n(\hat{s}_{\hat{m}}) + \mathbb{E}[\gamma_n(\hat{s}_{\hat{m}})] + \text{pen}(\hat{m}) \leq \bar{\gamma}_n(\hat{s}_m) + \mathbb{E}[\gamma_n(s_m)] + \text{pen}(m),$$

et

$$\mathbb{E}[\gamma_n(\hat{s}_m)] \leq \mathbb{E}[\gamma_n(s_m)] + \bar{\gamma}_n(\hat{s}_m) - \bar{\gamma}_n(s_m) + \text{pen}(m) - \text{pen}(\hat{m}).$$

De plus, le contraste empirique est défini tel que $\mathbb{E}[\gamma_n(t)] = \|s - t\|^2 - \|s\|^2$, ainsi

$$\|s - \hat{s}_m\|^2 \leq \|s - s_m\|^2 + \text{pen}(m) + \bar{\gamma}_n(\hat{s}_m) - \bar{\gamma}_n(s_m) - \text{pen}(\hat{m}),$$

Nous en déduisons alors que la pénalité doit être assez grande pour contrebalancer les variations de $\bar{\gamma}_n(\hat{s}_m) - \bar{\gamma}_n(s_m)$ mais raisonnablement grande puisque que nous voulons que $\|s - s_m\|^2 + \text{pen}(m)$ reste de l'ordre de $\mathbb{E}[\|s - \hat{s}_m\|^2]$. C'est pourquoi nous avons besoin d'inégalités de concentration afin d'analyser la déviation de $\bar{\gamma}_n(\hat{s}_m) - \bar{\gamma}_n(s_m)$.

Pour plus de détails, le lecteur pourra se référer à [Massart \(2003\)](#).

Inégalité de Talagrand

Les inégalités de Talagrand permettent de contrôler les déviations de suprema de processus empiriques autour de leur espérance. Le résultat suivant est une conséquence directe de l'inégalité de Talagrand donnée dans [Klein and Rio \(2005\)](#).

Lemme 1.1.1. Soient X_1, \dots, X_n des variables aléatoires *i.i.d.* et

$$r_n(f) = \frac{1}{n} \sum_{k=1}^n \left(f(X_k) - \mathbb{E}[f(X_k)] \right)$$

pour f appartenant à un certain ensemble dénombrable \mathcal{F} de fonctions mesurables uniformément bornées. Alors pour $\xi^2 > 0$,

$$\mathbb{E} \left[\left\{ \sup_{f \in \mathcal{F}} |r_n(f)|^2 - 2(1 + 2\xi^2)H^2 \right\}_+ \right] \leq \frac{4}{K_1} \left(\frac{v}{n} e^{-K_1 \xi^2 \frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi}{7\sqrt{2}} \frac{nH}{M_1}} \right)$$

avec $C(\xi) = (\sqrt{1 + \xi^2} - 1) \wedge 1$ et $K_1 = \frac{1}{6}$, M_1 , H et v sont telles que

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \quad \mathbb{E} \left[\sup_{f \in \mathcal{F}} |r_n(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \text{Var}[f(X_1)] \leq v.$$

Inégalité de Bernstein

L'inégalité de Bernstein permet de majorer la probabilité de déviation d'un processus empirique recentré. Elle s'énonce de la manière suivante :

Lemme 1.1.2. Soient X_1, \dots, X_n des variables aléatoires *i.i.d.* telles que $\text{Var}(X_1) \leq v^2$ et $|X_1| \leq b$ p.s. Soit $S_n = \sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, alors pour $\epsilon > 0$

$$\mathbb{P} [|S_n - \mathbb{E}[S_n]| \geq n\epsilon] \leq 2 \max \left(\exp \left(-\frac{n\epsilon^2}{4v^2} \right), \exp \left(-\frac{n\epsilon}{4b} \right) \right).$$

1.2 Modèle de convolution

Le but de cette thèse est de proposer des méthodes d'estimation adaptative dans des modèles bruités et tout particulièrement dans le modèle avec erreur additive appelé modèle de convolution. Nous allons maintenant mettre en lumière le problème de convolution et les stratégies de déconvolution associées.

Avant d'entrer dans les détails du traitement statistique du problème, arrêtons-nous d'abord sur quelques motivations issues d'applications. Les problèmes de déconvolution apparaissent dans de nombreux champs des statistiques non-paramétriques comme dans les problèmes d'estimation de densité avec des erreurs de mesures additives ou multiplicatives, de régression avec erreur de mesure sur les variables ou encore le traitement du signal. Ces modélisations se retrouvent en économie, en biologie, en médecine ou encore en reconstruction d'image.

Par exemple en économétrie, le modèle de volatilité stochastique est largement utilisé, dans la modélisation des séries temporelles financières. En effet, l'analyse des investissements financiers repose largement sur la modélisation de la volatilité des prix des actifs. Le lien avec le modèle de convolution a été fait par [Comte \(2004\)](#) en montrant que le modèle à volatilité stochastique peut aussi s'écrire comme un modèle de régression avec erreurs sur les variables.

En bio-informatique, les techniques de puce à ADN sont devenues très populaires ces dernières années. Une biopuce consiste en un ensemble de molécules d'ADN fixées sur plusieurs milliers de rangées sur une petite surface. Le fonctionnement de la puce repose sur le fait qu'un gène présent dans l'échantillon ARN reforme spontanément sa double hélice avec son homologue se trouvant sur la puce à ADN. Les endroits où ses liaisons se forment deviennent alors fluorescentes et un scanner à micro rayons est utilisé pour lire les intensités sur la puce. Cependant le processus entier pour obtenir les intensités fluorescentes dans une étude de puce à ADN est soumis aux erreurs de mesure.

Pour plus d'exemples, le lecteur pourra se référer à l'ouvrage de [Meister \(2009\)](#).

1.2.1 Problème et modèle statistique

Le problème de déconvolution consiste à estimer une fonction f alors que nous n'avons qu'un accès indirect à travers une autre fonction h comme suit

$$h(x) = f \star G(x) = \int f(x - y) dG(y). \quad (1.5)$$

Nous observons alors une version bruitée de f par l'opérateur de convolution G . La stratégie serait d'estimer h puis d'inverser l'opérateur de G . Cependant, cette opération ne peut être réalisée directement car nous ne connaissons pas de version analytique de l'opérateur inverse de G . C'est pourquoi nous allons étudier ce problème dans le domaine des fréquences en utilisant des résultats d'analyse de Fourier. Pour cela, nous nous restreignons au cas continu, c'est-à-dire que G admet une densité g par rapport à la mesure de Lebesgue

$$h(x) = f \star G(x) = \int f(x - y)g(y) dy.$$

Dans la suite, nous supposons que le support d'intégration est la droite réelle. Nous définissons la transformée de Fourier g^* d'une fonction g comme

$$g^*(t) = \int e^{itu} g(u) du, \quad t \in \mathbb{R}, \quad (1.6)$$

la transformée de Fourier inverse est alors

$$g(x) = \frac{1}{2\pi} \int e^{-itx} g^*(t) dt, \quad x \in \mathbb{R},$$

L'intérêt de la transformée de Fourier est qu'elle transforme le produit de convolution en un simple produit, en effet

$$h = f \star G \Leftrightarrow h^* = f^* g^*. \quad (1.7)$$

Ainsi nous observons le modèle de convolution suivant

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \quad (1.8)$$

où les $(X_j)_{j \in \llbracket 1, n \rrbracket}$ sont des variables aléatoires indépendantes et de même loi (i.i.d.) de densité f , les $(\varepsilon_j)_{j \in \llbracket 1, n \rrbracket}$ sont aussi i.i.d. de densité f_ε et les deux suites sont indépendantes. La densité des $(Y_j)_{j \in \llbracket 1, n \rrbracket}$ est notée f_Y . L'hypothèse clé est l'indépendance entre les deux séquences $(X_j)_{j \in \llbracket 1, n \rrbracket}$ et $(\varepsilon_j)_{j \in \llbracket 1, n \rrbracket}$. En effet, des contextes où les $(X_j)_{j \in \llbracket 1, n \rrbracket}$ ne sont pas indépendantes mais faiblement mélangeantes pourraient être envisagés (voir Masry (1993, 2002) ou Comte et al. (2008)). Ainsi le Modèle (1.8), sous les hypothèses d'indépendance, correspond bien à une convolution des densités f et f_ε comme décrit par l'Équation (1.5).

Dans le problème classique de déconvolution, le but est d'estimer la densité f des $(X_j)_{j \in \llbracket 1, n \rrbracket}$. Dans certains cas, la fonction de répartition ou encore le quantile sont estimés (voir Dattner et al. (2011); Dattner and Reiser (2013); Dattner et al. (2016)). La loi du bruit f_ε est selon les cas supposée connue ou inconnue. Évidemment le problème devient plus difficile quand la loi du bruit est supposée inconnue, cependant l'hypothèse de loi connue est peu réaliste dans les applications.

Partant du Modèle (1.8), les différentes étapes du problème de déconvolution peuvent être schématisées en trois points :

- ▷ Estimation de f_Y^* à partir des observations, notée \hat{f}_Y^* .
- ▷ Calcul de \hat{f}_Y^* et division par f_ε^* (lorsque c'est possible) pour obtenir l'estimateur \hat{f}^* .
- ▷ Régularisation de \hat{f}^* pour que sa transformée de Fourier inverse pour que \hat{f} existe.

Nous allons maintenant entrer dans les détails des méthodes de déconvolution en explicitant les différentes étapes du problème statistique pour l'estimation de densité. Comme décrit dans la Section 1.1, notre but est de construire un estimateur ayant de bonnes propriétés statistiques au sens minimax optimal puis de trouver un estimateur adaptatif qui atteint cette vitesse minimax grâce à une procédure complètement dictée par les données.

Ainsi nous remarquons que pour une densité la transformée de Fourier (1.6) s'écrit sous forme d'espérance

$$f_Y^*(t) = \int e^{itu} f_Y(u) du = \mathbb{E} \left[e^{itY_1} \right].$$

Cette expression correspond aussi à la fonction caractéristique de la variable aléatoire Y_1 . Cette écriture permet de proposer un estimateur basé sur la méthode des moments. En effet, nous observons directement les $(Y_j)_{j \in \llbracket 1, n \rrbracket}$ et pouvons donc proposer un estimateur fonctionnel classique de sa fonction caractéristique :

$$\hat{f}_Y^*(t) = \frac{1}{n} \sum_{j=1}^n e^{itY_j}. \quad (1.9)$$

De plus, cet estimateur a de très bonnes propriétés. Il est clair qu'il est sans biais

$$\mathbb{E}[\hat{f}_Y^*(t)] = f_Y^*(t), \quad t \in \mathbb{R}.$$

Et son risque quadratique est d'ordre $1/n$, en effet

$$\mathbb{E} \left[|\hat{f}_Y^*(t) - f_Y^*(t)|^2 \right] = \mathbb{V}\text{ar} \left[\hat{f}_Y^*(t) \right] = \frac{1}{n} \mathbb{V}\text{ar} \left[e^{itY_1} \right] = \frac{1}{n} \left(1 - |f_Y^*(t)|^2 \right). \quad (1.10)$$

D'après la formule (1.7) nous avons

$$f_Y^*(t) = f^*(t) f_\varepsilon^*(t).$$

Si nous supposons que f_ε^* ne s'annule pas alors f^* peut s'estimer comme

$$\hat{f}^*(t) = \frac{1}{n} \sum_{j=1}^n \frac{e^{itY_j}}{f_\varepsilon^*(t)} = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)}, \quad t \in \mathbb{R}, \quad (1.11)$$

nous pouvons prendre la transformée de Fourier inverse et obtenir

$$\hat{f}_{\text{naïf}}(t) = \frac{1}{2\pi} \int e^{-itu} \hat{f}^*(u) du = \frac{1}{2\pi} \int e^{-itu} \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} du.$$

Cet estimateur semble à première vue très intéressant car héritant des propriétés de \hat{f}_Y^* il est sans biais. Cependant, nous parlons d'estimateur *naïf* car ni \hat{f}_Y^* ni $1/f_\varepsilon^*$ ne sont intégrables sur \mathbb{R} . Ainsi il n'est pas correctement défini. L'estimateur a besoin d'être régularisé, par exemple en restreignant l'intervalle d'intégration à un compact où l'estimateur est intégrable ou encore en introduisant une troncature au dénominateur.

Ainsi la section suivante est dédiée à la présentation de différentes méthodes résolvant ce problème d'intégration. Nous présentons des techniques de déconvolution quand la loi des $(\varepsilon_j)_{j \in [1, n]}$ est connue puis lorsque la loi n'est plus connue. Nous passons en revue les résultats les plus importants de la littérature et expliquons de quelle manière nous avons contribué à améliorer des techniques de déconvolution à bruit connu et inconnu.

1.2.2 Méthodes de déconvolution à bruit connu

Résultats existants

Méthodes à noyaux. Dans ce paragraphe, nous expliquons la méthode d'estimation de densité à noyau dans la cas du modèle de convolution (1.8). Pour plus de détails sur l'estimation fonctionnelle à noyau, voir [Tsybakov \(2009\)](#). Nous nous attardons volontairement sur cette méthode et les travaux y afférant car c'est essentiellement avec ceux-ci que nous comparerons les résultats obtenus dans cette thèse.

Commençons par définir un estimateur de f_Y noté \tilde{f}_Y à partir d'un noyau noté K utilisant les observations $(Y_j)_{j \in [1, n]}$

$$\tilde{f}_Y(x) = \frac{1}{nh} \sum_{j=1}^n K \left(\frac{x - Y_j}{h} \right), \quad x \in \mathbb{R}$$

avec $h > 0$ le paramètre de lissage et $K : \mathbb{R} \rightarrow \mathbb{R}$ un noyau vérifiant $K \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$ afin que $\tilde{f}_Y \in \mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$. Cette condition assure l'existence de la transformée de Fourier de \tilde{f}_Y . Suivant le schéma de déconvolution de la Section précédente, il nous faut calculer la

transformée de Fourier de \tilde{f}_Y , ainsi pour tout réel t

$$\begin{aligned}\tilde{f}_Y^*(t) &= \int e^{itu} \tilde{f}_Y(u) du = \int e^{itu} \frac{1}{nh} \sum_{j=1}^n K\left(\frac{u - Y_j}{h}\right) du \\ &= \frac{1}{nh} \sum_{j=1}^n \int e^{it(hv + Y_j)} K(v) dv = \frac{1}{n} \sum_{j=1}^n e^{itY_j} \int e^{ithv} K(v) dv = \hat{f}_Y^*(t) K^*(th),\end{aligned}$$

avec \hat{f}_Y^* défini par l'Équation (1.9). D'après (1.7) un estimateur de f^* s'exprime comme

$$\tilde{f}^*(t) = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)} K^*(th) = \hat{f}^*(t) K^*(th). \quad (1.12)$$

Ce dernier estimateur diffère de l'estimateur introduit en (1.11) par la présence de la transformée de Fourier du noyau K . Précédemment, nous avons pointé le problème d'intégrabilité de cet estimateur. Nous allons maintenant tirer partie de la présence du noyau. Pour que l'estimateur défini en (1.12) soit intégrable, il faut trouver un noyau à support compact et borné. Pour cela, nous proposons d'utiliser le noyau *sinus cardinal* défini par $K(x) = \sin(\pi x)/\pi x$ qui a pour transformée de Fourier $K^*(t) = \mathbb{1}_{[-\pi, \pi]}(t)$. Ainsi pour ce noyau, nous obtenons

$$\hat{f}^*(t) = \frac{\hat{f}_Y^*(t)}{f_\varepsilon^*(t)} \mathbb{1}_{[-\pi/h, \pi/h]}(t).$$

L'estimateur est désormais borné et à support compact dès que f_ε^* n'est pas nulle, nous pouvons alors en prendre la transformée de Fourier inverse et obtenir un estimateur de f

$$\hat{f}_h(x) = \frac{1}{2\pi} \int_{-\pi/h}^{\pi/h} e^{-itx} \frac{\hat{f}_Y^*(x)}{f_\varepsilon^*(x)} dx. \quad (1.13)$$

Ainsi les premiers travaux à avoir adopté cette méthode à noyau sur la transformée de Fourier puis à en prendre la transformée de Fourier inverse sont [Carroll and Hall \(1988\)](#) et [Stefanski and Carroll \(1990\)](#). Suite à ces travaux, le problème de déconvolution a été étudié de manière intensive par entre autres : [Zhang \(1990\)](#), [Fan \(1991\)](#), [Efromovich \(1997\)](#), [Delaigle and Gijbels \(2004, 2006\)](#), [Meister \(2004\)](#) et [Butucea \(2004\)](#). Il a aussi été récemment utilisé par [Comte and Lacour \(2013\)](#) dans un cadre multidimensionnel.

Lien avec l'estimation par projection. Cet estimateur peut aussi être vu comme un estimateur par projection en posant $m = 1/h$ dans les bornes de l'intégrale de l'Équation (1.13). Cet estimateur est utilisé dans [Comte et al. \(2006\)](#). En effet, en construisant une base d'ondelette à partir d'une fonction mère définie comme $\psi(x) = \sin(\pi x)/\pi x$ et en posant $\psi_{m,j} = \sqrt{D_m} \psi(D_m x - j)$, pour $D_m = 2^m$, nous obtenons la base de Shannon. Ainsi $(\psi_{m,j})_{j \in \mathbb{Z}}$ forme une base orthonormale de $\mathbb{L}^2(\mathbb{R})$ dont la transformée de Fourier admet un support inclus dans $[-\pi D_m, \pi D_m]$.

Le problème réside désormais en un choix adéquat de la fenêtre h ou de la dimension m . Nous pouvons remarquer que cette correspondance entre l'estimateur à noyau et l'estimateur par projection nous autorise à prendre des dimensions non entières. L'intérêt de voir cet estimateur comme un estimateur par projection est qu'il permet la construction d'un estimateur adaptatif à partir des techniques de sélection modèle présentées en Section 1.1.3. De plus, l'estimateur à noyau et l'estimateur par projection ont les mêmes vitesses de convergence. Dans la suite, nous adopterons donc l'angle d'estimation par projection.

Nous proposons maintenant une brève étude du risque quadratique intégré afin d'en déduire les vitesses classiques en déconvolution.

Proposition 1.2.1. Dans le Modèle (1.8), si $f_\varepsilon^* \neq 0$, \hat{f}_m est défini par (1.13) et $m = 1/h$ alors

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{\Delta(m)}{n}, \quad (1.14)$$

avec $\Delta(m) = 1/(2\pi) \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} du$.

Démonstration. Avant d'étudier le risque définissons

$$f_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-itx} \frac{f_Y^*(x)}{f_\varepsilon^*(x)} dx.$$

À l'aide de la formule de Plancherel et du théorème Pythagore, décomposons le risque en un terme de biais plus un terme de variance

$$\begin{aligned} \|f - \hat{f}_m\|^2 &= \frac{1}{2\pi} \|f^* - \hat{f}_m^*\|^2 = \frac{1}{2\pi} \|f^* - f_m^* + f_m^* - \hat{f}_m^*\|^2 \\ &= \frac{1}{2\pi} \|f^* \mathbb{1}_{[-\pi m, \pi m]^c} + (f_m^* - \hat{f}_m^*) \mathbb{1}_{[-\pi m, \pi m]}\|^2 \\ &= \frac{1}{2\pi} \left(\|f^* \mathbb{1}_{[-\pi m, \pi m]^c}\|^2 + \|f_m^* - \hat{f}_m^*\|^2 \right) \\ &= \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2. \end{aligned}$$

Nous avons ainsi la décomposition biais-variance suivante

$$\mathbb{E}\|f - \hat{f}_m\|^2 = \|f - f_m\|^2 + \mathbb{E}\|f_m - \hat{f}_m\|^2.$$

Étudions le terme de variance :

$$\begin{aligned} \mathbb{E}\|f_m - \hat{f}_m\|^2 &= \frac{1}{2\pi} \mathbb{E}\|f_m^* - \hat{f}_m^*\|^2 = \frac{1}{2\pi} \mathbb{E} \left[\int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} - \frac{f_Y(u)}{f_\varepsilon^*(u)} \right|^2 du \right] \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y(u)|^2 \right] du \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \text{Var} [\hat{f}_Y^*(u)] du. \end{aligned}$$

D'après la formule (1.10)

$$\text{Var} [\hat{f}_Y^*(u)] = \frac{1}{n} \left(1 - |f_Y^*(u)|^2 \right) \leq \frac{1}{n}.$$

Finalement, la variance est majorée par

$$\mathbb{E}\|f_m - \hat{f}_m\|^2 \leq \frac{1}{n} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} du = \frac{\Delta(m)}{n}.$$

D'où le résultat annoncé. \square

Comme expliqué en début de chapitre pour déduire les vitesses de convergence, il faut pouvoir calculer les ordres de grandeur du biais et de la variance. Nous supposons que f appartient à un espace de fonction analytique défini comme suit

$$\mathcal{A}_{r,s}(L) = \left\{ f \geq 0, \int f = 1 \quad \text{et} \quad \int |f^*(u)|^2 (u^2 + 1)^s e^{2a|u|^r} du \leq L \right\} \quad (1.15)$$

avec $L > 0$, $r \geq 0$, $a > 0$, $s \in \mathbb{R}$ et $s > 1/2$ si $r = 0$. Si $r = 0$ alors $\mathcal{A}_{r,s}(L)$ correspond à un espace de Sobolev. L'ordre de grandeur du biais pour cet espace de régularité est

$$\|f - f_m\|^2 = \frac{1}{2\pi} \int_{[-\pi m, \pi m]^c} |f^*(u)|^2 du \leq \frac{L}{2\pi} ((\pi m)^2 + 1)^{-s} e^{-2a|\pi m|^r}.$$

Pour la loi du bruit f_ε , nous supposons qu'il existe des constantes positives k_0 , k'_0 , γ , μ , et δ telles que pour tout réel x

$$k_0(x^2 + 1)^{-\gamma/2} e^{-\mu|x|^\delta} \leq |f_\varepsilon^*(x)| \leq k'_0(x^2 + 1)^{-\gamma/2} e^{-\mu|x|^\delta}. \quad (1.16)$$

Si $\delta = 0$, la fonction est dite ordinairement régulière ou *ordinary smooth* car sa transformée de Fourier a une décroissance polynomiale. Si $\mu > 0$ et $\delta > 0$, la fonction est dite super régulière ou *supersmooth* car elle a une décroissance exponentielle. On peut citer comme exemple de fonctions *supersmooth*, la densité de la loi Gaussienne qui a pour paramètres $\gamma = 0$ et $\delta = 2$ ou encore la densité de la loi de Cauchy avec $\gamma = 0$ et $\delta = 1$. Comme fonction *ordinary smooth*, on peut citer la densité de la loi de Laplace avec $\delta = \mu = 0$ et $\gamma = 2$.

	$\delta = 0$	$\delta > 0$
$r = 0$	$n^{\frac{-2s}{2\gamma+2s+1}}$	$(\log n)^{-2s/\gamma}$
$r > 0$	$\frac{(\log n)^{\frac{2\gamma+1}{r}}}{n}$	

TABLE 1.1 – Vitesse de convergence de l'estimateur \hat{f}_m pour le risque quadratique intégré.

Les vitesses obtenues après résolution du compromis biais variance de la borne de risque de l'Équation (1.14) sont reportées dans le Tableau 1.1. Pour le cas où la densité à estimer f et celle des erreurs f_ε sont *supersmooth* nous ne précisons pas les vitesses. En effet dans ce cas les vitesses de convergence sont souvent données de façon implicite (voir Butucea and Tsybakov (2008a,b)), excepté dans des cas très particulier. Lacour (2006) donne dans le cas de deux fonctions super régulières des vitesses de convergence exactes et explicites. Nous pouvons noter que les vitesses dépendent alors de l'intervalle $]k/(k+1), (k+1)/(k+2)[$ avec k un entier auquel appartient le ratio r/γ ou γ/r .

En premier lieu remarquons que la difficulté des problèmes de déconvolution réside essentiellement dans la régularité de la densité du bruit f_ε . Plus la fonction f_ε est régulière, plus le problème est difficile. Plus la loi du bruit est régulière, plus la vitesse d'estimation est lente. En effet, le modèle de convolution étant un cas particulier de problème inverse, le degré de difficulté (*ill-posedness*) du problème dépend de la régularité des transformées de Fourier de f et f_ε . Plus f_ε^* décroît rapidement, plus il est difficile de reconstruire f à partir des données car pour de petites valeurs de f_ε^* même un fort changement de f^* peut ne pas entraîner un changement important de f_Y^* . De plus, si f^* décroît rapidement alors les vitesses de convergence obtenues sont meilleures. En effet dans la première ligne du tableau ($r = 0$), pour un bruit de type Laplace, c'est-à-dire ordinairement régulier ou *ordinary smooth*, la vitesse est polynomiale alors que pour un bruit de type Gaussien, c'est-à-dire super régulier ou *supersmooth*, la vitesse est logarithmique. Ce qui illustre notre propos. En revanche, si les deux fonctions sont de transformée de Fourier Laplace ou Gaussienne,

les vitesses peuvent redevenir bien meilleures voire polynomiales (case vide). Évidemment il est notable que les vitesses de convergence sont plus lentes que les vitesses classiques non-paramétriques d'estimation de densité.

Les premiers travaux portant sur l'étude des vitesses de convergence dans le modèle de convolution se consacraient à calculer les vitesses de convergence pour des lois particulières et à mettre en évidence des vitesses de convergence différentes selon les régularités des fonctions. Ainsi [Carroll and Hall \(1988\)](#) sont parmi les premiers à avoir étudié la vitesse de convergence d'un estimateur à noyau en déconvolution. Sous l'hypothèse que f admette un nombre fini de dérivées (cas $r = 0$) et que le bruit soit une variable Gaussienne ($\delta = 2$) alors la vitesse minimax pour le risque ponctuel est logarithmique. [Stefanski and Carroll \(1990\)](#) montre la consistance de l'estimateur ponctuel et donnent une approximation du risque quadratique intégré. Une procédure par validation croisée est proposée pour choisir le paramètre de lissage du noyau. [Zhang \(1990\)](#) utilise aussi des méthodes de l'analyse de Fourier pour calculer les bornes supérieures de ses estimateurs ainsi que les bornes inférieures correspondants au risque quadratique intégré. L'auteur calcule les vitesses optimales pour les lois Gaussiennes et Cauchy. Il montre que ses estimateurs sont asymptotiquement normaux.

[Fan \(1991\)](#) généralise les vitesses optimales dans le modèle de convolution en montrant que les différentes vitesses sont liées au type de décroissance de la transformée de Fourier des erreurs. Ainsi l'auteur suppose que f_ε vérifie l'Équation (1.16) et f appartient à une classe de Hölder. Il démontre que les vitesses qu'il obtient sont optimales au sens *minimax* pour le risque quadratique ponctuel et en déduit les vitesses optimales pour la fonction de répartition. [Fan \(1993\)](#) prouve les bornes inférieures pour le risque \mathbb{L}^p sous les mêmes conditions de régularité que [Fan \(1991\)](#). Or les classes de Hölder sont contenues dans les classes de Sobolev qui contiennent également les fonctions *ordinary smooth* et *supersmooth*, cela implique que les vitesses de la première ligne du Tableau 1.1 sont *minimax* optimales. [Butucea \(2004\)](#) établit la vitesse minimax de convergence quand f est *supersmooth* et le bruit *ordinary smooth* pour le risque quadratique ponctuel et donne une majoration du risque quadratique intégré. [Butucea \(2004\)](#) montre aussi que dans certains cas les estimateurs à noyaux atteignent de meilleures vitesses que les estimateurs par ondelettes étudiés dans [Pensky and Vidakovic \(1999\)](#). Les bornes supérieures pour le risque \mathbb{L}^2 sont obtenues pour les estimateurs par ondelettes sur des classes similaires à $\mathcal{A}_{r,s}(L)$ définies par (1.15). Cependant l'auteur montre que les estimateurs à noyaux atteignent les mêmes vitesses de convergence pour le risque ponctuel et intégré. Enfin, [Butucea and Tsybakov \(2008a,b\)](#) étudient le cas où f et f_ε sont *supersmooth* pour le risque quadratique ponctuel et intégré. Néanmoins, les vitesses de convergence sont données de manière implicite. Cependant, les régularités associées au problème sont inconnues c'est pourquoi des estimateurs adaptatifs sont construits. [Efromovich \(1997\)](#) étudie le cas le plus défavorable quand les erreurs sont super régulières. Ainsi il construit un estimateur par projection adaptatif pour le risque \mathbb{L}^2 sur des classes de Sobolev qui est asymptotiquement efficace. [Comte et al. \(2006\)](#) établissent une procédure par sélection modèle, à la Birgé-Massart, qui permet de construire un estimateur adaptatif de f et d'en déduire une borne non asymptotique du risque \mathbb{L}^2 . Leur estimateur atteint automatiquement les vitesses optimales sauf dans un cas où une perte logarithmique intervient, due à l'adaptation. Nous pouvons aussi citer les travaux plus récents de [Comte and Lacour \(2013\)](#) dans un cadre multidimensionnel pour le risque \mathbb{L}^2 et [Rebelles \(2015\)](#) pour des risques \mathbb{L}^p avec $1 \leq p \leq \infty$. Dans ces deux travaux, la sélection modèle est faite à l'aide d'une méthode de type Goldenshluger-Lepski. [Goldenshluger \(1999\)](#) généralise les vitesses minimax aux vitesses adaptatives pour des classes de

Sobolev \mathbb{L}^p plus générales pour des fonctions du bruit régulières et super régulières. Pour la déconvolution de densité appartenant à la classe des fonctions analytiques ($r = 1$) avec un bruit de type Cauchy, [Tsybakov \(2000\)](#) montre qu'une perte en log est inévitable lorsque l'on passe à l'estimation adaptative avec le risque \mathbb{L}^2 .

Un autre objet d'intérêt dans la littérature est l'estimation de la fonction de répartition (f.d.r). Nous pouvons citer les travaux de [Dattner et al. \(2011\)](#) et [Dattner and Reiser \(2013\)](#) qui, pour le premier, étudient l'estimation adaptative de la f.d.r. de f en supposant qu'elle appartient à un espace de Sobolev et que f_ε est ordinairement régulière; le deuxième, considère quant à lui le même problème quand les erreurs sont supposées super régulières. Ce problème est aussi étudié par [Hall and Lahiri \(2008\)](#) et [Söhl and Trabs \(2012\)](#).

Autres méthodes. Premièrement, nous pouvons citer les méthodes d'ondelettes qui sont tout particulièrement adaptées à l'étude du risque quadratique intégré contrairement à l'estimation à noyau qui permet de faire de l'estimation ponctuelle et de l'estimation globale. L'estimation à partir d'une base d'ondelette a été faite par [Pensky and Vidakovic \(1999\)](#) qui construisent des estimateurs par ondelettes linéaires et non linéaires basés sur des ondelettes de type Meyer. Les estimateurs sont asymptotiquement optimaux et s'adaptent si f_ε est super régulière. Cependant [Butucea and Tsybakov \(2008a,b\)](#) ont montré que ces estimateurs étaient sous optimaux. Nous pouvons aussi citer [Pensky \(2002\)](#) qui sous l'hypothèse que f et f_ε sont *supersmooth* construit un estimateur par ondelettes non linéaires de f à partir d'ondelettes à support borné. L'auteure montre que cet estimateur s'adapte localement à la régularité de f et a ainsi de meilleures performances que les estimateurs basés sur des ondelettes de type Meyer si f est irrégulière.

Les méthodes précédentes supposent que la transformée de Fourier du bruit ne peut jamais s'annuler. Cependant il s'agit d'une hypothèse forte car elle exclut de nombreuses lois dont la loi uniforme. En effet, la transformée de Fourier pour une loi uniforme sur $[-1, 1]$ est $\sin t/t$ et a donc une infinité de zéros de la forme $k\pi$ avec $k \in \mathbb{Z}$. Il existe aussi des lois symétriques avec un support borné dont la transformée de Fourier s'annule. Ces situations correspondent à de nombreuses applications. En effet, [Hu and Ridder \(2004\)](#) démontre que dans les applications économiques l'hypothèse selon laquelle f_ε^* ne s'annule jamais n'est pas raisonnable. De plus en astronomie, il est aussi démontré que les variables d'intérêt sont mesurées avec une erreur additive uniforme. [van Es \(2011\)](#) propose une solution spécifique quand le bruit est uniforme en mettant en œuvre des méthodes de maximum de vraisemblance non-paramétrique. [Meister and Neumann \(2010\)](#) se concentrent sur l'estimation de densité dans le modèle de convolution quand la transformée de Fourier de la loi des erreurs s'annule et oscille. Leur estimateur atteint la même vitesse de convergence que dans un modèle où l'erreur aurait les mêmes queues de distributions, serait régulière et sa transformée de Fourier n'aurait pas de zéros.

Plus généralement, des méthodes avec un paramètre *ridge* peuvent être utilisées comme dans [Delaigle et al. \(2008\)](#). L'estimateur considéré est alors de la forme suivante

$$\hat{f}_{\text{ridge}}^*(t) = \frac{1}{n} \sum_{j=1}^n e^{itY_j} \frac{f_\varepsilon^*(-t)}{\max(|f_\varepsilon^*(t)|^2, \rho_n(t))}$$

Si $\max(|f_\varepsilon^*(t)|^2, \rho_n(t)) = |f_\varepsilon^*(t)|^2$ alors $\hat{f}^*(t) = \hat{f}_{\text{ridge}}^*(t)$. Au contraire si $|f_\varepsilon^*(t)|^2$ devient trop petit, alors le paramètre *ridge* de régularisation entre en jeu, en général il est posé comme

$$\rho_n(t) = |t|^\zeta n^{-\eta},$$

avec $\eta > 0$. Si $\zeta > 1/2$, peu importe la fonction f_ε^* , l'estimateur \hat{f}_{ridge}^* est intégrable. Sa transformée de Fourier inverse peut alors être calculée.

Principaux résultats des Chapitres 2 à 4

La principale contribution de cette thèse au modèle de convolution avec bruit connu est son étude particulière quand les variables X et ε sont positives. La variable ε n'est plus supposée centrée et n'est plus interprétée comme un bruit. Par souci de différenciation avec cette interprétation, nous réécrivons le modèle comme suit

$$Z_j = X_j + Y_j, \quad j = 1, \dots, n, \quad (1.17)$$

Ce modèle intervient en analyse de survie où X modélise la durée d'infection d'une maladie et Y sa durée d'incubation. Cette modélisation est utilisée dans les recherches sur le SIDA. Dans ce cas X et Y sont bien à support \mathbb{R}^+ puisqu'elles représentent des durées.

Nous supposons toujours que les $(X_j)_{j \in [1, n]}$ sont i.i.d. de densité inconnue f et de fonction de survie inconnue S_X avec $S_X(x) = \mathbb{P}[X > x]$. Les $(Y_j)_{j \in [1, n]}$ sont aussi i.i.d. de densité connue g et de fonction de survie S_Y . Toutes les variables aléatoires sont supposées positives. Nous notons h la densité des $(Z_j)_{j \in [1, n]}$ et S_Z sa fonction de survie. De plus, les $(X_j)_{j \in [1, n]}$ et les $(Y_j)_{j \in [1, n]}$ sont supposées indépendantes. Le but est donc d'estimer la densité f et la fonction de survie S_X des $(X_j)_{j \in [1, n]}$ quand seules les $(Z_j)_{j \in [1, n]}$ sont observées. Pointons tout de suite que l'estimation de S_X ne s'appuie pas sur celle de f . Nous n'intégrons pas un estimateur de f pour proposer un estimateur de S_X . Une stratégie particulière est appliquée à la fonction de survie. Ainsi dans le Chapitre 2 nous montrons que l'hypothèse de positivité des variables aléatoires est d'une importance capitale dans la stratégie d'estimation. De plus, cette méthode a pour autre avantage de pouvoir traiter des cas où la densité f admet une discontinuité en zéro.

Décrivons maintenant la méthode spécifique que nous développons pour l'estimation de la densité et de la survie quand les variables X et Y dans le Modèle (1.17) ont un support \mathbb{R}^+ . Nous supposons tout au long de ce travail que g appartient à $\mathbb{L}^2(\mathbb{R}^+)$ et soit $f \in \mathbb{L}^2(\mathbb{R}^+)$ quand nous étudions l'estimation de f , soit $S_X \in \mathbb{L}^2(\mathbb{R}^+)$ quand nous étudions l'estimation de la survie S_X . Dans les deux cas, nous utilisons une méthode de projection par pénalisation comme décrite dans la Section 1.1.3. C'est pourquoi l'idée est de décomposer f sur une base orthonormale de $\mathbb{L}^2(\mathbb{R}^+)$, appropriée à notre problème, notée $(\varphi_k)_{k \geq 0}$,

$$f(x) = \sum_{k \geq 0} b_k(f) \varphi_k(x), \quad x \in \mathbb{R}^+$$

où $b_k(f)$ représente la k -ième composante de f dans la base orthonormée puis d'estimer les m premiers coefficients $b_0(f), \dots, b_{m-1}(f)$. Ainsi pour tenir compte de la positivité des variables aléatoires, nous introduisons la base de Laguerre définie comme

$$k \in \mathbb{N}, x \geq 0, \quad \varphi_k(x) = \sqrt{2} L_k(2x) e^{-x} \quad \text{avec} \quad L_k(x) = \sum_{j=0}^k \binom{k}{j} \frac{(-x)^j}{j!}$$

qui est bien une base de $\mathbb{L}^2(\mathbb{R}^+)$. Pour des fonctions à support \mathbb{R}^+ , le produit de convolution s'écrit

$$h(x) = \int_0^x f(u) g(x-u) du,$$

et ce qui rend l'utilisation de la base de Laguerre pertinente est la relation suivante

$$\int_0^x \varphi_k(u) \varphi_j(x-u) du = 2^{-1/2} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x))$$

(voir formule 22.13.14 dans Abramowitz and Stegun (1964)). À partir de cette dernière propriété en décomposant les fonctions f , g et h sur la base de Laguerre, nous pouvons définir une transformation linéaire des coefficients de la densité f à partir des coefficients de h . Plus précisément, si nous notons respectivement \vec{h}_m et \vec{f}_m les vecteurs de dimension m avec respectivement les coordonnées $b_k(f)$ et $b_k(h)$, pour $k = 0, 1, \dots, m-1$, nous montrons que

$$\vec{h}_m = \mathbf{G}_m \vec{f}_m$$

avec \mathbf{G}_m une matrice triangulaire inférieure dépendant seulement des coefficients de g dans la base de Laguerre. Donc comme g est connue, \mathbf{G}_m aussi. Ainsi nous pouvons calculer les m premiers coefficients de f dans la base de Laguerre à partir des coefficients de g (qui sont connus) et de ceux de h qui peuvent être estimés à partir des observations $(Z_j)_{j \in \llbracket 1, m \rrbracket}$ car $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$. Nous appliquons le même raisonnement pour l'estimation de la fonction de survie en projetant directement S_X sur la base. Cela nous permet d'obtenir directement la décomposition de S_X sur la base et d'en déduire un estimateur. L'idée clé pour obtenir un estimateur « direct » de la survie est de réécrire S_Z comme la convolution de S_X avec g plus un autre terme, ici connu, ainsi

$$S_Z = S_X \star g + S_Y.$$

À partir de cette dernière égalité, la procédure d'estimation sur la base de Laguerre se déduit. À notre connaissance, c'est une nouvelle stratégie d'estimation de la fonction de survie dans un problème de déconvolution. Enfin pour la sélection modèle, il suffit de minimiser les deux critères pénalisés suivants, respectivement pour la densité et la survie,

$$\begin{aligned} & - \|\hat{f}_m\|^2 + \frac{\kappa_1}{n} \left(2m\varrho^2 \left(\mathbf{G}_m^{-1} \right) \wedge \log(n) (\|h\|_\infty \vee 1) \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \right) \\ & - \|\hat{S}_{X,m}\|^2 + \frac{\kappa_2 \mathbb{E}[Z_1]}{n} \varrho^2 \left(\mathbf{G}_m^{-1} \right) \log n \end{aligned}$$

avec κ_1 et κ_2 deux constantes numériques, \hat{f}_m et $\hat{S}_{X,m}$ respectivement des estimateurs des projections de f et S_X sur la base de Laguerre restreinte à ses m premières fonctions de base et $\varrho^2 \left(\mathbf{G}_m^{-1} \right)$ est égale à la plus grande valeur propre en valeur absolue de ${}^t \mathbf{G}_m^{-1} \mathbf{G}_m^{-1}$ et $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$ est égale au carré de la norme de Frobenius (i.e. la somme des coefficients au carré de la matrice). Les estimateurs construits sont alors adaptatifs et convergent à vitesses respectives de $n^{-\frac{s}{s+2r}}$ et $n^{-\frac{s+1}{s+2r+1}}$ où s code la régularité de f et r la régularité de g . Les espaces de régularité sont particuliers au problème et à la base de Laguerre et sont largement commentés dans le Chapitre 2. De plus, nous montrons que des vitesses de l'ordre $(\log n)^\alpha/n$ avec $\alpha > 0$ peuvent être obtenues pour des familles de densité de type Gamma ou Beta.

Dans un deuxième temps, dans le Chapitre 3, nous proposons d'estimer les fonctionnelles linéaires de la forme $\vartheta(f) = \langle \psi, f \rangle = \mathbb{E}[\psi(X_1)]$ quand seules les $(Z_j)_{j \in \llbracket 1, m \rrbracket}$ sont observées dans le Modèle (1.17) et ψ est connue. Si f et ψ appartiennent à $\mathbb{L}^2(\mathbb{R}^+)$, alors elles sont décomposables sur la base de Laguerre, ce qui donne

$$\vartheta(f) = \langle \psi, f \rangle = \sum_{k=0}^{\infty} b_k(\psi) b_k(f).$$

Comme il est supposé que la fonction ψ est connue, ses coefficients $b_k(\psi)$ sont connus et peuvent être exprimés explicitement comme $b_k(\psi) = \langle \psi, \varphi_k \rangle$. Or dans les paragraphes précédents, nous avons défini un estimateur des coefficients de f dans le modèle de convolution

sur \mathbb{R}^+ . Cependant il est clair qu'un estimateur de $\vartheta(f)$ ne peut être $\hat{\vartheta}(f) = \sum_{k=0}^{\infty} b_k(\psi)\hat{b}_k$, puisque que cette somme n'est pas nécessairement convergente, ni calculable en pratique. Donc il faut tronquer la somme. Nous définissons $\vartheta_m(f)$ par

$$\vartheta_m = \vartheta_m(f) = \sum_{k=0}^{m-1} b_k(\psi)b_k(f),$$

et ψ_m la projection de ψ sur l'espace $\text{Vec}(\varphi_0, \dots, \varphi_{m-1})$. Remarquons alors

$$\langle \psi_m, f_m \rangle = \vartheta_m = \langle \psi, f_m \rangle = \langle \psi_m, f \rangle.$$

ϑ_m est estimée comme

$$\hat{\vartheta}_m = \vartheta(\hat{f}_m) = \langle \psi_m, \hat{f}_m \rangle = \sum_{k=0}^{m-1} b_k(\psi)\hat{b}_k.$$

La sélection de modèle est effectuée par méthode de Goldenshluger et Lepksi (expliquée dans la Section 1.1.3) en minimisant la quantité suivante

$$\hat{m} = \arg \min_{m \in \mathcal{M}} \{A(m) + V(m)\}$$

avec $V(m)$ un terme de l'ordre de la variance de l'estimateur, défini par

$$V(m) = \kappa(\|h\|_{\infty} \vee 1) \max_{1 \leq k \leq m} \frac{\|\mathbf{t}\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \log n}{n},$$

où $\mathbf{t}\vec{\psi}_k = \mathbf{t}(\psi_0, \dots, \psi_{k-1})$ et

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+.$$

Pour une certaine contrainte sur la collection de modèles considérée, l'estimateur adaptatif vérifie bien une inégalité de type oracle.

En plus d'établir une stratégie d'estimation adaptative pour des fonctionnelles linéaires de la densité f dans le modèle de convolution sur \mathbb{R}^+ , nous calculons différentes vitesses de convergence pour l'estimateur en fonction de la régularité des fonctions ψ , f et g . Par exemple, si f est un mélange Gamma sa fonctionnelle linéaire est estimée avec une vitesse quasi paramétrique de l'ordre de $(\log n)^{2r}/n$ où r code la régularité de la densité g du processus additif Y . Des vitesses paramétriques sont obtenues sous certaines conditions de sommabilités des coefficients de ψ et/ou g dans la base de Laguerre. Nous retrouvons entre autre le fait que la transformée de Laplace peut être estimée à vitesse paramétrique dans le modèle (1.17) quand la densité de g est connue.

Enfin, nous appliquons cette méthode au problème d'estimation ponctuelle de densité dans le Modèle (1.17). L'intérêt de ce dernier développement est l'étude du choix de la constante de calibration dans la méthode de [Goldenshluger and Lepski \(2011\)](#) inspirée de [Lacour and Massart \(2015\)](#). Leur méthode est pensée pour l'estimation à noyau de densité. Nous l'appliquons à l'estimation par projection.

Enfin, pour conclure cette partie, nous nous intéressons, dans le Chapitre 4, au problème d'agrégation de densité dans le modèle de convolution sur \mathbb{R}^+ . À notre connaissance il existe peu de travaux traitant d'agrégation dans les problèmes inverses et tout particulièrement les problèmes de déconvolution. Nous pouvons citer [Hebiri et al. \(2014\)](#) qui considèrent le problème d'agrégation d'estimateurs pour les problèmes inverses linéaires.

Partant du fait que la base de Laguerre peut être paramétrée, se pose alors la question du choix du paramètre. Ce problème a été laissé de côté dans le Chapitre 2 en fixant le paramètre. Un intérêt majeur de ce dernier développement est qu'il permet la mise en place d'une procédure d'agrégation linéaire d'estimateurs construits dans différentes bases de Laguerre. Une procédure d'agrégation consiste à créer un dictionnaire composé de nombreux estimateurs que l'on agrège pour construire un estimateur dont le risque quadratique intégré est au moins aussi bon que la meilleure combinaison linéaire du dictionnaire (au sens de la perte quadratique). La mise en place de la procédure d'agrégation repose sur les résultats du Chapitre 3.

Ce dernier chapitre est un travail en cours.

1.2.3 Méthodes de déconvolution à bruit inconnu

Résultats existants

Nous nous replaçons désormais dans le cadre du Modèle (1.8) où les variables aléatoires considérées sont distribuées sur \mathbb{R} . Nous supposons néanmoins que la loi du bruit est inconnue du statisticien. Neumann (1997) est l'un des premiers à étudier ce cas en supposant que nous avons un accès indirect au bruit en observant un échantillon dit préliminaire $(\varepsilon_{-j})_{j \in \llbracket 1, M \rrbracket}$ supposé indépendant de $(Y_j)_{j \in \llbracket 1, n \rrbracket}$. Cet échantillon préliminaire permet d'estimer sa fonction caractéristique

$$\hat{f}_\varepsilon^*(x) = \frac{1}{M} \sum_{j=1}^M e^{ix\varepsilon_{-j}}, \quad x \in \mathbb{R}. \quad (1.18)$$

Cet estimateur ne peut être substitué directement dans l'Équation (1.11). En effet, \hat{f}_ε^* est un excellent estimateur fonctionnel de la fonction caractéristique de ε car il converge avec une vitesse paramétrique de \sqrt{M} . Toujours est-il que nous voulons estimer $1/f_\varepsilon^*$, si nous remplaçons cette dernière quantité par $1/\hat{f}_\varepsilon^*$ alors de trop petites valeurs de \hat{f}_ε^* perturbe grandement l'estimation de f . Une estimation raisonnable de $1/f_\varepsilon^*$ n'est plus possible dès que le dénominateur est plus petit que sa vitesse de convergence. Donc si nous nous contentons de cet estimateur alors l'estimation de f est instable. Nous introduisons alors comme Neumann (1997) une version tronquée de l'estimateur quotient

$$\frac{1}{\tilde{f}_\varepsilon^*(x)} = \frac{\mathbb{1}_{\{|\hat{f}_\varepsilon^*(x)| \geq M^{-1/2}\}}}{\hat{f}_\varepsilon^*(x)}.$$

Un estimateur de la projection de f est alors donné \hat{f}_m défini comme

$$\tilde{f}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} du. \quad (1.19)$$

Étudions à nouveau le risque mais cette fois ci quand le bruit est inconnu, pour en déduire les vitesses de convergence.

Proposition 1.2.2. Pour $f_\varepsilon^* \neq 0$, \tilde{f}_m défini par (1.19) vérifie

$$\mathbb{E}\|f - \tilde{f}_m\|^2 \leq \|f - f_m\|^2 + (4C_1 + 2) \frac{\Delta(m)}{n} + 4C_1 \frac{\Delta_f(m)}{M} \quad (1.20)$$

avec $\Delta_f(m) = 1/(2\pi) \int_{-\pi m}^{\pi m} |f^*(u)/f_\varepsilon^*(u)|^2 du$ et C_1 une constante numérique définie dans le lemme 1.2.3 énoncé ci-dessous.

Remarquons que les deux premiers termes à droite de l'Équation (1.20) sont les mêmes que ceux de l'Équation (1.14), c'est-à-dire le biais au carré et le terme de variance quand la loi du bruit est connue. Le terme supplémentaire est dû à l'estimation de la fonction caractéristique de ε . De plus comme $|f^*(u)| \leq 1$, si $M \geq n$ alors $\Delta_f(m) \leq \Delta(m)$, il n'y a alors pas de perte comparé au modèle avec f_ε connu. Cette démarche est utilisée par [Comte and Lacour \(2011\)](#) qui établissent une procédure d'estimation adaptative du choix de la dimension m pour $M \geq n^{1+\eta}$ avec $\eta > 0$. Dans le paragraphe suivant, nous expliquerons comment dans le Chapitre 5 toute contrainte sur M par rapport à n est supprimée, c'est-à-dire que M est autorisé à être inférieur ou égal à n .

L'ordre du terme additionnel $\Delta_f(m)/M$ est obtenu grâce au lemme clé suivant permettant de contrôler en espérance l'écart entre $1/\tilde{f}_\varepsilon^*$ et $1/f_\varepsilon^*$.

Lemme 1.2.3. ([Neumann, 1997](#), p.310)

$$\mathbb{E} \left| \frac{\mathbb{1} \left\{ |\hat{f}_\varepsilon^*(x)| \geq M^{-1/2} \right\}}{\hat{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^2 \leq C_1 \left(\frac{M^{-1}}{|f_\varepsilon^*(x)|^4} \wedge \frac{1}{|f_\varepsilon^*(x)|^2} \right).$$

Démonstration de la Proposition 1.2.2. Tout d'abord d'après Pythagore :

$$\|f - \tilde{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \tilde{f}_m\|^2.$$

Pour le terme de variance, d'après Plancherel on a

$$\begin{aligned} \|f_m - \tilde{f}_m\|^2 &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u)}{\hat{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^2 du \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u)}{\hat{f}_\varepsilon^*(u)} - \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} + \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^2 du \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \hat{f}_Y^*(u)R(u) + \frac{\hat{f}_Y^*(u) - f_Y^*(u)}{f_\varepsilon^*(u)} \right|^2 du \\ &\leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} |\hat{f}_Y^*(u)R(u)|^2 du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Y^*(u) - f_Y^*(u)}{f_\varepsilon^*(u)} \right|^2 du \end{aligned}$$

avec $R(u) = \frac{1}{\hat{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)}$. D'où

$$\mathbb{E} \|f_m - \tilde{f}_m\|^2 \leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} \mathbb{E} \left[|\hat{f}_Y^*(u)R(u)|^2 \right] du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y^*(u)|^2 \right] du.$$

Or d'après l'Équation (1.10)

$$\mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y^*(u)|^2 \right] \leq \frac{1}{n}$$

et par hypothèse $(\varepsilon_{-j})_{j \in [1, M]}$ est indépendant de $(Y_j)_{j \in [1, n]}$

$$\begin{aligned} \int_{-\pi m}^{\pi m} \mathbb{E} \left[|\hat{f}_Y^*(u)R(u)|^2 \right] du &= \int_{-\pi m}^{\pi m} \mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y^*(u) + f_Y^*(u)|^2 |R(u)|^2 \right] du \\ &\leq 2 \int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 \mathbb{E} \left[|R(u)|^2 \right] du + 2 \int_{-\pi m}^{\pi m} \mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y^*(u)|^2 |R(u)|^2 \right] du \\ &\leq 2 \int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 \mathbb{E} \left[|R(u)|^2 \right] du + \frac{2}{n} \int_{-\pi m}^{\pi m} \mathbb{E} \left[|R(u)|^2 \right] du. \end{aligned}$$

Enfin en appliquant le lemme 1.2.3

$$\begin{aligned}
& \mathbb{E} \|f_m - \tilde{f}_m\|^2 \\
& \leq \frac{2}{\pi} \int_{-\pi m}^{\pi m} |f_Y^*(u)|^2 \mathbb{E} [|R(u)|^2] \, du + \frac{2}{\pi} \frac{1}{n} \int_{-\pi m}^{\pi m} \underbrace{\mathbb{E} [|R(u)|^2]}_{\leq C_1 |f_\varepsilon^*(u)|^{-2}} \, du + \frac{1}{\pi} \frac{1}{n} \int_{-\pi m}^{\pi m} |f_\varepsilon^*(u)|^{-2} \, du \\
& \leq \frac{2}{\pi} \int_{-\pi m}^{\pi m} |f^*(u)|^2 |f_\varepsilon^*(u)|^2 \underbrace{\mathbb{E} [|R(u)|^2]}_{\leq C_1 M^{-1} |f_\varepsilon^*(u)|^{-4}} \, du + \frac{2}{\pi} \frac{1}{n} \int_{-\pi m}^{\pi m} C_1 |f_\varepsilon^*(u)|^{-2} \, du + \frac{2}{n} \Delta(m) \\
& \leq 4C_1 \frac{\Delta_f(m)}{M} + (4C_1 + 2) \frac{\Delta(m)}{n}.
\end{aligned}$$

Donc

$$\mathbb{E} \|f - \tilde{f}_m\|^2 \leq \|f - f_m\|^2 + 4C_1 \frac{\Delta_f(m)}{M} + (4C_1 + 2) \frac{\Delta(m)}{n}.$$

D'où le résultat de la Proposition 1.2.2. \square

Sous les mêmes hypothèses de régularité que dans la section précédente, nous obtenons les vitesses de convergences reportées dans le Tableau 1.2. Aux vitesses du Tableau 1.1 viennent s'ajouter un terme supplémentaire dépendant de la taille de l'échantillon préliminaire du bruit M et de la régularité de f_ε^* . Nous renvoyons aussi à [Comte and Lacour \(2011\)](#) pour le détail des vitesses de convergence dans le cas où les densité f et f_ε sont super régulières.

	$\delta = 0$	$\delta > 0$
$r = 0$	$n^{\frac{-2s}{2\gamma+2s+1}} + M^{-(1 \wedge \frac{s}{\gamma})}$	$(\log n)^{-2s/\delta} + (\log M)^{-2s/\delta}$
$r > 0$	$\frac{(\log n)^{\frac{2\gamma+1}{r}}}{n} + \frac{1}{M}$	

TABLE 1.2 – Vitesse de convergence de l'estimateur \tilde{f}_m pour le risque quadratique intégré.

[Neumann \(1997\)](#), en plus de construire un estimateur, considère le cas où la densité de f et f_ε sont régulières et montre que la borne de risque qu'il obtient est optimale au sens minimax. Cette vitesse correspond à celle de la première case du Tableau 1.2. Ensuite, nous pouvons noter que dans le cas où la densité à estimer est super régulière et le bruit régulier, l'estimation de la transformée de Fourier des erreurs n'altère pas la vitesse de convergence de l'estimateur final.

Bien qu'il existe une importante littérature concernant l'estimation de densité lorsque la loi du bruit est connue dans le modèle de convolution, ce problème à loi inconnue des erreurs a été peu étudié. La sélection de modèle est faite par [Comte and Lacour \(2011\)](#). Nous pouvons citer [Efromovich \(1997\)](#) et [Johannes and Schwarz \(2013a\)](#) pour le cas des données circulaires. [Johannes and Schwarz \(2013a\)](#) considèrent le problème de déconvolution circulaire quand la loi du bruit est inconnue. Ils supposent avoir à disposition un échantillon préliminaire du bruit. Les auteurs construisent un estimateur qui atteint le risque

minimax. Ils s'intéressent au problème d'adaptation en mettant en place une procédure en deux temps. D'abord ils établissent une procédure en supposant la loi des erreurs connue puis ils remplacent les quantités associées par leurs équivalents empiriques. [Johannes \(2009\)](#) construit un estimateur de f et ses dérivées en appliquant un cut off spectral dans le domaine Fourier. Il détermine les vitesses de convergence associées à cet estimateur pour f_ε connue et inconnue en supposant que f satisfait diverses hypothèses de régularité. L'estimateur proposé est optimal au sens minimax pour f_ε connue et inconnue si f appartient à un espace de Sobolev et f_ε est ordinairement ou super régulière. [Meister \(2004\)](#) étudie le problème d'une mauvaise spécification de la loi des erreurs. En effet, des hypothèses de type paramétrique sont souvent formulées : par exemple en imposant des conditions sur le comportement asymptotique de la transformée de Fourier de f . L'auteur montre que le MISE converge vers une certaine fonctionnelle qui dans certains cas peut être infinie. Enfin nous pouvons encore citer [Diggle and Hall \(1993\)](#) qui étudient le cas $M \geq n$.

Nous pouvons encore citer [Dattner et al. \(2016\)](#) qui étudient le problème de l'estimation de quantile, via des méthodes de Lepski, dans le cas du modèle de convolution à bruit connu et inconnu. Leur méthode est basée sur la déconvolution de densité. L'estimateur obtenu est minimax sous certaines hypothèses minimales. Les auteurs obtiennent aussi au passage la vitesse optimale de convergence de la fonction de répartition.

Une alternative à cette approche, c'est-à-dire l'accès à un échantillon préliminaire, est d'étudier des données répétées comme suit

$$Y_{j,k} = X_j + \varepsilon_{j,k}, \quad j = 1, \dots, n \quad \text{and} \quad k = 2. \quad (1.21)$$

Par différence nous obtenons, les égalités suivantes

$$\forall u \in \mathbb{R}, \quad \mathbb{E} \left[e^{iu(Y_{1,1} - Y_{1,2})} \right] = \mathbb{E} \left[e^{iu(\varepsilon_{1,1} - \varepsilon_{1,2})} \right] = \left| \mathbb{E} \left[e^{iu\varepsilon_{11}} \right] \right|^2 = |f_\varepsilon^*(u)|^2.$$

Cette dernière relation permet de construire un estimateur de $|f_\varepsilon^*(u)|^2$. Des conditions de symétrie de la densité f_ε^* peuvent être imposées pour supprimer le module. Ainsi dans le cas de données répétées, nous pouvons citer : avec loi symétrique des erreurs [Delaigle et al. \(2008\)](#) et [Comte et al. \(2013\)](#) ; avec une loi asymétrique [Li and Vuong \(1998\)](#) et [Comte and Kappus \(2015\)](#).

[Neumann \(2007\)](#) établit une nouvelle méthode d'estimation dans le modèle de convolution à partir de données répétées quand la loi des erreurs est inconnue. Il prouve la consistance de l'estimation sous des hypothèses minimales sur l'ensemble des zéros des transformées de Fourier. [Delaigle et al. \(2008\)](#) justifient l'utilisation de données répétées par l'augmentation de ce type de mesure. Les auteurs développent des méthodes non-paramétriques pour le problème de déconvolution de densité et de régression avec erreur de mesure sur les variables. Ils montrent que même si le nombre de données répétées est petit, l'estimateur a les mêmes performances que si la loi de f_ε était connue. [Meister and Neumann \(2010\)](#) proposent un estimateur de la densité basé sur des données répétées avec des erreurs indépendantes. Ils s'intéressent en particulier au cas où f_ε^* s'annule. Les résultats obtenus sont similaires à ceux de [Meister \(2004\)](#). Les auteurs montrent aussi le caractère minimax pour l'estimateur de f.d.r. et du support de f . [Comte and Kappus \(2015\)](#) étudient le problème de déconvolution avec des données répétées dans le cas où f_ε n'est pas symétrique. Les auteurs montrent que leur méthodologie conduit à de meilleures vitesses de convergence que celles qui existent dans la littérature.

Enfin plusieurs auteurs ont mis en place une approche semi-paramétrique. Nous pouvons citer pour le cas d'une erreur Gaussienne avec une variance inconnue : [Meister \(2006\)](#) et

Matias (2002). Dans les travaux de Butucea and Matias (2005), Butucea et al. (2008) et Meister (2006) le même modèle est considéré sous l'hypothèse que $f^* > c > 0$ ainsi f est identifiable.

Principaux résultats des Chapitres 5 à 7

La contribution de cette thèse aux problèmes de déconvolution avec bruit inconnu est l'estimation adaptative de densité. Pour le rendre le problème identifiable nous avons besoin soit d'observations préliminaires de ε (Modèle (1.8)), soit d'observations répétées (Modèle (1.21)). Nous considérons les deux modèles en parallèle. Ces résultats sont développés dans le Chapitre 5. Cette méthodologie se révélant riche est aussi appliquée à d'autres modèles dans les Chapitres 6 et 7.

Cette partie s'appuie sur les résultats de Kappus (2014) développés pour l'étude des processus de Lévy lorsque l'échantillon préliminaire du bruit est de même taille que l'échantillon des observations $(Y_j)_{j \in [1, n]}$ dans le cas du Modèle (1.8). En effet, l'estimation des processus de Lévy est intrinsèquement liée au modèle de convolution avec bruit inconnu. Cela est dû à l'utilisation à la formule de Lévy-Khintchine qui amène à l'estimation d'un quotient similaire à celui de l'Équation (1.7). Ainsi le but du Chapitre 5 est de généraliser ces méthodes au modèle de convolution avec bruit inconnu à des tailles quelconques d'échantillons. Pour cela, nous proposons une procédure adaptative avec une pénalité qui prend en compte l'aléa sur les deux échantillons comme mis en évidence par l'Équation (1.20). Ce travail améliore donc les résultats de Comte and Lacour (2011) en élargissant le cadre de l'étude théorique et en supprimant leur hypothèse semi-paramétrique sur la régularité de la densité du bruit f_ε .

Comme dans le paragraphe précédent, les résultats reposent sur la modification de l'estimateur \hat{f}_ε^* . Notre particularité est de modifier le seuil en introduisant une certaine fonction de poids w comme suit.

Définition 1.2.4. Pour $\kappa_p > 0$, on pose :

$$\check{f}_\varepsilon^*(x) = \begin{cases} \hat{f}_\varepsilon^*(x) & \text{si } |\hat{f}_\varepsilon^*(x)| \geq k_M(u) := \kappa_p (\log M)^{1/2} w(x)^{-1} M^{-1/2}, \\ \kappa_p (\log M)^{1/2} w(x)^{-1} M^{-1/2} & \text{sinon,} \end{cases} \quad (1.22)$$

avec $w(x) = (\log(e + |x|))^{-\frac{1}{2}-\delta}$, $\delta > 0$ et \hat{f}_ε^* défini par l'Équation (1.18).

La fonction de poids w est introduite à l'origine par Neumann and Reiß (2009) et les considérations présentées dans ce travail combinées avec les idées données dans Kappus (2014), développées pour le cadre particulier des processus de Lévy, jouent un rôle majeur dans nos arguments. En effet, comme le terme de pénalité dans la procédure de sélection de modèle fait intervenir l'estimateur de la fonction caractéristique au dénominateur, l'inégalité oracle dépend alors d'un contrôle précis de la déviation \hat{f}_ε^* autour de f_ε^* sur \mathbb{R} . Ainsi dans Neumann and Reiß (2009), il est montré que la distance entre ces deux objets pondérée par w est petite sur \mathbb{R} . Enfin il y aura une perte d'ordre logarithmique en comparaison au terme de variance de l'Équation (1.20) dû à l'utilisation d'inégalité de déviation de type Bernstein. Ainsi l'estimateur du ratio introduit par l'Équation (1.22) satisfait le lemme suivant.

Lemme 1.2.5. Soit un entier $q \geq 1$, il existe une constante C_q telle que

$$\mathbb{E} \left| \frac{1}{\check{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^{2q} \leq C_q \left(\frac{\log M w(x)^{-2} M^{-q}}{|f_\varepsilon^*(x)|^{4q}} \wedge \frac{1}{|f_\varepsilon^*(x)|^{2q}} \right)$$

Ce lemme est similaire au Lemme 1.2.3 mais diffère par l'introduction d'un nouveau seuil. Ainsi nous obtenons une procédure adaptative pour l'estimation de la densité f dans les cas où nous disposons d'un échantillon préliminaire du bruit ou de données répétées. Nous notons que les pénalités associées au problème sont aléatoires car il faut estimer les termes de variance de l'Équation(1.20) comme le montrent les deux pénalités suivantes, associées respectivement au modèle avec un échantillon préliminaire et au modèle avec des données répétées (NS= noise sample, RD= repeated data)

$$\begin{aligned}\widehat{\text{pen}}_{\text{NS}}(m) &:= \widehat{\text{pen}}_{1,\text{NS}}(m) + \widehat{\text{pen}}_{2,\text{NS}}(m), \\ &:= \frac{96}{3\pi} \int_{-\pi m}^{\pi m} \frac{k_n^2(u)}{|\tilde{f}_{\varepsilon,\text{NS}}^*(u)|^2} du + \frac{12}{3\pi} \int_{-\pi m}^{\pi m} \frac{k_M^2(u) |\hat{f}_{Y,\text{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\text{NS}}^*(u)|^4} du,\end{aligned}$$

et

$$\begin{aligned}\widehat{\text{pen}}_{\text{RD}}(m) &:= \widehat{\text{pen}}_{1,\text{RD}}(m) + \widehat{\text{pen}}_{2,\text{RD}}(m), \\ &:= \frac{96}{3\pi} \int_{-\pi m}^{\pi m} \frac{k_n^2(u)}{|\tilde{f}_{\varepsilon,\text{RD}}^*(u)|^2} du + \frac{12}{9\pi} \int_{-\pi m}^{\pi m} \frac{k_n^2(u) |\hat{f}_{Y,\text{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\text{RD}}^*(u)|^6} du.\end{aligned}$$

De plus, la collection de modèle dépend d'une contrainte sur la pénalité ce qui implique que cette dernière est aussi aléatoire. Pour ce choix de pénalité nous obtenons des inégalités oracles qui montrent que le compromis biais variance est réalisé automatiquement. Tous ces résultats sont développés dans le Chapitre 5.

Dans un deuxième temps, nous étendons les résultats du Chapitre 5 au cas des modèles mixtes, dans le Chapitre 6, décrits comme suit

$$Y_{k,j} = \alpha_k + \beta_k t_j + \varepsilon_{k,j}, \quad k = 1, \dots, N \quad \text{and} \quad j = 1, \dots, J.$$

Ces modèles sont proposés en pharmacologie pour tenir compte des variations des quantités mesurées non seulement temporellement mais aussi en fonction des individus. Les coefficients de ces modèles sont donc représentés comme des réalisations individuelles de variables aléatoires. Les modèles mixtes correspondent à des données longitudinales pouvant se ramener à des données répétées.

Le but est d'estimer la densité des $(\alpha_k)_{k \in \llbracket 1, N \rrbracket}$ notée f_α et la densité des $(\beta_k)_{k \in \llbracket 1, N \rrbracket}$ notée f_β lorsque la loi du bruit est inconnue. Nous adoptons alors deux stratégies adaptées respectivement à l'estimation de f_α et f_β . Les deux stratégies reposent sur une manière différente de construire un estimateur de la fonction caractéristique de ε . Les deux approches reposent sur le point de vue des données répétées évoquées plus tôt dans cette section.

Pour f_α remarquons que si $t_1 = 0$ nous retrouvons le modèle classique de convolution, la densité du bruit est alors estimée en remarquant que

$$U_k = Y_{k,4} - Y_{k,3} - (Y_{k,2} - Y_{k,1}) = \varepsilon_{k,4} - \varepsilon_{k,3} - \varepsilon_{k,2} + \varepsilon_{k,1},$$

sa transformée de Fourier est alors sous hypothèse de symétrie de ε

$$f_U^*(x) = \mathbb{E} \left[e^{ixU} \right] = |f_\varepsilon^*(x)|^4 = (f_\varepsilon^*(x))^4.$$

Nous pouvons donc estimer cette dernière quantité, en prendre la puissance 1/4 pour avoir un estimateur de f_ε^* et développer la même méthodologie que dans le Chapitre 5 en appliquant le Lemme 1.2.5. Nous en déduisons alors une procédure adaptative.

Pour f_β sans perte de généralités nous supposons que J est pair. Pour $1 \leq j \leq J/2$, nous transformons les données de la manière suivante

$$Z_{k,j} = \frac{Y_{k,2j} - Y_{k,2j-1}}{\Delta} = \beta_k + \frac{\varepsilon_{k,2j} - \varepsilon_{k,2j-1}}{\Delta} = \beta_k + \frac{\eta_{k,j}}{\Delta}, \quad \eta_{k,j} = \varepsilon_{k,2j} - \varepsilon_{k,2j-1}.$$

De plus, nous supposons que les $(\beta_k)_{k \in \llbracket 1, N \rrbracket}$ sont indépendants des $(\varepsilon_{k,j})$ donc des $(\eta_{k,j})_{k \in \llbracket 1, N \rrbracket}$, nous pouvons alors écrire $f_{Z_1} = f_\beta \star f_{\eta_{k,j}/\Delta}$ ce qui implique $f_{Z_j}^*(x) = f_\beta^*(x) |f_\varepsilon^*(\frac{x}{\Delta})|^2$. Contrairement à l'estimation de f_α , nous n'avons pas besoin d'estimer la transformée de Fourier de f_ε^* mais seulement $|f_\varepsilon^*|^2$. Nous pouvons donc construire un estimateur de $|f_\varepsilon^*|^2$ en remarquant l'égalité suivante

$$\frac{U_k}{\Delta} = Z_{k,2} - Z_{k,1} = \frac{1}{\Delta} (\varepsilon_{k,4} - \varepsilon_{k,3} - \varepsilon_{k,2} + \varepsilon_{k,1})$$

ce qui implique

$$f_{\frac{U}{\Delta}}^*(x) = \mathbb{E} \left[e^{ixU/\Delta} \right] = |f_\varepsilon^*(\frac{x}{\Delta})|^4.$$

Enfin en établissant un lemme similaire au Lemme 1.2.5 pour $|f_\varepsilon^*|^2$ et son estimateur, nous appliquons la même stratégie adaptative que dans le Chapitre 5. L'étude complète est menée dans le Chapitre 6.

Enfin, cette méthode est encore appliquée dans le Chapitre 7 pour l'estimation de la densité de somme de variables aléatoires observées avec du bruit. Le but est d'estimer la ℓ -ime convolée de f quand nous observons le Modèle (1.8). Nous supposons, dans un premier temps, que la densité du bruit est connu et définissons des estimateurs non adaptatifs pour lesquels nous obtenons des majorations du MISE. De plus, nous montrons que sous certaines conditions de régularité de f et f_ε , la vitesse paramétrique $1/n$ peut être atteinte par notre estimateur. Dans un deuxième temps, nous construisons des estimateurs adaptatifs inspirés du Chapitre 5 aux performances proches des estimateurs non adaptatifs à une perte logarithmique près. Enfin, ces résultats sont étendus au cas où la loi du bruit n'est pas connue.

1.3 Perspectives

Le but de cette thèse est de proposer des méthodes d'estimation non-paramétrique adaptative dans les modèles bruités avec un accent particulier pour l'étude du modèle de convolution. À peu près toute expérience peut être réalisée avec une erreur de mesure ce qui entrerait dans notre champ de recherche.

Dans la première partie de cette thèse, nous nous sommes attachés à développer des procédures adaptatives par pénalisation ou par méthode de Goldenshluger and Lepski (2011) dans le modèle de convolution avec des v.a. positives et en supposant l'erreur additive connue. Cette dernière hypothèse peut se montrer restrictive dans certains contextes. Ainsi il faudrait établir une procédure d'estimation adaptative ne reposant pas sur la connaissance de la loi des variables $(Y_j)_{j \in \llbracket 1, n \rrbracket}$. Pour cela nous pouvons supposer, comme décrit en Section 1.2.3, avoir accès à un échantillon préliminaire du bruit (Y_{-M}, \dots, Y_{-1}) , indépendant de $(X_j)_{j \in \llbracket 1, n \rrbracket}$ et $(Y_j)_{j \in \llbracket 1, n \rrbracket}$. Grâce à cet échantillon, nous pouvons aisément construire un estimateur $\hat{\mathbf{G}}_m$ de la matrice \mathbf{G}_m intervenant dans l'estimation des coefficients de f dans la base de Laguerre. Plus précisément pour estimer les coefficients de f , dans le cas où la loi

du bruit est connue, il faut inverser la matrice \mathbf{G}_m . Il nous faut donc définir un estimateur de cette inverse comme suit par exemple

$$\tilde{\mathbf{G}}_m^{-1} = \begin{cases} \widehat{\mathbf{G}}_m^{-1} & \text{si } \min \text{sp}(\widehat{\mathbf{G}}_m) \geq M^{-1/2} \\ 0 & \text{sinon,} \end{cases}$$

où $\text{sp}(\widehat{\mathbf{G}}_m)$ représente l'ensemble des valeurs propres en valeur absolue de la matrice $\widehat{\mathbf{G}}_m$. Le résultat clé de ce travail serait alors de trouver un contrôle adéquat de $\mathbb{E}[\rho^2(\mathbf{G}_m^{-1} - \tilde{\mathbf{G}}_m^{-1})]$ similaire au Lemme 1.2.3 où $\tilde{\mathbf{G}}_m^{-1}$ définit un estimateur pertinent de \mathbf{G}_m^{-1} .

Une seconde piste serait d'envisager des contextes dits mélangeants, comme souligné en Section 1.2.1, des contextes où les v.a. $(X_j)_{j \in \llbracket 1, n \rrbracket}$ ne sont pas indépendantes peuvent être étudiées dans le contexte du modèle de convolution sur \mathbb{R}^+ . Nous pourrions aussi envisager un modèle avec erreur sur les variables

$$\begin{cases} Y_j &= \mu(X_j) + \epsilon_j \\ Z_j &= Y_j + \eta_j, \end{cases}$$

où nous observerions les couples $(X_j, Z_j)_{j \in \llbracket 1, n \rrbracket}$.

Enfin dans le Chapitre 4 de la première partie, nous tentons d'établir une procédure d'agrégation linéaire pour adapter le paramètre d'échelle de la base de Laguerre aux données. Pour compléter ce travail, il serait intéressant d'envisager, si possible, une procédure de [Goldenshluger and Lepski \(2011\)](#) pour choisir simultanément le paramètre d'échelle de la base de Laguerre et la dimension de l'espace d'approximation de la densité f associé. La difficulté du problème réside dans la dépendance entre le paramètre d'échelle et les termes de biais et de variance. En effet, il n'est pas clair par exemple que le biais diminue quand le paramètre d'échelle augmente ou l'inverse. Cependant, [Rebelles \(2015\)](#) a réussi à établir une règle de sélection de paramètres sans connaître leurs influences sur le biais et la variance.

Une deuxième perspective en rapport avec les modèles de survie serait de mélanger des problèmes de mesure de type censure avec des erreurs de mesure additive. Ainsi nous considérerions le modèle suivant :

$$Y_j = X_j \wedge C_j + \varepsilon_j, \quad j = 1, \dots, n,$$

avec $(X_j)_{j \in \llbracket 1, n \rrbracket}$, $(\varepsilon_j)_{j \in \llbracket 1, n \rrbracket}$ et $(C_j)_{j \in \llbracket 1, n \rrbracket}$ i.i.d et mutuellement indépendantes. Les variables C_j représentant la censure. D'abord nous étudierons le problème à loi du bruit connue en supposant que les v.a. sont distribuées sur \mathbb{R} . Le but serait d'estimer la fonction de hasard de X définie par $h_X = f/(1 - F)$ sur un certain compact. Ce modèle a été notamment envisagé par [Comte et al. \(2015b\)](#) qui proposent une stratégie quotient pour estimer la fonction de hasard h_X , c'est-à-dire que l'estimateur est obtenu en faisant le rapport entre un estimateur de la densité au numérateur et un estimateur de la fonction de survie au dénominateur. Néanmoins, comme expliqué précédemment, il est naturel dans le modèle de convolution d'utiliser des méthodes de Fourier. Ce type de méthodes requièrent que les transformées de Fourier des fonctions de la base choisie aient un support compact, tandis que la fonction de hasard ne peut être estimée que sur un compact car elle n'est pas de carré de intégrable. La difficulté du problème réside donc dans cette dualité car ces deux cadres semblent opposés. Il s'agit d'un travail en cours visant à améliorer les résultats de [Comte et al. \(2015b\)](#) non garanties par une inégalité oracle.

Récemment [Lepski and Willer \(2015\)](#) et [Duval \(2015\)](#) ont considéré un modèle mêlant les observations directes et le modèle de convolution. Dans le cas du problème de densité le

modèle s'écrit comme


$$h = (1 - \alpha)f + \alpha f \star g.$$

Duval (2015) suppose que $\alpha \in [0, 1/2)$ et Lepski and Willer (2015) que $\alpha \in [0, 1]$. Dans ce modèle, le signal est observé soit directement, il n'y a pas de bruit, soit avec une erreur additive. Cependant, nous ignorons si à quelle catégorie appartiennent les observations. Lepski and Willer (2015) établissent les bornes inférieures asymptotiques pour le risque minimax pour un problème de convolution modérément mal posé, c'est-à-dire pour une certaine régularité de la fonction g avec une discontinuité en α . Ils ne proposent pas d'estimateur de la densité f dans ce modèle, ainsi ils ne calculent pas la borne supérieure du risque. Duval (2015) considère une méthode de point fixe pour estimer la densité f ne reposant pas sur des méthodes d'analyse de Fourier. L'estimateur proposé est consistant et en ajoutant quelques restrictions sur α cet estimateur atteint la borne inférieure démontrée par Lepski and Willer (2015).

Ce modèle est associé aux observations suivantes :

$$Z_j = X_j + \epsilon_j Y_j, \quad j = 1, \dots, n,$$

avec $(X_j)_{j \in \llbracket 1, n \rrbracket}$, $(Y_j)_{j \in \llbracket 1, n \rrbracket}$ et $(\epsilon_j)_{j \in \llbracket 1, n \rrbracket}$ i.i.d et mutuellement indépendantes. Les variables $(\epsilon_j)_{j \in \llbracket 1, n \rrbracket}$ sont distribuées selon une loi de Bernoulli de paramètre α connu, c'est-à-dire $\mathbb{P}[\epsilon_1 = 1] = \alpha$. Si les ϵ_j sont observées alors f peut être estimée avec les $(1 - \alpha)n Z_j$ où il n'y a pas de bruit. L'intérêt de ce modèle est son étude quand les ϵ_j ne sont pas observées. Les deux articles cités précédemment supposent que les variables sont distribuées sur \mathbb{R}^d , ainsi nous pourrions étudier ce modèle quand les variables aléatoires sont positives.



Estimation adaptative dans le modèle de
convolution sur \mathbb{R}^+

2

Adaptive deconvolution on the nonnegative real line

Ce chapitre est une version modifiée de l'article *Adaptive deconvolution on the nonnegative real line* soumis pour publication.

RÉSUMÉ. Dans ce chapitre, nous étudions le problème d'estimation adaptative de densité ou de fonction de survie dans le modèle additif défini par $Z = X + Y$ avec X indépendant de Y quand les deux variables aléatoires sont positives. Notre but est d'estimer la distribution de X (densité ou fonction de survie) à partir de n observations indépendantes de Z en supposant que la distribution de Y est connue. Ce problème peut être vu comme le problème statistique de déconvolution qui a été étudiés dans de nombreux cas avec des approches de type Fourier. Cependant dans le cas présent les variables aléatoires ont la particularité d'être distribuées sur \mathbb{R}^+ . Sachant cela nous proposons un nouvel angle d'attaque en construisant des estimateurs par projection sur la base de Laguerre. Nous présentons des majorations du risque quadratique intégré pour nos estimateurs de la densité et de la fonction de survie. Ensuite nous établissons une procédure de sélection de modèles par pénalisation pour un choix adéquat de la dimension de l'espace de projection. Les procédures sont illustrées sur des données simulées et comparées aux performances d'estimateurs classiques en déconvolution de type Fourier. Notre procédure atteint des vitesses de convergence plus rapides que celles des méthodes Fourier pour l'estimation de densité à support \mathbb{R}^+ .

ABSTRACT. In this chapter we consider the problem of adaptive density or survival function estimation in an additive model defined by $Z = X + Y$ with X independent of Y , when both random variables are nonnegative. Our goal is to recover the distribution of X (density or survival function) through n independent observations of Z , assuming that the distribution of Y is known. This issue can be seen as the classical statistical problem of deconvolution which has been tackled in many cases using Fourier-type approaches. Nonetheless, in the present case the random variables have the particularity to be \mathbb{R}^+ -supported. Knowing that, we propose a new angle of attack by building a projection estimator with an appropriate Laguerre basis. We present upper bounds on the mean squared integrated risk of our density and survival function estimators. We then describe a nonparametric data driven strategy for selecting a relevant projection space. The procedures are illustrated with simulated data and compared to the performances of more classical deconvolution setting using a Fourier approach. Our procedure achieves faster convergence rates than Fourier methods for estimating \mathbb{R}^+ -supported functions.

Contents

2.1	Introduction	35
2.1.1	Model	35
2.1.2	Bibliographical context	35
2.1.3	General strategy	36
2.1.4	Plan of the paper	37
2.2	Statistical model and estimation procedure	38
2.2.1	Notations	38
2.2.2	Laguerre basis	38
2.2.3	Projection estimator of the density function	38
2.2.4	Projection estimator of the survival function	39
2.3	Bounds on the \mathbb{L}^2 risk	40
2.3.1	Upper bounds	40
2.3.2	Rates of convergence	41
2.3.3	In what context does the Laguerre procedure improve the Fourier approach ?	43
2.4	Model selection	45
2.4.1	Adaptive density estimation	45
2.4.2	Adaptive survival function estimation	46
2.5	Illustrations	47
2.5.1	Simulation setting	47
2.5.2	Practical estimation procedure	48
2.5.3	Simulation results	49
2.6	Concluding remarks	50
2.7	Proofs	52
2.7.1	Proof of Section 2.3	52
2.7.2	Proof of Theorem 2.4.1	54
2.7.3	Proof of Theorem 2.4.2 and Corollary 2.4.3	60
2.8	Appendix	65

2.1 Introduction

2.1.1 Model

It occurs frequently that a variable of interest X is not directly observed ; instead we have at hand observations of Z , equal to the sum of X and another random variable Y . In many contexts, Y may be a measurement error, and as such, symmetric or centered. But we can also, in reliability fields, observe the sum of the lifetimes of two components, the second one being well known. In survival analysis, X can be the time of infection of a disease and Y the incubation time, and this happens in the so called back calculation problems in AIDS research. In this last two cases, distributions of X and Y are \mathbb{R}^+ -supported.

More formally, we consider the following model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \quad (2.1)$$

where the X_i 's are independent identically distributed (i.i.d.) nonnegative variables with unknown density f and unknown survival function S_X where $S_X(x) = \mathbb{P}[X > x]$. The Y_i 's are also i.i.d. nonnegative variables with known density g and survival function S_Y . We denote by h the density of the Z_i 's and S_Z its survival function. Moreover the X_i 's and the Y_i 's are assumed to be independent. Our target is the estimation of the density f along with the survival function S_X of the X_i 's when the Z_i 's are observed. We want to emphasize that the estimation of the survival function S_X does not rely on the estimation of the density. A particular strategy is established. Thus we are going to show that the assumption of nonnegativity of the random variables is of huge importance for the estimation strategy.

2.1.2 Bibliographical context

The assumptions imply that, in Model (2.1), $h(x) = (f \star g)(x)$ where $(\varphi \star \psi)(x) = \int \varphi(x - u)\psi(u) du$ denotes the convolution product. This setting matches the setting of convolution models which is classical in nonparametric statistics. [Groeneboom and Wellner \(1992\)](#) have introduced the problem of one-sided error in the convolution model under monotonicity of the cumulative distribution function (c.d.f.). They derive nonparametric maximum likelihood estimators (NPMLE) of the c.d.f. Some particular cases have been tackled as uniform or exponential deconvolution by [Groeneboom and Jongbloed \(2003\)](#) and [Jongbloed \(1998\)](#) who propose NPMLE of the c.d.f. of the X_i 's, which have explicit expressions. For other cases [van Es et al. \(1998\)](#) circumvent the lack of explicit expression for the NPMLE by proposing an isotonic inverse estimator. More recently, [van Es \(2011\)](#), in the uniform deconvolution problem, proposes a density estimator and an estimator of the c.d.f. using kernel estimators and inversion formula. All this works are focused on deriving the asymptotic normality of monotonic estimators which are not always explicit. Here we adopt a very different point view since we are concerned by adaptive estimators of the survival and density functions. We do not focus on the monotonicity of our estimators but it can be noted that monotonic transformation of estimators defined in [Chernozhukov et al. \(2009\)](#) can be applied. Those techniques do not degrade theoretical results. In this paper, our method subsumes the existing ones and in this way unifies the approach to tackle the problem of nonnegative variables in the convolution model in a global setting.

Model (2.1) is also related to the field of mixture models. We can cite [Roueff and Rydén \(2005\)](#) and [Rebafka and Roueff \(2015\)](#) who want to estimate mixtures of Exponential and Gamma, which are included in the present framework. [Rebafka and Roueff \(2015\)](#) use Legendre polynomials to derive their estimators and obtain procedure relative to a compact interval. We do not have this constraint in our method thanks to the Laguerre basis. Besides

we will show that for the mixture gamma case, we can derive excellent convergence rates for our estimators.

More broadly the problem of nonnegative variables appears in actuarial or insurance models. Recently, in a financial context, some papers as [Jirak et al. \(2014\)](#) or [Reiß and Selk \(2013\)](#) have addressed the problem of one-sided errors. The first authors are interested in the optimal adaptive estimation in nonparametric regression when the errors are not assumed to be centered anymore, and typically with Exponential density.

Concerning the convolution literature, the problem of recovering the signal distribution f when it is observed with an additive noise with known error distribution, has been extensively studied. In this context, the noise is more likely to be centered and thus not one-sided. Rates of convergence and their optimality for kernel estimators have been studied in [Carroll and Hall \(1988\)](#), [Stefanski \(1990\)](#), [Stefanski and Carroll \(1990\)](#), [Fan \(1991\)](#) and [Efromovich \(1997\)](#). For the study of sharp asymptotic optimality, we can cite [Butucea \(2004\)](#), [Butucea and Tsybakov \(2008a,b\)](#). For the most part, the adaptive bandwidth selection in deconvolution models has been addressed with a known error distribution, see for example [Pensky and Vidakovic \(1999\)](#) for wavelet strategy, [Delaigle and Gijbels \(2004\)](#) for bandwidth selection, [Comte et al. \(2006\)](#) for projection strategies with penalization, or [Meister \(2009\)](#) and references therein. Concerning the estimation of the c.d.f. in the convolution model, some papers can be found as [Zhang \(1990\)](#), [Fan \(1991\)](#), [Hall and Lahiri \(2008\)](#), [Dattner et al. \(2011\)](#), [Dattner and Reiser \(2013\)](#), [Dattner et al. \(2016\)](#). They all present pointwise estimation procedures since the distribution function is not square integrable on \mathbb{R} . Note that the assumption is not so strong for survival functions on \mathbb{R}^+ . The last two papers consider the pointwise estimation of the c.d.f. when the error distribution is unknown under the assumption that the tail of the characteristic function of the measurement error distribution has a certain decay: polynomial or exponential. These estimators reach the optimal rates under the condition that the target function belongs to a Sobolev space.

All these works suppose that the variables X_i 's and Y_i 's are \mathbb{R} -supported. Therefore they are still valid when the variables are \mathbb{R}^+ -supported and we will provide some convergence rates of these estimators and compare them with our estimators to show that a more specific solution performs better.

2.1.3 General strategy

Let us describe now our specific method for the estimation of the density and survival functions when the random variables X and Y in Model (2.1) are \mathbb{R}^+ -supported. We assume all along the paper that g belongs to $\mathbb{L}^2(\mathbb{R}^+)$ and either $f \in \mathbb{L}^2(\mathbb{R}^+)$ when the estimation of f is under study, or $S_X \in \mathbb{L}^2(\mathbb{R}^+)$, when we want to recover the survival function. In both cases, we use a penalized projection method (see [Birgé and Massart \(1997\)](#)). The idea is to decompose the density function f on an appropriate orthonormal basis on $\mathbb{L}^2(\mathbb{R}^+)$, $(\varphi_k)_{k \geq 0}$,

$$f(x) = \sum_{k \geq 0} b_k(f) \varphi_k(x)$$

where $b_k(f)$ represents the k -th component of f in the orthonormal basis and to estimate the m first ones $b_0(f), \dots, b_{m-1}(f)$. To deal with the particularity of nonnegative variables we introduce the Laguerre basis defined by

$$k \in \mathbb{N}, x \geq 0, \quad \varphi_k(x) = \sqrt{2} L_k(2x) e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}. \quad (2.2)$$

This basis has already been used to estimate a nonnegative function f in [Comte et al. \(2015a\)](#). These authors consider a regression model defined by $Y_i = f \star g(t_i) + \varepsilon_i$ where Y_i is observed, t_i are deterministic times of observation, ε_i is subgaussian and g is known. We can also cite [Vareschi \(2015\)](#) in a similar context with unknown g . For \mathbb{R}^+ -supported functions, the convolution product writes

$$h(x) = \int_0^x f(u)g(x-u) du \quad (2.3)$$

and what makes the Laguerre basis relevant, in the previous works and in ours, is the relation

$$\int_0^x \varphi_k(u)\varphi_j(x-u) du = 2^{-1/2} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x)). \quad (2.4)$$

(see formula 22.13.14 in [Abramowitz and Stegun \(1964\)](#)). From this property, by decomposing f , g and h on the Laguerre basis, we are able to define a linear transformation of the coefficients of the density function f to obtain those of h . More precisely, if we denote by \vec{h}_m and \vec{f}_m m -dimensional vectors with coordinates $b_k(f)$ and $b_k(h)$, $k = 0, 1, \dots, m-1$ respectively, we prove

$$\vec{h}_m = \mathbf{G}_m \vec{f}_m \quad (2.5)$$

where \mathbf{G}_m is a lower triangular invertible matrix depending on the coefficients of g . As g is known, so is \mathbf{G}_m . Thus we can recover the m first coefficients of f , from those of g which are known and those of h which can be estimated from the Z_i 's since $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$. We then derive the same reasoning for the survival function estimation. Let us point out that we do not integrate the estimator of f to obtain an estimator of S_X . Our idea is to directly project S_X on the Laguerre basis. This enables us to obtain directly the decomposition of S_X on the Laguerre basis and thus its estimator. To our knowledge this a new strategy for the survival function estimation in a deconvolution setting.

2.1.4 Plan of the paper

The estimators are precisely defined and illustrated in [Section 2.2](#).

We develop in [Section 2.3](#) a study of the integrated risk of the estimators of the density and survival function. We discuss the resulting rates of convergence of these two estimators. For that we introduce subspaces of $\mathbb{L}^2(\mathbb{R}^+)$, called Laguerre-Sobolev spaces with index $s > 0$ which are defined in [Bongioanni and Torrea \(2009\)](#). This enables us to determine the order of the squared bias terms. This, together with variance order, provides rates of convergence of the estimators of f belonging to a Laguerre-Sobolev space. We also obtain rates of convergence for estimators of survival function.

In [Section 2.4](#), we establish a data driven choice by penalization of the dimension m in our two models and oracle inequalities. For the estimation of the density and survival functions, the methods rely mostly on the fact that we are able to build nested models since the first $m-1$ coordinates \vec{h}_m and \vec{f}_m are the same as those of \vec{h}_{m-1} and \vec{f}_{m-1} . Finally we illustrate these procedures with some simulations and compare our results to those of [Comte et al. \(2006\)](#) in the case of the density estimation.

To sum up this paper is organized as follows. In [Section 2.2](#), we give the notations, specify the statistical model and estimation procedures for f and S_X . In [Section 2.3](#), we present upper bounds of the \mathbb{L}^2 integrated risk and derive the corresponding rates of convergence. In [Section 2.4](#), we propose a new adaptive procedure by penalization for the density and survival functions. Besides the theoretical properties of the adaptive estimators are studied. In [Section 2.5](#), we lead a study of the adaptive estimators through simulation

experiments. Numerical results are then presented and compared to the performances in a more classical deconvolution setting using a Fourier approach. The results show that our procedure works significantly better than Fourier methods for estimating \mathbb{R}^+ -supported functions. In the concluding Section 2.6 we give further possible developments or extensions of the method. All the proofs are postponed to Section 2.7.

2.2 Statistical model and estimation procedure

2.2.1 Notations

For two real numbers a and b , we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For two functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the \mathbb{L}^2 norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx$. Let d be an integer, for two vectors \vec{u} and \vec{v} belonging to \mathbb{R}^d , we denote $\|\vec{u}\|_{2,d}$ the Euclidean norm defined by $\|\vec{u}\|_{2,d}^2 = {}^t\vec{u}\vec{u}$ where ${}^t\vec{u}$ is the transpose of \vec{u} . The scalar product between \vec{u} and \vec{v} is $\langle \vec{u}, \vec{v} \rangle_{2,d} = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$. We introduce the spectral norm of a matrix \mathbf{A} : $\varrho^2(\mathbf{A}) = \lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ where $\lambda_{\max}(\mathbf{A})$ is the largest eigenvalue of A in absolute value. We denote by $\|\mathbf{A}\|_F = \sum_{i,j} a_{ij}^2$ the Frobenius norm of a matrix \mathbf{A} .

2.2.2 Laguerre basis

The Laguerre polynomials L_k defined by (2.2) are orthonormal with respect to the weight function $x \mapsto e^{-x}$ on \mathbb{R}^+ . In other words, $\int_{\mathbb{R}^+} L_k(x)L_{k'}(x)e^{-x} dx = \delta_{k,k'}$ where $\delta_{k,k'}$ is the Kronecker symbol. Hence $(\varphi_k)_{k \geq 0}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. We remind that the Laguerre basis verifies the following inequality for all integer k

$$\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_{\infty} \leq \sqrt{2}. \quad (2.6)$$

We also introduce the space $\mathcal{S}_m = \text{Span}\{\varphi_0, \dots, \varphi_{m-1}\}$.

2.2.3 Projection estimator of the density function

For a function p in $\mathbb{L}^2(\mathbb{R}^+)$, we denote

$$p(x) = \sum_{k \geq 0} b_k(p)\varphi_k(x) \quad \text{where} \quad b_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u) du.$$

Thus if $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, f and g admit a development on the Laguerre basis. Since X and Y are independent and nonnegative variables, we have a convolution relation between h , f and g . Starting from Equation (2.3) and using the decomposition on the Laguerre basis, the following equality holds

$$h(x) = \sum_{k=0}^{\infty} \sum_{j=0}^{\infty} b_k(f)b_j(g) \int_0^x \varphi_k(u)\varphi_j(x-u) du. \quad (2.7)$$

So we decompose h on the Laguerre basis as $\sum_{k=0}^{\infty} b_k(h)\varphi_k(x)$ and apply Equation (2.4) to (2.7). We get

$$\sum_{k=0}^{\infty} b_k(h)\varphi_k(x) = \sum_{k=0}^{\infty} \varphi_k(x) \left(2^{-1/2}b_k(f)b_0(g) + \sum_{l=0}^{k-1} 2^{-1/2}(b_{k-l}(g) - b_{k-l-1}(g))b_l(f) \right).$$

We finally obtain an infinite triangular system of linear equations. We can write for any m that $\vec{h}_m = \mathbf{G}_m \vec{f}_m$ where \mathbf{G}_m is the lower triangular Toeplitz matrix with elements

$$[\mathbf{G}_m]_{i,j} = \begin{cases} 2^{-1/2} b_0(g) & \text{if } i = j, \\ 2^{-1/2} (b_{i-j}(g) - b_{i-j-1}(g)) & \text{if } j < i, \\ 0 & \text{otherwise,} \end{cases} \quad (2.8)$$

(see Comte et al. (2015a)). And as

$$b_0(g) = \int_{\mathbb{R}^+} g(u) \varphi_0(u) du = \sqrt{2} \int_{\mathbb{R}^+} g(u) e^{-u} du = \sqrt{2} \mathbb{E}[e^{-Y}] > 0.$$

\mathbf{G}_m is invertible. The principle of a projection method for estimation is to reduce the question of estimating f to the one of estimating f_m the projection of f on \mathcal{S}_m . Clearly

$$f_m(x) = \sum_{k=0}^{m-1} b_k(f) \varphi_k(x)$$

and $b_k(f)$ for $k = 0, \dots, m-1$ can be estimated by Equation (2.5) since $\mathbf{G}_m^{-1} \vec{h}_m = \vec{f}_m$. So, as $b_k(h) = \mathbb{E}[\varphi_k(Z_1)]$, the projection of f on \mathcal{S}_m can be estimated by

$$\hat{f}_m(x) = \sum_{k=0}^{m-1} \hat{b}_k(f) \varphi_k(x) \quad \text{with} \quad \hat{f}_m = \mathbf{G}_m^{-1} \hat{h}_m \quad \text{and} \quad \hat{b}_k(h) = \frac{1}{n} \sum_{i=1}^n \varphi_k(Z_i). \quad (2.9)$$

Let us notice that if $Y = 0$ a.s. then $g = \delta_0$, and we have for any integer k $b_k(g) = \varphi_k(0) = \sqrt{2}$. This implies $\mathbf{G}_m = \mathbf{I}_m$, with \mathbf{I}_m the identity matrix. Therefore if there is no additional noise, we are able to estimate f_m directly from the observations. It means that in Equation (2.9) we have $Z_i = X_i$ and

$$\mathbf{G}_m = \mathbf{I}_m \quad \text{and} \quad \hat{f}_m = \hat{h}_m. \quad (2.10)$$

2.2.4 Projection estimator of the survival function

We have to point out that in the case of the projection estimation of the survival function, the estimation of the coefficients in the Laguerre basis is slightly different from the previous section. Let us consider for example $b_k(S_Z)$ the k -th coefficient of S_Z

$$\begin{aligned} b_k(S_Z) &= \int_{\mathbb{R}^+} S_Z(u) \varphi_k(u) du = \int_{\mathbb{R}^+} \varphi_k(u) \left(\int_u^{+\infty} h(v) dv \right) du \\ &= \int_{\mathbb{R}^+} \left(\int_0^v \varphi_k(u) du \right) h(v) dv = \mathbb{E}[\Phi_k(Z_1)] \end{aligned}$$

with Φ_k a primitive of φ_k defined as $\Phi_k(x) = \int_0^x \varphi_k(u) du$. We can notice that

$$\Phi_k(x) = \sqrt{2} \int_0^x \sum_{j=0}^k (-2)^j \binom{k}{j} \frac{u^j}{j!} e^{-u} du = \sqrt{2} \sum_{j=0}^k \frac{(-2)^j}{j!} \binom{k}{j} \gamma(j+1, x). \quad (2.11)$$

where γ is the lower incomplete gamma function defined by formula 6.5.2. in Abramowitz and Stegun (1964). In order to apply the same method as for the density estimation,

let us see how convolution is modified for survival functions. Let $z \geq 0$, by definition $S_Z(z) = \mathbb{P}(Z > z)$, and

$$\begin{aligned} S_Z(z) &= \mathbb{P}(X + Y > z) = \iint \mathbf{1}_{x+y>z} f(x) \mathbf{1}_{x \geq 0} g(y) \mathbf{1}_{y \geq 0} dx dy \\ &= \int \left(\int_{z-y}^{+\infty} f(x) dx \right) g(y) \mathbf{1}_{y \geq 0} \mathbf{1}_{z-y \geq 0} dy + \int \left(\int_0^{+\infty} f(x) dx \right) g(y) \mathbf{1}_{y \geq 0} \mathbf{1}_{z-y \leq 0} dy \\ &= \int_0^z S_X(z-y) g(y) dy + \int_z^{+\infty} g(y) dy \\ &= \int_0^z S_X(z-y) g(y) dy + S_Y(z) = S_X \star g(z) + S_Y(z). \end{aligned}$$

We can notice that we have one more term: the survival function of Y . Nevertheless similarly to the density estimation the coefficients of S_X , $b_k(S_X)$ can also be represented as a solution of an infinite triangular system of linear equations as follows

$$\begin{aligned} S_Z(z) - S_Y(z) &= \sum_{k \geq 0} (b_k(S_Z) - b_k(S_Y)) \varphi_k(z) \\ &= 2^{-1/2} \sum_{k=0}^{\infty} \varphi_k(x) \left(b_k(S_X) b_0(g) + \sum_{l=0}^k (b_{(k-l)}(g) - b_{(k-l-1)}(g)) b_l(S_X) \right). \end{aligned}$$

Now let us define, $S_{X,m}$ the projection of S_X on the space \mathcal{S}_m

$$S_{X,m}(x) = \sum_{k=0}^{m-1} b_k(S_X) \varphi_k(x).$$

Thus, with \mathbf{G}_m defined by Equation (2.8) and Φ_k defined by (2.11), the projection estimator of $S_{X,m}$ on the Laguerre basis is given by

$$\begin{aligned} \hat{S}_{X,m}(x) &= \sum_{k=0}^{m-1} \hat{b}_k(S_X) \varphi_k(x) \\ \text{with } \hat{\vec{S}}_{X,m} &= \mathbf{G}_m^{-1} \left(\hat{\vec{S}}_{Z,m} - \vec{S}_{Y,m} \right) \quad \text{and} \quad \hat{b}_k(S_Z) = \frac{1}{n} \sum_{i=1}^n \Phi_k(Z_i), \end{aligned} \quad (2.12)$$

where $\vec{S}_{Y,m}$ is known since $b_k(S_Y) = \mathbb{E}[\Phi_k(Y_1)]$ and g is known.

Remark 1. It is worth mentioning that here we do not integrate the estimator of the density \hat{f}_m to estimate the survival function.

2.3 Bounds on the \mathbb{L}^2 risk

In this section, we study the integrated risk of our estimators.

2.3.1 Upper bounds

Before stating any results, let us remind that

$$h(x) = \int f(u) g(x-u) \mathbf{1}_{u \geq 0} \mathbf{1}_{x-u \geq 0} du = \int_0^x f(u) g(x-u) du.$$

By Cauchy-Schwarz inequality, we have $\forall x \in \mathbb{R}^+$, $h(x) \leq \|f\| \|g\|$. Thus for f and $g \in \mathbb{L}^2(\mathbb{R}^+)$, it yields that $\|h\|_{\infty} \leq \|f\| \|g\| < \infty$. Moreover, if $\|g\|_{\infty} < \infty$ then we can bound $\|h\|_{\infty}$ by $\|g\|_{\infty}$. Since g is known in our model, it can be interesting to switch the two quantities in a simulation setting for instance.

Proposition 2.3.1. *If f and $g \in \mathbb{L}^2(\mathbb{R}^+)$, for \mathbf{G}_m defined by (2.8) and \hat{f}_m defined by (2.9), the following result holds*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n} \varrho^2(\mathbf{G}_m^{-1}) \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2. \quad (2.13)$$

This result can easily be applied to the estimation of the density when $Y = 0$:

Corollary 2.3.2. *For $f \in \mathbb{L}^2(\mathbb{R}^+)$, in the model without noise defined by (2.10) we get*

$$\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + (2 \wedge \|h\|_\infty) \frac{m}{n}. \quad (2.14)$$

Finally, we derive the following upper bound for the projection estimator of the survival function.

Proposition 2.3.3. *If S_X and $g \in \mathbb{L}^2(\mathbb{R}^+)$ and $\mathbb{E}[Z_1] < \infty$, for \mathbf{G}_m defined by (2.8) and $\hat{S}_{X,m}$ defined by (2.12), the following result holds*

$$\mathbb{E}\|S_X - \hat{S}_{X,m}\|^2 \leq \|S_X - S_{X,m}\|^2 + \frac{\mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1}). \quad (2.15)$$

Lemma 2.3.4. *$m \mapsto \varrho^2(\mathbf{G}_m^{-1})$ is nondecreasing.*

Remark 2. The terms of the right-hand side of Equations (2.13), (2.14) and (2.15) correspond to a squared bias and variance term. Indeed the first one gets smaller when m gets larger and vice versa for the other one thanks to Lemma 2.3.4.

2.3.2 Rates of convergence

In order to derive the corresponding rates of convergence of the estimators \hat{f}_m and $\hat{S}_{X,m}$ respectively defined by (2.9) and (2.12), we need to evaluate the smoothness of the signal along with the order of $\varrho^2(\mathbf{G}_m^{-1})$. In the first place, we assume that f belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{ f : \mathbb{R}^+ \rightarrow \mathbb{R}, f \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s b_k^2(f) \leq L < +\infty \right\} \quad \text{with } s \geq 0 \quad (2.16)$$

where $b_k(f) = \langle f, \varphi_k \rangle$. [Bongioanni and Torrea \(2009\)](#) have introduced Laguerre-Sobolev space but the link with the coefficients of a function on a Laguerre basis was done by [Comte and Genon-Catalot \(2015\)](#). Indeed, let s be an integer, for $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ and $f \in \mathbb{L}^2(\mathbb{R}^+)$, we have that

$$\sum_{k \geq 0} k^s b_k^2(f) < +\infty$$

is equivalent to the fact that f admits derivatives up to order $s - 1$ with $f^{(s-1)}$ absolutely continuous and for $0 \leq k \leq s - 1$, $x^{(k+1)/2} \sum_{j=0}^{k+1} \binom{k+1}{j} f^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$. For more details we refer to section 7 of [Comte and Genon-Catalot \(2015\)](#). Now for $f \in W^s(\mathbb{R}^+, L)$ defined by (2.16),

$$\|f - f_m\|^2 = \sum_{k=m}^{\infty} b_k^2(f) = \sum_{k=m}^{\infty} b_k^2(f) k^s k^{-s} \leq L m^{-s}.$$

Before deriving the order of the spectral norm of \mathbf{G}_m^{-1} , we can already give the rate of convergence in the forward problem.

Proposition 2.3.5. *In the model without noise defined by (2.10), suppose that f belongs to $W^s(\mathbb{R}^+, L)$ defined by (2.16) and let $m_{opt} \propto n^{1/(s+1)}$, then the following holds*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \hat{f}_{m_{opt}}\|^2 \leq C_1(s, L) n^{-s/(s+1)}$$

where $C_1(s, L)$ is a positive constant.

Secondly in the deconvolution problem, we must evaluate the variance term of Equations (2.13) and (2.15) which means assess the order of $\varrho^2(\mathbf{G}_m^{-1})$ and $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$. Let us notice that Equation (2.13) also admits the following upper bound

$$\mathbb{E} \|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{(2 \vee \|h\|_{\infty})}{n} \left(m \varrho^2(\mathbf{G}_m^{-1}) \wedge \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \right). \quad (2.17)$$

Now recall that we have the following equivalence between the spectral and Frobenius norms:

$$\frac{1}{\sqrt{m}} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}} \leq \varrho(\mathbf{G}_m^{-1}) \leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}} \quad (2.18)$$

Now let us assess the order of these norms in function of m . Comte et al. (2015a) show that under the following conditions on the density g , we can recover the order of the Frobenius norm and the spectral norm of \mathbf{G}_m^{-1} . First we define an integer $r \geq 1$ such that

$$\frac{d^j}{dx^j} g(x) \Big|_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \dots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

And we make the two following assumptions:

(C1) $g \in \mathbb{L}^1(\mathbb{R}^+)$ is r times differentiable and $g^{(r)} \in \mathbb{L}^1(\mathbb{R}^+)$.

(C2) The Laplace transform defined by $G(z) = \mathbb{E}[e^{-zY}]$ of g has no zero with non negative real parts except for the zeros of the form $\infty + ib$.

Remark 3. A Gamma distribution of parameter p and θ verifies these **(C1)**-**(C2)** for $r = p$ ($r = 1$ for an Exponential). On the contrary an Inverse Gamma distribution does not satisfy **(C1)** because there exists no r such that the derivative is different from 0 in 0.

Lemma 2.3.6 (Comte et al. (2015a)). *If Assumptions **(C1)**-**(C2)** are true, then there exists a positive constants C_{ϱ} and C'_{ϱ} such that*

$$\begin{aligned} C'_{\varrho} m^{2r} &\leq \varrho^2(\mathbf{G}_m^{-1}) \leq C_{\varrho} m^{2r}, \\ C'_{\varrho} m^{2r} &\leq \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2 \leq C_{\varrho} m^{2r}. \end{aligned}$$

Thus under Assumptions **(C1)**-**(C2)** the spectral and Frobenius norms have the same order. Gathering Equations (2.17) and (2.18) the minimum of the two norms is the Frobenius norm.

Proposition 2.3.7. *Assume that f belongs to $W^s(\mathbb{R}^+, L)$ defined by (2.16), that Assumptions **(C1)**-**(C2)** are fulfilled and let $m_{opt} \propto n^{1/(s+2r)}$, then*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \|f - \hat{f}_{m_{opt}}\|^2 \leq C_2(s, L, C_{\varrho}) n^{-s/(s+2r)}.$$

where $C_2(s, L, C_{\varrho})$ is a positive constant.

Proposition 2.3.8. *Assume that S_X belongs to $W^{s+1}(\mathbb{R}^+, L)$ defined by (2.16), that Assumptions (C1)-(C2) are fulfilled and let $m_{opt} \propto n^{1/(s+2r+1)}$, then*

$$\sup_{S_X \in W^{s+1}(\mathbb{R}^+, L)} \mathbb{E} \|S_X - \widehat{S}_{X,m}\|^2 \leq C_3(s, L, C_\rho) n^{-(s+1)/(s+2r+1)}.$$

with $C_3(s, L, C_\rho)$ is a positive constant.

Remark 4. We clearly see that in Propositions 2.3.5, 2.3.7 and 2.3.8 the value of m that permits to compute the rate of convergence of the estimator depends on the regularity of the function under estimation. So the solution of the best compromise between the squared bias and the variance depends on unknown quantities L and s . That is why we consider the problem of data driven selection of m . Our goal is then to find a procedure that does not require prior information on f nor S_X and whose risk automatically reaches the optimal rate.

2.3.3 In what context does the Laguerre procedure improve the Fourier approach ?

In this section we want to emphasize that for at least certain classes of functions the Laguerre method achieves better rates of convergence than Fourier method which are optimal if f belongs to a Sobolev class. Thus we consider Gamma distributions and mixed Gamma distributions which belong to a Sobolev class and are said to be ordinary smooth. We choose two different distributions for Y : $\Gamma(q, \mu)$ and $\beta(a, b)$ for $a, b \in \mathbb{N}^*$ such that $b > a$. The results are reported in Table 2.1.

f	g	$\Gamma(q, \mu)$	$\beta(a, b) \ b > a \geq 1$
$\Gamma(p, \theta)$	Laguerre	$\frac{(\log n)^{2q}}{n}$	$\frac{(\log n)^{2a}}{n}$
	Fourier	$n^{-(2p-1)/(2q+2p-1)}$	$n^{-(2p-1)/(2a+2p-1)}$
$\mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})$	Laguerre	$\frac{(\log n)^{2q}}{n}$	$\frac{(\log n)^{2a}}{n}$
	Fourier	$n^{-(2p^*-1)/(2q+2p^*-1)}$ $p^* = \min_i \alpha_i$	$n^{-(2p^*-1)/(2a+2p^*-1)}$ $p^* = \min_i \alpha_i$

Table 2.1: Rates of convergence for the MISE

To derive the rates of convergence of Table 2.1 we have to do a bias variance tradeoff in the Laguerre and Fourier settings. First let us compute the squared bias where the signal X is a Gamma distribution $\Gamma(p, \theta)$ with $p \in \mathbb{N}^*$ and $\theta > 0$ in the Laguerre procedure and in the Fourier setting. In the Laguerre procedure we first need to compute the coefficients

in the Laguerre basis of the signal

$$\begin{aligned} b_k(f) &= \int_{\mathbb{R}^+} \theta^p u^{p-1} e^{-\theta u} \sqrt{2} e^{-u} L_k(2u) \, du = \sqrt{2} \theta^p \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}^+} u^{p+j-1} e^{-(1+\theta)u} \, du \\ &= \frac{\sqrt{2} \theta^p}{(1+\theta)^p} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{(1+\theta)^j} \frac{(p+j-1)!}{j!} = \frac{\sqrt{2} \theta^p}{(1+\theta)^p} \frac{d^{p-1}}{dx^{p-1}} \left[x^{p-1} (1-x)^k \right] \Big|_{x=2/(1+\theta)}. \end{aligned} \quad (2.19)$$

It leads to

$$\|f - f_m\|^2 = \sum_{k \geq m} \left(\frac{\sqrt{2} \theta^p}{(1+\theta)^p} \frac{d^{p-1}}{dx^{p-1}} \left[x^{p-1} (1-x)^k \right] \Big|_{x=2/(1+\theta)} \right)^2 \leq C(p, \theta) \left(\frac{\theta-1}{\theta+1} \right)^{2m} m^{2(p-1)}$$

with $0 < C(p, \theta) < \infty$. For $X \sim \Gamma(p, \theta)$ with $\theta > 0$, we can note that the squared bias decays exponentially. It can also be noted that if $\theta = 1$ then the bias is null. The rate of convergence, in this case, reaches the parametric rate in the Laguerre setting. Moreover if $f \in \mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})$ defined by

$$\mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda}) = \left\{ f = \sum_{i=1}^p \lambda_i \gamma_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^p \lambda_i = 1 \quad \text{and} \quad \gamma_i \sim \Gamma(\alpha_i, \theta_i) \right\}, \quad (2.20)$$

the squared bias decays exponentially

$$\|f - f_m\|^2 = \sum_{k \geq m} \left(\sum_{i=1}^p \lambda_i b_k(\gamma_i) \right)^2 \leq \sum_{k \geq m} \sum_{i=1}^p \lambda_i (b_k(\gamma_i))^2 \leq \sum_{i=1}^p \lambda_i C(\alpha_i, \theta_i) \left(\frac{\theta_i - 1}{\theta_i + 1} \right)^{2m} m^{2(\alpha_i - 1)}.$$

In the Fourier procedure, according to [Comte et al. \(2006\)](#), the squared bias is computed as follows

$$\|f - f_m\|^2 = \frac{1}{2\pi} \|f^* - f_m^*\|^2 = \frac{1}{2\pi} \int_{|t| \geq \pi m} |f^*(t)|^2 \, dt = \frac{1}{2\pi} \int_{|t| \geq \pi m} (1 + \theta^2 t^2)^{-p} \, dt \leq C m^{-2p+1}.$$

It yields that if $f \in \mathcal{M}\Gamma(p, \vec{\alpha}, \vec{\theta}, \vec{\lambda})$ the order of the squared bias is $\max_{1 \leq i \leq p} m^{-2p_i+1}$. We notice in this setting, contrary to the Laguerre setting, that the bias is polynomial.

Secondly let us compute the order of the variance. If $g \sim \Gamma(q, \mu)$, for the Laguerre procedure the variance term is $\|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2/n$ and is upper bounded $2C_\rho m^{2q}/n$ and if $g \sim \beta(a, b)$ it is upper bounded by $2C_\rho m^{2a}/n$ for $b > a$, from [Remark 3](#). We find the same orders of variance for the Fourier setting after computing $1/(2\pi n) \int_{-\pi m}^{\pi m} |g^*(t)|^{-2} \, dt$. Thus we can derive the rates of convergence summarized in [Table 2.1](#).

For the Laguerre procedure we see that the rates of convergence in both cases have order $(\log n)^\alpha/n$ with α depending on the model. While for the Fourier procedure we find classical rates of convergence of the deconvolution setting which are slower. For instance in the double Gamma case we have a rate of convergence $n^{-p/(2q+p+1)}$. Thus in the context of nonnegative variables of Gamma type, we recover faster rates of convergence with our Laguerre method than with a Fourier procedure. We can extend those results to the case of Exponential and Gamma mixtures. This context fits fields of survival analysis and duration models.

Remark 5. Lower bounds in deconvolution problems on the real line have been studied in [Fan \(1991\)](#) and [Butucea and Tsybakov \(2008a,b\)](#), yet those results cannot be extended to

the setting of this paper since we do not consider the same spaces of regularity. Otherwise we can cite [Vareschi \(2015\)](#) who proves lower bounds in the context of a Laplace regression model. But this methodology cannot be applied in our context.

Recently, [Belomestny et al. \(2016\)](#) study the problem of the lower bound in our model without noise. In their paper they explain how the difficulty lies in the fact that the density alternative proposal is a density on \mathbb{R}^+ , and particularly that this density is indeed nonnegative. Due to this constraint, they are able to prove if s is an integer that the rate $n^{-s/(s+1)}$ is nearly optimal up to a log factor depending of a positive constant ϵ that can be chosen arbitrarily.

2.4 Model selection

The aim of this section is to provide an integer m that enables us to compute an estimator of the unknown density or survival function with the \mathbb{L}^2 risk as close as possible to the oracle risk $\inf_m \mathbb{E} \|f - \hat{f}_m\|^2$ or $\inf_m \mathbb{E} \|S_X - \hat{S}_{X,m}\|^2$. We follow the model selection paradigm (see [Birgé and Massart \(1997\)](#), [Birgé \(1999\)](#), [Massart \(2003\)](#)) and choose the dimension of projection spaces m as the minimizer of a penalized criterion.

2.4.1 Adaptive density estimation

We add the two following assumptions:

$$(A1) \quad \mathcal{M}_n^{(1)} = \left\{ 1 \leq m \leq d_1, m \varrho^2(\mathbf{G}_m^{-1}) \leq \frac{n}{\log n} \right\}, \text{ where } d_1 < n \text{ may depend on } n.$$

$$(A2) \quad \forall b > 0, \quad \sum_{m \in \mathcal{M}_n^{(1)}} \varrho^2(\mathbf{G}_m^{-1}) e^{-bm} < C(b) < \infty.$$

We define the penalty as

$$\text{pen}_1(m) = \frac{\kappa_1}{n} \left(2m \varrho^2(\mathbf{G}_m^{-1}) \wedge \log(n) (\|h\|_\infty \vee 1) \|\mathbf{G}_m^{-1}\|_F^2 \right)$$

where κ_1 is a numerical constant see our comment below.

Theorem 2.4.1. *If f and $g \in \mathbb{L}^2(\mathbb{R}^+)$, let us suppose that (A1)-(A2) are true. Let $\hat{f}_{\hat{m}_1}$ be defined by (2.9) and*

$$\hat{m}_1 = \arg \min_{m \in \mathcal{M}_n^{(1)}} \left\{ -\|\hat{f}_m\|^2 + \text{pen}_1(m) \right\}$$

with pen_1 defined by (2.4.1), then there exists a positive numerical constant κ_0 such that $\kappa_1 \geq \kappa_0$ then

$$\mathbb{E} \|f - \hat{f}_{\hat{m}_1}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(1)}} \left\{ \|f - f_m\|^2 + \text{pen}_1(m) \right\} + \frac{C}{n}, \quad (2.21)$$

where C depends on $\|f\|$ and $\|g\|$.

The constant κ_0 can be estimated from the proof, but in practice, values obtained from the theory are generally too large. Thus the constant is calibrated by simulations. Once chosen, it remains fixed for all simulation experiments.

The oracle inequality (2.21) establishes a non asymptotic oracle bound. It shows that the squared bias variance tradeoff is automatically made up to a multiplicative constant. We have shown in Section 2.3 that the rates of convergence in deconvolution problems are

intricate and depend on the regularity types of the function f under estimation and the noise density g . As shown in Section 2.3, the best compromise between square bias and variance orders in Equation (2.13) yields to a certain value of m function of n and depending on unknown quantities, and thus cannot be implemented. That is why Equation (2.21) is of high interest: rates of convergence are reached without requiring to be specified in the framework.

Remark 6. Note it is common in the literature to assume that the distributions belong to a certain semi-parametric model which is not the case in this paper. In the deconvolution setting with a Fourier approach, papers as Comte et al. (2006) for instance, assume that the Fourier transform of the target and error densities have a particular decay behavior. Here this is replaced the spectral norm of the matrix \mathbf{G}_m^{-1} . We can notice that if $\varrho^2(\mathbf{G}_m^{-1}) = O(m^\alpha)$ then for any α , Assumption (A2) is true. It is satisfied for instance for Gamma distributions. Similarly if $\varrho^2(\mathbf{G}_m^{-1}) = O(m^\alpha e^{\lambda m^\beta})$ then it is enough that $\beta < 1$ to ensure (A2). If Assumption (A2) is released then an adaptive procedure can still be obtained with the associated penalty $\text{pen}(m) = \kappa \log(n)(2m\varrho^2(\mathbf{G}_m^{-1}) \wedge \|h\|_\infty \|\mathbf{G}_m^{-1}\|_{\text{F}}^2)/n$.

2.4.2 Adaptive survival function estimation

In this particular framework, we make the two following assumptions:

$$\text{(B1)} \quad \mathcal{M}_n^{(2)} = \left\{ 1 \leq m \leq d_2, \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} \leq C \right\}, \text{ where } d_2 < n \text{ may depend on } n \text{ and } C > 0.$$

$$\text{(B2)} \quad 0 < \mathbb{E}[Z_1^3] < \infty.$$

We define the penalty as

$$\text{pen}_2(m) = \frac{\kappa_2 \mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1}) \log n \quad (2.22)$$

Theorem 2.4.2. *If S_X and $g \in \mathbb{L}^2(\mathbb{R}^+)$, let us suppose that (B1)-(B2) are true. Let $\widehat{S}_{X, \widehat{m}_2}$ be defined by (2.12) and*

$$\widehat{m}_2 = \arg \min_{m \in \mathcal{M}_n^{(2)}} \left\{ -\|\widehat{S}_{X,m}\|^2 + \text{pen}_2(m) \right\}$$

with pen_2 defined by (2.22), then there exists a positive numerical constant κ'_0 such that $\kappa_2 \geq \kappa'_0$ then

$$\mathbb{E} \|S_X - \widehat{S}_{X, \widehat{m}_2}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \left\{ \|S_X - S_{X,m}\|^2 + \text{pen}_2(m) \right\} + \frac{C}{n},$$

where C is a constant depending on $\mathbb{E}[Z_1^3]$.

We can also notice that in the penalty associated with this procedure a logarithmic term appears while it was not in the upper bound of Equation (2.15). Such logarithms often appear in adaptive procedures.

Comments after Theorem 2.4.1 still hold. This oracle inequality shows that the squared bias variance tradeoff is automatically made. Asymptotically, this ensures that the rates of convergence are reached up to a $\log n$ factor. To our knowledge this the first time that a global adaptive procedure of the survival function is considered. This result rests upon the

particularity of the Laguerre basis which enables to extend the adaptive estimation of the density function to the survival function.

Nevertheless this estimation cannot be computed directly since the penalty depends on the expectation of Z . A solution is to prove an oracle inequality for a random penalty associated to (2.22) which is made in the next corollary.

Corollary 2.4.3. *If S_X and $g \in \mathbb{L}^2(\mathbb{R}^+)$, let us suppose that (B1)-(B2) are true. Let $\widehat{S}_{X, \widetilde{m}_2}$ be defined by (2.12) and*

$$\widetilde{m}_2 = \arg \min_{m \in \mathcal{M}_n^{(2)}} \left\{ -\|\widehat{S}_{X,m}\|^2 + \widehat{\text{pen}}_2(m) \right\} \quad (2.23)$$

$$\widehat{\text{pen}}_2(m) = \frac{2\kappa_2 \bar{Z}_n}{n} \varrho^2 \left(\mathbf{G}_m^{-1} \right) \log n \quad \text{where} \quad \bar{Z}_n = \frac{1}{n} \sum_{i=1}^n Z_i, \quad (2.24)$$

then there exists a positive numerical constant κ_2 such that

$$\mathbb{E} \|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \left\{ \|S_X - S_{X,m}\|^2 + \text{pen}_2(m) \right\} + \frac{C}{n}$$

where C is a constant depending on $\mathbb{E}[Z_1]$, $\mathbb{E}[Z_1^3]$ and $\text{Var}[Z_1]$.

2.5 Illustrations

The whole implementation is conducted using R software. The integrated squared errors $\|f - \widehat{f}_{\widehat{m}}\|^2$ and $\|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2$ are computed via a standard approximation and discretization (over 300 points) of the integral on an interval of \mathbb{R} respectively denoted by I_f and I_S . Then the mean integrated squared errors (MISE) $\mathbb{E} \|f - \widehat{f}_{\widehat{m}_1}\|^2$ and $\mathbb{E} \|S_X - \widehat{S}_{X, \widetilde{m}_2}\|^2$ are computed as the empirical mean of the approximated ISE over 200 simulation samples.

2.5.1 Simulation setting

The performance of the procedure is studied for the seven following distributions for X . All the densities are normalized with unit variance except the Pareto distribution which has infinite variance.

- ▷ Exponential $\mathcal{E}(1)$, $I_f = [0, 5]$, $I_S = [0, 10]$.
- ▷ Gamma distribution : $2 \cdot \Gamma(4, \frac{1}{4})$, $I_f = [0, 10]$, $I_S = [0, 5]$.
- ▷ Gamma distribution : $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$, $I_f = [0, 13]$, $I_S = [0, 5]$.
- ▷ Rayleigh distribution with $\sigma^2 = 2/(4 - \pi)$, $f(x) = \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$, $I_f = [0, 5]$, $I_S = [0, 25]$.
- ▷ Weibull, $X/\sqrt{\Gamma(4) - \Gamma(1 + 2/3)^2}$, $f(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k} \mathbf{1}_{x \geq 0}$, with $k = \frac{3}{2}$ and $\lambda = 1$, $I_f = [0, 5]$, $I_S = [0, 5]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I_f = [0, 26]$, $I_S = [0, 15]$.
- ▷ Chi-squared distribution with 10 degrees of freedom, $\chi^2(10)/\sqrt{20}$, $I_f = [0, 10]$, $I_S = [0, 10]$.

- ▷ Pareto distribution with shape parameter $\alpha = 2$ and scale parameter $x_m = 1$, $I_f = [0, 5]$, $I_S = [0, 10]$.

Exponential and Weibull distributions are often used in survival and failure analysis. The Gamma distribution is also often used in insurance modelization. The Rayleigh distribution arises in wind velocity analysis for instance.

In the simulation, the variance σ^2 of the error distribution g takes the values 0, 1/10 and 1/4. The case where the variance σ^2 is null, which corresponds to the case $Y = 0$, is used as a benchmark for the quality of the estimation in the model with noise. We are not aware of any other specific global method of deconvolution on the nonnegative real line. In that case for the density function, we use our procedure with $\mathbf{G}_m = \mathbf{I}_m$. Concerning the survival function, we simply compute the empirical estimator $S_n(x) = n^{-1} \sum_{i=1}^n \mathbb{1}\{X_i > x\}$ (since $Y = 0$) which reaches the parametric rate of convergence.

We then choose a Gamma distribution for the error distribution which verifies **(C1)**-**(C3)** for $r = 2$:

- ▷ Gamma noise: $\Gamma(2, \frac{1}{\sqrt{20}})$ and $\Gamma(2, \frac{1}{\sqrt{8}})$.

Thus the first Gamma distribution has a variance 1/10 and the second 1/4. We refer to Equation (2.19) for the computation of the matrix \mathbf{G}_m .

2.5.2 Practical estimation procedure

The adaptive procedure is then implemented as follows:

- ▷ For $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$, compute $-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)$.
- ▷ Choose \hat{m} such that $\hat{m} = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\}$.
- ▷ And compute $\hat{f}_{\hat{m}}(x) = \sum_{k=0}^{\hat{m}-1} \hat{b}_k(f) \varphi_k(x)$.

The procedure is given for the density estimation. For the survival case the three steps are the same with the right quantities associated to the problem and described in Section 2.4.2. Besides, the penalties are chosen according to Theorem 2.4.1 and Corollary 2.4.3. The constant calibrations were done with intensive preliminary simulations. We take $\kappa_1 = 0.03$ and $\kappa_2 = 0.065$. It can be noted that the values of κ_1 and κ_2 are much smaller than what comes in theory. We consider the two following model collections $\mathcal{M}_n^{(1)} = \{m \in \llbracket 1, n-1 \rrbracket, 1 \leq m \leq \lfloor n^{1/5} \rfloor\}$ and $\mathcal{M}_n^{(2)} = \{m \in \llbracket 1, n-1 \rrbracket, 1 \leq m \leq \lfloor (n \log n)^{1/4} \rfloor\}$ for the density and survival function estimation.

In order to measure the performances of our procedure (density estimation) to the literature, we also compute the MISE obtained when using Fourier deconvolution approach. More precisely, we apply the procedure of Comte et al. (2006). It corresponds to a projection method with a \mathbb{R} -supported sinus cardinal basis or kernel. Besides this procedure is minimax optimal if f belongs to a Sobolev class in the case of a known ordinary smooth error distribution. We therefore compute the following estimator and penalty. Let g^* be the Fourier transform of g defined as $g^*(x) = \int e^{iux} g(u) du$. For a Gamma distribution of parameter p and θ , its Fourier transform is $g^*(u) = (1 - iu\theta)^{-p}$. We compute

$$\hat{f}_{\text{Fo},m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{n^{-1} \sum_{j=1}^n e^{iuZ_j}}{g^*(u)} du \quad \text{pen}_{\text{Fo}}^{(1)}(m) = \frac{\kappa_{\text{Fo}}^{(1)}}{2\pi n} \int_{-\pi m}^{\pi m} \frac{du}{|g^*(u)|^2}$$

We select m by minimizing $-\|\hat{f}_{\text{Fo},m}\|^2 + \text{pen}_{\text{Fo}}^{(1)}(m)$. If $Y = 0$, we set $g^* \equiv 1$ and $\text{pen}_{\text{Fo}}^{(2)}(m) = \kappa_{\text{Fo}}^{(2)}m/n$. The model collection is $\{m/10 : m \in \mathbb{N}, 1 \leq m \leq 50\}$. After calibration we find $\kappa_{\text{Fo}}^{(1)} = 41$ and $\kappa_{\text{Fo}}^{(2)} = 5$. We consider two different penalties since in the case of the model without noise, the estimator $\hat{f}_{\text{Fo},m}$ of f_m can be computed directly without approximating the integral.

Both procedures (Laguerre and Fourier) are fast.

2.5.3 Simulation results

		$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
		200	2000	200	2000	200	2000
Exponential	Lag	0.874	0.118	0.833	0.090	0.699	0.109
	Fou	3.950	3.234	6.002	3.359	11.228	3.889
Gamma $\Gamma(4, 1/4)$	Lag	0.243	0.048	1.181	0.076	1.872	0.384
	Fou	0.585	0.076	2.027	0.250	8.497	0.860
Gamma $\Gamma(20, 1/2)$	Lag	0.332	0.035	1.348	0.540	4.180	0.916
	Fou	0.521	0.059	1.917	0.245	2.728	1.092
Rayleigh	Lag	0.287	0.044	0.844	0.068	1.008	0.083
	Fou	0.498	0.073	1.546	0.248	7.523	0.797
Weibull	Lag	1.714	0.275	8.518	1.529	13.768	3.543
	Fou	7.004	6.611	8.751	6.839	14.237	7.421
Mixed Gamma	Lag	0.333	0.032	1.568	0.359	2.806	0.365
	Fou	0.488	0.062	1.038	0.204	8.317	0.829
Chi-squared	Lag	0.357	0.037	0.443	0.260	2.861	0.315
	Fou	0.542	0.069	1.887	0.250	8.135	0.844
Pareto	Lag	10.72	10.78	12.53	10.50	13.89	10.83
	Fou	12.42	6.54	18.70	9.38	30.06	26.28

Table 2.2: Results of simulation: MISE $\mathbb{E}(\|f - \hat{f}_m\|^2) \times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$.

The results are given in Tables 2.2 and 2.3. For both tables, the values of the MISE are multiplied by 100 for each case and computed from 200 simulated data. In Table 2.2 the abbreviations *Lag* and *Fou* correspond respectively to the Laguerre method and Fourier method of Comte et al. (2006). First we see that the risk decreases when the sample size increases. Likewise, the risk increases when the variance of the noise increases. If $Y = 0$ i.e. $\sigma^2 = 0$, we see that the Laguerre deconvolution has better performances than the Fourier deconvolution. For instance, when $n = 2000$ the MISE in the Fourier setting is almost systematically twice larger than the Laguerre for the Gamma, Rayleigh, mixed Gamma and Chi-squared distributions between the Laguerre and Fourier methods. For the Exponential density estimation, the ratio of the MISE of Fourier divided by Laguerre is of 1 to 3 and for the Weibull distribution 1 to 33. If σ^2 equals 1/10 or 1/4, we can make the same kind of remarks noticing that the ratio is worse for the Fourier deconvolution: 1 to 18.

Let us concentrate on the Pareto distribution. This distribution contrarily to the others does not have a density close to the Laguerre basis. In the model without noise, we see that the Fourier procedure is better. For small sample size the risks are very close. When the sample size increases, the risk of the Laguerre estimator decreases very slowly while the risk of the Fourier estimator is divided by two. On the other hand, in the deconvolution setting the Laguerre procedure performs better than the Fourier method especially for $\sigma^2 = 1/4$.

Thus the results point out the relevance of a specific method for nonnegative variables in a deconvolution problem.

In Table 2.3, the first two columns correspond to the estimation with the empirical estimator of the survival function if we observe directly the data. The estimation is very good: this was expected since the estimator converges to the true function with rate \sqrt{n} . Yet for the estimation of the Exponential distribution we note that the penalization procedure always beats the empirical estimator S_n . It is also the case for the density estimation. It is explained by the fact that the Exponential density with parameter 1 corresponds to the first function of the basis. We notice that the risk decreases when the sample size increases. For the Exponential distribution, it is divided by 10, by 3.5 for the Gamma distribution, by 4.5 for the Rayleigh distribution, by 13 for the Chi-squared distribution. And risk increases when the variance of the noise increases.

	$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
	200	2000	200	2000	200	2000
Exponential	0.262	0.022	0.122	0.014	0.134	0.012
Gamma	0.263	0.024	0.688	0.203	1.363	0.233
Rayleigh	0.813	0.115	0.878	0.199	1.336	0.297
Chi-squared	0.310	0.027	1.313	0.117	1.445	0.679

Table 2.3: Results of simulation: $\text{MISE } \mathbb{E} \left(\|S_X - \hat{S}_{X, \tilde{m}_2}\|^2 \right) \times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$.

We also illustrate the results with some figures. Figure 2.1 and 2.2 display the results of the data driven estimation respectively for the mixed Gamma and the Gamma $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$ for the Laguerre and Fourier methods. We can observe some oscillations near the origin for the Laguerre procedure, while for the Fourier method we can see that the estimators are a little bit shifted from the true density. For both methods the sample size n needs to be large enough to estimate the two modes of the mixed Gamma.

2.6 Concluding remarks

This paper deals with the estimation of nonnegative variables in a deconvolution setting with a known error distribution. First we have considered the adaptive estimation of the density f of the X_i 's in a deconvolution setting and deduced a procedure when there is no additional noise. Secondly we have tackled the problem of the adaptive estimation of the survival function which is new to our knowledge, in a global estimation setting on \mathbb{R}^+ . Moreover we have illustrated the performances of our new procedure and compared it, when

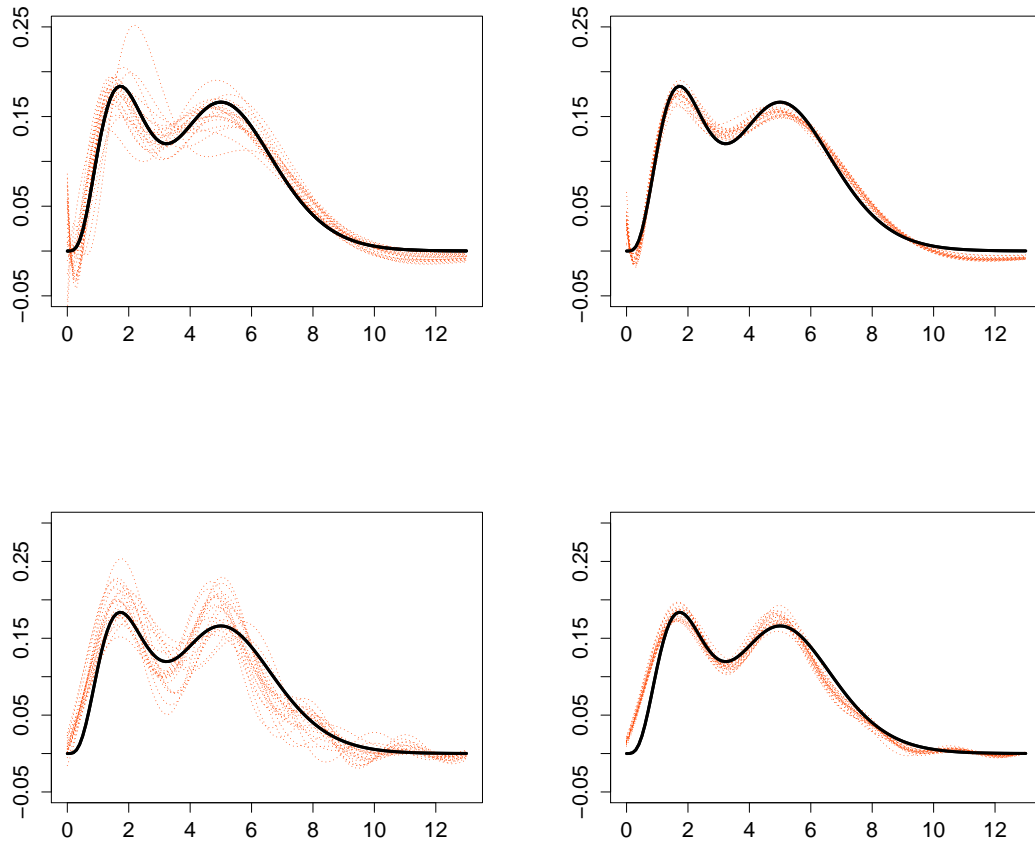


Figure 2.1: Estimation of the mixed Gamma density with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (bottom left for $n = 200$ and bottom right $n = 2000$), with $\sigma^2 = 1/10$.

it is possible to the performances of the Fourier procedure described in [Comte et al. \(2006\)](#). Our procedure outperforms the previous one in the simulations. These results show that the Laguerre procedure is worthy of interest when the variables are nonnegative.

The assumption of the knowledge of the error distribution is often not realistic in applications. Nevertheless this would require additional information on the error distribution. In the deconvolution literature with unknown error distribution it is assumed that we have access to a preliminary sample of the noise, see for instance [Neumann \(1997\)](#). Thanks to this preliminary observation we could estimate the coefficients of the matrix \mathbf{G}_m since we could provide unbiased estimators of the coefficients of the matrix which are the coefficients of the distribution g on the Laguerre basis. [Vareschi \(2015\)](#), in a Laplace regression model, considers this problem ; he assumes that a perturbation of the coefficients of the matrix \mathbf{G}_m are observed instead of preliminary sample drawn from g . At last, in our model we would need to control the deviation of the spectral norm of $\hat{\mathbf{G}}_m^{-1}$ around \mathbf{G}_m^{-1} .

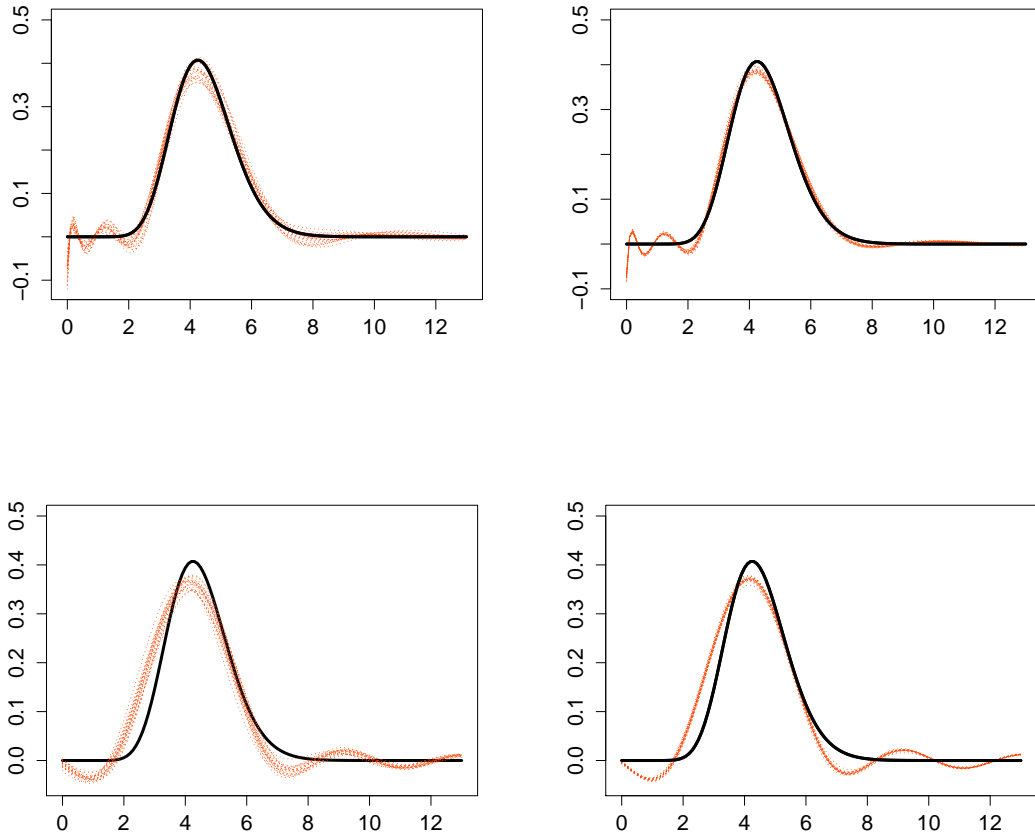


Figure 2.2: Estimation of the Gamma density $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$ with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (bottom left for $n = 200$ and bottom right $n = 2000$), with $\sigma^2 = 1/10$.

2.7 Proofs

2.7.1 Proof of Section 2.3

Proof of Proposition 2.3.1. According to the Pythagorean theorem, we have

$$\|f - \hat{f}_m\|^2 = \|f - f_m + f_m - \hat{f}_m\|^2 = \|f - f_m\|^2 + \|f_m - \hat{f}_m\|^2.$$

The first term corresponds to the bias term of Equation (2.13). Let us study the second term: using the decomposition on the orthonormal Laguerre basis, we have

$$\|f_m - \hat{f}_m\|^2 = \sum_{k=0}^{m-1} (b_k(f) - \hat{b}_k(f))^2.$$

First we apply (2.6) and get

$$\begin{aligned}\mathbb{E}\|f_m - \hat{f}_m\|^2 &= \mathbb{E}\|\mathbf{G}_m^{-1}(\hat{h}_m - \vec{h}_m)\|_{2,m}^2 \leq \varrho^2(\mathbf{G}_m^{-1})\mathbb{E}\|\hat{h}_m - \vec{h}_m\|_{2,m}^2 \\ &\leq \varrho^2(\mathbf{G}_m^{-1})\mathbb{E}\left[\sum_{j=1}^m \left(\frac{1}{n} \sum_{i=1}^n \varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_1)]\right)^2\right] \\ &\leq \frac{\varrho^2(\mathbf{G}_m^{-1})}{n} \sum_{j=1}^m \mathbb{E}[\varphi_j^2(Z_1)] \leq \varrho^2(\mathbf{G}_m^{-1}) \frac{2m}{n}.\end{aligned}$$

Secondly, we can notice that

$$\begin{aligned}\mathbb{E}\|\mathbf{G}_m^{-1}(\hat{h}_m - \vec{h}_m)\|_{2,m}^2 &= \mathbb{E}\sum_{k=1}^m \left(\sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} (b_{j-1}(h) - \hat{b}_{j-1}(h))\right)^2 \\ &= \mathbb{E}\sum_{k=1}^m \sum_{j'=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} (b_{j-1}(h) - \hat{b}_{j-1}(h)) [\mathbf{G}_m^{-1}]_{k,j'} (b_{j'-1}(h) - \hat{b}_{j'-1}(h)) \\ &= \sum_{k=1}^m \sum_{j'=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} [\mathbf{G}_m^{-1}]_{k,j'} \mathbb{E}\left[(b_{j-1}(h) - \hat{b}_{j-1}(h))(b_{j'-1}(h) - \hat{b}_{j'-1}(h))\right],\end{aligned}$$

and since $b_{j-1}(h) - \hat{b}_{j-1}(h) = (1/n) \sum_{i=1}^n (\varphi_{j-1}(Z_i) - \mathbb{E}[\varphi_{j-1}(Z_i)])$, it yields that

$$\begin{aligned}\mathbb{E}\|\mathbf{G}_m^{-1}(\hat{h}_m - \vec{h}_m)\|_{2,m}^2 &= \frac{1}{n} \sum_{k=1}^m \sum_{j=1}^m \text{Var}\left[[\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1)\right] \leq \frac{1}{n} \mathbb{E}\left[\left(\sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(Z_1)\right)^2\right] \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \int_{\mathbb{R}^+} \left(\sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j} \varphi_{j-1}(u)\right)^2 du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{1 \leq j, j' \leq m} [\mathbf{G}_m^{-1}]_{k,j} [\mathbf{G}_m^{-1}]_{k,j'} \int_{\mathbb{R}^+} \varphi_{j-1}(u) \varphi_{j'-1}(u) du \\ &\leq \frac{\|h\|_\infty}{n} \sum_{k=1}^m \sum_{j=1}^m [\mathbf{G}_m^{-1}]_{k,j}^2 = \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2.\end{aligned}$$

In the end we get: $\mathbb{E}\|f - \hat{f}_m\|^2 \leq \|f - f_m\|^2 + \frac{2m}{n} \varrho^2(\mathbf{G}_m^{-1}) \wedge \frac{\|h\|_\infty}{n} \|\mathbf{G}_m^{-1}\|_{\mathbb{F}}^2$. \square

Proof of Proposition 2.3.3. As in the previous proof, we can write that

$$\|S_X - \hat{S}_{X,m}\|^2 = \|S_X - S_{X,m}\|^2 + \|S_{X,m} - \hat{S}_{X,m}\|^2.$$

We can notice that

$$\|S_{X,m} - \hat{S}_{X,m}\|^2 = \left\| \hat{\vec{S}}_{X,m} - \vec{S}_{X,m} \right\|_{2,m}^2 = \left\| \mathbf{G}_m^{-1} \left(\hat{\vec{S}}_{Z,m} - \vec{S}_{Z,m} \right) \right\|_{2,m}^2$$

Then we repeat the same scheme as in the proof of Proposition 2.3.1 and we get

$$\mathbb{E}\|S_{X,m} - \hat{S}_{X,m}\|^2 \leq \frac{1}{n} \varrho^2(\mathbf{G}_m^{-1}) \sum_{j=0}^{m-1} \mathbb{E}[\Phi_j^2(Z_1)].$$

Yet

$$\sum_{j=0}^{m-1} \Phi_j^2(Z_1) = \sum_{j=0}^{m-1} \left(\int \varphi_j(u) \mathbf{1}_{0 \leq u \leq Z_1} du \right)^2 = \sum_{j=0}^{m-1} \langle \varphi_j, \mathbf{1}_{\cdot \leq Z_1} \rangle^2 \leq \|\mathbf{1}_{\cdot \leq Z_1}\|_{\mathbb{R}^+}^2 = Z_1 \quad (2.25)$$

which implies $\mathbb{E} \left[\sum_{j=0}^{m-1} \Phi_j^2(Z_1) \right] \leq \mathbb{E}[Z_1]$.

In the end: $\mathbb{E} \|S_X - \widehat{S}_{X,m}\|^2 \leq \|S_X - S_{X,m}\|^2 + \frac{\mathbb{E}[Z_1]}{n} \varrho^2(\mathbf{G}_m^{-1})$. \square

Proof of Lemma 2.3.4. To see that the spectral norm grows with the dimension m , recall that for a matrix \mathbf{A} of dimension m the spectral norm can be written as $\varrho^2(\mathbf{A}) = \max_{\|\vec{u}\|_{2,m}=1} \|\mathbf{A}\vec{u}\|_{2,m}^2$. Now consider $\vec{u}_m = \arg \max_{\|\vec{u}\|_{2,m}=1} \|\mathbf{T}_m \vec{u}\|_{2,m}$ with \mathbf{T}_m a lower triangular matrix and \mathbf{T}_m a submatrix of \mathbf{T}_{m+1} . We put $\vec{v}_{m+1} = (\vec{u}_m \ 0) = (u_1, \dots, u_m, 0)$ thus $\|\vec{v}_{m+1}\|_{2,m+1}^2 = 1$. Then we get

$$\|\mathbf{T}_{m+1} \vec{v}_{m+1}\|_{2,m+1}^2 = \|\mathbf{T}_m \vec{u}_m\|_{2,m}^2 + \sum_{i=1}^m \left([\mathbf{T}_{m+1}]_{m+1,i} u_i \right)^2 \geq \varrho^2(\mathbf{T}_m).$$

It yields

$$\varrho^2(\mathbf{T}_{m+1}) = \max_{\|\vec{v}\|_{2,m+1}=1} \|\mathbf{T}_{m+1} \vec{v}\|_{2,m+1}^2 \geq \varrho^2(\mathbf{T}_m).$$

\square

2.7.2 Proof of Theorem 2.4.1

First for $m \in \mathcal{M}_n^{(1)}$, let us define the associated subspaces $\mathcal{S}_{d_1}^m \subseteq \mathbb{R}^{d_1}$

$$\mathcal{S}_{d_1}^m = \left\{ \vec{t}_m \in \mathbb{R}^{d_1} / \vec{t}_m = {}^t(b_0(t), b_1(t), \dots, b_{m-1}(t), 0, \dots, 0) \right\}.$$

This space is defined to give nested models. When we increase the dimension from m to $m+1$ we only compute one more coefficient. Then for any $\vec{t} \in \mathbb{R}^{d_1}$, we define the following contrast for the density estimation

$$\gamma_n(\vec{t}) = \|\vec{t}\|_{2,d_1}^2 - 2 \langle \vec{t}, \mathbf{G}_{d_1}^{-1} \widehat{h}_{d_1} \rangle_{2,d_1}.$$

Let us notice that for $\vec{t}_m \in \mathcal{S}_{d_1}^m$, thanks to the null coordinates of \vec{t}_m and the lower triangular form of \mathbf{G}_{d_1} and \mathbf{G}_m , we have

$$\langle \vec{t}_m, \mathbf{G}_{d_1}^{-1} \widehat{h}_{d_1} \rangle_{2,d_1} = \langle \vec{t}_m, \mathbf{G}_m^{-1} \widehat{h}_m \rangle_{2,m} = \langle \vec{t}_m, \widehat{f}_m \rangle_{2,m}.$$

So we clearly have that

$$\widehat{f}_m = \arg \min_{\vec{t}_m \in \mathcal{S}_{d_1}^m} \gamma_n(\vec{t}_m).$$

Now let $m, m' \in \mathcal{M}_n^{(1)}$, $\vec{t}_m \in \mathcal{S}_{d_1}^m$ and $\vec{s}_{m'} \in \mathcal{S}_{d_1}^{m'}$. Denote $m^* = m \vee m'$. Notice that

$$\gamma_n(\vec{t}_m) - \gamma_n(\vec{s}_{m'}) = \|\vec{t}_m - \widehat{f}\|_{2,d_1}^2 - \|\vec{s}_{m'} - \widehat{f}\|_{2,d_1}^2 - 2 \langle \vec{t}_m - \vec{s}_{m'}, \mathbf{G}_{d_1}^{-1} (\widehat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$$

and due to orthonormality of Laguerre basis, for any m we have the following relations between the \mathbb{L}^2 norm and the Euclidean norms,

$$\|\hat{f}_m - f\|^2 = \|\hat{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (b_j(f))^2 \quad \text{and} \quad \|f_m - f\|^2 = \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \sum_{j=d_1}^{\infty} (b_j(f))^2 \quad (2.26)$$

We set $\nu_n(\vec{t}) = \langle \vec{t}, \mathbf{G}_{d_1}^{-1}(\hat{h}_{d_1} - \vec{h}_{d_1}) \rangle_{2,d_1}$ for $\vec{t} \in \mathbb{R}^{d_1}$.

According to the definition of \hat{m} , for any m in the model collection $\mathcal{M}_n^{(1)}$, we have the following inequality

$$\gamma_n(\hat{f}_{\hat{m}}) + \text{pen}_1(\hat{m}) \leq \gamma_n(\vec{f}_m) + \text{pen}_1(m).$$

It yields that

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 - \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 - 2\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) \leq \text{pen}_1(m) - \text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) + \text{pen}_1(m) - \text{pen}_1(\hat{m}).$$

Let us notice that $\nu_n(\hat{f}_{\hat{m}} - \vec{f}_m) = \|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \nu_n \left(\frac{\hat{f}_{\hat{m}} - \vec{f}_m}{\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}} \right)$ and due to the relation $2ab \leq a^2/4 + 4b^2$, we have the following inequalities

$$\begin{aligned} \|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1} \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m}) \\ &\leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{4}\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m}) \end{aligned}$$

where $\mathcal{B}(m, \hat{m}) = \left\{ \vec{t}_{m \vee \hat{m}} \in \mathcal{S}_{d_1}^{m \vee \hat{m}}, \|\vec{t}_{m \vee \hat{m}}\|_{2,d_1} = 1 \right\}$. Now notice that

$$\|\hat{f}_{\hat{m}} - \vec{f}_m\|_{2,d_1}^2 \leq 2\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + 2\|\vec{f}_m - \vec{f}\|_{2,d_1}^2$$

we then have

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq \|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 + \frac{1}{2}\|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 4 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) + \text{pen}_1(m) - \text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - \vec{f}\|_{2,d_1}^2 \leq 3\|\vec{f}_m - \vec{f}\|_{2,d_1}^2 + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m}).$$

Using Equation (2.26), we have

$$\|\hat{f}_{\hat{m}} - f\|^2 - \sum_{j=d_1}^{\infty} (b_j(f))^2 \leq 3 \left(\|f - f_m\|^2 - \sum_{j=d_1}^{\infty} (b_j(f))^2 \right) + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m})$$

which implies

$$\|\hat{f}_{\hat{m}} - f\|^2 \leq 3\|f - f_m\|^2 + 2\text{pen}_1(m) + 8 \sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - 2\text{pen}_1(\hat{m}) \quad (2.27)$$

Now let p_1 be a function such that for any m, m' , we have : $4p_1(m, m') \leq \text{pen}_1(m) + \text{pen}_1(m')$.

$$\begin{aligned} \|\hat{f}_{\hat{m}} - f\|^2 &\leq 3\|f - f_m\|^2 + 4\text{pen}_1(m) + 8 \left[\sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} \nu_n^2(\vec{t}) - p_1(m, \hat{m}) \right]_+ \\ &\leq 3\|f - f_m\|^2 + 4\text{pen}_1(m) + 8 \sum_{m' \in \mathcal{M}_n^{(1)}} \left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} \nu_n^2(\vec{t}) - p_1(m, m') \right\}_+ \end{aligned}$$

We now use the following result which ensures the validity of Theorem 2.4.1.

Proposition 2.7.1. *Under the assumptions of Theorem 2.4.1, there exists a constant $C_1 > 0$ depending on $\|h\|_\infty$ such that for*

$$p_1(m, m') = \frac{\kappa_1}{n} \min \left(2(m \vee m') \varrho^2 \left(\mathbf{G}_{m \vee m'}^{-1} \right), \log n (\|h\|_\infty \vee 1) \|\mathbf{G}_{m \vee m'}^{-1}\|_{\mathbb{F}}^2 \right),$$

we have

$$\mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} \nu_n^2(\vec{t}) - p(m, m') \right\}_+ \right] \leq \frac{C_1}{n}.$$

In the end:

$$\mathbb{E} \|f - \hat{f}_{\hat{m}}\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(1)}} \left\{ \|f - f_m\|^2 + \text{pen}_1(m) \right\} + \frac{C_1}{n},$$

as soon as $\kappa_1 \geq 294$. □

Proof of Proposition 2.7.1. To prove Proposition 2.7.1, we apply a Talagrand inequality. So we need to determine H , M_1 and v defined as

$$\begin{aligned} \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \|\langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(\cdot) \rangle_{2, d_1}\|_\infty &\leq M_1, \quad \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})| \right] \leq H, \\ \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \text{Var} \left[\langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2, d_1} \right] &\leq v. \end{aligned}$$

where $m^* = m \vee m'$.

• Let us start with the empirical process, first let us notice that

$$\begin{aligned} &\mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})|^2 \right] \\ &= \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{d_1}(Z_i) - \mathbb{E}[\vec{\varphi}_{d_1}(Z_i)]) \right) \rangle_{2, d_1} \right|^2 \right] \\ &= \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \rangle_{2, m^*} \right|^2 \right] \end{aligned}$$

We now apply Cauchy-Schwarz inequality and get

$$\begin{aligned}
& \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\nu_n(\vec{t}_{m^*})|^2 \right] \\
& \leq \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left\| \vec{t}_{m^*} \right\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \right\|_{2, m^*}^2 \right] \\
& \leq \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \mathbb{E} \left[\sum_{j=0}^{m^*-1} \left(\frac{1}{n} \sum_{i=1}^n (\varphi_j(Z_i) - \mathbb{E}[\varphi_j(Z_i)]) \right)^2 \right] \\
& \leq \frac{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \sum_{j=0}^{m^*-1} \text{Var}(\varphi_j(Z_1)) \leq \frac{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \sum_{j=0}^{m^*-1} \mathbb{E}[\varphi_j^2(Z_1)] \leq \frac{2m^*}{n} \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right).
\end{aligned}$$

Moreover as in proof of Proposition 2.3.1, we can notice that

$$\mathbb{E} \left[\left\| \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\varphi}_{m^*}(Z_i) - \mathbb{E}[\vec{\varphi}_{m^*}(Z_i)]) \right) \right\|_{2, m^*}^2 \right] \leq \frac{(\|h\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{n}.$$

We then set $H := \sqrt{\frac{2m^*}{n} \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \wedge \frac{(\|h\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{n}}$.

Before deriving a bound for the term of variance, let us remind that for any $x \in \mathbb{R}^+$, we have

$$h(x) = \int f(u)g(x-u)\mathbf{1}_{u \geq 0}\mathbf{1}_{x-u \geq 0} du = \int_0^x f(u)g(x-u) du.$$

Then we apply Cauchy-Schwarz inequality and get $\forall x \in \mathbb{R}^+$, $h(x) \leq \|f\| \|g\|$. If f and $g \in \mathbb{L}^2(\mathbb{R}^+)$ is fulfilled, we clearly have $\|h\|_\infty \leq \|f\| \|g\| < \infty$.

• Now for the term of variance, let $\vec{t}_{m^*} \in \mathcal{B}(m, m')$. By definition we have the following equalities

$$\begin{aligned}
& \mathbb{E} \left[\left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2, d_1} \right|^2 \right] = \mathbb{E} \left[\left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(Z_1) \rangle_{2, m^*} \right|^2 \right] \\
& = \mathbb{E} \left[\left| \sum_{j=0}^{m^*-1} b_j(t) \sum_{k=0}^{m^*-1} \left[\mathbf{G}_{m^*}^{-1} \right]_{jk} \varphi_k(Z_1) \right|^2 \right] = \int_{\mathbb{R}^+} \left| \sum_{0 \leq k, j \leq m^*-1} b_j(t) \left[\mathbf{G}_{m^*}^{-1} \right]_{jk} \varphi_k(u) \right|^2 h(u) du.
\end{aligned}$$

which implies

$$\begin{aligned}
& \mathbb{E} \left[\left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(Z_1) \rangle_{2, d_1} \right|^2 \right] \\
& \leq \|h\|_\infty \int_{\mathbb{R}^+} \sum_{0 \leq j, j', k, k' \leq m^*-1} b_j(t) b_{j'}(t) \left[\mathbf{G}_{m^*}^{-1} \right]_{jk} \left[\mathbf{G}_{m^*}^{-1} \right]_{j'k'} \varphi_k(u) \varphi_{k'}(u) du \\
& \leq \|h\|_\infty \sum_{0 \leq j, j', k, k' \leq m^*-1} b_j(t) b_{j'}(t) \left[\mathbf{G}_{m^*}^{-1} \right]_{jk} \left[\mathbf{G}_{m^*}^{-1} \right]_{j'k'} \delta_{k, k'} \\
& \leq \|h\|_\infty \sum_{0 \leq j, j', k \leq m^*-1} b_j(t) b_{j'}(t) \left[\mathbf{G}_{m^*}^{-1} \right]_{jk} \left[\mathbf{G}_{m^*}^{-1} \right]_{j'k} \\
& \leq \|h\|_\infty \vec{t}_{m^*} \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{t}_{m^*} \leq \|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \|\vec{t}_{m^*}\|_{2, m^*}^2 \leq \|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right).
\end{aligned}$$

So we set $v := \|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)$.

• Now applying Cauchy-Schwarz inequality

$$\begin{aligned} & \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_1}^{-1} \vec{\varphi}_{d_1}(x) \rangle_{2, d_1} \right| = \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x) \rangle_{2, m^*} \right| \\ & \leq \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \left\| \vec{t}_{m^*} \right\|_{2, m^*} \left\| \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x) \right\|_{2, m^*} \leq \sup_{x \in \mathbb{R}^+} \left\| \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x) \right\|_{2, m^*} \\ & \leq \sup_{x \in \mathbb{R}^+} \sqrt{t \vec{\varphi}_{m^*}(x) \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{\varphi}_{m^*}(x)} \leq \sqrt{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \sup_{x \in \mathbb{R}^+} \sum_{j=0}^{m^*-1} \varphi_j^2(x)} \leq \sqrt{2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}. \end{aligned}$$

We take $M_1 = \sqrt{2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}$.

• We can now apply Talagrand's inequality (see Appendix 2.8.1).

i) First case : $2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \wedge \|h\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2 = 2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)$.

For $\xi^2 = 1/2$ with $K_1 = 1/6$, Talagrand's inequality implies

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\nu_n(\vec{t})|^2 - 4 \frac{2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \right\}_+ \right] \\ & \leq C \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} e^{-K_1 \frac{m^*}{\|h\|_\infty}} + \frac{m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n^2} e^{-c_1 \sqrt{n}}. \end{aligned}$$

Yet under Assumption **(A2)**, we have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} e^{-K_1 \frac{m^*}{\|h\|_\infty}} \leq \frac{C}{n}.$$

Moreover according to Assumptions **(A1)** and **(A2)**, we also have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n^2} e^{-c_1 \sqrt{n}} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{e^{-c_1 \sqrt{n}}}{n} \leq \frac{C}{n}.$$

ii) Second case : $2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \wedge \|h\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2 = \|h\|_\infty \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2$.

For $\xi^2 = \mathbf{a}/K_1 \log n$ with $\mathbf{a} = 7^2/K_1$, Talagrand's inequality implies

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\nu_n(\vec{t})|^2 - 2 \left(1 + \frac{\mathbf{a}}{K_1} \log n \right) \frac{(\|h\|_\infty \vee 1) \|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{n} \right\}_+ \right] \\ & \leq C \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \exp \left(-\mathbf{a} \log(n) \frac{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)} \right) \\ & \quad + \frac{m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{C^2(\xi^2) n^2} \exp \left(-\frac{2K_1}{7\sqrt{2}} C(\xi^2) \sqrt{\frac{\mathbf{a}}{K_1}} \sqrt{\log(n)} n \frac{(\|h\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}} \right). \quad (2.28) \end{aligned}$$

The first summand of Equation (2.28) implies that

$$\begin{aligned}
& \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \exp\left(-\mathfrak{a} \log(n) \frac{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}{\varrho^2(\mathbf{G}_{m^*}^{-1})}\right) \\
& \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \exp(-\mathfrak{a} \log(n)) \quad (\text{Equivalence between norms}) \\
& \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{\|h\|_\infty \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^\mathfrak{a}} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \|h\|_\infty \frac{1}{n^\mathfrak{a}} \quad (\text{Assumption (A1)}) \\
& = |\mathcal{M}_n^{(1)}| \frac{\|h\|_\infty}{n^\mathfrak{a}} = \frac{\|h\|_\infty}{n^{\mathfrak{a}-1}} \leq \frac{\|h\|_\infty}{n}
\end{aligned}$$

since $\mathfrak{a} > 2$.

Now let us bound the second summand of Equation (2.28). Assumption (A1) gives that $m^* \varrho^2(\mathbf{G}_{m^*}^{-1}) \leq n/\log n$, yet

$$\varrho(\mathbf{G}_{m^*}^{-1}) \geq \sqrt{2}/a_0(g) = 1/\mathbb{E}[e^{-Y}] \geq 1,$$

where we use that the operator norm is greater than the spectral radius. Then it implies that $m^* \leq m^* \varrho^2(\mathbf{G}_{m^*}^{-1})$. This remark combined with the equivalence between norms (see Eq. (2.18)) yields

$$-n \frac{(\|h\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}} \leq -\frac{n}{\sqrt{2n}} \frac{(\|h\|_\infty \vee 1)}{\sqrt{m^*}} \leq -\frac{1}{\sqrt{2}} \frac{n \sqrt{\log n}}{\sqrt{n} \sqrt{n}} \leq -\frac{\sqrt{\log n}}{\sqrt{2}}.$$

Secondly we have $-C(\xi^2) = -\left(\sqrt{1 + \frac{\mathfrak{a}}{K_1} \log n} - 1 \wedge 1\right) \leq -1$, which implies that $C(\xi^2) = 1$ since $C(\xi^2) \leq 1$ by definition. This condition is verified as soon as $n \geq e^{K_1/\mathfrak{a}}$, in other words for $n \geq 2$. Then we get that

$$\begin{aligned}
& \exp\left(-\frac{2K_1}{7\sqrt{2}} C(\xi^2) \sqrt{\frac{\mathfrak{a}}{K_1}} \sqrt{\log(n)} n \frac{(\|h\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}}\right) \leq \exp\left(-\frac{2K_1}{7\sqrt{2}} \sqrt{\frac{\mathfrak{a}}{K_1}} \log n \frac{1}{\sqrt{2}}\right) \\
& \leq \exp\left(-\frac{\sqrt{\mathfrak{a}K_1}}{7} \log n\right) \leq \frac{1}{n}
\end{aligned}$$

since $\mathfrak{a} = 7^2/K_1$. Thus, invoking again Assumption (A1), the second summand of Equation (2.28) can be bounded as follows

$$\begin{aligned}
& \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{C(\xi^2) n^2} \exp\left(-\frac{2K_1}{7\sqrt{2}} C(\xi^2) \sqrt{\log(n)} n \frac{(\|h\|_\infty \vee 1) \sqrt{\|\mathbf{G}_{m^*}^{-1}\|_{\mathbb{F}}^2}}{\sqrt{n} \sqrt{2m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}}\right) \\
& \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2(\mathbf{G}_{m^*}^{-1})}{C(\xi^2) n^2} \frac{1}{n} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{1}{C(\xi^2) n^2} = \frac{|\mathcal{M}_n^{(1)}|}{C(\xi^2) n^2} = \frac{1}{C(\xi^2) n} \leq \frac{1}{n}.
\end{aligned}$$

In the end we have the desired result. \square

Proof of Remark 6. We have to prove that Proposition 2.7.1 is still valid although Assumption (A2) is no longer true. We set $\xi^2 = 2 \log n / K_1$, $H^2 = 2m^* \varrho^2 (\mathbf{G}_{m^*}^{-1}) / n$, $v = 2m^* \varrho^2 (\mathbf{G}_{m^*}^{-1})$, $M_1 = \sqrt{2m^* \varrho^2 (\mathbf{G}_{m^*}^{-1})}$. Under (A2), we have

$$\sum_{m' \in \mathcal{M}_n^{(1)}} \frac{v}{n} \exp \left(-K_1 \xi^2 n \frac{H^2}{v} \right) = \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{2m^* \varrho^2 (\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^2} \leq \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{2}{n^2} \leq \frac{2|\mathcal{M}_n^{(1)}|}{n^2} \leq \frac{2}{n}.$$

For $C(\xi^2)$ defined in Lemma 2.8.1, we get

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{M_1^2}{K_1 C^2(\xi^2) n^2} \exp \left(-\frac{2K_1 C(\xi^2) \xi n H}{7\sqrt{2} M_1} \right) \\ &= \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2 (\mathbf{G}_{m^*}^{-1})}{K_1 C^2(\xi^2) n^2} \exp \left(-\frac{2K_1 C(\xi^2) \xi \sqrt{n}}{7\sqrt{2}} \right) \\ &\leq C_2 \sum_{m' \in \mathcal{M}_n^{(1)}} \frac{m^* \varrho^2 (\mathbf{G}_{m^*}^{-1})}{n^2} \exp \left(-C_1 \sqrt{n \log n} \right) \\ &\leq \frac{C_3 |\mathcal{M}_n^{(1)}|}{n} \exp \left(-C_1 \sqrt{n \log n} \right) \leq \frac{C_3}{n}. \end{aligned}$$

□

2.7.3 Proof of Theorem 2.4.2 and Corollary 2.4.3

For any $\vec{t} \in \mathbb{R}^{d_2}$, we define the following contrast for the survival function estimation

$$\delta_n(\vec{t}) = \|\vec{t}\|_{2, d_2}^2 - 2 \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left(\widehat{\vec{S}}_{Z, d_2} - \vec{S}_{Y, d_2} \right) \rangle_{2, d_2}$$

and we also have

$$\langle \vec{t}_m, \mathbf{G}_{d_2}^{-1} \left(\widehat{\vec{S}}_{Z, d_2} - \vec{S}_{Y, d_2} \right) \rangle_{2, d_2} = \langle \vec{t}_m, \mathbf{G}_m^{-1} \left(\widehat{\vec{S}}_{Z, m} - \vec{S}_{Y, m} \right) \rangle_{2, m} = \langle \vec{t}_m, \widehat{\vec{S}}_{X, m} \rangle_{2, m}$$

which yields that

$$\widehat{\vec{S}}_{X, m} = \arg \min_{\vec{t}_m \in \mathcal{S}_{d_2}^m} \delta_n(\vec{t}_m).$$

Proof of Theorem 2.4.2

The beginning of the proof is the same as the proof of Theorem 2.4.1 with the quantities associated to the survival function estimation. Then we start from Equation (2.27) with $\nu_n(\vec{t})$ replacing the following empirical process $\zeta_n(\vec{t}) := \zeta_n^{(1)}(\vec{t}) + \zeta_n^{(2)}(\vec{t})$ where,

$$\begin{aligned} \zeta_n^{(1)}(\vec{t}) &:= \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \left(\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \rangle_{2, d_2} \\ \zeta_n^{(2)}(\vec{t}) &:= \langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \left(\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} - \mathbb{E} \left[\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} \right] \right) \right) \rangle_{2, d_2}. \end{aligned}$$

So we have the following inequality

$$\begin{aligned} \|\widehat{S}_{X,\widehat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X,m}\|^2 + 2\text{pen}_2(m) \\ &\quad + 16 \left(\sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) - 2\text{pen}_2(\widehat{m}). \end{aligned} \quad (2.29)$$

Now let q be a function such that for any m, m' , we have : $4q(m, m') \leq \text{pen}_2(m) + \text{pen}_2(m')$.

$$\begin{aligned} \|\widehat{S}_{X,\widehat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X,m}\|^2 + 4\text{pen}_2(m) + 16 \left[\sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, \widehat{m}) \right]_+ \\ &\quad + 16 \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \\ &\leq 3\|S_X - S_{X,m}\|^2 + 4\text{pen}_2(m) + 16 \sum_{m' \in \mathcal{M}_n^{(2)}} \left\{ \sup_{\vec{t} \in \mathcal{B}(m,m')} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, m') \right\}_+ \\ &\quad + 16 \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2. \end{aligned}$$

We now use the following result which ensures the validity of Theorem 2.4.2.

Proposition 2.7.2. *Under the assumptions of Theorem 2.4.2, then there exists a universal constant $C > 0$ such that for $q(m, m') = \kappa_2 \varrho^2 \left(\mathbf{G}_{m \vee m'}^{-1} \right) \mathbb{E}[Z_1] \frac{\log n}{n}$*

$$\begin{aligned} (i) \quad &\mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m,m')} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, m') \right\}_+ \right] \leq \frac{C}{n} \\ (ii) \quad &\mathbb{E} \left[\sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right] \leq \frac{\mathbb{E}[Z_1^3]}{n}. \end{aligned}$$

Finally, $\mathbb{E}\|\widehat{S}_{X,\widehat{m}} - S_X\|^2 \leq 4 \inf_{m \in \mathcal{M}_n^{(2)}} \{\|S_X - S_{X,m}\|^2 + \text{pen}_2(m)\} + \frac{C}{n}$. □

Proof of Proposition 2.7.2. To prove (i), we apply a Talagrand inequality. So we need to determine H , M_1 and v .

• Let us start with the empirical process, first let us notice that

$$\begin{aligned} &\mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m,m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ &= \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m,m')} \left| \left\langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \left(\vec{\Phi}_{d_2}(Z_i) \mathbb{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[\vec{\Phi}_{d_2}(Z_i) \mathbb{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \right\rangle_{2,d_2} \right|^2 \right] \\ &= \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m,m')} \left| \left\langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n \left(\vec{\Phi}_{m^*}(Z_i) \mathbb{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} \left[\vec{\Phi}_{m^*}(Z_i) \mathbb{1}_{Z_i \leq \sqrt{n}} \right] \right) \right) \right\rangle_{2,m^*} \right|^2 \right]. \end{aligned}$$

We now apply Cauchy-Schwarz inequality and get

$$\begin{aligned} & \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ & \leq \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left\| \vec{t}_{m^*} \right\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} [\vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}}]) \right) \right\|_{2, m^*}^2 \right] \\ & \leq \mathbb{E} \left[\left\| \mathbf{G}_{m^*}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} [\vec{\Phi}_{m^*}(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}}]) \right) \right\|_{2, m^*}^2 \right]. \end{aligned}$$

It follows that, with Equation (2.25)

$$\begin{aligned} & \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t}_{m^*})|^2 \right] \\ & \leq \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \mathbb{E} \left[\sum_{j=0}^{m^*-1} \left(\frac{1}{n} \sum_{i=1}^n (\Phi_j(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}} - \mathbb{E} [\Phi_j(Z_i) \mathbf{1}_{Z_i \leq \sqrt{n}}]) \right)^2 \right] \\ & \leq \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \sum_{j=0}^{m^*-1} \text{Var} \left(\Phi_j(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right) \\ & \leq \frac{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \sum_{j=0}^{m^*-1} \mathbb{E} \left[\Phi_j^2(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right] \leq \frac{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \mathbb{E} \left[Z_1 \mathbf{1}_{Z_1 \leq \sqrt{n}} \right]. \end{aligned}$$

We then set $H := \sqrt{\frac{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}{n} \mathbb{E} [Z_1]}$.

• Now for the term of variance, let $\vec{t}_{m^*} \in \mathcal{B}(m, m')$

$$\begin{aligned} & \mathbb{E} \left[\left| \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \rangle_{2, d_2} \right|^2 \right] = \mathbb{E} \left[\left| \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \rangle_{2, m^*} \right|^2 \right] \\ & \leq \mathbb{E} \left[\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left\| \vec{t}_{m^*} \right\|_{2, m^*}^2 \left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right\|_{2, m^*}^2 \right] \\ & \leq \mathbb{E} \left[\left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(Z_1) \mathbf{1}_{Z_1 \leq \sqrt{n}} \right\|_{2, m^*}^2 \right] \leq \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \mathbb{E} \left[Z_1 \mathbf{1}_{Z_1 \leq \sqrt{n}} \right]. \end{aligned}$$

So we set $v := \mathbb{E} [Z_1] \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)$.

• First notice again

$$\sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, d_2} = \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \sup_{x \in \mathbb{R}^+} \langle \vec{t}_{m^*}, \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, m^*}$$

Now applying Cauchy-Schwarz inequality and using Equation (2.25) again

$$\begin{aligned} & \sup_{\vec{t}_{m^*} \in \mathcal{B}(m, m')} \left\| \langle \vec{t}_{m^*}, \mathbf{G}_{d_2}^{-1} \vec{\Phi}_{d_2}(x) \mathbf{1}_{x \leq \sqrt{n}} \rangle_{2, d_2} \right\|_{\infty} \\ & \leq \sup_{x \in \mathbb{R}^+} \left\| \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq \sqrt{n}} \right\|_{2, m^*} \leq \sup_{x \in \mathbb{R}^+} \sqrt{\vec{\Phi}_{m^*}(x) {}^t \mathbf{G}_{m^*}^{-1} \mathbf{G}_{m^*}^{-1} \vec{\Phi}_{m^*}(x) \mathbf{1}_{x \leq \sqrt{n}}} \\ & \leq \sqrt{\varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right) \sup_{x \in \mathbb{R}^+} \sum_{j=0}^{m^*-1} \Phi_j^2(x) \mathbf{1}_{x \leq \sqrt{n}}} \leq \sqrt{\sqrt{n} \varrho^2 \left(\mathbf{G}_{m^*}^{-1} \right)}. \end{aligned}$$

We take $M_1 = \sqrt{\sqrt{n}\varrho^2(\mathbf{G}_{m^*}^{-1})}$.

We apply Talagrand's inequality for $\xi^2 = \frac{2}{K_1} \log(n)$. We get

$$\begin{aligned} & \mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2(\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \\ & \leq \frac{4}{K_1} \frac{\mathbb{E}[Z_1] \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} + \frac{98n\varrho^2(\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi)\xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \end{aligned}$$

which implies that

$$\begin{aligned} & \sum_{m' \in \mathcal{M}_n^{(2)}} \mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2(\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \\ & \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{4}{K_1} \frac{\mathbb{E}[Z_1] \varrho^2(\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} + \frac{98\sqrt{n}\varrho^2(\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi)\xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}}. \end{aligned}$$

Thus we have under Assumption **(B1)**

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\varrho^2(\mathbf{G}_{m^*}^{-1})}{n} e^{-K_1 \xi^2} \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\varrho^2(\mathbf{G}_{m^*}^{-1})}{n} \frac{1}{n^2} \leq \frac{|\mathcal{M}_n^{(2)}|}{n^2} \leq \frac{1}{n}.$$

And

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \frac{\sqrt{n}\varrho^2(\mathbf{G}_{m^*}^{-1})}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi)\xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \leq \sum_{m' \in \mathcal{M}_n^{(2)}} \frac{C\sqrt{n}}{(\log n)^2 n} e^{-\frac{2K_1 C(\xi)\xi}{7\sqrt{2}} \sqrt{\mathbb{E}[Z_1]} n^{1/4}} \leq \frac{C}{n}.$$

So for $q(m, m') = \kappa_2 \varrho^2(\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n}$ we just showed that

$$\sum_{m' \in \mathcal{M}_n^{(2)}} \mathbb{E} \left[\left\{ \sup_{\vec{t} \in \mathcal{B}(m, m')} |\zeta_n^{(1)}(\vec{t})|^2 - \kappa_2 \varrho^2(\mathbf{G}_{m^*}^{-1}) \mathbb{E}[Z_1] \frac{\log n}{n} \right\}_+ \right] \leq \frac{C}{n}.$$

Now we prove (ii). We have, using **(A2)**,

$$\begin{aligned} & \mathbb{E} \left[\sup_{\vec{t} \in \mathcal{B}(m, \hat{m})} |\zeta_n^{(2)}(\vec{t}_{m^*})|^2 \right] \\ & = \mathbb{E} \left[\sup_{\vec{t} \in \mathcal{B}(d_2, d_2)} \left| \left\langle \vec{t}, \mathbf{G}_{d_2}^{-1} \left(\frac{1}{n} \sum_{i=1}^n (\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}} - \mathbb{E}[\vec{\Phi}_{d_2}(Z_i) \mathbf{1}_{Z_i > \sqrt{n}}]) \right) \right\rangle_{2, d_2} \right|^2 \right] \\ & \leq \frac{\varrho^2(\mathbf{G}_{d_2}^{-1})}{n} \mathbb{E} \left[Z_1 \mathbf{1}_{Z_1 > \sqrt{n}} \right] \leq \frac{\mathbb{E}[Z_1^3]}{n}. \end{aligned}$$

In the end we have the desired result. \square

Proof of Corollary 2.4.3

The beginning of the proof is the same as the proof of Theorem 2.4.2 except that we consider \tilde{m}_2 (defined by Equation (2.23)) instead of \hat{m}_2 and $\widehat{\text{pen}}_2$ (defined by Equation (2.24)) instead of pen_2 . Starting from Equation (2.29) we have

$$\begin{aligned} \|\widehat{S}_{X,\widehat{m}} - S_X\|^2 &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left(\sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) \\ &\quad - 2\widehat{\text{pen}}_2(\tilde{m}_2) \\ &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left(\sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 + \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 \right) \\ &\quad - 2\text{pen}_2(\tilde{m}_2) + 2\text{pen}_2(\tilde{m}_2) - 2\widehat{\text{pen}}_2(\tilde{m}_2) \\ &\leq 3\|S_X - S_{X,m}\|^2 + 2\widehat{\text{pen}}_2(m) + 16 \left\{ \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(1)}(\vec{t}))^2 - q(m, \tilde{m}_2) \right\}_+ \\ &\quad + 2\text{pen}_2(m) + 16 \sup_{\vec{t} \in \mathcal{B}(m,\widehat{m})} (\zeta_n^{(2)}(\vec{t}))^2 + 2 \left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+. \end{aligned}$$

We now apply the following Proposition which ensures the validity of Corollary 2.4.3

Proposition 2.7.3. *Under the Assumptions of Corollary 2.4.3, the following holds*

$$\mathbb{E}[\widehat{\text{pen}}_2(m)] = 2\text{pen}_2(m) \quad \text{and} \quad \mathbb{E} \left[\left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] \leq \frac{C}{n}.$$

Finally, $\mathbb{E}\|S_X - \widehat{S}_{X,\tilde{m}_2}\|^2 \leq \inf_{m \in \mathcal{M}_n^{(2)}} \{\|S_X - S_{X,m}\|^2 + \text{pen}_2(m)\} + \frac{C}{n}$. \square

Proof of Proposition 2.7.3. First let us notice the following

$$\mathbb{E}[\widehat{\text{pen}}_2(m)] = 2\kappa_2 \mathbb{E}[\bar{Z}_n] \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} = 2\kappa_2 \mathbb{E}[Z_1] \frac{\varrho^2(\mathbf{G}_m^{-1}) \log n}{n} = 2\text{pen}_2(m).$$

For the second inequality, let us introduce the following favorable set:

$$\Lambda = \left\{ \left| \mathbb{E}[Z_1] - \bar{Z}_n \right| \leq \mathbb{E}[Z_1]/2 \right\},$$

which yields

$$\begin{aligned} \mathbb{E} \left[\left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] &= \mathbb{E} \left[\left\{ 2\kappa_2 \left(\frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \frac{\log n}{n} \right\}_+ \mathbf{1}_\Lambda \right] \\ &\quad + \mathbb{E} \left[\left\{ 2\kappa_2 \left(\frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \frac{\log n}{n} \right\}_+ \mathbf{1}_{\Lambda^c} \right]. \end{aligned}$$

Yet on the set Λ , $\mathbb{E}[Z_1]/2 - \bar{Z}_n \leq 0$ which yields

$$\begin{aligned} \mathbb{E} \left[\left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] &= \mathbb{E} \left[2\kappa_2 \left(\frac{\mathbb{E}[Z_1]}{2} - \bar{Z}_n \right) \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \mathbf{1}_{\Lambda^c} \right] \frac{\log n}{n} \\ &\leq \mathbb{E} \left[2\kappa_2 \left| \mathbb{E}[Z_1] - \bar{Z}_n \right| \varrho^2(\mathbf{G}_{\tilde{m}_2}^{-1}) \mathbf{1}_{\Lambda^c} \right] \frac{\log n}{n}. \end{aligned}$$

Now we apply Cauchy-Schwarz

$$\mathbb{E} \left[\left| \mathbb{E}[Z_1] - \bar{Z}_n \right| \mathbf{1}_{\Lambda^c} \right] \leq \sqrt{\mathbb{E} \left[\left| \mathbb{E}[Z_1] - \bar{Z}_n \right|^2 \right]} \sqrt{\mathbb{P}[\Lambda^c]} = \sqrt{\text{Var}[\bar{Z}_n]} \sqrt{\mathbb{P} \left[\left| \mathbb{E}[Z_1] - \bar{Z}_n \right| \geq \frac{\mathbb{E}[Z_1]}{2} \right]}.$$

We apply Markov inequality then Rosenthal inequality

$$\mathbb{E} \left[\left| \mathbb{E}[Z_1] - \bar{Z}_n \right| \mathbf{1}_{\Lambda^c} \right] \leq \frac{\sqrt{\text{Var}[Z_1]}}{\sqrt{n}} \sqrt{\frac{\mathbb{E} \left[\left| \mathbb{E}[Z_1] - \bar{Z}_n \right|^2 \right]}{\mathbb{E}[Z_1]^2}} \leq \frac{\sqrt{\text{Var}[Z_1]}}{\mathbb{E}[Z_1]n}.$$

Moreover under Assumption **(B1)**

$$\mathbb{E} \left[\left\{ \text{pen}_2(\tilde{m}_2) - \widehat{\text{pen}}_2(\tilde{m}_2) \right\}_+ \right] \leq \frac{C}{n} \mathbb{E} \left[\varrho^2 \left(\mathbf{G}_{\tilde{m}_2}^{-1} \right) \frac{\log n}{n} \right] \leq \frac{C'}{n}$$

which ends the proof. \square

2.8 Appendix

Lemma 2.8.1. (*Talagrand's inequality*) Let Y_1, \dots, Y_n be i.i.d. variables and

$$r_n(f) = \frac{1}{n} \sum_{k=1}^n \left(f(Y_k) - \mathbb{E}[f(Y_k)] \right)$$

for f belonging to some countable set \mathcal{F} of uniformly bounded measurable functions. Then for $\xi^2 > 0$,

$$\mathbb{E} \left[\left\{ \sup_{f \in \mathcal{F}} |r_n(f)|^2 - 2(1 + 2\xi^2)H^2 \right\}_+ \right] \leq \frac{4}{K_1} \left(\frac{v}{n} e^{-K_1 \xi^2 \frac{nH^2}{v}} + \frac{98M_1^2}{K_1 n^2 C^2(\xi^2)} e^{-\frac{2K_1 C(\xi) \xi nH}{7\sqrt{2} M_1}} \right)$$

with constants $C(\xi^2) = (\sqrt{1 + \xi^2} - 1) \wedge 1$ and $K_1 = \frac{1}{6}$, M_1 , H and v are such that

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \leq M_1, \quad \mathbb{E} \left[\sup_{f \in \mathcal{F}} |r_n(f)| \right] \leq H, \quad \sup_{f \in \mathcal{F}} \text{Var}(f(Y_1)) \leq v.$$

3

Adaptive deconvolution of linear functionals on the nonnegative real line

Ce chapitre est une version modifiée de l'article *Adaptive deconvolution of linear functionals on the nonnegative real line* accepté pour publication à *Journal of Statistical Planning and Inference*.

RÉSUMÉ. Dans ce chapitre, nous étudions le modèle de convolution $Z = X + Y$ avec X de densité f , indépendante de Y , quand ces deux variables aléatoires sont positives, comme dans le Chapitre 2. Notre but est d'estimer des fonctionnelles linéaires de f telles que $\langle \psi, f \rangle$ pour une fonction connue ψ en supposant que la loi de Y est connue et seul Z est observé. Nous proposons un estimateur de $\langle \psi, f \rangle$ s'appuyant sur un estimateur par projection de f sur la base de Laguerre, puis donnons une majoration du risque quadratique intégré et déduisons les vitesses de convergence en fonction des régularités de f , g et ψ . Ensuite nous établissons une procédure de sélection de modèle inspirée de [Goldenshluger and Lepski \(2011\)](#) pour sélectionner un espace d'approximation pertinent. Cette méthodologie est adaptée pour l'estimation ponctuelle de f . Nous illustrons aussi les bonnes performances de cette nouvelle méthode grâce à des simulations. Nous testons une nouvelle approche pour choisir la constante de calibration de la procédure de Goldenshluger et Lepski en s'inspirant d'idées développées dans [Lacour and Massart \(2015\)](#).

ABSTRACT. In this chapter, the convolution model $Z = X + Y$ with X of unknown density f , independent of Y , when both random variables are nonnegative as in Chapter 2. Our goal is to estimate linear functionals of f such as $\langle \psi, f \rangle$ for a known function ψ assuming that the distribution of Y is known and only Z is observed. We propose an estimator of $\langle \psi, f \rangle$ based on a projection estimator of f on Laguerre spaces, present upper bounds on the quadratic risk and derive the rate of convergence in function of the smoothness of f , g and ψ . Then we propose a nonparametric data driven strategy, inspired [Goldenshluger and Lepski \(2011\)](#) method to select a relevant projection space. This methodology is then adapted to pointwise estimation of f . We illustrate the good performance of the new method through simulations. We also test a new approach for choosing the tuning parameter in Goldenshluger-Lepski data driven estimators following ideas developed in [Lacour and Massart \(2015\)](#).

Contents

3.1	Introduction	69
3.2	Statistical model and estimation procedure	71
3.2.1	Notations	71
3.2.2	Laguerre basis	71
3.2.3	Projection estimator of the linear functional	71
3.2.4	Upper bound on the mean squared error and rates of convergence	74
3.3	Model selection and adaptive estimation	78
3.4	Particular case of pointwise estimation	80
3.4.1	Theoretical results	80
3.4.2	Illustrations	82
3.5	Proofs	88
3.5.1	Proofs of Section 3.2	88
3.5.2	Proofs of Section 3.3	90
3.6	Appendix 1	97
3.7	Appendix 2	97

3.1 Introduction

In many experiments statisticians do not observe directly the variable of interest X ; instead they have at hand observations of Z , equal to the sum of X and another random variable Y . In various situations, Y can modelize a measurement error, and as such, is assumed to be symmetric or centered. But we can also, in reliability fields, observe the sum of the lifetimes of two components, the second one being well known. In survival analysis, X can be the time of infection of a disease and Y the incubation time, and this happens in the so called back calculation problems in AIDS research. In these last two cases, distributions of X and Y are \mathbb{R}^+ -supported. In this situation, Y is not considered as a noise but as an additional nuisance process. Indeed a noise distribution is assumed to be centered, which is not the case anymore. Thus we consider the following model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \quad (3.1)$$

where the X_i 's are independent identically distributed (i.i.d.) nonnegative random variables (r.v.) with unknown density f . The Y_i 's are also i.i.d. nonnegative variables with known density g . We denote by h the density of the Z_i 's. Moreover the X_i 's and the Y_i 's are assumed to be independent, they are not observed. Our goal is to estimate linear functionals of f defined by $\vartheta(f) = \langle \psi, f \rangle = \mathbb{E}[\psi(X_1)]$ with ψ a known function, from observations Z_1, \dots, Z_n .

Those assumptions imply that, in Model (3.1), $h(x) = (f \star g)(x)$ where $(\varphi \star \psi)(x) = \int \varphi(x-u)\psi(u) du$ denotes the convolution product. This setting matches convolution models which is a classical topic in nonparametric statistics. The problem of recovering the signal distribution f when it is observed with an additive noise with known error distribution on the real line, has been extensively studied, see for rates of convergence and their optimality for kernel estimators [Carroll and Hall \(1988\)](#) and [Fan \(1991\)](#), for wavelets strategy [Pensky and Vidakovic \(1999\)](#), for projection strategies with penalization [Comte et al. \(2006\)](#) and [Butucea and Tsybakov \(2008a,b\)](#) for the study of sharp asymptotic optimality; with an unknown error density, see [Neumann \(1997\)](#) for kernel estimator and minimax optimality, [Johannes \(2009\)](#) for minimax optimality under various regularity conditions on f , [Comte and Lacour \(2011\)](#) and [Kappus and Mabon \(2014\)](#) (which corresponds to Chapter 5) for projection strategies with penalization.

The problem of one-sided error in the convolution model has been first introduced by [Groeneboom and Wellner \(1992\)](#) under a constraint of monotonicity of the cumulative distribution function (c.d.f.). They concentrate on deriving nonparametric maximum likelihood estimators (NPMLE) of the c.d.f. Some particular cases have been tackled as uniform or exponential deconvolution by [Groeneboom and Jongbloed \(2003\)](#) and [Jongbloed \(1998\)](#) who propose NPMLE of the c.d.f. of the X_i 's, which have explicit expressions. For other cases [van Es et al. \(1998\)](#) circumvent the lack of explicit expression for the NPMLE by proposing an isotonic inverse estimator. In this paper, we do not use the approach of the NPLME.

Moreover Model (3.1) is also related to the field of mixture models, see [Roueff and Rydén \(2005\)](#) and [Rebafka and Roueff \(2015\)](#) who study in particular mixtures of Exponential and Gamma distributions, which are contained in our framework. These models play a major role in natural sciences phenomena of discharge or disexcitation as in radioactive decays, the electric discharge of a capacitor or the temperature difference between two objects.

On one hand the problem of estimating linear functionals in linear models has been widely studied especially in the setting of the white noise model, see [Cai and Low \(2003\)](#),

2005) and Laurent et al. (2008) for instance. On the other hand adaptive estimation of linear functionals has not been much studied in the convolution model. Butucea and Comte (2009), in the context of Model (3.1) when the variables X_i 's and Y_i 's are \mathbb{R} -supported, propose a general estimator of $\vartheta(f)$ using a spectral cut-off in the Fourier domain when the random variables are distributed on the real line. They apply it to the pointwise estimation of the density on the real line and prove that their losses in their adaptive procedure is optimal in the minimax sense. They do not prove it for their general estimator. They also apply their adaptive procedure to the pointwise Laplace transform estimation and the stochastic volatility model. Recently Pensky (2014) has improved their results by deriving minimax lower bounds for estimators of a general linear functional of the deconvolution density. The author even extends the techniques when ψ is not integrable or square integrable and considers the possibility that the vector of observations is sparse (i.e. has a lot of zeros).

In this paper, our goal is to establish a specific method for the estimation of linear functional when we know that the random variables are \mathbb{R}^+ -supported. In that way we can cite the work of Mabon (2014) (which corresponds to Chapter 2) who proposes a specific estimation for nonnegative variables in Model (3.1). The methodology, based on a penalized projection strategy in a Laguerre basis, allows to estimate the density and survival function of X . Moreover this work has showed that for certain classes of density, the Laguerre estimation in the convolution model (3.1) gives faster rates of convergence of the estimators than with the classical method based on Fourier inversion as obtained in Comte et al. (2006) for instance. In particular, it is verified for mixed Gamma distributions. The contribution of this paper is to extend the particular methodology developed for nonnegative variables to estimate linear functionals of f . Thanks to this new methodology, we can derive the pointwise estimation of the probability density function (p.d.f.) f , of the c.d.f of X and also its Laplace transform.

In Section 3.2, we explain how projection coefficient estimators in the Laguerre basis can be used to define estimators of linear functionals. Next we lead a theoretical study of estimators of $\vartheta(f)$ and derive upper bound on the quadratic risk and rates of convergence in function of the smoothness of f , g and ψ when these three functions belong to $\mathbb{L}^2(\mathbb{R}^+)$. We show that under some assumptions on g or ψ , the parametric rate of estimation can be achieved. In Section 3.3 we propose a nonparametric data driven strategy, following Goldenshluger and Lepski (2011) method, for selecting a relevant projection space. In Section 3.4, when f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$ we adapt the procedure to pointwise estimation of f in the setting of Model (3.1) in this case ψ is specified and do not belong to $\mathbb{L}^2(\mathbb{R}^+)$ anymore. The method is then illustrated through simulations. Following Lacour and Massart (2015), we apply a new procedure to choose the tuning parameter appearing in the penalization term of the data-driven estimator. This procedure is promising and shows good results.

To sum up the paper is organized as follows. In Section 3.2, we give the notations, specify the statistical model and estimation procedures for projection estimators of f and $\vartheta(f)$, upper bound on the pointwise mean squared error. In Section 3.3, we propose a new data-driven procedure for choosing the tuning parameter linear functionals of f which allows in particular to estimate the cumulative distribution function of X . In Section 3.4, we derive an adaptive procedure for the pointwise estimation of f and provide an empirical study on simulations. All the proofs are postponed to Section 3.5.

3.2 Statistical model and estimation procedure

3.2.1 Notations

For two real numbers a and b , we denote $a \vee b = \max(a, b)$ and $a \wedge b = \min(a, b)$. For two functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ belonging to $\mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the \mathbb{L}^2 norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\psi(x)dx$.

Let d be an integer, for two vectors \vec{u} and \vec{v} belonging to \mathbb{R}^d , we denote $\|\vec{u}\|_2$ the Euclidean norm defined by $\|\vec{u}\|_2^2 = {}^t\vec{u}\vec{u}$ where ${}^t\vec{u}$ is the transpose of \vec{u} . The scalar product between \vec{u} and \vec{v} is $\langle \vec{u}, \vec{v} \rangle_2 = {}^t\vec{u}\vec{v} = {}^t\vec{v}\vec{u}$. We introduce the square of the spectral norm of a matrix $\mathbf{A} : \varrho^2(\mathbf{A}) = \lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ where $\lambda_{\max}({}^t\mathbf{A}\mathbf{A})$ is the largest eigenvalue of ${}^t\mathbf{A}\mathbf{A}$ in absolute value. We also introduce the following operator norm: $\|\mathbf{A}\|_1 = \max_{1 \leq j \leq d} \sum_{i=1}^d |a_{ij}|$.

3.2.2 Laguerre basis

We define the Laguerre basis as

$$\forall k \in \mathbb{N}, \forall x \geq 0, \quad \varphi_k(x) = \sqrt{2}L_k(2x)e^{-x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}. \quad (3.2)$$

The Laguerre polynomials L_k defined by Equation (3.2) are orthonormal with respect to the weight function $x \mapsto e^{-x}$ on \mathbb{R}^+ . In other words, $\int_{\mathbb{R}^+} L_k(x)L_{k'}(x)e^{-x} dx = \delta_{k,k'}$ where $\delta_{k,k'}$ is the Kronecker symbol. Thus $(\varphi_k)_{k \geq 0}$ is an orthonormal basis of $\mathbb{L}^2(\mathbb{R}^+)$. We can also notice that the Laguerre basis verifies the following inequality for any integer k

$$\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| = \|\varphi_k\|_{\infty} \leq \sqrt{2}. \quad (3.3)$$

We also introduce the space $\mathcal{S}_m = \text{Span}\{\varphi_0, \dots, \varphi_{m-1}\}$. For a function p in $\mathbb{L}^2(\mathbb{R}^+)$, we note

$$p(x) = \sum_{k \geq 0} a_k(p)\varphi_k(x) \quad \text{where} \quad a_k(p) = \int_{\mathbb{R}^+} p(u)\varphi_k(u) du.$$

According to formula 22.13.14 in [Abramowitz and Stegun \(1964\)](#), what makes the Laguerre basis relevant in our deconvolution setting is the relation

$$\varphi_k \star \varphi_j(x) = \int_0^x \varphi_k(u)\varphi_j(x-u) du = 2^{-1/2} (\varphi_{k+j}(x) - \varphi_{k+j+1}(x)) \quad (3.4)$$

where \star stands for the convolution product.

3.2.3 Projection estimator of the linear functional

Statistical procedure

The goal is to estimate $\vartheta(f) = \langle \psi, f \rangle = \mathbb{E}[\psi(X_1)]$ when only the Z_i 's are observed for $i = 1, \dots, n$. If f and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$ then they can be decomposed on the Laguerre basis, it yields that

$$\vartheta(f) = \langle \psi, f \rangle = \sum_{k=0}^{\infty} a_k(\psi)a_k(f).$$

Since ψ is known and the φ_k 's are known, coefficients $a_k(\psi)$ can be explicitly computed. Therefore we have to define estimators \hat{a}_k of $a_k(f)$ but an estimator of $\vartheta(f)$ would not be

$\widehat{\vartheta}(f) = \sum_{k=0}^{\infty} a_k(\psi) \widehat{a}_k$. Indeed the last sum is not necessarily convergent nor computable in practice, so we need to truncate it. Let us consider $\vartheta_m(f)$ defined by

$$\vartheta_m = \vartheta_m(f) = \sum_{k=0}^{m-1} a_k(\psi) a_k(f).$$

Let us define ψ_m and f_m respectively the projection of ψ and f on the space \mathcal{S}_m . It yields that

$$\langle \psi_m, f_m \rangle = \vartheta_m = \langle \psi, f_m \rangle = \langle \psi_m, f \rangle.$$

Then we can estimate ϑ_m by

$$\widehat{\vartheta}_m = \vartheta(\widehat{f}_m) = \langle \psi_m, \widehat{f}_m \rangle = \sum_{k=0}^{m-1} a_k(\psi) \widehat{a}_k, \quad (3.5)$$

with $\widehat{f}_m = \sum_{k=0}^{m-1} \widehat{a}_k \varphi_k$. The coefficients \widehat{a}_k are computed as follows

$${}^t(\widehat{a}_0 \dots \widehat{a}_{m-1}) := \widehat{f}_m = \mathbf{G}_m^{-1} \widehat{h}_m, \quad \text{with} \quad \widehat{h}_m = {}^t(\widehat{a}_0(Z) \dots \widehat{a}_{m-1}(Z)) \quad (3.6)$$

where $\widehat{a}_k(Z) = (1/n) \sum_{i=1}^n \varphi_k(Z_i)$ for $k = 0, \dots, m-1$ are estimators of $a_k(h)$ and

$$[\mathbf{G}_m] := \left([\mathbf{G}_m]_{i,j} \right)_{0 \leq i,j \leq m-1} = \begin{cases} 2^{-1/2} a_0(g) & \text{if } i = j, \\ 2^{-1/2} (a_{i-j}(g) - a_{i-j-1}(g)) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \quad (3.7)$$

We know from Chapter 2 that formula (3.4) implies that $\widehat{\vartheta}_m$ is an unbiased estimator of ϑ_m , see details therein, i.e. $\mathbb{E} \widehat{\vartheta}_m = \vartheta_m$.

Before deriving the statistical properties of our estimator let us give some examples to show that the method is feasible. Since the vector \widehat{h}_m is computed directly from the data, let us concentrate on the matrix \mathbf{G}_m and derive, if possible, an explicit expression of its coefficients.

Example 1. If g is an Exponential distribution of parameter $\lambda > 0$, then for all integer k we have

$$\begin{aligned} a_k(g) &= \int_{\mathbb{R}^+} g(u) \varphi_k(u) \, du = \sqrt{2} \lambda \sum_{j=0}^k \binom{k}{j} \frac{(-1)^j}{j!} \int_{\mathbb{R}^+} (2u)^j e^{-(1+\lambda)u} \, du \\ &= \frac{\sqrt{2} \lambda}{\lambda + 1} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{(1+\lambda)^j} = \frac{\sqrt{2} \lambda}{\lambda + 1} \frac{(\lambda - 1)^k}{(\lambda + 1)^k}. \end{aligned}$$

We can compute the coefficients of the matrix as

$$a_0(g) = \frac{\sqrt{2} \lambda}{\lambda + 1} \quad \text{and} \quad a_{i-j}(g) - a_{i-j-1}(g) = -2\sqrt{2} \lambda \frac{(\lambda - 1)^{i-j-1}}{(\lambda + 1)^{i-j+1}} \quad \text{if } j < i.$$

In that case, we can also inverse the matrix as follows

$$[\mathbf{G}_m^{-1}]_{i,j} = \frac{\lambda + 1}{\lambda} \begin{cases} 1 & \text{if } i = j, \\ 2 & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \quad (3.8)$$

Thus we can explicitly compute the coefficients \hat{a}_k with formula (3.6) and $\hat{\vartheta}_m$ is easy to obtain by (3.5). Extension to the case of a Gamma distribution of parameters p an integer, $p \geq 1$ and $\lambda > 0$, can be obtained by using the more general formula

$$\begin{aligned} a_k(g) &= \int_{\mathbb{R}^+} \lambda^p u^{p-1} e^{-\lambda u} \sqrt{2} e^{-u} L_k(2u) du = \sqrt{2} \lambda^p \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{j!} \int_{\mathbb{R}^+} u^{p+j-1} e^{-(1+\lambda)u} du \\ &= \frac{\sqrt{2} \lambda^p}{(1+\lambda)^p} \sum_{j=0}^k \binom{k}{j} \frac{(-2)^j}{(1+\lambda)^j} \frac{(p+j-1)!}{j!} = \frac{\sqrt{2} \lambda^p}{(1+\lambda)^p} S_{p-1,k} \left(\frac{2}{1+\lambda} \right). \end{aligned} \quad (3.9)$$

with

$$S_{p-1,k}(x) = \frac{1}{(p-1)!} \frac{d^{p-1}}{dx^{p-1}} \left[x^{p-1} (1-x)^k \right].$$

Discussion: noticeable linear functionals

We propose to point out some interesting linear functionals. Some can be directly obtained with the previous methodology since they can be written as an inner product of f with a square integrable function. Others do not fit this particular assumption but can nonetheless be estimated. As said in the introduction, in Chapter 2 we study the problem of global estimation in Model (3.1) and proposes adaptive estimators of the probability density and survival functions. Thanks to the particular procedure for linear functionals, we can derive the pointwise estimation of the c.d.f., p.d.f and Laplace transform.

a) Cumulative distribution function.

Let us define $\psi(x) = \mathbb{1}_{[0,c]}(x)$ for $c > 0$. The function ψ belongs to $\mathbb{L}^2(\mathbb{R}^+)$. This definition of ψ enables us to consider the estimation of the c.d.f. of f denoted by F . Indeed we can write the c.d.f. as

$$F(c) = \int_0^c f(u) du = \int_0^\infty \mathbb{1}_{[0,c]}(u) f(u) du = \langle \psi, f \rangle.$$

The methodology of Subsection 3.2.3 can therefore be applied.

b) Probability density function.

Let x_0 be a nonnegative real number, we have

$$f(x_0) = \int_0^\infty f(u) \delta_{\{x_0\}}(u) du = \langle \delta_{\{x_0\}}(\cdot), f \rangle = \langle \psi, f \rangle,$$

with $\psi(x) = \delta_{\{x_0\}}(x)$ and δ stands for the Dirac measure. Obviously in that case ψ is not square integrable. This case is studied in detail in Section 3.4.

c) Laplace transform.

Let t be a positive real number and consider the function $\psi_t(x) = e^{-tx}$ which is square integrable, we can write the Laplace transform of X as follows

$$\begin{aligned} \mathbb{E}e^{-tX} &= \int_0^\infty e^{-tu} f(u) du = \int_0^\infty e^{-tu} \sum_{k \geq 0} a_k(f) \varphi_k(u) du \\ &= \sum_{k \geq 0} a_k(f) \int_0^\infty e^{-tu} \varphi_k(u) du = \sum_{k \geq 0} a_k(f) a_k(\psi_t). \end{aligned}$$

We can apply the procedure of Subsection 3.2.3. Moreover we explain in Subsection 3.2.4 how to estimate the Laplace transform with a parametric rate.

3.2.4 Upper bound on the mean squared error and rates of convergence

Risk bound

Let us notice that if $f, g \in \mathbb{L}^2(\mathbb{R}^+)$ then h is bounded. Indeed

$$\|h\|_\infty = \sup_{x \geq 0} |f \star g(x)| \leq \|f\| \|g\|.$$

Then we can state the following Proposition:

Proposition 3.2.1. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, for \mathbf{G}_m defined by Equation (3.7) and $\hat{\vartheta}_m$ defined by Equation (3.5), the following result holds*

$$\mathbb{E} \left[(\vartheta - \hat{\vartheta}_m)^2 \right] \leq (\vartheta - \vartheta_m)^2 + \|h\|_\infty \frac{\|\mathbf{t}\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2}{n}. \quad (3.10)$$

The terms on the right-hand side of Equation (3.10) correspond to a squared bias variance decomposition. We can notice by applying the Cauchy-Schwarz inequality that

$$(\vartheta - \vartheta_m)^2 = \left(\sum_{k \geq m} a_k(\psi) a_k(f) \right)^2 \leq \|\psi - \psi_m\|^2 \|f - f_m\|^2 = \sum_{k \geq m} a_k^2(\psi) \sum_{k \geq m} a_k^2(f). \quad (3.11)$$

Therefore the first term gets smaller when m increases. On the contrary noticing that

$$\|\mathbf{t}\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq \|\psi\|^2 \varrho^2(\mathbf{G}_m^{-1}),$$

we obtain a bound on the variance which is nondecreasing with m (see Lemma 2.3.4 in Chapter 2). In that case we have to tackle the usual bias variance trade-off.

We can note that if there exist some function ψ and/or some density function g such that $\|\mathbf{t}\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2$ is bounded by a constant not depending on m , then we can estimate the resulting linear functionals with a parametric rate. First let us consider the general rates of convergence.

General rates of convergence

In order to derive general rates of convergence of estimators $\hat{\vartheta}_m$ defined by Equation (3.5), we need to evaluate the smoothness of the signal along with the order of $\varrho^2(\mathbf{G}_m^{-1})$. In the first place, we assume that f belongs to a Laguerre-Sobolev space defined as

$$W^s(\mathbb{R}^+, L) = \left\{ p : \mathbb{R}^+ \rightarrow \mathbb{R}, p \in \mathbb{L}^2(\mathbb{R}^+), \sum_{k \geq 0} k^s a_k^2(p) \leq L < +\infty \right\} \quad \text{with } s \geq 0 \quad (3.12)$$

where we recall that $a_k(p) = \langle p, \varphi_k \rangle$. Bongioanni and Torrea (2009) have introduced Laguerre-Sobolev spaces but the link with the coefficients of a function on a Laguerre basis was done by Comte and Genon-Catalot (2015) when s is an integer. Indeed, let s be an integer, $f \in W^s(\mathbb{R}^+, L)$ is equivalent to the fact that f admits derivatives up to order $s-1$ with $f^{(s-1)}$ absolutely continuous and for $0 \leq k \leq s-1$, $x^{(k+1)/2} \sum_{j=0}^{k+1} \binom{k+1}{j} f^{(j)}(x) \in \mathbb{L}^2(\mathbb{R}^+)$.

For more details we refer to section 7 of [Comte and Genon-Catalot \(2015\)](#).

Thanks to those spaces we can evaluate the bias order. Let $f \in W^s(\mathbb{R}^+, L)$ and $\psi \in W^\alpha(\mathbb{R}^+, L')$ with $\alpha, s > 1$, starting from Equation (3.11) we have

$$(\vartheta - \vartheta_m)^2 \leq \sum_{k \geq m} a_k^2(\psi) k^{-\alpha} k^\alpha \sum_{k \geq m} a_k^2(f) k^{-s} k^s \leq LL' m^{-(s+\alpha)}.$$

Then we must evaluate the variance term of Equation (3.10) which means assess the order of $\varrho^2(\mathbf{G}_m^{-1})$. [Comte et al. \(2015a\)](#) show that under the following conditions on the density g , we can recover the order of the spectral norm of \mathbf{G}_m^{-1} . First we define an integer $r \geq 1$ such that

$$\frac{d^j}{dx^j} g(x) \Big|_{x=0} = \begin{cases} 0 & \text{if } j = 0, 1, \dots, r-2 \\ B_r \neq 0 & \text{if } j = r-1. \end{cases}$$

And we make the two following assumptions:

- (C1) $g \in \mathbb{L}^1(\mathbb{R}^+)$ is r times differentiable and $g^{(r)} \in \mathbb{L}^1(\mathbb{R}^+)$.
(C2) The Laplace transform of g , defined by $G(z) = \mathbb{E}[e^{-zY}]$, has no zero with nonnegative real parts except for the zeros of the form $\infty + ib$.

For instance if Y follows a Gamma distribution of parameters p and λ , then its density g verifies these three conditions for $r = p$. Especially an Exponential distribution satisfies those assumptions for $r = 1$. According to [Comte et al. \(2015a\)](#), under Assumptions (C1)-(C2) there exist some positive constants C_ϱ and C'_ϱ such that

$$C'_\varrho m^{2r} \leq \varrho^2(\mathbf{G}_m^{-1}) \leq C_\varrho m^{2r}.$$

Then the order of the variance can be derived as follows

$$\|\mathfrak{t}\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq \|\mathfrak{t}\vec{\psi}_m\|_2^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \|\psi\|^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \|\psi\|^2 C_\varrho m^{2r}. \quad (3.13)$$

Proposition 3.2.2. *Assume that $\psi \in W^\alpha(\mathbb{R}^+, L')$ and $f \in W^s(\mathbb{R}^+, L)$ with $\alpha, s > 1$, that g satisfies Assumptions (C1)-(C2) and let $m_{opt} \propto n^{1/(s+\alpha+2r)}$, then there exists a positive constant c_1 depending on $L, L', s, r, \alpha, \|\psi\|, C_\varrho$ such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \left[(\vartheta - \hat{\vartheta}_{m_{opt}})^2 \right] \leq c_1 n^{-(s+\alpha)/(s+\alpha+2r)}.$$

Thus the choice $m = m_{opt}$ enables us to compute the rate of convergence of the estimator. However this choice depends on the regularity of the unknown function f and cannot be used in practice. This is why we will look for another way of performing the compromise between the squared bias and the variance.

Almost parametric rate: mixed Gamma densities

We want to point out that the Laguerre method enables us to obtain fast rates of convergence on the class of mixed Gamma densities noted $\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$ and defined by

$$\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda}) = \left\{ f = \sum_{i=1}^q \alpha_i \gamma_i \geq 0, \gamma_i \sim \Gamma(p_i, \lambda_i), \lambda_i > 0, p_i \in \mathbb{N}, \right. \\ \left. q \in \mathbb{N} \cup \{\infty\}, \alpha_i \in \mathbb{R}, \sum_{i=1}^q |\alpha_i| < \infty, \sum_{i=1}^q \alpha_i = 1 \right\}, \quad (3.14)$$

Let us assume that f belongs to $\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$. We obtain from (3.9) (see detail in Chapter 2):

$$\begin{aligned} (\vartheta - \vartheta_m)^2 &= \left(\sum_{k \geq m} a_k(\psi) \sum_{i=1}^q \alpha_i a_k(\gamma_i) \right)^2 \leq \sum_{k \geq m} a_k^2(\psi) \sum_{k \geq m} \sum_{i=1}^q |\alpha_i| a_k^2(\gamma_i) \\ &\leq \|\psi\|^2 \sum_{i=1}^q |\alpha_i| C(p_i, \lambda_i) \left(\frac{\lambda_i - 1}{\lambda_i + 1} \right)^{2m} m^{2(p_i-1)} \leq C \|\psi\|^2 m^{2(p^*-1)} \rho^{2m}. \end{aligned}$$

with $\rho = \max_i |(\lambda_i - 1)/(\lambda_i + 1)| \in (0, 1)$ and $p^* = \max_i p_i$. Thus the squared bias decays exponentially. And if we assume that g verifies (C1)-(C2) according to Equation (3.13), the order of the variance is less than $C_\varrho \|\psi\|^2 m^{2r}$. Therefore we obtain the following result

Proposition 3.2.3. *Assume that $f \in \mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})$, that g satisfies Assumptions (C1)-(C2) and let $m_{opt} = c \log n / |\log \rho|$ with $c \geq 1$, then there exists a positive constant c_2 depending on $q, \vec{\alpha}, \vec{p}, \vec{\lambda}, \alpha, \|\psi\|, C_\varrho$ such that*

$$\sup_{f \in \mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\lambda})} \mathbb{E} \left[(\vartheta - \hat{\vartheta}_{m_{opt}})^2 \right] \leq c_2 \frac{(\log n)^{2r}}{n}.$$

So for the classes of mixed Gamma densities, the Laguerre procedure enables us to obtain almost parametric rates of convergence, which is remarkable in the deconvolution model. As previously, the choice m_{opt} cannot be performed in practice, since f is unknown.

Parametric rate

In this section, we propose to find conditions ensuring that the parametric rate can be achieved.

Proposition 3.2.4. *If $\psi(x) = \alpha e^{-x}$ with $\alpha \in \mathbb{R}$, then*

$$\mathbb{E} \left[(\vartheta - \hat{\vartheta}_1)^2 \right] = \frac{\alpha^2}{a_0^2(g)} \frac{\text{Var}[\varphi_0(Z_1)]}{n}. \quad (3.15)$$

This result is interesting in the sense that it does not require any additional assumptions on the density g . It allows us to recover the Laplace transform of X at the point $t = 1$ with a parametric rate. Note that, as Model (3.1) implies that $\mathbb{E}[e^{-Z_i}] = \mathbb{E}[e^{-X_i}] \mathbb{E}[e^{-Y_i}]$ and the distribution of Y is assumed to be known, we can estimate the Laplace transform with a parametric rate with $(1/n) \sum_{i=1}^n e^{-X_i} / \mathbb{E}[e^{-Y_1}]$, without knowing anything about Laguerre basis. This formula shows that the Laplace transform can be estimated at a parametric rate at any point $t > 0$. It can also be achieved with the Laguerre basis but the procedure is a little bit more complicated than with a plug-in estimator.

In fact, the Laguerre basis can be parametrized. In the present work for sake of clarity we consider that this scale parameter is equal to 1. More generally we can write the basis as follows

$$\forall k \in \mathbb{N}, \forall x \geq 0, \quad \varphi_k^\tau(x) = \sqrt{2\tau} L_k(2\tau x) e^{-\tau x} \quad \text{with } \tau > 0.$$

For $\tau > 0$, if we take the Laguerre basis $(\varphi_k^\tau)_{k \geq 0}$ and proceed exactly as in the proof of Proposition 3.2.4, we find that the functions proportional to $\exp(-\tau x)$ satisfy Equation (3.15) for the corresponding scale parameter τ of the basis. The extension is straightforward. So for $\psi(x) = \alpha \exp(-\tau x)$, the parametric rate is still achieved. We can use this methodology to estimate the Laplace transform of X at any point.

The following two results are more elaborate.

Proposition 3.2.5. *If $\sum_{j \geq 2} |a_{j-1}(g) - a_{j-2}(g)| / a_0(g) < 1$ then $\varrho(\mathbf{G}_m^{-1})$ is upper bounded by a constant independent of m and the parametric rate is achieved.*

We can notice that Exponential distributions with parameter $\lambda = 1$ satisfy this Proposition. From Example 1, we have the coefficients of $g \sim \mathcal{E}(\lambda)$, which yields for $\lambda = 1$

$$\sum_{j \geq 2} \frac{|a_{j-1}(g) - a_{j-2}(g)|}{a_0(g)} = 0,$$

and for $\lambda \neq 1$

$$\begin{aligned} \sum_{j \geq 2} \frac{|a_{j-1}(g) - a_{j-2}(g)|}{a_0(g)} &= \sum_{j \geq 2} 2 \frac{|\lambda - 1|^{j-2}}{|\lambda + 1|^{j-1}} = \frac{2}{\lambda + 1} \sum_{j \geq 2} \frac{|\lambda - 1|^{j-2}}{|\lambda + 1|^{j-2}} \\ &= \frac{2}{\lambda + 1} \frac{1}{1 - (\lambda - 1)/(\lambda + 1)} = 1. \end{aligned}$$

Proposition 3.2.6. *If $g \sim \mathcal{E}(\lambda)$ with $\lambda > 0$ and $\sum_{k=0}^{m-1} (\sum_{j=k}^{m-1} |a_j(\psi)|)^2 < \infty$, then there exists a positive constant C depending on λ such that*

$$\mathbb{E} \left[(\vartheta - \hat{\vartheta}_m)^2 \right] \leq (\vartheta - \vartheta_m)^2 + \frac{C}{n}.$$

Propositions 3.2.5 and 3.2.6 give conditions ensuring that the variance can be upper bounded by C/n whatever the value of m . Then to derive a parametric rate in this case we need to choose m_n large enough such that $(\vartheta - \vartheta_{m_n})^2 = \mathcal{O}(n^{-1})$.

Let us give some examples of functions ψ to illustrate Proposition 3.2.6.

Example 2. Let us consider here that ψ is the density $\Gamma(p, \mu)$. According to Equation (3.9) we have

$$a_j(\psi) = \frac{\sqrt{2}\mu^p}{(1 + \mu)^p} S_{p-1,j} \left(\frac{2}{1 + \mu} \right)$$

The term $S_{p-1,j}(x)$ has the following order for in $x \in [0, 1]$

$$|S_{p-1,j}(x)| \leq 2^p |1 - x|^{j-p+1}, \quad \text{for } j \geq p - 1$$

(see detail in Appendix 3.6), it yields

$$\begin{aligned} \sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 &\leq \frac{2^{2p+1} \mu^{2p}}{(\mu + 1)^{2p}} \sum_{k=p-1}^{m-1} \left(\sum_{j=k}^{m-1} \left| \frac{\mu - 1}{\mu + 1} \right|^{j-p+1} \right)^2 \\ &\leq \frac{2^{2p+1} \mu^{2p}}{(\mu + 1)^{2p}} \sum_{k=p-1}^{m-1} \left| \frac{\mu - 1}{\mu + 1} \right|^{2k} \left(\sum_{\ell=0}^{m-1-k} \left| \frac{\mu - 1}{\mu + 1} \right|^\ell \right)^2 \\ &\leq \frac{2^{2p+1} \mu^{2p}}{(\mu + 1)^{2p}} \frac{(\mu + 1)^4}{2\mu(\mu + 1 - |\mu - 1|)^2} \end{aligned}$$

Thus condition $\sum_{k=0}^{m-1} (\sum_{j=k}^{m-1} |a_j(\psi)|)^2 < \infty$ of Proposition 3.2.6 is fulfilled. Therefore we conclude that if g is $\mathcal{E}(\lambda)$ and ψ a finite mixture of Gamma distributions, then the parametric rate is achieved.

Example 3. Let α an integer such that $\alpha > 1$ and assume $\psi \in W^\alpha(\mathbb{R}^+, L')$, it yields

$$\left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 = \left(\sum_{j=k}^{m-1} j^{\alpha/2} |a_j(\psi)| j^{-\alpha/2} \right)^2 \leq \sum_{j=k}^{m-1} a_j^2(\psi) j^\alpha \sum_{j=k}^{m-1} j^{-\alpha} \leq L' \frac{\alpha k^{-\alpha+1}}{\alpha-1},$$

then

$$\sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2 \leq \frac{L'\alpha}{\alpha-1} \sum_{k=0}^{m-1} k^{-(\alpha-1)} < \infty \quad \text{if } \alpha > 1.$$

In conclusion any function ψ such that $\sum_{k \geq 0} k^\alpha a_k^2(\psi) \leq L' < +\infty$ with $\alpha > 1$ satisfies the assumption of Proposition 3.2.6.

For instance $\psi(x) = (1+x)^{-\beta} \mathbf{1}_{x \geq 0}$ with $\beta > (\alpha+1)/2$ belongs to $W^\alpha(\mathbb{R}^+, L')$ (see detail in Appendix 3.7).

3.3 Model selection and adaptive estimation

To build data-driven estimators of linear functionals of the density f , we propose a new selection strategy in the spirit of Goldenshluger and Lepski (2011). First we add the following assumption on the model collection:

$$\text{(A)} \quad \mathcal{M} = \left\{ m \in \{1, \dots, n\}, m \leq n/\log n \text{ and } \max_{1 \leq k \leq m} \|\mathbf{t}\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \leq n \right\}.$$

The model selection is done by minimizing the following quantity

$$\hat{m} = \arg \min_{m \in \mathcal{M}} \{A(m) + V(m)\} \quad (3.16)$$

where $V(m)$ has the order of the variance in the right-hand side of Equation (3.10) and is defined by

$$V(m) = \kappa(\|h\|_\infty \vee 1) \max_{1 \leq k \leq m} \frac{\|\mathbf{t}\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \log n}{n},$$

with κ a numerical constant which will be chosen later (see comment below). The term of variance is slightly modified to ensure that V is nondecreasing with m . The term $A(m)$ has the order of the bias and is based on the comparison of the estimators built in the previous section as follows

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+.$$

We can now state the following oracle inequality:

Theorem 3.3.1. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption (A) is true. Let $\hat{\vartheta}_{\hat{m}}$ be defined by Equations (3.5) and (3.16). Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that*

$$\mathbb{E} \left[(\vartheta - \hat{\vartheta}_{\hat{m}})^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + V(m) \right\} + \frac{C}{n}. \quad (3.17)$$

where C^{ad} is a positive numerical constant and C depends only on $\|h\|_\infty$.

The oracle inequality (3.17) establishes a non asymptotic oracle bound. It shows that the squared bias variance tradeoff is automatically made up to a loss of logarithmic order and a multiplicative constant. Let us notice that $C^{ad} = 7$ and $\kappa_0 = 384$ would suit. Let us comment the specific squared bias term which differs from the squared bias established in Proposition 3.2.1. The term $\sup_{k>m} \{(\vartheta_k - \vartheta_m)^2\}$ is defined such that it is non-decreasing in m and we take for convention $\vartheta_\infty = \vartheta$. Usually in Lepski's method, it is assumed that the p.d.f. belongs to some regularity space indexed by β and the squared bias is expressed as a function of β and m which is non-decreasing in m . Nevertheless it is a strong assumption since we do not know the coefficient β depending on f . Our approach enables us to circumvent this assumption.

Theorem 3.3.1 is derived under mild assumptions. Indeed we do not need to assume that ψ or f belong to some semi-parametric space to derive the theorem. Concerning the model selection procedure, it is simpler in its formulation than those of Laurent et al. (2008) and Butucea and Comte (2009) thanks to the Goldenshluger and Lepski (2011) method. Butucea and Comte (2009) obtain as a consequence of their general adaptive procedure the pointwise estimation of the density on the real line, the stochastic volatility model and the pointwise Laplace transform estimation when X is a positive random variable, even when the Laplace transform of the noise is infinite. In the next section, we also show that a procedure for the pointwise estimation on the nonnegative real line is possible. A direct application of Theorem 3.3.1 is the estimation of the c.d.f. of X as pointed out in Subsection 3.2.4, which is a novelty compared to Butucea and Comte (2009). Pensky (2014) also develops an adaptive procedure based on Lepski's method. This paper clearly improves previous papers and covers a lot of cases (c.d.f, pointwise, moments) for which lower bounds are provided.

Some comments for practical use are in order. Indeed in the variance term $V(m)$, there are two quantities which deserve some explanations: κ and $\|h\|_\infty$. First note that when κ gets larger, so do $V(m)$ and thus the bound (3.17). It follows from the proof that $\kappa = 384$ 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. There is still an unknown term in the penalty, $\|h\|_\infty$, that needs to be estimated. We have to check that we can derive an oracle inequality when this term is estimated, which is done in the following Corollary.

Beforehand let

$$\hat{h}_D(x) = \sum_{k=0}^{D-1} \hat{a}_k(Z) \varphi_k(x), \quad (3.18)$$

with $\hat{a}_k(Z) = (1/n) \sum_{i=1}^n \varphi_k(Z_i)$. We can see that \hat{h}_D is an unbiased estimator of $h_D(x) = \sum_{k=0}^{D-1} a_k(h) \varphi_k(x)$.

Corollary 3.3.2. *For f, g and $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption (A) is true. Let $\hat{\vartheta}_{\tilde{m}}$ be defined by (3.5) and*

$$\tilde{m} = \arg \min_{m \in \mathcal{M}} \left\{ \tilde{A}(m) + \tilde{V}(m) \right\}$$

with

$$\begin{aligned} \tilde{A}(m) &= \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - \tilde{V}(m') \right\}_+ \\ \tilde{V}(m) &= \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \|\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n} \end{aligned}$$

where \hat{h}_D is given by (3.18) and D satisfies $\log n \leq D \leq \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^3}$. Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that

$$\mathbb{E} \left[(\vartheta - \hat{\vartheta}_m^{\sim})^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} \{(\vartheta_k - \vartheta_m)^2\} + V(m) \right\} + \frac{C}{n}.$$

where C^{ad} is a positive numerical constant and C depends only on $\|h\|_\infty$.

Note that the constraint on D is fulfilled for n large enough as soon as $D \simeq \sqrt{n}$ for instance. In this sense Corollary 3.3.2 has rather an asymptotic flavor.

3.4 Particular case of pointwise estimation

3.4.1 Theoretical results

For pointwise estimation of f , we take $\psi(x) = \delta_{\{x_0\}}(x)$ for any nonnegative x_0 , where δ stands for the Dirac measure. It yields that $\vartheta = f(x_0)$. In this case Equation (3.5) becomes

$$\hat{f}_m(x_0) = \sum_{k=0}^{m-1} \hat{a}_k \varphi_k(x_0), \quad (3.19)$$

since $a_k(\psi) = \langle \delta_{\{x_0\}}, \varphi_k \rangle = \varphi_k(x_0)$. We can notice that ψ is not square integrable.

Proposition 3.4.1. *For $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, for \mathbf{G}_m defined by Equation (3.7) and $\hat{f}_m(x_0)$ defined by Equation (3.19) for any nonnegative x_0 , the following result holds*

$$\mathbb{E} \left[(f(x_0) - \hat{f}_m(x_0))^2 \right] \leq (f(x_0) - f_m(x_0))^2 + \|h\|_\infty \frac{\|{}^t\vec{\varphi}_m(x_0) \mathbf{G}_m^{-1}\|_2^2}{n},$$

where ${}^t\vec{\varphi}_m(x_0) = (\varphi_0(x_0) \dots \varphi_{m-1}(x_0))$.

We can notice that the squared bias and the variance have not the same order as in the case where ψ is square integrable. Let us assume as in Section 3.2.4 that f belongs to $W^s(\mathbb{R}^+, L)$, applying Equation (3.3) and the Cauchy-Schwarz inequality we get the following order for the squared bias

$$\begin{aligned} (f(x_0) - f_m(x_0))^2 &= \left(\sum_{k \geq m} a_k(f) \varphi_k(x_0) \right)^2 \leq 2 \left(\sum_{k \geq m} |a_k(f)| \right)^2 = 2 \left(\sum_{k \geq m} |a_k(f)| k^{s/2} k^{-s/2} \right)^2 \\ &\leq 2 \sum_{k \geq m} |a_k(f)|^2 k^s \sum_{k \geq m} k^{-s} \leq \frac{2sL}{s-1} m^{-s+1} \end{aligned}$$

for $s > 1$. Moreover applying Equation (3.3), we get

$$\|{}^t\vec{\varphi}_m(x_0) \mathbf{G}_m^{-1}\|_2^2 \leq \|{}^t\vec{\varphi}_m(x_0)\|_2^2 \varrho^2(\mathbf{G}_m^{-1}) \leq \sum_{k=0}^{m-1} |\varphi_k(x_0)|^2 \varrho^2(\mathbf{G}_m^{-1}) \leq 2m\varrho^2(\mathbf{G}_m^{-1}). \quad (3.20)$$

Proposition 3.4.2. *Assume that $f \in W^s(\mathbb{R}^+, L)$ with $s > 1$, that Assumptions (C1)-(C2) are true and let $m_{opt} \propto n^{1/(s+2r)}$, then there exists a positive constant c_2 depending on L, s, r, C_ϱ such that*

$$\sup_{f \in W^s(\mathbb{R}^+, L)} \mathbb{E} \left[(f(x_0) - f_{m_{opt}}(x_0))^2 \right] \leq c_2 n^{-(s-1)/(s+2r)}.$$

Note that this rate is different from the global one: according to Chapter 2 for the global estimation of the density f with the \mathbb{L}^2 -risk, the rate of convergence is of order $n^{-s/(s+2r+1)}$. As in Section 3.2.4, we can point out that faster rates of convergence of order $\log n/n$ can be obtained for functions of Gamma type defined by Equation (3.14).

The data-driven strategy is done according to the one developed in Section 3.3 for the particular case of pointwise estimation, with the specific variance term according to this model. It yields that

$$\begin{aligned} \tilde{A}(m, x_0) &= \sup_{m' \in \mathcal{M}} \left\{ (\hat{f}_{m'}(x_0) - \hat{f}_{m' \wedge m}(x_0))^2 - \tilde{V}(m', x_0) \right\}_+ \quad \text{with} \\ \tilde{V}(m, x_0) &= \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \frac{\|\hat{\varphi}_k(x_0) \mathbf{G}_k^{-1}\|_2^2 \log n}{n} \end{aligned} \quad (3.21)$$

where κ is a numerical constant, \hat{h}_D is the estimator of the projection of h on \mathcal{S}_D given by (3.18) and D satisfies $\log n \leq D \leq \frac{\|h\|_\infty n}{128\sqrt{2}(\log n)^2}$ and

$$\tilde{m} = \arg \min_{m \in \mathcal{M}} \left\{ \tilde{A}(m, x_0) + \tilde{V}(m, x_0) \right\}. \quad (3.22)$$

Theorem 3.4.3. *For $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption (A) is true. Let $\hat{f}_{\tilde{m}}(x_0)$ be defined by Equations (3.19)-(3.22). Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that*

$$\mathbb{E} \left[(f(x_0) - \hat{f}_{\tilde{m}}(x_0))^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k > m \\ k \in \mathbb{N} \cup \{+\infty\}}} (f_k(x_0) - f_m(x_0))^2 + V(m, x_0) \right\} + \frac{C}{n}.$$

where C^{ad} is a positive numerical constant, C depends only on $\|h\|_\infty$ and $V(m, x_0) = \kappa(\|h\|_\infty \vee 1) \max_{1 \leq k \leq m} \|\hat{\varphi}_k(x_0) \mathbf{G}_k^{-1}\|_2^2 \log n/n$.

Applying Equation (3.3) note that

$$\sup_{k > m} |f_k(x_0) - f_m(x_0)|^2 = \sup_{k > m} \left| \sum_{j=m}^{k-1} a_j(f) \varphi_k(x_0) \right|^2 \leq 2 \left(\sum_{j \geq m} |a_j(f)| \right)^2$$

and using Equation (3.20) yields

$$V(m, x_0) \leq 2\kappa \|h\|_\infty m \varrho^2(\mathbf{G}_m^{-1}) \frac{\log n}{n},$$

since $m \mapsto \varrho^2(\mathbf{G}_m^{-1})$ is nondecreasing with m (see Lemma 2.3.4 in Chapter 2). These two bounds are independent of x_0 which proves that the risk, may tend uniformly to 0 for an adequate choice $m = m(n)$, when n tends to infinity.

3.4.2 Illustrations

The whole implementation is conducted using R software. The mean squared error (MSE) $\mathbb{E}(f(x_0) - \hat{f}_{\tilde{m}}(x_0))^2$ is computed as the empirical mean over 500 simulation samples.

Simulation setting

The performance of the procedure is studied for the three following distributions for X . Moreover all the densities are normalized with unit variance. Each density is computed over an interval I .

- ▷ Gamma distribution : $\Gamma(4, \frac{1}{2})$, $I = [0, 8]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [0, 13]$.
- ▷ Log-normal distribution : $\mathcal{LN}(1, 1)$, $I = [0, 15]$.

We then choose a Gamma distribution for the error distribution which verifies **(C1)**-**(C2)** for $r = 2$.

- ▷ Gamma noise: $\Gamma(2, \frac{1}{\sqrt{20}})$ and $\Gamma(2, \frac{1}{\sqrt{8}})$.

Thus the first Gamma distribution has a variance 1/10 and the second 1/4. We refer to Example 1 for the computation of the matrix \mathbf{G}_m .

We also compute estimators where there is no additional noise, i.e. $\sigma^2 = 0$. The procedure is applied with $\mathbf{G}_m = \mathbf{I}_m$.

Implementation of the estimation procedure

We consider the following model collection $\mathcal{M}_n = \{m \in \mathbb{N}, 0 \leq m \leq \lfloor n/\log n \rfloor\}$. The pointwise estimation at a point x_0 is computed according to the following adaptive procedure:

- ▷ For $m \in \mathcal{M}_n$, compute $\tilde{m} = \arg \min_{m \in \mathcal{M}_n} \{ \tilde{A}(m, x_0) + \tilde{V}(m, x_0) \}$, with
 - $\tilde{A}(m, x_0) = \sup_{m' \in \mathcal{M}_n} \{ (\hat{f}_{m'}(x_0) - \hat{f}_{m' \wedge m}(x_0))^2 - \tilde{V}(m', x_0) \}_+$
 - $\tilde{V}(m, x_0) = \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{0 \leq k \leq m-1} \| {}^t \tilde{\varphi}_k(x_0) \mathbf{G}_k^{-1} \|_2^2 \log n/n$ with \hat{h}_D defined by (3.18).
- ▷ Compute $\hat{f}_{\tilde{m}}(x_0) = \sum_{k=0}^{\tilde{m}-1} \hat{a}_k \varphi_k(x_0)$.

In addition we compute the integrated squared risk over an interval I of 300 equidistant points, by computing

$$\frac{\text{length}(I)}{300} \sum_{k=1}^{300} (f(x_k) - \hat{f}_{\tilde{m}_k}(x_k))^2$$

and $(x_k)_k$ a subdivision of I , which gives an approximation of the integrated risk for our method. Thus we can measure the performances of the pointwise estimation by comparing our method with the method of Chapter 2 who proposes adaptive estimators of the density

f associated with the \mathbb{L}^2 -risk.

One of the major issues in data-driven strategies is the choice of the constant κ . To detect this value in the model selection paradigm, we can use the slope heuristics for instance based on the works of [Birgé and Massart \(2007\)](#). [Baudry et al. \(2012\)](#) developed a method based on slope heuristics or on dimension jump. We cannot use those methods in the present work since we use Goldenshluger-Lepski methodology for the automatic selection of an estimator among a collection.

The results of [Goldenshluger and Lepski \(2011\)](#) ensure that if the variance term is large enough, the data-driven estimator is almost as efficient as the best one in the collection. In the proof we derive a theoretical value of κ defined by Equation (3.21). Nonetheless it is well known that theoretical constants in $\tilde{V}(m, x_0)$ are often too large in practice. Thus for practical implementation, the good behavior of the estimator is based on the calibration of the constant κ . Few studies have been conducted on the constant calibration in Goldenshluger-Lepski methodology to see if there exists a critical value such that, if $V(m) > V_0$ the risk is quasi optimal.

To find this value in the setting of Goldenshluger-Lepski method, we follow the recent work of [Lacour and Massart \(2015\)](#) who show that, for density estimation based on kernel estimators with Goldenshluger-Lepski type bandwidth selection, if the variance term is chosen too small then the procedure fails. They are the first to give a minimal penalty for Goldenshluger-Lepski methodology. In our case, we do not necessarily expect the same behavior of the variance term since we are in the particular case of the deconvolution model but we expect to find a critical value of κ which characterizes a failure of the procedure.

To detect such a value, we propose to observe the evolution of the choice \tilde{m} for different values of κ . It is clear that to challenge the global estimation, our pointwise procedure must be able to adapt to the point and to choose different \tilde{m} for different points.

To avoid over fitting, we choose two distributions that we will not estimate in the following: $\beta(4, 5) \times 5$ and $\chi^2(10)/\sqrt{20}$. The first one has variance $50/81$ and has a compact support $[0, 5]$. The other one is normalized with unit variance and its support is \mathbb{R}^+ . Only here in our procedure, we replace $\tilde{V}(m)$ by its uniform bound $2\kappa m \|\hat{h}_D\|_\infty \varrho^2(\mathbf{G}_m^{-1}) \log n/n$. Indeed with this approach it is easier to detect stable ranges where the choice of \tilde{m} does not change. Otherwise we observe stronger oscillations of the choice \tilde{m} and it is difficult to isolate the good values of κ from the bad ones.

In Figure 3.1, we apply this methodology for a wide range of κ and in Figure 3.2 we concentrate on smaller range of κ which corresponds to the neighborhood of the κ of the black thin line with circles. In Figure 3.1, we also draw the data driven choice for $\kappa = 0$ and $\kappa = 20$ which respectively corresponds to the bold black and the bold red lines. For $\kappa = 0$, the selected dimension is systematically the largest as possible while for $\kappa = 20$ it is mainly the smallest. We choose to concentrate on the black thin line with circles because it is the curve which moves the most and is the farthest from the two curves $\kappa = 0$ and $\kappa = 20$ which represent extreme situations. In Figure 3.2, we see that for both distributions the curves have the same behavior. Finally according to this methodology we choose a small value of $\kappa = 3 \cdot 10^{-4}$.

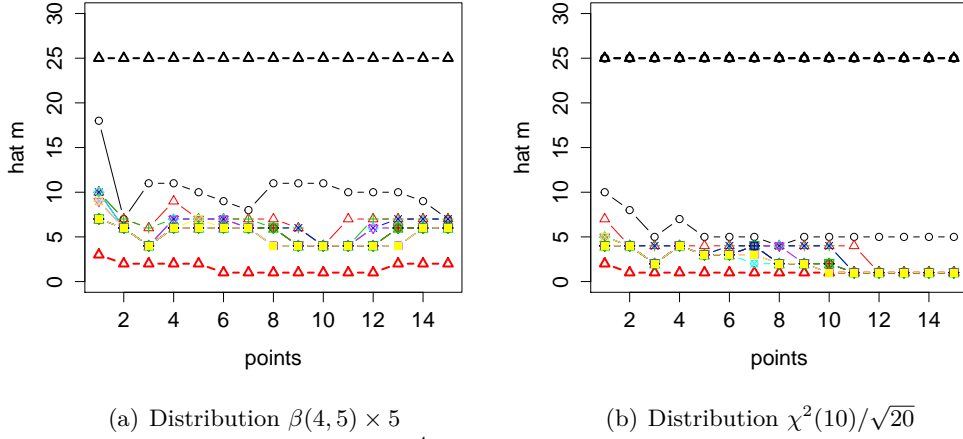


Figure 3.1: \hat{m} in function of $\kappa \in [10^{-4}, 10]$ (15 equidistant calibration constants, some are superposed) plus $\kappa = 0$ (top) and $\kappa = 20$ (bottom), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0, 5]$ for the β and $[0, 10]$ for the χ^2 .

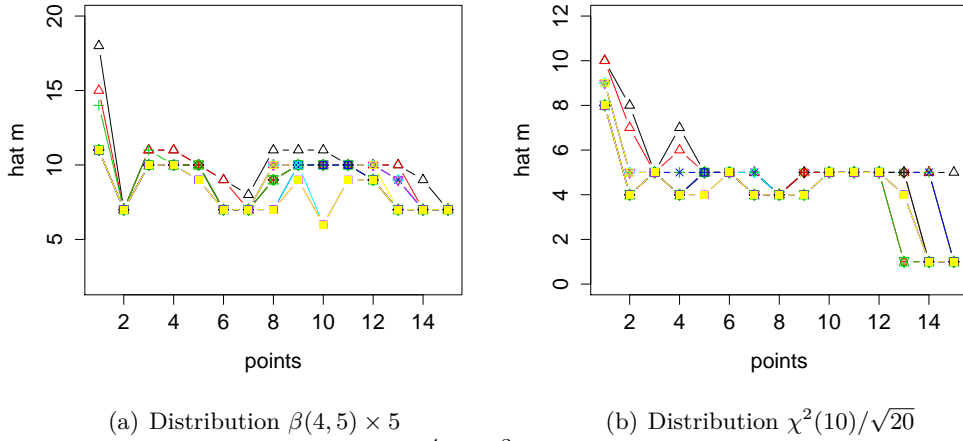


Figure 3.2: \hat{m} in function of $\kappa \in [10^{-4}, 10^{-3}]$ (15 equidistant calibration constants, some are superposed), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0, 5]$ for the β and $[0, 10]$ for the χ^2 .

In a second step we also propose to test another calibration suggested by [Lacour and Massart \(2015\)](#) in their concluding remark section. The idea is to consider

$$\bar{A}(m, x_0) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - \kappa(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \|\mathbf{t}\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|^2 \log n/n \right\}_+$$

$$\bar{m} = \arg \min_{m \in \mathcal{M}} \left\{ \bar{A}(m, x_0) + \kappa'(2\|\hat{h}_D\|_\infty \vee 1) \max_{1 \leq k \leq m} \|\mathbf{t}\vec{\varphi}_k(x_0)\mathbf{G}_k^{-1}\|_2^2 \log n/n \right\}.$$

Then, it is easy to prove that [Theorem 3.3.1](#) holds for $\hat{\vartheta}_{\bar{m}}$ as soon as $\kappa' \geq \kappa$. The authors suggest to find an adequate value of κ for $\kappa = \kappa'$ and then to put $\kappa' = 2\kappa$. This method is illustrated in the next section.

Simulation results

In Table 3.1 we report the results of the simulation (averaged values of the integrated squared errors multiplied by 100) for the global estimator computed according to Chapter 2. The results of the pointwise estimation are presented in Table 3.2. It reports averaged values of the approximation of the integrated risk for our method of pointwise estimation. First remarks: increasing the sample size improves the quality of the estimation and increasing the variance degrades the estimation but in an acceptable way for both methods.

If we compare the methods, we can see for the mixed Gamma distribution that the pointwise estimation is better than the global estimation. The risk is approximately divided by 2. For the Gamma and the log-normal, the estimation is clearly better with the global method where there is no additional noise. Otherwise the two methods are equivalent. We can note for the pointwise estimation that when the sample size increases, the results are very close to those without noise ($\sigma^2 = 0$).

In Table 3.2 we compute the risk associated to the oracles. We see that the oracles are very good, approximately divided by 30 compared to data driven estimators. It shows that the pointwise estimation could be excellent if the data driven strategy could perform adequately. Nonetheless in Section 3.3, we have seen that the bias variance compromise was made up to a logarithmic loss. It may also explain the gap between oracles and data driven procedure results. We also compare the quality of the estimation when $\kappa' = \kappa$ or $\kappa' = 2\kappa$. The results are always and slightly better when $\kappa' = \kappa$. Our examples do not permit us to highlight the idea of dissociation of the constant presented in Lacour and Massart (2015) for the Goldenshluger-Lepski strategy. Nonetheless this detection method of the constant κ leads to very good results.

We also illustrate the results with some figures. Figures 3.3-3.4 display the results of the data driven estimation respectively for the Gamma and the mixed Gamma. For each Figure we present the Goldenshluger-Lepski estimator with the calibration proposed by Lacour and Massart (2015) for $\kappa' = \kappa$. First we see that the risk decreases when the sample size increases. For the Gamma, we see that data driven estimators are very close to oracle estimators which is not really the case for the mixed Gamma especially for $n = 2000$. We can observe some oscillations near the origin. Thus the pointwise estimation does not seem to radically improve the estimation near the origin.

	$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
	200	2000	200	2000	200	2000
Gamma	0.313	0.045	0.625	0.037	0.752	0.289
Mixed Gamma	0.334	0.031	1.304	0.342	2.738	0.348
Log normal	0.347	0.043	0.684	0.129	0.829	0.246

Table 3.1: Results of simulation for the global estimation based on Chapter 2. MISE $\mathbb{E}(\|f - \hat{f}_m\|^2) \times 100$ averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$.

	$\sigma^2 = 0$		$\sigma^2 = \frac{1}{10}$		$\sigma^2 = \frac{1}{4}$	
	200	2000	200	2000	200	2000
<i>oracle</i>						
Gamma	0.064	0.004	0.016	0.008	0.056	0.016
Mixed Gamma	0.007	0.005	0.026	0.013	0.003	0.013
Log normal	0.090	0.0006	0.045	0.002	0.105	0.045
$\kappa' = 2\kappa$						
Gamma	0.464	0.080	0.504	0.064	1.176	0.144
Mixed Gamma	0.442	0.299	0.455	0.286	0.624	0.299
Log normal	0.390	0.165	0.480	0.165	0.960	0.180
$\kappa' = \kappa$						
Gamma	0.352	0.064	0.416	0.064	0.976	0.088
Mixed Gamma	0.429	0.338	0.442	0.286	0.572	0.286
Log normal	0.345	0.150	0.495	0.165	0.885	0.180

Table 3.2: Results of simulation for the pointwise estimation. Approximation of the integrated risk multiplied by 100 averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$. $oracle = \text{length}(I)/300 \sum_{k=1}^{300} \min_{m \in \mathcal{M}} \mathbb{E}|f(x_0) - f_m(x_0)|^2$

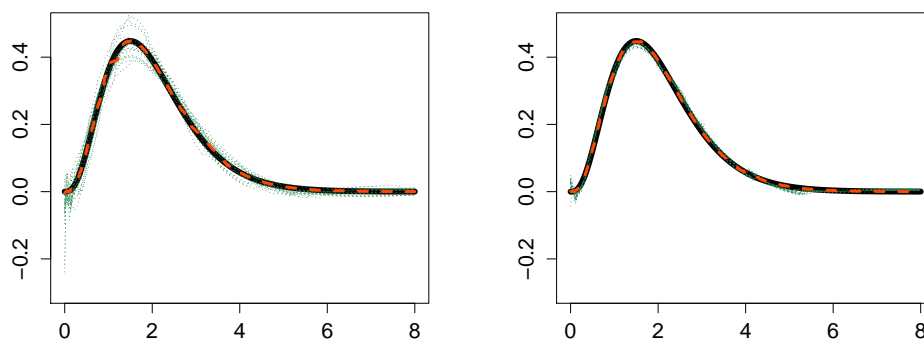


Figure 3.3: Estimation of the Gamma density: left for $n = 200$ and right $n = 2000$; $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.

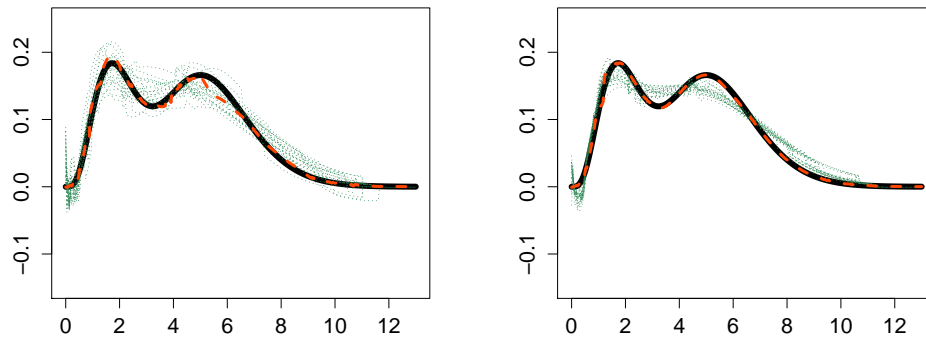


Figure 3.4: Estimation of the Mixed Gamma density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.

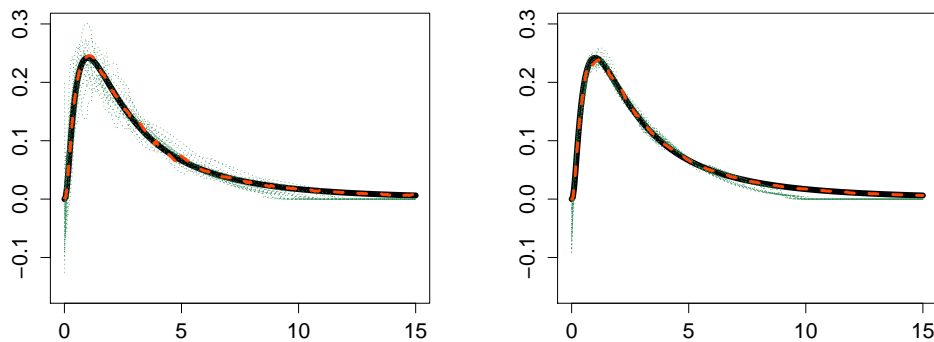


Figure 3.5: Estimation of the log Normal density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.

3.5 Proofs

3.5.1 Proofs of Section 3.2

Proof of Proposition 3.2.1

We have the following bias variance decomposition:

$$\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 = (\vartheta - \vartheta_m)^2 + \mathbb{E}(\vartheta_m - \hat{\vartheta}_m)^2.$$

The first term corresponds to the squared bias term of Equation (3.10). Let us study the second term, by definition we have

$$(\vartheta_m - \hat{\vartheta}_m)^2 = \langle \psi, f_m - \hat{f}_m \rangle^2 = \langle \vec{\psi}_m, \vec{f}_m - \hat{\vec{f}}_m \rangle_2^2 = \langle \vec{\psi}_m, \mathbf{G}_m^{-1}(\vec{h}_m - \hat{\vec{h}}_m) \rangle_2^2.$$

Then we get

$$\begin{aligned} \mathbb{E}(\vartheta_m - \hat{\vartheta}_m)^2 &= \mathbb{V}\text{ar}[\hat{\vartheta}_m] = \mathbb{V}\text{ar} \left[\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k(\psi) [\mathbf{G}_m^{-1}]_{k,l} \hat{a}_l(Z) \right] \\ &= \mathbb{V}\text{ar} \left[\frac{1}{n} \sum_{i=1}^n \sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k(\psi) [\mathbf{G}_m^{-1}]_{k,l} \varphi_l(Z_i) \right] \\ &= \frac{1}{n} \mathbb{V}\text{ar} \left[\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k(\psi) [\mathbf{G}_m^{-1}]_{k,l} \varphi_l(Z_1) \right] \\ &\leq \frac{1}{n} \mathbb{E} \left[\left(\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k(\psi) [\mathbf{G}_m^{-1}]_{k,l} \varphi_l(Z_1) \right)^2 \right] \\ &\leq \frac{1}{n} \int_{\mathbb{R}^+} \left(\sum_{k=0}^{m-1} \sum_{l=0}^{m-1} a_k(\psi) [\mathbf{G}_m^{-1}]_{k,l} \varphi_l(u) \right)^2 h(u) du \leq \frac{\|h\|_\infty}{n} \|\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2. \quad \square \end{aligned}$$

Proof of Proposition 3.2.4

According to Equation (3.10), we just have to find the eigenvectors of the matrix \mathbf{G}_m^{-1} . Indeed if ${}^t\vec{\psi}_m$ is an eigenvector of \mathbf{G}_m^{-1} associated with eigenvalue λ_m , it yields that $\|{}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 = \lambda_m^2 \|{}^t\vec{\psi}_m\|_2^2$ and since we assumed that $\psi \in \mathbb{L}^2(\mathbb{R}^+)$, the variance is upper bounded $\lambda_m^2 \|\psi\|_2^2$. First note that

- For any vector \vec{u} , $\|{}^t\mathbf{G}_m^{-1}\vec{u}\|_2^2 = \|{}^t\vec{u}\mathbf{G}_m^{-1}\|_2^2$.
- If we have ${}^t\mathbf{G}_m\vec{u} = \alpha\vec{u}$ with $\alpha \neq 0$, then $(1/\alpha)\vec{u} = {}^t\mathbf{G}_m^{-1}\vec{u}$.

Therefore let us look for the eigenvalues of ${}^t\mathbf{G}_m$.

Since ${}^t\mathbf{G}_m$ is a triangular matrix, its eigenvalues are its diagonal elements. Thus the only eigenvalue of ${}^t\mathbf{G}_m$ is $a_0(g)/\sqrt{2}$ associated with the eigenvectors belonging to $\mathcal{S}_0 = \text{Span}\{\varphi_0\}$. Then the matrix ${}^t\mathbf{G}_m^{-1}$ has for eigenvalue $\sqrt{2}/a_0(g)$ associated with the same space of eigenvectors.

Let $\psi \in \mathcal{S}_0$, i.e. $\psi(x) = \alpha e^{-x} \mathbb{1}_{\mathbb{R}^+}(x)$ with $\alpha \in \mathbb{R}$, it yields that ${}^t\vec{\psi}_m = (\alpha/\sqrt{2} \ 0 \ \dots \ 0)$ and ${}^t\vec{\psi}_m \mathbf{G}_m^{-1} = (\alpha/a_0(g) \ 0 \ \dots \ 0)$ which implies $\|{}^t\mathbf{G}_m^{-1}\vec{\psi}_m\|_2^2 = \|{}^t\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 = \alpha^2/a_0^2(g) = \vartheta_1$. Thus the variance term does not depend on the dimension m of the model as announced.

Now for the bias term, let us notice that $\vartheta = \langle \psi, f \rangle = \sum_{k \geq 0} a_k(\psi) a_k(f) = \alpha a_0(f)/\sqrt{2}$ which implies that $\mathbb{E}[\hat{\vartheta}_1] = \vartheta_1 = \vartheta$. So the estimator $\hat{\vartheta}_1$ of ϑ , in that case, is unbiased.

Finally, we have that

$$\mathbb{E}(\vartheta - \hat{\vartheta}_1)^2 = \mathbb{E}(\vartheta_1 - \hat{\vartheta}_1)^2 = \text{Var}[\hat{\vartheta}_1] = \frac{1}{n} \text{Var} \left[a_0(\psi) [\mathbf{G}_m^{-1}]_{1,1} \varphi_0(Z_1) \right] = \frac{\alpha^2}{a_0^2(g)} \frac{\text{Var}[\varphi_0(Z_1)]}{n}.$$

□

Proof of Proposition 3.2.5

The matrix \mathbf{G}_m admits only one eigenvalue $a_0(g)/\sqrt{2}$ with multiplicity m , it yields that

$$\left(\mathbf{G}_m - \frac{a_0(g)}{\sqrt{2}} \mathbf{I}_m \right)^m = 0$$

which implies that

$$\left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right)^m = 0.$$

Writing that

$$\mathbf{I}_m = \mathbf{I}_m - \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right)^m = \left(\mathbf{I}_m - \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right) \right) \sum_{k=0}^{m-1} \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right)^k,$$

therefore

$$\mathbf{G}_m^{-1} = \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{m-1} \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right)^k.$$

It yields that

$$\varrho(\mathbf{G}_m^{-1}) \leq \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{m-1} \left(\varrho \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right) \right)^k.$$

Yet using the fact that for any matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$ (see Chapter 5 in [Horn and Johnson \(1990\)](#)),

$$\varrho(\mathbf{A}) \leq \sqrt{\|\mathbf{A}\|_1 \|\mathbf{A}\|_\infty}$$

where $\|\cdot\|_1$ and $\|\cdot\|_\infty$ are respectively the maximum column sum matrix and the maximum row sum matrix. Besides noticing that since \mathbf{G}_m^{-1} is a triangular Toeplitz matrix we have that $\|\mathbf{G}_m^{-1}\|_1 = \|\mathbf{G}_m^{-1}\|_\infty$. So we get

$$\varrho \left(\mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right) \leq \left\| \mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right\|_1$$

Moreover let us notice that

$$\left\| \mathbf{I}_m - \frac{\sqrt{2}}{a_0(g)} \mathbf{G}_m \right\|_1 = \sum_{j \geq 2}^m \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right| \leq \sum_{j \geq 2} \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right|.$$

Thus if

$$q = \sum_{j \geq 2} \left| \frac{a_{j-1}(g) - a_{j-2}(g)}{a_0(g)} \right| < 1,$$

it yields that

$$\varrho(\mathbf{G}_m^{-1}) \leq \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{m-1} q^k \leq \frac{\sqrt{2}}{a_0(g)} \sum_{k=0}^{\infty} q^k = \frac{\sqrt{2}}{a_0(g)} \frac{1}{1-q}$$

then $\varrho(\mathbf{G}_m^{-1})$ is upper bounded by a constant independent of m . □

Proof of Proposition 3.2.6

Let the coefficients of \mathbf{G}_m^{-1} be defined by Equation (3.8), it yields

$$\begin{aligned} \|\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 &= \left(\frac{\lambda+1}{\lambda}\right)^2 \sum_{k=0}^{m-1} \left(a_k(\psi) + \frac{2}{\lambda+1} \sum_{j=k+1}^{m-1} a_j(\psi) \right)^2 \\ &\leq \left(\frac{\lambda+1}{\lambda}\right)^2 \left(1 \vee \frac{2}{\lambda+1}\right)^2 \sum_{k=0}^{m-1} \left(\sum_{j=k}^{m-1} |a_j(\psi)| \right)^2. \end{aligned}$$

Then if $\sum_{k=0}^{m-1} (\sum_{j=k}^{m-1} |a_j(\psi)|)^2 < \infty$, we get that $\|\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2 \leq C$ where C does not depend on m . \square

3.5.2 Proofs of Section 3.3

Proof of Theorem 3.3.1

By definition of \hat{m} , the following inequalities hold

$$\begin{aligned} (\vartheta - \hat{\vartheta}_{\hat{m}})^2 &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\hat{\vartheta}_m - \hat{\vartheta}_{m \wedge \hat{m}})^2 + 3(\hat{\vartheta}_{m \wedge \hat{m}} - \hat{\vartheta}_{\hat{m}})^2 \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(A(\hat{m}) + V(m) + A(m) + V(\hat{m})) \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m)) \end{aligned}$$

Taking expectation, we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\hat{m}})^2 \leq 3\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 + 6\mathbb{E}[A(m)] + 6V(m) \quad (3.23)$$

with

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+ \quad (3.24)$$

Proposition 3.5.1. *Under Assumptions of Theorem 3.3.1, it holds true that*

$$\mathbb{E}[A(m)] \leq \sup_{\substack{k>m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \frac{C}{n}.$$

Starting from Equation (3.23), applying Proposition 3.2.1 and Proposition 3.5.1 we get

$$\begin{aligned} \mathbb{E}(\vartheta - \hat{\vartheta}_{\hat{m}})^2 &\leq 3(\vartheta - \vartheta_m)^2 + \|h\|_\infty \frac{\|\vec{\psi}_m \mathbf{G}_m^{-1}\|_2^2}{n} + 6 \sup_{\substack{k>m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \frac{C}{n} + 6V(m) \\ &\leq 3(\vartheta - \vartheta_m)^2 + 6 \sup_{\substack{k>m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \left(\frac{1}{\kappa} + 6 \right) V(m) + \frac{C}{n} \\ &\leq 6 \sup_{\substack{k>m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + 7V(m) + \frac{C}{n}, \end{aligned}$$

since $\kappa \geq 1$ and given that $\vartheta = \vartheta_\infty = \lim_{m \rightarrow \infty} \vartheta_m$ if $f, g \in \mathbb{L}^2(\mathbb{R}^+)$ Finally by taking the infimum over all $m \in \mathcal{M}$ we get the desired result

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\hat{m}})^2 \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\substack{k>m \\ k \in \mathbb{N} \cup \{+\infty\}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + V(m) \right\} + \frac{C}{n}.$$

\square

Proof of Proposition 3.5.1. First let us notice the following upper bound

$$A(m) = \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \hat{\vartheta}_{m' \wedge m})^2 - V(m') \right\}_+ \leq 3(D_1 + D_2 + D_3)$$

where

$$\begin{aligned} D_1 &= \sup_{m' \in \mathcal{M}} \left\{ (\vartheta_{m'} - \vartheta_{m \wedge m'})^2 \right\} \\ D_2 &= \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \\ D_3 &= \sup_{m' \in \mathcal{M}} \left\{ (\hat{\vartheta}_{m \wedge m'} - \vartheta_{m \wedge m'})^2 - \frac{V(m')}{6} \right\}_+. \end{aligned}$$

• First consider D_1 , we have

$$D_1 = \sup_{m' \in \mathcal{M}} \left\{ (\vartheta_{m'} - \vartheta_{m \wedge m'})^2 \right\} = \sup_{\substack{k > m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\}. \quad (3.25)$$

• Now consider D_2

$$\begin{aligned} \mathbb{E}[D_2] &\leq \sum_{m' \in \mathcal{M}} \mathbb{E} \left[\left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \right] \\ &\leq \sum_{m' \in \mathcal{M}} \int_{\mathbb{R}^+} \mathbb{P} \left[\left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \geq u \right] du \\ &\leq \sum_{m' \in \mathcal{M}} \int_{\mathbb{R}^+} \mathbb{P} \left[|\hat{\vartheta}_{m'} - \vartheta_{m'}| \geq \sqrt{\frac{V(m')}{6} + u} \right] du. \end{aligned} \quad (3.26)$$

To handle this probability, we will apply Bernstein's inequality. First we need to identify the empirical process, by definition we have

$$\begin{aligned} \hat{\vartheta}_{m'} &= \sum_{k=0}^{m'-1} a_k(\psi) \hat{a}_k = \sum_{k=0}^{m'-1} a_k(\psi) \left(\sum_{i=0}^{m'-1} [\mathbf{G}_{m'}^{-1}]_{k,i} \frac{1}{n} \sum_{j=1}^n \varphi_i(Z_j) \right) \\ &= \frac{1}{n} \sum_{j=1}^n \sum_{k=0}^{m'-1} \sum_{i=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(Z_j) = \frac{1}{n} \sum_{j=1}^n \omega_{m'}(Z_j), \end{aligned}$$

which yields $\hat{\vartheta}_{m'} - \vartheta_{m'} = \frac{1}{n} \sum_{j=1}^n (\omega_{m'}(Z_j) - \mathbb{E}[\omega_{m'}(Z_j)])$. We need to ensure that there exist b and v such that $|\omega_{m'}(Z_j)| \leq b$ a.s. and $\text{Var}[\omega_{m'}(Z_j)] \leq v$. First notice

$$\begin{aligned} |\omega_{m'}(Z_j)| &= \left| \sum_{0 \leq i, k \leq m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(Z_j) \right| \leq \sum_{i=0}^{m'-1} \left| \varphi_i(Z_j) \sum_{k=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \right| \\ &\leq \sqrt{2} \sum_{i=0}^{m'-1} \left| \sum_{k=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \right| \leq \sqrt{2} \|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_1 := b. \end{aligned}$$

Then we have

$$\begin{aligned} \text{Var}[\omega_{m'}(Z_j)] &\leq \mathbb{E}[\omega_{m'}^2(Z_j)] \leq \mathbb{E} \left[\left(\sum_{k=0}^{m'-1} \sum_{i=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(Z_j) \right)^2 \right] \\ &\leq \int_{\mathbb{R}^+} \left(\sum_{k=0}^{m'-1} \sum_{i=0}^{m'-1} a_k(\psi) [\mathbf{G}_{m'}^{-1}]_{k,i} \varphi_i(u) \right)^2 h(u) du \leq \|h\|_\infty \|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_2^2 := v. \end{aligned}$$

At last applying Bernstein inequality, we get

$$\begin{aligned}
& \mathbb{P} \left[|\hat{\vartheta}_{m'} - \vartheta_{m'}| \geq \sqrt{\frac{V(m')}{6} + u} \right] \\
& \leq 2 \left\{ \exp \left(-\frac{n}{4v} \left(\frac{V(m')}{6} + u \right) \right) \vee \exp \left(-\frac{n}{4b} \sqrt{\frac{V(m')}{6} + u} \right) \right\} \\
& \leq 2 \left\{ \exp \left(-\frac{n}{4v} \left(\frac{V(m')}{6} + u \right) \right) \vee \exp \left(-\frac{n\alpha}{4b} \sqrt{\frac{V(m')}{6} + u} - \frac{n(1-\alpha)}{4b} \sqrt{\frac{V(m')}{6} + u} \right) \right\} \\
& \leq 2 \left\{ \exp \left(-\frac{n}{4v} \left(\frac{V(m')}{6} + u \right) \right) \vee \exp \left(-\frac{n\alpha}{4b} \sqrt{\frac{V(m')}{6}} \right) \exp \left(-\frac{n(1-\alpha)}{4b} \sqrt{u} \right) \right\},
\end{aligned}$$

with $\alpha \in [0, 1]$. Yet we get

$$\frac{n}{4v} \frac{V(m')}{6} \geq \frac{\kappa \log n}{24} \geq p \log n$$

as soon as $\kappa \geq 24p$.

By definition we have

$$\begin{aligned}
\frac{n\alpha}{4b} \sqrt{\frac{V(m')}{6}} &= \frac{n\alpha}{4\sqrt{2} \|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_1} \sqrt{\frac{\kappa(\|h\|_\infty \vee 1) \max_{1 \leq k \leq m'} \|\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 \log n}{6n}} \\
&= \alpha \sqrt{\frac{\kappa(\|h\|_\infty \vee 1)}{96}} \sqrt{\frac{\max_{1 \leq k \leq m'} \|\vec{\psi}_k \mathbf{G}_k^{-1}\|_2^2 n \log n}{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_1^2}}
\end{aligned}$$

And as $m' \in \mathcal{M}$, $m' \leq n/\log n$ and according to the following inequalities for any $\vec{u} \in \mathbb{R}^k$

$$\|\vec{u}\|_2 \leq \|\vec{u}\|_1 \leq \sqrt{k} \|\vec{u}\|_2,$$

it yields using **(A)**

$$\frac{n\alpha}{4b} \sqrt{\frac{V(m')}{6}} \geq \alpha \sqrt{\frac{\kappa(\|h\|_\infty \vee 1)}{96}} \sqrt{\frac{n \log n}{m'}} \geq \alpha \sqrt{\frac{(\|h\|_\infty \vee 1)}{96}} \log n \geq \alpha \sqrt{\frac{\|h\|_\infty}{96}} \log n.$$

Thus

$$\frac{n\alpha}{4b} \sqrt{\frac{V(m')}{6}} \geq p \log n$$

as soon as $\kappa \geq 96p^2/\alpha^2$.

Now putting these results into Equation (3.26), we get

$$\begin{aligned}
\mathbb{E}[D_2] &\leq 2 \sum_{m' \in \mathcal{M}} \int_{\mathbb{R}^+} \exp \left(-\frac{nu}{4v} - p \log n \right) \vee \exp \left(-\frac{n(1-\alpha)}{4b} \sqrt{u} - p \log n \right) du \\
&\leq 2 \sum_{m' \in \mathcal{M}} n^{-p} \int_{\mathbb{R}^+} \exp \left(-\frac{nu}{4v} \right) \vee \exp \left(-\frac{n(1-\alpha)}{4b} \sqrt{u} \right) du \\
&\leq 8n^{-p} \sum_{m' \in \mathcal{M}} \max \left(\|h\|_\infty \frac{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_2^2}{n}, 8 \frac{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_1^2}{n^2(1-\alpha)^2} \right) \\
&\leq C(\alpha, \|h\|_\infty) n^{-p} \sum_{m' \in \mathcal{M}} \max \left(\frac{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_2^2}{n}, \frac{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_1^2}{n^2} \right).
\end{aligned}$$

Since $m' \in \mathcal{M}$, we get

$$\mathbb{E}[D_2] \leq C(\alpha, \|h\|_\infty) n^{-p} \sum_{m' \in \mathcal{M}} \frac{\|\vec{\psi}_{m'} \mathbf{G}_{m'}^{-1}\|_2^2}{n} \leq C(\alpha, \|h\|_\infty) n^{-p} \text{Card}(\mathcal{M}).$$

Thus for $p = 2$, we have

$$\mathbb{E}[D_2] \leq \frac{C}{n}. \quad (3.27)$$

• Finally noticing that

$$\begin{aligned} D_3 &= \max \left(\sup_{m' > m} \left\{ (\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m')}{6} \right\}_+, \sup_{m' \leq m} \left\{ (\hat{\vartheta}_{m'} - \vartheta_{m'})^2 - \frac{V(m')}{6} \right\}_+ \right) \\ &\leq \max \left(\sup_{m' > m} \left\{ (\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m')}{6} \right\}_+, D_2 \right) \\ &\leq \max \left(\sup_{m' > m} \left\{ (\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m)}{6} \right\}_+, D_2 \right) \\ &\leq \left\{ (\hat{\vartheta}_m - \vartheta_m)^2 - \frac{V(m)}{6} \right\}_+ + D_2 \leq 2D_2, \end{aligned}$$

where we used the fact that $m \mapsto V(m)$ is nondecreasing. Now applying the same reasoning as for D_2 it yields that

$$\mathbb{E}[D_3] \leq \frac{C}{n}. \quad (3.28)$$

Now gathering Equations (3.25), (3.27) and (3.28) into Equation (3.24), we get

$$\mathbb{E}[A(m)] \leq \sup_{\substack{k > m \\ k \in \mathcal{M}}} \left\{ (\vartheta_k - \vartheta_m)^2 \right\} + \frac{C}{n}.$$

□

Proof of Corollary 3.3.2

By definition of \tilde{m} , the following inequalities hold

$$\begin{aligned} (\vartheta - \hat{\vartheta}_{\tilde{m}})^2 &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\hat{\vartheta}_m - \hat{\vartheta}_{m \wedge \tilde{m}})^2 + 3(\hat{\vartheta}_{m \wedge \tilde{m}} - \hat{\vartheta}_{\tilde{m}})^2 \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 3(\tilde{A}(\tilde{m}) + \tilde{V}(m) + \tilde{A}(m) + \tilde{V}(\tilde{m})) \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(\tilde{A}(m) + \tilde{V}(m)) \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m)) + 6(\tilde{V}(m) - V(m)) \\ &\quad + 6 \sup_{\substack{m' > m \\ m, m' \in \mathcal{M}}} \left\{ V(m') - \tilde{V}(m') \right\}_+ \\ &\leq 3(\vartheta - \hat{\vartheta}_m)^2 + 6(A(m) + V(m)) + 12 \sup_{\substack{m' \geq m \\ m, m' \in \mathcal{M}}} \left| V(m') - \tilde{V}(m') \right|. \end{aligned}$$

Taking expectation, we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_{\tilde{m}})^2 \leq 3\mathbb{E}(\vartheta - \hat{\vartheta}_m)^2 + 6(\mathbb{E}[A(m)] + V(m)) + 12\mathbb{E} \left[\sup_{\substack{m' \geq m \\ m, m' \in \mathcal{M}}} \left| V(m') - \tilde{V}(m') \right| \right].$$

To bound the first two terms on the r.h.s. of the previous inequality we apply Proposition 3.2.1 and Proposition 3.5.1 we get

$$\mathbb{E}(\vartheta - \hat{\vartheta}_m^{\sim})^2 \leq 3 \sup_{\substack{k>m \\ k \in \mathbb{N} \cup \{+\infty\}}} \{(\vartheta_k - \vartheta_m)^2\} + 48V(m) + \frac{C}{n} + 12\mathbb{E} \left[\sup_{\substack{m' \geq m \\ m, m' \in \mathcal{M}}} |V(m') - \tilde{V}(m')| \right]. \quad (3.29)$$

To control the fluctuation of the random penalty $\tilde{V}(m')$ around the true one, we need to introduce the following set

$$\Lambda = \left\{ \left| \|\hat{h}_D\|_\infty - \|h\|_\infty \right| < \frac{\|h\|_\infty}{2} \right\}.$$

On one hand, on the set Λ we have that $\|\hat{h}_D\|_\infty \leq (3/2)\|h\|_\infty$, it yields that

$$\begin{aligned} & \mathbb{E} \left[\sup_{\substack{m' > m \\ m, m' \in \mathcal{M}}} |V(m') - \tilde{V}(m')| \mathbf{1}_\Lambda \right] \\ &= \mathbb{E} \left[\left((\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1) \right) \mathbf{1}_\Lambda \right] \kappa \max_{1 \leq k \leq m} \|\psi_k^\vec{\cdot} \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n} \\ &\leq 4\kappa (\|h\|_\infty \vee 1) \max_{1 \leq k \leq m} \|\psi_k^\vec{\cdot} \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n} = 4V(m). \end{aligned}$$

On the set Λ^c , on the other hand using that $m' \in \mathcal{M}$ and Assumption (A)

$$\begin{aligned} & \mathbb{E} \left[\sup_{\substack{m' > m \\ m, m' \in \mathcal{M}}} |V(m') - \tilde{V}(m')| \mathbf{1}_{\Lambda^c} \right] \\ &= \mathbb{E} \left[\left((\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1) \right) \mathbf{1}_{\Lambda^c} \right] \kappa \max_{1 \leq k \leq m} \|\psi_k^\vec{\cdot} \mathbf{G}_k^{-1}\|_2^2 \frac{\log n}{n} \\ &\leq \kappa \log n \mathbb{E} \left[\left((\|h\|_\infty \vee 1) - (2\|\hat{h}_D\|_\infty \vee 1) \right) \mathbf{1}_{\Lambda^c} \right] \\ &\leq \kappa \log n \mathbb{E} [(\|h\|_\infty \vee 1) \mathbf{1}_{\Lambda^c}] + 2\kappa \log n \mathbb{E} [(\|\hat{h}_D\|_\infty \vee 1) \mathbf{1}_{\Lambda^c}]. \end{aligned}$$

Noticing that $\|\hat{h}_D\|_\infty \leq \left\| \sum_{k=0}^{D-1} \varphi_k \right\|_\infty \leq 2D \leq 2n$, we get

$$\mathbb{E} \left[\sup_{\substack{m' > m \\ m, m' \in \mathcal{M}}} |V(m') - \tilde{V}(m')| \mathbf{1}_{\Lambda^c} \right] \leq \kappa \log n (\|h\|_\infty \vee 1) \mathbb{E} [\mathbf{1}_{\Lambda^c}] + 2\kappa n \log n \mathbb{E} [\mathbf{1}_{\Lambda^c}].$$

We now apply the following Lemma proved just after

Lemma 3.5.2. For $p > 0$, $\log n \leq D \leq \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^p}$ and n large enough

$$\mathbb{P} \left[\left| \|\hat{h}_D\|_\infty - \|h\|_\infty \right| \geq \frac{\|h\|_\infty}{2} \right] \leq \frac{2D}{n^p}.$$

It yields for $p = 2$

$$\log n \mathbb{E} [\mathbf{1}_{\Lambda^c}] \leq 2 \log n \frac{D}{n^2} \leq 2 \log n \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^2} \frac{1}{n^2} \leq \frac{\sqrt{2}}{128} \frac{\|h\|_\infty}{n}$$

and similarly for $p = 3$

$$n \log n \mathbb{E} [\mathbf{1}_{\Lambda^c}] \leq \frac{\sqrt{2}}{128} \frac{\|h\|_\infty}{n}.$$

Finally, we have

$$\mathbb{E} \left[\sup_{\substack{m' \geq m \\ m, m' \in \mathcal{M}}} |V(m') - \tilde{V}(m')| \right] \leq 4V(m) + \frac{C}{n}. \quad (3.30)$$

In conclusion, gathering Equations (3.29) and (3.30) gives the desired result. \square

Proof of Lemma 3.5.2. Let us notice that

$$\begin{aligned} \mathbb{P} \left[\left| \|\hat{h}_D\|_\infty - \|h\|_\infty \right| \geq \|h\|_\infty/2 \right] &\leq \mathbb{P} \left[\|\hat{h}_D - h\|_\infty \geq \|h\|_\infty/2 \right] \\ &\leq \mathbb{P} \left[\|\hat{h}_D - h_D\|_\infty \geq \|h\|_\infty/4 \right] \\ &\quad + \mathbb{P} \left[\|h_D - h\|_\infty \geq \|h\|_\infty/4 \right] := P_1 + P_2. \end{aligned} \quad (3.31)$$

• First consider P_1 , the following inequalities hold true.

$$\begin{aligned} P_1 &= \mathbb{P} \left[\sup_{x \in \mathbb{R}^+} \left| \sum_{k=0}^{D-1} (\hat{a}_k(Z) - a_k(h)) \varphi_k(x) \right| \geq \|h\|_\infty/4 \right] \\ &\leq \mathbb{P} \left[\sum_{k=0}^{D-1} |\hat{a}_k(Z) - a_k(h)| \geq \|h\|_\infty/(4\sqrt{2}) \right] \\ &\leq \sum_{k=0}^{D-1} \mathbb{P} \left[|\hat{a}_k(Z) - a_k(h)| \geq \|h\|_\infty/(4\sqrt{2}D) \right] \\ &\leq \sum_{k=0}^{D-1} \mathbb{P} \left[\left| \frac{1}{n} \sum_{i=1}^n (\varphi_k(Z_i) - \mathbb{E}[\varphi_k(Z_i)]) \right| \geq \|h\|_\infty/(4\sqrt{2}D) \right] \end{aligned}$$

In order to apply Bernstein's inequality, let us notice that $|\varphi_k(Z_1)| \leq \sqrt{2}$ and $\text{Var}[\varphi_k(Z_1)] \leq 2$ which yields that

$$P_1 \leq 2D \left(\exp \left(-\frac{n}{8} \frac{\|h\|_\infty}{4\sqrt{2}D} \right) \vee \exp \left(-\frac{n}{4\sqrt{2}} \sqrt{\frac{\|h\|_\infty}{4\sqrt{2}D}} \right) \right).$$

Taking $D \leq \frac{\|h\|_\infty}{128\sqrt{2}} \frac{n}{(\log n)^p}$, we get that

$$P_1 \leq \frac{2D}{n^p}.$$

• Now let us prove that for n large enough $P_2 = 0$. According to Chapter 2 we can write the following series of equalities,

$$\begin{aligned}
h_D(x) &= \sum_{k=0}^{D-1} a_k(h) \varphi_k(x) \\
&= \sum_{k=0}^{D-1} \varphi_k(x) \left(2^{-1/2} a_k(f) a_0(g) + \sum_{l=0}^{k-1} 2^{-1/2} (a_{k-l}(g) - a_{k-l-1}(g)) a_l(f) \right) \\
&= \sum_{k=0}^{D-1} \sum_{j=0}^{D-1} a_k(f) a_j(g) \int_0^x \varphi_k(u) \varphi_j(x-u) du = \int_0^x f_D(x-u) g_D(u) du.
\end{aligned}$$

Thus we have

$$\begin{aligned}
|h_D(x) - h(x)| &= \left| \int_0^x f_D(x-u) g_D(u) du - \int_0^x f(x-u) g(u) du \right| \\
&= \left| \int_0^x (f_D - f)(x-u) g_D(u) du - \int_0^x f(x-u) (g - g_D)(u) du \right| \\
&\leq \left| \int_0^x (f_D - f)(x-u) g_D(u) du \right| + \left| \int_0^x f(x-u) (g - g_D)(u) du \right| \\
&\leq \|f_D - f\| \|g\| + \|f\| \|g - g_D\|.
\end{aligned}$$

Since $f, g \in \mathbb{L}^2(\mathbb{R}^+)$, we get

$$\|g - g_D\|^2 = \sum_{k \geq D} a_k^2(g) \xrightarrow{D \rightarrow \infty} 0 \quad \text{and} \quad \|f - f_D\|^2 = \sum_{k \geq D} a_k^2(f) \xrightarrow{D \rightarrow \infty} 0$$

Then there exists a certain n_0 such that $\|f_D - f\| \|g\| + \|f\| \|g - g_D\| \leq \varepsilon$. Besides $D \geq \log n$ and $n \geq n_0$ ensure that $\|f_D - f\| \|g\| + \|f\| \|g - g_D\| \leq \varepsilon$. Then starting from Equation (3.31), we get

$$P_2 \leq \mathbb{P} \left[\varepsilon \geq \frac{\|h\|_\infty}{4} \right] = 0$$

In the end

$$\mathbb{P} \left[\left| \|\hat{h}_D\|_\infty - \|h\|_\infty \right| \geq \frac{\|h\|_\infty}{2} \right] \leq \frac{2D}{n^p}.$$

□

3.6 Appendix 1

Let $x \in [0, 1]$, we have

$$\begin{aligned}
S_{p,j}(x) &= \frac{1}{p!} \frac{d^p}{dx^p} [x^p(1-x)^j] \\
&= \frac{1}{p!} \sum_{k=0}^p \left(\binom{p}{k} \frac{d^{p-k}}{dx^{p-k}} [x^p] \frac{d^k}{dx^k} [(1-x)^j] \right) \\
&= \frac{1}{p!} \sum_{k=0}^p \binom{p}{k} \frac{p!}{k!} x^k (-1)^k \frac{j!}{(j-k)!} (1-x)^{j-k} \\
&= (1-x)^{j-p} \sum_{k=0}^p \binom{p}{k} \frac{x^k}{k!} (-1)^k \frac{j!}{(j-k)!} (1-x)^{p-k} \\
&= (1-x)^{j-p} \sum_{k=0}^p \binom{p}{k} \binom{j}{k} \frac{x^k}{k!} (-1)^k (1-x)^{p-k} \\
&\leq |1-x|^{j-p} \max_{0 \leq k \leq p} \binom{j}{k} (|x| + |1-x|)^p \\
&\leq 2^p |1-x|^{j-p}.
\end{aligned}$$

3.7 Appendix 2

According to [Comte and Genon-Catalot \(2015\)](#), proving that a function ψ belongs to a Laguerre-Sobolev defined by (3.12), with α an integer, is equivalent to prove that

$$\|x^{\alpha/2}[\psi(x)e^x]^{(\alpha)}e^{-x}\|^2 = \sum_{k \geq \alpha} k(k-1)\dots(k-\alpha+1)a_k^2(\psi) \leq L.$$

And we have for ψ defined as in Example 3 that

$$\begin{aligned}
\|x^{\alpha/2}[\psi(x)e^x]^{(\alpha)}e^{-x}\|^2 &= \left\| x^{\alpha/2} \sum_{j=0}^{\alpha} \binom{\alpha}{j} (-1)^j \beta(\beta+1)\dots(\beta+j-1)(1+x)^{-\beta-j} \right\|^2 \\
&= \int_0^{\infty} x^{\alpha} \left| \sum_{j=0}^{\alpha} \binom{\alpha}{j} \beta(\beta+1)\dots(\beta+j-1)(1+x)^{-\beta-j} \right|^2 dx \\
&= \sum_{i=0}^{\alpha} \sum_{j=0}^{\alpha} \binom{\alpha}{i} \binom{\alpha}{j} \beta(\beta+1)\dots(\beta+j-1)\beta(\beta+1)\dots(\beta+i-1) \\
&\quad \int_0^{\infty} x^{\alpha} (1+x)^{-2\beta-i-j} dx.
\end{aligned}$$

Yet $x \mapsto x^{\alpha}(1+x)^{-2\beta-i-j}$ is integrable for all i, j if $\beta > (\alpha+1)/2$.

4

Density aggregation on the nonnegative real line in the convolution model

Ce chapitre est un travail en cours s'appuyant sur les résultats des Chapitres 2 et 3.

RÉSUMÉ. Dans ce chapitre, notre but est de développer une procédure permettant de trouver la meilleure combinaison linéaire de K estimateurs de densité dans le modèle de convolution sur \mathbb{R}^+ . Nous considérons le modèle de convolution $Z = X + Y$ avec X de densité f , indépendante de Y , quand ces deux variables aléatoires sont positives. Dans le Chapitre 2, nous avons établi une nouvelle procédure d'estimation dans le modèle de convolution sous hypothèse que les variables aléatoires sont positives. Ce travail repose sur des estimateurs par projection sur la base de Laguerre. Cependant cette base dépend d'un paramètre d'échelle et ce degré de liberté de la base n'est pas utilisé car il est fixé pendant toute l'étude du problème. Ainsi nous proposons une procédure d'agrégation linéaire pour pallier à une approche par sélection de modèle prenant en compte la structure inconnue des données. Comme dans [Rigollet and Tsybakov \(2007\)](#), nous prouvons des inégalités oracles sharp pour le risque quadratique intégré sous de faibles hypothèses. Ces résultats reposent essentiellement sur ceux du Chapitre 3.

ABSTRACT. In this chapter, we are interested in finding the best linear combination of K estimators of a density in the convolution model on the nonnegative real line. We consider the model $Z = X + Y$ with X , of unknown density f , independent of Y , when both random variables are nonnegative. In Chapter 2, we provided a new estimation procedure in the convolution model under the assumption that the random variables are nonnegative. This work is based on projection estimators computed in Laguerre basis. This basis is defined up to a scale parameter yet this degree of freedom is not taken into account and the scale parameter is arbitrarily fixed. Thus we propose a linear aggregation procedure as an alternative to model selection for taking advantage of that unknown structure in the data. Following [Rigollet and Tsybakov \(2007\)](#), we prove sharp oracle inequality for the mean squared risk of the aggregate under mild conditions. This works relies mainly on the results developed in Chapter 3.

Contents

4.1	Introduction	101
4.1.1	Motivations	101
4.1.2	Aim and model	102
4.1.3	Notations	103
4.2	Classical aggregation problem	104
4.3	Linear aggregation of Laguerre estimators	105
4.3.1	Estimation of the weights of aggregation	106
4.3.2	Oracle inequality for the aggregate	108
4.3.3	Discussion	108
4.4	Illustration	110
4.5	Proofs	112
4.5.1	Proof of Lemma 4.3.5	112
4.5.2	Proof of Theorem 4.3.3	112
4.5.3	Proof of Corollary 4.3.4	113

4.1 Introduction

4.1.1 Motivations

A series of papers deals with the Laguerre basis of $\mathbb{L}^2(\mathbb{R}^+)$, see for a regression context [Abramovich et al. \(2013\)](#) and [Comte et al. \(2015a\)](#); for density estimation, [Comte and Genon-Catalot \(2015\)](#) in mixed Poisson models, [Mabon \(2014, 2016\)](#) (see Chapters 2 and 3) in convolution models. The Laguerre basis is defined by

$$k \in \mathbb{N}, x \geq 0, \tau > 0, \quad \varphi_k^\tau(x) = \sqrt{2\tau} L_k(2\tau x) e^{-\tau x} \quad \text{with} \quad L_k(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} \frac{x^j}{j!}.$$

But in those papers the choice of the scale parameter τ is not discussed from a theoretical point of view. Let us illustrate in a simple case how the scale parameter τ can affect the estimation. For instance let us consider the classical framework where we observe n i.i.d. r.v. distributed on the nonnegative real line: X_1, \dots, X_n . Our goal is to estimate the unknown density f of the X_i 's. Let us assume that the underlying density is an Exponential distribution of parameter $\lambda > 0$, then we have that

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}_{x \geq 0} = \frac{\sqrt{\lambda}}{\sqrt{2}} \varphi_0^\lambda(x) \mathbb{1}_{x \geq 0}. \quad (4.1)$$

According to (4.1), the density of an Exponential distribution is proportional to the first function of the Laguerre basis with a scale parameter λ . More generally any function $p \in \mathbb{L}^2(\mathbb{R}^+)$ can be expanded on the Laguerre basis as

$$p = \sum_{k \geq 0} a_k^\tau(p) \varphi_k^\tau \quad \text{with} \quad a_k^\tau(p) = \langle p, \varphi_k^\tau \rangle.$$

Since in our example we directly observe the X_i 's, we can easily compute an estimator of the coefficients by $\hat{a}_k^\tau = (1/n) \sum_{i=1}^n \varphi_k^\tau(X_i)$. Obviously, the \hat{a}_k^τ are unbiased estimators of $a_k^\tau(f)$ the k -th component of the projection of f . Now let us consider an estimator of the projection of f on the m first functions of the basis with parameter τ defined as $\sum_{k=0}^{m-1} \hat{a}_k^\tau \varphi_k^\tau$. We can easily derive that

$$\mathbb{E} \left[\left\| f - \sum_{k=0}^{m-1} \hat{a}_k^\tau \varphi_k^\tau \right\|^2 \right] = \left\| f - \sum_{k=0}^{m-1} a_k^\tau(f) \varphi_k^\tau \right\|^2 + \mathbb{E} \left[\left\| \sum_{k=0}^{m-1} (a_k^\tau(f) - \hat{a}_k^\tau) \varphi_k^\tau \right\|^2 \right],$$

notice that

$$\mathbb{E} \left[\left\| \sum_{k=0}^{m-1} (a_k^\tau(f) - \hat{a}_k^\tau) \varphi_k^\tau \right\|^2 \right] \leq \frac{1}{n} \sum_{k=0}^{m-1} |\varphi_k^\tau(X_i)|^2$$

and since $\sup_{x \in \mathbb{R}^+} |\varphi_k(x)| \leq \sqrt{2\tau}$, then the following upper bound on the MISE holds

$$\mathbb{E} \left[\left\| f - \sum_{k=0}^{m-1} \hat{a}_k^\tau \varphi_k^\tau \right\|^2 \right] \leq \left\| f - \sum_{k=0}^{m-1} a_k^\tau(f) \varphi_k^\tau \right\|^2 + \frac{2\tau m}{n}. \quad (4.2)$$

For $f \sim \mathcal{E}(\lambda)$ and $m \geq 1$, if we apply the procedure with the Laguerre basis of parameter $\tau = \lambda$, there is no bias according to Equation (4.2). And consequently the resulting estimator achieves the parametric rate while for $\tau \neq \lambda$ the rate should be slower. This example shows that in a nonparametric estimation problem the parametric rate can be achieved. Thus the choice of the parameter is not without consequences. What is even more noteworthy is that

in the convolution model described in Chapter 2 we still have a bias variance decomposition where the bias is null for $m = 1$, which implies that this estimator achieves the parametric rate. So the convergence rate of the estimator can be strongly linked to the choice of the scale parameter of the basis.

Thus noticing that the parameter of the basis can affect the quality of the estimation in the right or the wrong way, Comte et al. (2015a), for instance, propose a practical way of choosing τ according to the data. Nevertheless, this methodology should involve a heavy preliminary simulation study. Comte and Genon-Catalot (2015) is an example of work where the parameter is chosen according to the model. In Chapters 2 and 3, estimation strategies are provided with $\tau = 1$.

In this chapter, we want to complete the work of Mabon (2014) (Chapter 2) who describes a new reliable approach to estimate density function in the convolution model if the r.v. are \mathbb{R}^+ -supported. Indeed a specific solution for nonnegative r.v. has been proposed to estimate the distribution (density and survival function) of a random variable observed through an additive perturbation. The method is based on a penalized projection strategy involving the Laguerre basis. Chapter 2 has showed that faster rates of convergence can be achieved with this method rather than with the one existing in the convolution literature, in particular when considering distributions of Gamma type. Nevertheless that work does not deal with the choice of an adequate parameter τ .

An idea would be to choose τ according to the data X_1, \dots, X_n through a model selection procedure *à la* Lepski for instance. Nevertheless the influence of τ in the bias is not clear that is why we rather consider an aggregation strategy. Aggregation procedure is an alternative to model selection for taking advantage of unknown structure in the data. Moreover it also exhibits all the good properties of model selection in the sense that we can derive oracle inequalities. In the case of direct observations, we can apply the work of Rigollet and Tsybakov (2007) who study the problem of linear and convex aggregation of density estimators. A solution would be to compute several estimators of f with different scale parameter in the Laguerre basis to obtain a better estimator.

The general framework of aggregation is explained in Nemirovski (2000) and in Tsybakov (2013). Optimal rates of aggregation in density and regression estimation have been derived for instance in Tsybakov (2003), Lounici (2007), Rigollet and Tsybakov (2007). Most of the results are obtained with direct observations. Recently the problem of aggregation in linear inverse problems particularly in regression were addressed by Hebiri et al. (2014).

4.1.2 Aim and model

In this note, we study the problem of estimating an unknown density f given n independent observation Z_1, \dots, Z_n drawn from $h = f * g$ where $*$ denotes the convolution product. Thus we consider the following model

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n, \quad (4.3)$$

where the X_i 's are independent identically distributed (i.i.d.) nonnegative random variables with unknown density f . The Y_i 's are also i.i.d. nonnegative random variables (r.v.) with known density g . We denote by h the density of the Z_i 's. Moreover the X_i 's and the Y_i 's are assumed to be independent. The considered random variables are real-valued. All along the paper we will assume that f and g belong to $\mathbb{L}^2(\mathbb{R}^+)$.

We refer to Chapters 2 and 3 for a survey of the convolution literature and examples where the random variables are nonnegative.

Given K functions $(\hat{f}_1, \dots, \hat{f}_K)$, called dictionary, we would like to build a function \tilde{f} almost as good as the best linear combination of $\hat{f}_1, \dots, \hat{f}_K$. Our aim is to derive oracle inequalities, such that

$$\mathbb{E} \left[\|f - \tilde{f}\|^2 \right] \leq \inf_{\lambda \in \mathbb{R}^K} \mathbb{E} \left[\|f - \hat{f}_\lambda\|^2 \right] + \Delta_{n,K}, \quad (4.4)$$

where $\hat{f}_\lambda = \sum_{k=1}^K \lambda_k \hat{f}_k$ and $\Delta_{n,K}$ is a sufficiently small remainder term that does not depend on f . By “best”, we mean that we want to build an estimator \tilde{f} with the smallest *excess risk* or *regret*, i.e. the quantity

$$\mathbb{E} \left[\|f - \tilde{f}\|^2 \right] - \inf_{\lambda \in \mathbb{R}^K} \mathbb{E} \left[\|f - \hat{f}_\lambda\|^2 \right]$$

should be small. Thus $\Delta_{n,K}$ is the “price” to pay for aggregation. In an aggregation procedure, we are interested in bounding the excess risk by a remainder as small as possible.

Besides let us notice that with the sample $\mathbb{Z}^n = (Z_1, \dots, Z_n)$ we have to estimate the dictionary functions $(\hat{f}_1, \dots, \hat{f}_K)$ and the weights of aggregation $(\hat{\lambda}_1, \dots, \hat{\lambda}_K)$. Thus to avoid some difficulties in the estimation of the aggregate \tilde{f} , a sample splitting scheme is generally applied. The sample \mathbb{Z}^n is divided into two independent subsamples: \mathbb{Z}^{n_1} used to compute the estimators of the dictionary and \mathbb{Z}^{n_2} used to aggregate them and build \tilde{f} , respectively of sizes n_1 and n_2 such that $n = n_1 + n_2$. Therefore Equation (4.4) is equivalent to

$$\mathbb{E}_{n_1} \mathbb{E}_{n_2} \left[\|f - \tilde{f}\|^2 \right] \leq \inf_{\lambda \in \mathbb{R}^K} \mathbb{E}_{n_1} \left[\|f - \hat{f}_\lambda\|^2 \right] + \Delta_{n,K}, \quad (4.5)$$

where \mathbb{E}_{n_i} denotes the expectation w.r.t the distribution of \mathbb{Z}^{n_i} . Then to obtain Equation (4.5), it is enough to prove that for any fixed functions f_1, \dots, f_K and \tilde{f} an estimator of the best linear combination of f_1, \dots, f_K that

$$\mathbb{E}_{n_2} \left[\|f - \tilde{f}\|^2 \right] \leq \inf_{\lambda \in \mathbb{R}^K} \|f - f_\lambda\|^2 + \Delta_{n,K}, \quad (4.6)$$

with $f_\lambda = \sum_{k=1}^K \lambda_k f_k$. In the sequel, we will consider fixed functions to estimate the weights of the linear aggregate and derive an oracle inequality of type (4.6). To conclude, we want to develop the following specific procedure:

- ▷ Compute K Laguerre estimators with K different scale parameters according to the methodology of Chapter 2.
- ▷ Estimate the best linear combination of these K Laguerre estimators.
- ▷ Prove an oracle inequality for our estimator of the best linear combination.
- ▷ Check that the procedure improves the initial problem of estimating the density in the convolution model.

4.1.3 Notations

For two real numbers a and b , we denote $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ and $a_+ = \max(a, 0)$. For two functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{R}$ belonging to $\mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the \mathbb{L}^2 norm of φ defined by $\|\varphi\|^2 = \int |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int \varphi(x) \psi(x) dx$. Let d be an integer, for two vectors \vec{u} and \vec{v} belonging to \mathbb{R}^d , we denote $\|\vec{u}\|_{2,d}$ the Euclidean norm defined by $\|\vec{u}\|_{2,d}^2 = {}^t \vec{u} \vec{u}$ where ${}^t \vec{u}$ is the transpose of \vec{u} . The scalar product between \vec{u} and \vec{v} is $\langle \vec{u}, \vec{v} \rangle_{2,d} = {}^t \vec{u} \vec{v} = {}^t \vec{v} \vec{u}$. We introduce the spectral norm of a matrix \mathbf{A} : $\varrho^2(\mathbf{A})$ is the largest eigenvalue of ${}^t \mathbf{A} \mathbf{A}$ in absolute value.

4.2 Classical aggregation problem

In this part, we explain precisely the now classical aggregation problem of density when the r.v. $(X_i)_{i \in \llbracket 1, n \rrbracket}$ are directly observed and i.i.d. Assume that we have at hand $K \geq 2$ fixed functions compiled in the dictionary

$$\mathcal{D} = \{f_1, \dots, f_K\}, \quad (4.7)$$

and define a combination as

$$f_\lambda = \sum_{k=1}^K \lambda_k f_k, \quad \text{with } \lambda = (\lambda_1, \dots, \lambda_K) \in \Lambda.$$

We want to build an estimator \tilde{f} called the *aggregate* by taking the best possible linear combination of the elements in \mathcal{D} . The definition of Λ yields to different kind of aggregation:

- ▷ Model selection: $\Lambda = \{\mathbf{e}_k, k = 1, \dots, K\}$ where \mathbf{e}_k are the canonical basis vectors in \mathbb{R}^K . The goal of model selection type aggregation is to build an estimator \tilde{f} , which is approximately as good as the best among the initial estimators.
- ▷ Convex aggregation: $\Lambda = [0, 1]^K, \sum_{k=1}^K \lambda_k = 1\}$. The goal of convex aggregation is to build an estimator \tilde{f} , which is approximately as good as the best convex combination of the initial estimators.
- ▷ Linear aggregation: $\Lambda \in \mathbb{R}^K$. The goal of linear aggregation is to build an estimator \tilde{f} , which is approximately as good as the best linear combination of the initial estimators.

In the sequel, we will assume that $\Lambda = \mathbb{R}^K$ so we are interested in a linear aggregation procedure, i.e. that our goal is to derive oracle inequality of type

$$\mathbb{E} \left[\|f - \tilde{f}\|^2 \right] \leq \inf_{\lambda \in \mathbb{R}^K} \|f - f_\lambda\|^2 + \Delta_{n,K}$$

with $\tilde{f} = \sum_{k=1}^K \hat{\lambda}_k f_k$ and $\Delta_{n,K}$ a small remainder term. Following [Rigollet and Tsybakov \(2007\)](#), we notice that the problem of linear aggregation can be reduced to a problem of projection estimation. Thus we define $\mathcal{L} = \text{span}(f_1, \dots, f_K)$ and let $\phi_1, \dots, \phi_{K'}$ be an orthonormal basis of \mathcal{L} in $\mathbb{L}^2(\mathbb{R}^+)$, where $K' \leq K$. So the theoretical best combination of $(\phi_1, \dots, \phi_{K'})$ is defined by

$$f_{\mathcal{L}}^* = \arg \min_{f_w \in \mathcal{L}} \|f - f_w\|^2.$$

Thus the minimizer of $\|f - f_w\|^2$ is the orthogonal projection of f onto \mathcal{L} , which writes

$$f_{\mathcal{L}}^* = f_{w^*} = \sum_{k=1}^{K'} w_k^* \phi_k = \sum_{k=1}^{K'} \langle f, \phi_k \rangle \phi_k.$$

Clearly the basis functions $\phi_1, \dots, \phi_{K'}$ depend only on the dictionary \mathcal{D} defined by (4.7), hence they are known by the statistician. Thus the problem is reduced to estimate linear functionals of type $\langle f, \phi_k \rangle$ for $k \in \llbracket 1, K' \rrbracket$.

Here we consider the classical density estimation, so we directly observe the X_i 's and we can easily estimate the linear functionals $\langle f, \phi_k \rangle$ by

$$\hat{w}_k = \frac{1}{n} \sum_{i=1}^n \phi_k(X_i), \quad \text{for } k \in \llbracket 1, K' \rrbracket.$$

It yields that the aggregate is defined by

$$\tilde{f}_L = \sum_{k=1}^{K'} \hat{w}_k \phi_k. \quad (4.8)$$

Mainly we have just built an estimator by projection, it yields that according to Pythagoras' theorem we can write

$$\|f - \tilde{f}_L\|^2 = \|f - f_{\mathcal{L}}^* + f_{\mathcal{L}}^* - \tilde{f}_L\|^2 = \|f - f_{\mathcal{L}}^*\|^2 + \|f_{\mathcal{L}}^* - \tilde{f}_L\|^2.$$

In addition

$$\|f_{\mathcal{L}}^* - \tilde{f}_L\|^2 = \|w^* - \hat{w}\|_2^2 = \sum_{k=1}^{K'} (w_k^* - \hat{w}_k)^2,$$

taking the expectation it yields that for each k

$$\mathbb{E} \left[(w_k^* - \hat{w}_k)^2 \right] = \mathbb{E} \left[(\mathbb{E}[\hat{w}_k] - \hat{w}_k)^2 \right] = \text{Var}[\hat{w}_k] \leq \frac{1}{n} \mathbb{E}[\phi_k^2(X_1)] = \frac{1}{n} \int \phi_k^2(u) f(u) \, du \leq \frac{L}{n},$$

for $\|f\|_\infty \leq L < \infty$.

We just proved the following theorem.

Theorem 4.2.1. *Assume $f_1, \dots, f_K \in \mathbb{L}^2(\mathbb{R}^+)$, $\|f\|_\infty \leq L$ and \tilde{f}_L is defined by (4.8) then*

$$\mathbb{E} \|f - \tilde{f}_L\|^2 \leq \min_{\lambda \in \mathbb{R}^K} \|f - f_\lambda\|^2 + \frac{KL}{n}.$$

This result is proved for r.v. distributed on \mathbb{R}^d in [Rigollet and Tsybakov \(2007\)](#). Here we just develop the same reasoning for nonnegative r.v. Besides this Theorem is derived under mild conditions, therefore we can apply it for estimators build with the Laguerre procedure described in Chapter 2 when the r.v. are observed without noise.

4.3 Linear aggregation of Laguerre estimators

In this section, we want to derive a linear aggregation procedure in the convolution model on the nonnegative real line. Thus we assume that the dictionary \mathcal{D} is made up of Laguerre estimators as described in Chapter 2. That is, from now on, we assume that the dictionary functions belongs to $\mathbb{L}^2(\mathbb{R}^+)$. More precisely, we define the dictionary as

$$\mathcal{D} = \{f_1, \dots, f_K\}$$

with f_k in $\mathbb{L}^2(\mathbb{R}^+)$ for $k \in \llbracket 1, K \rrbracket$ and

$$f_k(x) = \sum_{\ell=0}^{m_k-1} a_\ell^{\tau_k}(f_k) \varphi_\ell^{\tau_k}(x). \quad (4.9)$$

This formulation of the dictionary matches exactly the aggregation idea of Laguerre estimators.

The aim is still to build a linear aggregation procedure except that now we observe the target random variables X_i with an independent additive nonnegative error Y_i which means that we have at hand the following observations:

$$Z_i = X_i + Y_i, \quad i = 1, \dots, n.$$

Our problem is the same as in the previous section which means estimate $f_{\mathcal{L}}^*$ defined by

$$f_{\mathcal{L}}^* = \arg \min_{f_w \in \mathcal{L}} \|f - f_w\|^2 \quad \text{with} \quad f_{\mathcal{L}}^* = \sum_{k=1}^{K'} w_k^* \phi_k.$$

The goal is to compute

$$\tilde{f}_w = \sum_{k=1}^{K'} \hat{w}_k \phi_k, \quad \hat{w} = (\hat{w}_1, \dots, \hat{w}_{K'}) \in \mathbb{R}^{K'}$$

where the difficulty lies in the estimation of \hat{w} since we do not observe the X_i 's but the Z_i 's. The estimator is different from the one defined in the previous section.

Thus the aim of this section is to precisely explain how we compute the weights of aggregation in a deconvolution problem on the nonnegative real line. As demonstrated previously, the heart of the problem is to find how the coefficients w_k of f_w in the basis $(\phi_1, \dots, \phi_{K'})$ should be estimated, which cannot directly be done in the convolution model. Thus the method is based on the results of Chapter 3 (Mabon (2016)) which tackles the problem of estimating linear functionals of the density f in Model (4.3). Under mild conditions, oracle inequality of adaptive estimators by penalization of linear functionals is derived. Those results are the key of our procedure. For more details we refer to Chapter 3. The methodology is as follows.

4.3.1 Estimation of the weights of aggregation

First since f_1, \dots, f_K belong to $\mathbb{L}^2(\mathbb{R}^+)$, so $\phi_1, \dots, \phi_{K'}$. And we also assume that $f \in \mathbb{L}^2(\mathbb{R}^+)$, it yields that they can be expanded in the Laguerre basis $(\varphi_j^{\tau_0})_{j \geq 0}$ for some given nonnegative τ_0 , it yields that

$$w_k^* = \langle \phi_k, f \rangle = \sum_{j=0}^{\infty} a_j^{\tau_0}(\phi_k) a_j^{\tau_0}(f).$$

Therefore to give an estimate of w_k^* we need to estimate the coefficients $a_j^{\tau_0}(f)$ and to appropriately truncate the sum. Let us define the truncated sum of w_k^* :

$$w_{k,m}^* = \sum_{j=0}^{m-1} a_j^{\tau_0}(\phi_k) a_j^{\tau_0}(f).$$

Thus an estimator of $w_{k,m}^*$ is given by

$$\hat{w}_{k,m} = \sum_{j=0}^{m-1} a_j^{\tau_0}(\phi_k) \hat{a}_j^{\tau_0}, \quad (4.10)$$

since the ϕ_k are known fixed functions and where the coefficients $\hat{a}_j^{\tau_0}$ are computed as follows

$${}^t(\hat{a}_0^{\tau}, \dots, \hat{a}_{m-1}^{\tau}) := \hat{f}_{\tau,m}^{\tau} = \mathbf{G}_{\tau,m}^{-1} \hat{h}_{\tau,m}^{\tau} \quad \text{with} \quad \hat{h}_{\tau,m}^{\tau} = {}^t(\hat{a}_0^{\tau}(Z), \dots, \hat{a}_{m-1}^{\tau}(Z))$$

where $\hat{a}_j^{\tau}(Z) = (1/n) \sum_{i=1}^n \varphi_j^{\tau}(Z_i)$ for $j \in \llbracket 0, m-1 \rrbracket$ are estimators of $a_j^{\tau}(h)$, and

$$([\mathbf{G}_{\tau,m}]_{i,j})_{0 \leq i,j \leq m-1} = \frac{1}{\sqrt{2\tau}} \begin{cases} a_0^{\tau}(g) & \text{if } i = j, \\ a_{i-j}^{\tau}(g) - a_{i-j-1}^{\tau}(g) & \text{if } j < i, \\ 0 & \text{otherwise.} \end{cases} \quad (4.11)$$

These linear functionals satisfy the assumptions of Proposition 3.2.1 p.74 in Chapter 3, we can then apply the result.

Proposition 4.3.1. *For f, g and $\phi_k \in \mathbb{L}^2(\mathbb{R}^+)$, for $\mathbf{G}_{\tau_0, m}$ defined by Equation (4.11) and $\hat{w}_{k, m}$ defined by Equation (4.10), the following result holds*

$$\mathbb{E} \left[|w_k^* - \hat{w}_{k, m}|^2 \right] \leq |w_k^* - w_{k, m}^*|^2 + \|h\|_\infty \frac{\|{}^t\vec{\phi}_{k, m} \mathbf{G}_{\tau_0, m}^{-1}\|_2^2}{n} \quad (4.12)$$

where ${}^t\vec{\phi}_{k, m} = (a_0^{\tau_0}(\phi_k), \dots, a_{m-1}^{\tau_0}(\phi_k))$.

Now we have to select a dimension m such that the squared risk of the estimator is near the minimum. We add the following assumption on the model collection:

$$\text{(A)} \quad \mathcal{M}_k = \left\{ m \in \{1, \dots, \lfloor n/\log n \rfloor\}, \max_{1 \leq \ell \leq m} \|{}^t\vec{\phi}_{k, \ell} \mathbf{G}_{\tau_0, \ell}^{-1}\|_2^2 \leq n \right\}.$$

The model selection is done by minimizing the following quantity

$$\hat{m}_k = \arg \min_{m \in \mathcal{M}_k} \{A_k(m) + V_k(m)\},$$

where $V_k(m)$ has the order of the variance in the right-hand side of Equation (4.12) and is defined by

$$V_k(m) = \kappa (\|h\|_\infty \vee 1) \frac{\log n}{n} \max_{1 \leq \ell \leq m} \|{}^t\vec{\phi}_{k, \ell} \mathbf{G}_{\tau_0, \ell}^{-1}\|_2^2,$$

with κ a numerical constant which will be chosen by calibration. The term of variance is slightly modified to ensure that V_k is nondecreasing with m . The term $A_k(m)$ has the order of the bias and is based on the comparison of the estimators of the linear functionals

$$A_k(m) = \sup_{m' \in \mathcal{M}_k} \left\{ (\hat{w}_{k, m'} - \hat{w}_{k, m' \wedge m})^2 - V_k(m') \right\}_+.$$

It yields that we estimate the weights of aggregation as

$$\tilde{w}_k = \hat{w}_{k, \hat{m}_k} = \sum_{j=0}^{\hat{m}_k - 1} a_j^{\tau_0}(\phi_k) \hat{a}_j^{\tau_0}. \quad (4.13)$$

We can now apply the oracle inequality of Theorem 2.4.1 p.45 in Chapter 3, for the aggregation weights:

Theorem 4.3.2. *For f, g and $\phi_k \in \mathbb{L}^2(\mathbb{R}^+)$, assume that Assumption (A) is true. Let \tilde{w}_k be defined by Equation (4.13). Then there exists a constant κ_0 such that for $\kappa \geq \kappa_0$ such that*

$$\mathbb{E} \left[(w^* - \tilde{w}_k)^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}_k} \left\{ \sup_{\substack{\ell > m \\ \ell \in \mathbb{N} \cup \{+\infty\}}} \left\{ (w_{k, \ell} - w_{k, m})^2 \right\} + V_k(m) \right\} + \frac{C}{n}.$$

where C^{ad} is a positive numerical constant and C depends only on $\|h\|_\infty$.

We can note that if $\sup_x g(x) < \infty$ then we can bound $\|h\|_\infty$ by $\|g\|_\infty$. If we keep $\|h\|_\infty$, we have to estimate it in practice. Yet we assume that g is known, so we can save time in a simulation setting.

At last we define an estimate of the linear aggregate as

$$\tilde{f}_w = \sum_{k=1}^{K'} \tilde{w}_k \phi_k,$$

where \tilde{w}_k are defined by Equation (4.13).

4.3.2 Oracle inequality for the aggregate

Theorem 4.3.3. *For f and g belonging to $\mathbb{L}^2(\mathbb{R}^+)$, for all dictionaries $f_1, \dots, f_K \in \mathbb{L}^2(\mathbb{R}^+)$, then there exists a positive constant \mathfrak{C} depending only on $\|h\|_\infty$ such that*

$$\mathbb{E}\|f - \tilde{f}_w\|^2 \leq \min_{\lambda \in \mathbb{R}^K} \|f - f_\lambda\|^2 + \mathfrak{C}K \left(\psi_1(n) + \frac{1}{n} \right), \quad (4.14)$$

where $\psi_1(n) = \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{m, \tau_0}\|^2 + \kappa(\|h\|_\infty \vee 1)(\log n/n)\varrho^2 \left(\mathbf{G}_{\tau_0, m}^{-1} \right) \right\}$.

The oracle inequality (4.14) establishes a non asymptotic oracle bound for the linear aggregate. Moreover this oracle inequality is sharp which means the leading constant is equal to 1.

Sample splitting and averaged aggregate

We use a jackknife procedure to aggregate our estimators to avoid arbitrary splitting which may influence the estimation. Thus we denote by \mathcal{S} a split of the initial sample of observations into two independent subsamples such that $\mathbb{Z}^n = (\mathbb{Z}^{n_1}, \mathbb{Z}^{n_2})$. We use the sample \mathbb{Z}^{n_1} to estimate the density f and \mathbb{Z}^{n_2} to estimate the weights of the aggregation. So we consider the linear aggregate $\tilde{f}^{\mathcal{S}}$ based on the validation sample $\mathbb{Z}_{\mathcal{S}}^{n_2}$ and the initial set of estimators $f_k = \hat{f}_{n_1, k}^{\mathcal{S}}$ for $k \in \llbracket 1, K \rrbracket$ where $\hat{f}_{n_1, k}^{\mathcal{S}}$ is computed with the trainig sample $\mathbb{Z}_{\mathcal{S}}^{n_1}$. At last we consider the averaged aggregate

$$\tilde{f}_w^{\mathcal{S}} = \frac{1}{\text{card}(\mathcal{S})} \sum_{\mathcal{S} \in \mathcal{S}} \tilde{f}_w^{\mathcal{S}}.$$

Corollary 4.3.4. *For f and g belonging to $\mathbb{L}^2(\mathbb{R}^+)$, let $n_1 < n$, $n/\log n \leq n_2 < n$ such that $n_2 = n - n_1$ and \mathcal{S} be an arbitrary set of splits. Assume that $\hat{f}_{n_1, 1}^{\mathcal{S}}, \dots, \hat{f}_{n_1, K}^{\mathcal{S}} \in \mathbb{L}^2(\mathbb{R}^+)$ for fixed $\mathbb{Z}_{\mathcal{S}}^{n_1}$, $\forall \mathcal{S} \in \mathcal{S}$. Then the averaged aggregate satisfies*

$$\mathbb{E} \left[\|f - \tilde{f}_w^{\mathcal{S}}\|^2 \right] \leq \min_{\lambda \in \mathbb{R}^K} \mathbb{E}_{n_1} \left[\left\| f - \sum_{k=1}^K \lambda_k \hat{f}_{n_1, k}^{\mathcal{S}} \right\|^2 \right] + \mathfrak{C}K \left(\psi_2(n) + \frac{\log n}{n} \right),$$

where $\psi_2(n) = \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{\tau_0, m}\|^2 + \kappa(\|h\|_\infty \vee 1)(\log n)^2/n\varrho^2 \left(\mathbf{G}_{\tau_0, m}^{-1} \right) \right\}$.

4.3.3 Discussion

As mentioned in the introduction, our target is to derive oracle inequalities of type (4.4). By identification to Equation (4.4), we have

$$\Delta_{n, K} = K \left(\psi_2(n) + \frac{\log n}{n} \right).$$

Thus our procedure is noteworthy if $\Delta_{n, K}$ is a small remainder term, in other words, if $\Delta_{n, K} \leq \min_{w \in \mathbb{R}^K} \|f - f_w\|^2$. For the second term $K \log nn^{-1}$ it is obvious that taking K of order $\log n$ would be enough to ensure that this term is small compared to $\min_{w \in \mathbb{R}^K} \|f - f_w\|^2$. The difficulty lies in the fact that $K\psi_2(n)$ is small enough or not.

First suppose that $f \in W^s(\mathbb{R}^+, L)$ where $W^s(\mathbb{R}^+, L)$ is a Laguerre-Sobolev space defined by Equation (2.16) in Chapter 2 and g satisfies Lemma 2.3.6, it yields that

$$\|f - f_w\|^2 = \mathcal{O} \left(n^{-s/s+2r} \right)$$

$$\inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{m, \tau_0}\|^2 + \kappa(\|h\|_\infty \vee 1) \log n \frac{\varrho^2 \left(\mathbf{G}_{\tau_0, m}^{-1} \right)}{n} \right\} = \mathcal{O} \left((n/(\log n)^2)^{-s/s+2r} \right)$$

In that case, $\Delta_{n,K}$ has the same order as $\|f - f_w\|^2$ for n large enough. It gives our result an asymptotic flavor. Then in that situation, we do not improve the estimation of the density f by applying a linear aggregation procedure but we do not degrade it also for n large enough.

Secondly suppose that f belongs to $\mathcal{M}\Gamma(q, \vec{\alpha}, \vec{p}, \vec{\mu})$ (a mixed Gamma family) defined by Equation (2.20) p.44 and g still satisfies Lemma 2.3.6 in Chapter 2, we get that

$$\|f - f_w\|^2 = \mathcal{O}\left(\frac{(\log n)^{2r+1}}{n}\right)$$

$$\inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{m,\tau_0}\|^2 + \kappa(\|h\|_\infty \vee 1)(\log n)^2 \frac{\varrho^2(\mathbf{G}_{\tau_0,m}^{-1})}{n} \right\} = \mathcal{O}\left(\frac{(\log n)^{2r+3}}{n}\right).$$

Thus in this case, we observe a logarithmic loss in the rate of convergence which is asymptotically negligible. The aggregation procedure does not improve the rate. Nonetheless, this is a very particular case since we know from Chapter 2 that mixed Gamma densities exhibit very good rates of convergence for a non-parametric problem. The rates are almost parametric. It may be expected that the procedure does not improve tremendously the rates when a Gamma type density is under estimation with the Laguerre procedure.

Perspective. A first trail would be to use the underlying structure of the functions of the dictionary to replace the bias term $\|f - f_{\tau_0,m}\|^2$ by a term which tends faster to zero. According to the proof of Theorem 4.3.3, this term should be $\|\phi_j - \phi_{j,m}\|^2$. We know that ϕ_j is a linear combination of f_1, \dots, f_K defined by Equation (4.9) and the dictionary functions are also linear combinations of Laguerre basis. To make use of the underlying Laguerre structure, which is of Gamma type, we need to compute explicitly ϕ_j . We state the following lemma.

Lemma 4.3.5. *If f_1, \dots, f_K are linearly independent then ϕ_1, \dots, ϕ_K defined by*

$$\phi_k = \frac{\tilde{\phi}_k}{\|\tilde{\phi}_k\|} \quad \text{with} \quad \tilde{\phi}_k = \sum_{j=1}^K [\mathbf{P}_K]_{k,j} f_j \quad \text{and} \quad \|\tilde{\phi}_k\|^2 = \lambda_k(f)$$

with \mathbf{P}_K an orthogonal matrix such that $\mathbf{M}_f \mathbf{P}_K = \mathbf{P}_K \Delta \mathbf{P}_K^t$ with $\Delta = \text{diag}(\lambda_1(f), \dots, \lambda_K(f))$ and $(\lambda_k(f))_{1 \leq k \leq K}$ eigenvalues of $\mathbf{M}_f = (\langle f_i, f_k \rangle)_{1 \leq i, j \leq K}$, is an orthonormal basis of $\mathcal{L} = \text{span}\{f_1, \dots, f_K\}$.

To apply this lemma to our dictionary \mathcal{D} we would have to assume that the scale parameters of the Laguerre base $(\tau_k)_{k \in [1, K]}$ are all different, i.e. $\tau_k \neq \tau_{k'}$ for $k \neq k'$. Next according to Equation (4.9), we would use the decomposition of f_1, \dots, f_K on the different Laguerre base to control the norm $\|\phi_j - \phi_{j,m}\|^2$ which should only depend on the different Laguerre base. Since the Laguerre basis contain mixed Gamma type functions this bias should decay exponentially as proved in Chapter 2 Section 2.3.3. This approach should require a control on various sums and is not easy to handle.

A second idea corresponds to a slight modification of the problem. The aim of this chapter is to aggregate Laguerre estimators since Laguerre functions have a particular shape such that they inherit a parametric structure of Gamma type. Besides, we know that Gamma type functions with a Laguerre estimation procedure exhibit almost parametric rates of convergence and if the scale parameter is equal to the Laguerre parameter the rate

is parametric. Thus the idea would be to “break” the Laguerre structure of the dictionary by considering only Gamma densities. We would use the following dictionary

$$\mathcal{D}_\gamma = \{\gamma_1, \dots, \gamma_K\},$$

with

$$\forall k \in \llbracket 1, K \rrbracket, \forall x > 0, \quad \gamma_k(x) = \frac{\beta_k^{\alpha_k}}{\Gamma(\alpha_k)} x^{\alpha_k-1} e^{-\beta_k x}, \quad \alpha_k, \beta_k > 0.$$

This idea is used by [De Canditiis and Pensky \(2016\)](#). This is rather natural to think of this alternative since if we fix $\beta_k = \beta$ and put $\alpha_k = 1, 2, \dots$, the linear span of the γ_k for $k \in \mathbb{N}$ matches $\mathbb{L}^2(\mathbb{R}^+)$. In addition a linear combination of γ_k for K large enough approaches any square integrable function with a small error. In fact, with this definition of the dictionary, it contains linear combinations of the Laguerre functions and hence it gives an approximation of $\mathbb{L}^2(\mathbb{R}^+)$. Moreover we can approximate a function with less functions than the Laguerre basis since we can change the parameters of the Gamma distributions. Nevertheless with this approach it yields to a completely deterministic dictionary, so the dictionary do not learn from the data.

4.4 Illustration

In this section, we illustrate the aggregation procedure for the direct case when the X_i 's are directly observed. Indeed the numerical procedure for the case of indirect observations is still unstable due the adaptive estimation of the weights of aggregation.

The whole implementation is conducted using R software. The integrated squared error $\|f - \tilde{f}_w^S\|^2$ is computed via a standard approximation and discretization (over 300 points) of the integral on an interval of \mathbb{R} denoted by I . Then the mean integrated squared error (MISE) $\mathbb{E}\|f - \tilde{f}_w^S\|^2$ is computed as the empirical mean of the approximated ISE over 200 simulation samples.

Implementation of the estimation procedure

Dictionary functions. The estimators of the dictionary are computed according to the following adaptive procedure developed in [Chapter 2](#)

- ▷ For $m \in \mathcal{M}_n = \{m_1, \dots, m_n\}$, compute $-\|\hat{f}_m^\tau\|^2 + \widehat{\text{pen}}_\tau(m)$.
- ▷ Choose \hat{m}_τ such that $\hat{m}_\tau = \arg \min_{m \in \mathcal{M}_n} \{-\|\hat{f}_m^\tau\|^2 + \widehat{\text{pen}}_\tau(m)\}$.
- ▷ Compute $\hat{f}_{\hat{m}_\tau}^\tau(x) = \sum_{k=0}^{\hat{m}_\tau-1} \hat{a}_k^\tau(f) \varphi_k^\tau(x)$.

The procedure is given for the estimation of the density f with different parameters of the Laguerre basis $\tau \in \{1/4, 1/2, 1, 2, 4\}$. Besides the penalty is chosen according to [Theorem 2.4.1](#) of [Chapter 2](#). The constant calibrations were done with intensive preliminary simulations. We take respectively the following constants $\kappa = \{0.07, 0.1, 0.09, 0.1, 0.08\}$ for $\tau = \{1/4, 1/2, 1, 2, 4\}$. We consider the following model collection $\mathcal{M}_n = \{m \in \{1, \dots, n-1\}\}$.

Splitting. We refer to [Corollary 4.3.4](#) for the splitting of the sample. We take 10 repetitions for the bootstrap because increasing the size does not improve significantly the estimation and the computing time of the procedure is already long.

Simulation setting

The performance of the procedure is studied for the seven following distributions for X . Moreover all the densities are normalized with unit variance.

- ▷ Exponential: $\mathcal{E}(1)$, $I = [0, 5]$.
- ▷ Gamma distribution: $\Gamma(2, \frac{1}{\sqrt{2}})$, $I = [0, 9]$.
- ▷ Mixed Gamma distribution: $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [0, 13]$.
- ▷ Rayleigh distribution: $\sigma^2 = 2/(4 - \pi)$, $f(x) = \frac{x}{\sigma^2}e^{-\frac{x^2}{2\sigma^2}}$, $I = [0, 5]$.

We can notice that all the distributions belong to a mixed Gamma class except the Rayleigh distribution. Besides the Exponential and the Gamma $\Gamma(4, \frac{1}{2})$ distributions have a null bias for one of the estimators of the dictionary.

Simulation results

	n	200	2000
$\mathcal{E}(1)$	aggregate	2.483	1.421
	best of the dictionary	0.648	0.065
	worst of the dictionary	2.370	0.252
Gamma $\Gamma(4, \frac{1}{2})$	aggregate	0.426	0.054
	best of the dictionary	0.228	0.025
	worst of the dictionary	0.803	0.090
Mixed Gamma	aggregate	0.387	0.047
	best of the dictionary	0.252	0.029
	worst of the dictionary	0.979	0.100
Rayleigh	aggregate	0.522	0.068
	best of the dictionary	0.282	0.035
	worst of the dictionary	0.910	0.098

Table 4.1: Results of simulations. Columns report the average values $100 \times \text{ISE} = 100 \times \|\tilde{f}_w^{\mathbb{S}} - f\|^2$ (aggregate) computed over 200 samples.

The results of the simulations are given in Table 4.1. We report the average values of the integrated squared errors multiplied by 100 computed over 200 simulated data sets. We give the MISE of the linear aggregate estimator, the best and worst estimators of the dictionary.

First we see that the risk decreases when the sample size increases for all estimators. We can note that in the dictionary there is a gap between the best and the worst estimators for all distributions. So the influence of the scale parameter of the Laguerre basis is empirically verified.

We see that for the Exponential distribution the aggregation procedure does not improve the estimation for $n = 200$. It even deteriorates the estimation in this particular case. It may be explained by the fact that for the Exponential distribution, there is an estimator in the dictionary that reaches the parametric rate. Thus the remainder term of Corollary 4.3.4 deteriorate the best rate. For the other distributions of Gamma type, the aggregate does not do better than the best estimator in the collection but its risk is close to it.

Last we have applied our procedure to a Rayleigh distribution which does not belong to mixed Gamma class. However the procedure still works when computing Laguerre estimators of its density. It should incite us to expand our strategy to all densities estimated with Laguerre estimators.

4.5 Proofs

4.5.1 Proof of Lemma 4.3.5

Let u and v belong to \mathcal{L} such that $u = \sum_{i=1}^K u_i f_i$ and $v = \sum_{j=1}^K v_j f_j$, it yields

$$\langle u, v \rangle_2 = {}^t u \mathbf{M}_f v,$$

where $\mathbf{M}_f = (\langle f_i, f_k \rangle)_{1 \leq i, k \leq K}$. Yet ${}^t u \mathbf{M}_f u = \|u\|_2^2 \geq 0$, thus \mathbf{M}_f is a semi-definite positive matrix. It implies that its eigenvalues $(\lambda_1(f), \dots, \lambda_K(f))$ are nonnegative. Besides we assumed that the f_k 's are linearly independent which implies that \mathbf{M}_f is invertible. Thus 0 is not an eigenvalue of \mathbf{M}_f . So \mathbf{M}_f is a positive-definite matrix.

We can now express the orthonormal basis ϕ_1, \dots, ϕ_K . The eigenvalues $(\lambda_k(f))_{1 \leq k \leq K}$ of \mathbf{M}_f satisfy $\lambda_k(f) \in (0, +\infty)$ for \mathbf{M}_f is a positive-definite matrix. This matrix is diagonalizable in an orthonormal basis which means there exists a matrix \mathbf{P}_K such that ${}^t \mathbf{P}_K \mathbf{P}_K = \mathbf{I}_K$ and ${}^t \mathbf{P}_K \mathbf{M}_f \mathbf{P}_K = \text{diag}(\lambda_1(f), \dots, \lambda_K(f))$. Let us put

$$\tilde{\phi}_j = \sum_{k=1}^K [\mathbf{P}_K]_{j,k} f_k,$$

by construction $\tilde{\phi}_j$ is orthogonal and $\|\tilde{\phi}_j\|^2 = \lambda_j(f)$. At last we consider the following orthonormal basis for $k = 1, \dots, K$: $\phi_k = \frac{\tilde{\phi}_k}{\|\tilde{\phi}_k\|}$. \square

4.5.2 Proof of Theorem 4.3.3

Applying Pythagora's theorem we get

$$\|f - \tilde{f}\|^2 = \|f - f_{\mathcal{L}}^*\|^2 + \|f_{\mathcal{L}}^* - \tilde{f}\|^2 = \min_{f_w \in \mathcal{L}} \|f - f_w\|^2 + \sum_{k=1}^K \left(w_k^* - \hat{w}_{k, \hat{m}_k} \right)^2.$$

Hence according to Theorem 3.1 in Mabon (2016), we get for any k

$$\mathbb{E} \left[\left(w_k^* - \hat{w}_{k, \hat{m}_k} \right)^2 \right] \leq C^{ad} \inf_{m \in \mathcal{M}} \left\{ \sup_{\ell > m} \left(w_{k, \ell}^* - w_{k, m}^* \right)^2 + \kappa (\|h\|_\infty \vee 1) \log n \frac{\varrho^2 \left(\mathbf{G}_{\tau_0, m}^{-1} \right)}{n} \right\} + \frac{C}{n}.$$

By definition we have

$$\begin{aligned} w_{k, \ell}^* - w_{k, m}^* &= \sum_{j=m}^{\ell-1} a_j^{\tau_0}(\phi_k) a_j^{\tau_0}(f) = \sum_{j=m}^{\ell-1} \langle \phi_k^{\tau_0}, \varphi_j \rangle a_j^{\tau_0}(f) = \langle \phi_k, \sum_{j=m}^{\ell-1} a_j^{\tau_0}(f) \varphi_j^{\tau_0} \rangle \\ &= \langle \phi_k, f_{\tau_0, \ell} - f_{\tau_0, m} \rangle. \end{aligned}$$

Then applying Cauchy-Schwarz, we get

$$\sup_{\ell > m} \left(w_{k, \ell}^* - w_{k, m}^* \right)^2 \leq \sup_{\ell > m} \|\phi_k\|^2 \|f_{\tau_0, \ell} - f_{\tau_0, m}\|^2 \leq \|f - f_{\tau_0, m}\|^2.$$

In the end we get

$$\begin{aligned} \mathbb{E}\|f - \tilde{f}\|^2 &\leq \min_{f_w \in \mathcal{L}} \|f - f_w\|^2 \\ &+ KC^{ad} \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{m, \tau_0}\|^2 + \kappa(\|h\|_\infty \vee 1) \log n \frac{\varrho^2(\mathbf{G}_{\tau_0, m}^{-1})}{n} \right\} + \frac{CK}{n} \end{aligned}$$

□

4.5.3 Proof of Corollary 4.3.4

First we put $\hat{f}_\lambda^{\mathcal{S}} = \sum_{k=1}^K \lambda_k \hat{f}_{n_1, k}^{\mathcal{S}}$. For any fixed $\mathcal{S} \in \mathbb{S}$ and $\mathbb{Z}_{1, \mathcal{S}}^{n_1}$, Theorem 4.3.3 yields

$$\begin{aligned} \mathbb{E}_{n_2} [\|f - \tilde{f}_{n_2}^{\mathcal{S}}\|^2] &\leq \inf_{\lambda \in \mathbb{R}^K} \|f - \hat{f}_\lambda^{\mathcal{S}}\|^2 \\ &+ KC^{ad} \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{\tau_0, m}\|^2 + \kappa(\|h\|_\infty \vee 1) \log(n_2) \frac{\varrho^2(\mathbf{G}_{\tau_0, m}^{-1})}{n_2} \right\} + \frac{CK}{n_2}, \end{aligned}$$

it implies

$$\begin{aligned} \mathbb{E} [\|f - \tilde{f}^{\mathcal{S}}\|^2] &= \mathbb{E}_{n_1} \mathbb{E}_{n_2} [\|\tilde{f}_{n_2}^{\mathcal{S}} - f\|^2] \\ &\leq \inf_{\lambda \in \mathbb{R}^K} \mathbb{E}_{n_1} [\|f - \hat{f}_\lambda^{\mathcal{S}}\|^2] \\ &+ KC^{ad} \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{\tau_0, m}\|^2 + \kappa(\|h\|_\infty \vee 1) \log(n_2) \frac{\varrho^2(\mathbf{G}_{\tau_0, m}^{-1})}{n_2} \right\} + \frac{CK}{n_2}. \end{aligned}$$

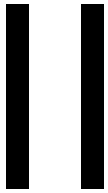
Applying Jensen's inequality, we get

$$\mathbb{E} [\|f - \tilde{f}_w^{\mathbb{S}}\|^2] \leq \frac{1}{\text{card}(\mathbb{S})} \sum_{\mathcal{S} \in \mathbb{S}} \mathbb{E} [\|f - \tilde{f}_w^{\mathcal{S}}\|^2]$$

since \mathcal{S} is fixed and $n/\log n \leq n_2 \leq n$, it yields

$$\begin{aligned} \mathbb{E} [\|f - \tilde{f}_w^{\mathbb{S}}\|^2] &\leq \inf_{\lambda \in \mathbb{R}^K} \mathbb{E}_{n_1} [\|f - \hat{f}_\lambda^{\mathcal{S}}\|^2] \\ &+ KC^{ad} \max_{1 \leq k \leq K'} \inf_{m \in \mathcal{M}_k} \left\{ \|f - f_{\tau_0, m}\|^2 + \kappa\|h\|_\infty (\log n)^2 \frac{\varrho^2(\mathbf{G}_{\tau_0, m}^{-1})}{n} \right\} + CK \frac{\log n}{n}. \end{aligned}$$

□



Estimation adaptative de densité dans le modèle de convolution avec bruit inconnu

5

Adaptive density estimation in deconvolution problems with unknown error distribution

Ce chapitre est une version modifiée de l'article *Adaptive density estimation in deconvolution problems with unknown error distribution* publié avec Johanna Kappus dans *Electronic Journal of Statistics* **8**(2): 2879-2904 (2014).

RÉSUMÉ. Dans ce chapitre, nous étudions le choix adaptatif d'un paramètre de lissage pour l'estimation de densité dans le problème de déconvolution quand la loi du bruit est inconnue. Pour que la densité cible soit identifiable, nous supposons qu'une information supplémentaire sur le bruit est disponible. Ainsi nous considérons en parallèle deux modèles : dans le premier nous supposons qu'un échantillon préliminaire du bruit est disponible, dans le deuxième nous supposons que la v.a. d'intérêt est observée de manière répétée avec des erreurs indépendantes. Nous introduisons des estimateurs avec un cutoff spectral et donnons une majoration du risque quadratique intégré. L'intérêt majeur de ce travail est le choix optimal de la fenêtre grâce à de la sélection de modèles par pénalisation conduisant à des inégalités oracles non-asymptotiques.

ABSTRACT. We investigate the data driven choice of the cutoff parameter in density deconvolution problems with unknown error distribution. To make the target density identifiable, one has to assume that some additional information on the noise is available. We consider two different models: the framework where some additional sample of the pure noise is available, as well as the model of repeated measurements, where the contaminated random variables of interest can be observed repeatedly, with independent errors. We introduce spectral cutoff estimators and present upper risk bounds. The focus of this work lies on the optimal choice of the bandwidth by penalization strategies, leading to non-asymptotic oracle bounds.

Contents

5.1	Introduction	119
5.2	Statistical model, estimation procedure and risk bounds	120
5.2.1	Notations	121
5.2.2	Statistical model and estimators	121
5.2.3	Upper risk bounds	123
5.3	Data driven bandwidth selection and oracle bounds	124
5.4	Illustrations	126
5.4.1	Practical estimation procedure	126
5.4.2	Comparison with Comte and Lacour (2011) and influence of M in the NS-model	127
5.4.3	Illustrations in the RD-model	128
5.4.4	Comparison with a kernel estimator	128
5.5	Concluding remarks	132
5.6	Proofs	133
5.6.1	Preliminaries	133
5.6.2	A technical auxiliary result	134
5.6.3	Proof of the oracle bounds	138

5.1 Introduction

This paper addresses the problem of the adaptive bandwidth selection via penalization in deconvolution problems with unknown error distribution. We study in parallel two different models. In both models, we assume that the random variables are real-valued.

Model 1. The random variable X of interest is perturbed by some additional additive error ε , independent of X , so the empirically accessible quantity is $Y = X + \varepsilon$. The distribution of the noise is assumed to be unknown. One observes n independent copies of Y :

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

In addition it is assumed that a sample $(\varepsilon_{-j})_{j=1, \dots, M}$ of the pure noise, independent of the Y_j , is available.

Model 2. The noisy random variable X can be measured repeatedly, with independent errors. The observations are then of the form,

$$Y_{j,k} = X_j + \varepsilon_{j,k}, \quad j = 1, \dots, n \quad \text{and} \quad k = 2.$$

All the X_j are assumed to be independent and identically distributed and all the $\varepsilon_{j,k}$ are independent and identically distributed and independent of the X_j . Again, it is assumed that the distribution of the $\varepsilon_{j,k}$ is unknown. In addition, the error terms are assumed to be centered.

Density deconvolution is a classical topic in nonparametric statistics and a large amount of literature on this subject has been published since the late 80s. Rates of convergence and their optimality have been studied, for example, in [Carroll and Hall \(1988\)](#), [Stefanski \(1990\)](#), [Stefanski and Carroll \(1990\)](#), [Fan \(1991\)](#) and [Efromovich \(1997\)](#). For the study of sharp asymptotic optimality, see [Butucea \(2004\)](#), [Butucea and Tsybakov \(2008a,b\)](#). Adaptive estimation for deconvolution problems has then been investigated by [Pensky and Vidakovic \(1999\)](#), who apply wavelet techniques, by [Comte et al. \(2006\)](#), who consider the adaptive bandwidth selection for projection estimators and by [Butucea and Comte \(2009\)](#) for linear functionals. We can also cite [Comte and Lacour \(2013\)](#) in a multivariate setting.

However, the above mentioned papers have been working under the assumption that the distribution of the errors is perfectly known, which is clearly not realistic in most fields of application. The systematic study of deconvolution with unknown error distribution in presence of an additional noise sample, which corresponds to Model 1, dates back to the late 90s. For the study of convergence rates, see [Neumann \(1997\)](#), [Johannes \(2009\)](#) or [Meister \(2009\)](#).

The rigorous study of adaptive procedures in a deconvolution model with unknown errors has only recently been addressed. We are aware of the work by [Comte and Lacour \(2011\)](#), by [Johannes and Schwarz \(2013a\)](#) who consider a model of circular deconvolution and by [Dattner et al. \(2016\)](#), who deal with adaptive quantile estimation via Lepski's method.

In comparison to the classical deconvolution model with known errors, the research on the model of repeated measurements has not been so intense. For the more general model of repeated measurements with skew error densities, rates of convergence have been studied in [Li and Vuong \(1998\)](#) and recently been improved in [Comte and Kappus \(2015\)](#). Consistent estimation under minimal a priori assumptions is investigated in [Neumann \(2007\)](#). For repeated measurements with symmetric error density, we refer to [Delaigle et al. \(2008\)](#) and [Comte et al. \(2013\)](#). This model has various applications in economics, see, for example, [Bonhomme and Robin \(2010\)](#), but also in a medical context, see [Comte et al. \(2013\)](#).

In the last mentioned paper, as well as in [Delaigle et al. \(2008\)](#), practical strategies for the adaptive bandwidth selection have been proposed, but a theoretical justification is not given which is a motivation for the rigorous study presented in this paper. Essential tools for our approach rely on considerations presented in [Kappus \(2014\)](#). There are also many common points with recent contributions by [Dattner et al. \(2016\)](#). In comparison to the last mentioned authors, the main difference lies in the fact that their approach is minimax and asymptotic whereas we are interested in non-asymptotic oracle bounds, thus following the model selection paradigm in the sense of [Birgé \(1999\)](#) and [Massart \(2003\)](#).

Model 1 corresponds, in many respects, to the situation which has been investigated in [Comte and Lacour \(2011\)](#). Let us clarify the essential differences: in a deconvolution model with estimated characteristic function of the errors, the risk bounds are determined by the size M of the noise sample, as well as the number n of observations of Y . For sample sizes $M \geq n$ the risk bounds correspond to the model with known error distribution. However, for $M < n$ the bound on the risk gets worse. The approach by [Comte and Lacour \(2011\)](#) is tailored for the case where M is, by a polynomial factor, larger than n , and cannot be extended to $M < n$. The additional considerations presented in the present work allow to handle the case of small noise samples. This seems to be of some practical relevance when one turns away from the classical measurement error model and regards, in some context of physics or biology, ε as a signal overlying some other signal X . In such a framework, the size of the noise sample will be determined by some extraneous influence and the assumption that $M > n$ may fail to hold true.

We want to emphasize another important difference between our reasoning and the arguments given in [Comte and Lacour \(2011\)](#), but also in [Dattner et al. \(2016\)](#). The last mentioned papers do always work under the standing assumption that the interesting density and error density belong to certain prescribed classes of functions. More precisely, it is assumed therein that the characteristic functions have an exponential or polynomial decay behavior. We are able to dispose completely of any of such semi-parametric assumptions.

This is motivated by arguments given in [Birgé \(1999\)](#). From a model selection point of view, rather than considering a family of parameter sets and aiming at building an estimator which is simultaneously asymptotically minimax, the target is to find the best estimator within a collection leading to non-asymptotic oracle inequalities. It is argued in [Birgé \(1999\)](#) that these considerations make sense without specifying any particular family of parameter sets. From this point of view, it is desirable to avoid, as far as possible, any a priori parametric assumptions and provide a fully general treatment.

This paper is organized as follows. In Section 5.2, we fix the notation and assumptions, introduce the estimators and present upper risk bounds. In Section 5.3, the data driven choice of the cutoff parameter is investigated. We introduce penalized criteria and derive non asymptotic oracle bounds for the corresponding estimators. In Section 5.4, we present some data examples to illustrate the practical performance of our estimator. All proofs are postponed to Section 5.6.

5.2 Statistical model, estimation procedure and risk bounds

In the present section we fix the statistical model and assumptions, introduce the estimators and recall, for the readers convenience, the non asymptotic risk bounds which have been presented in earlier publications on the subject. We start by introducing some notation which will be used throughout the rest of the text.

5.2.1 Notations

For two real numbers a and b , $a \vee b := \max(a, b)$ and $a \wedge b := \min(a, b)$. For two functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{C}$ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, $\|\varphi\|$ denotes the \mathbb{L}^2 -norm of φ , defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, and $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ , defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x)\overline{\psi(x)} dx$. The Fourier transform φ^* is defined by

$$\varphi^*(x) = \int e^{ixu} \varphi(u) du.$$

Besides, if φ^* belongs to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, then the function φ is the inverse Fourier transform of φ^* and can be written $\varphi(x) = 1/(2\pi) \int e^{-ixu} \varphi^*(u) du$. The convolution product $*$ is defined as $(\varphi * \psi)(x) = \int \varphi(x-u)\psi(u) du$. Lastly, we abbreviate Model 1 by NS-model, where NS stands for noise sample, and Model 2 by RD-model for replicate data.

5.2.2 Statistical model and estimators

In the situation of Model 1 and Model 2 defined in the introductory part, the target is to recover the density f of X from the data. In the sequel, we limit our considerations to the case where the number k of repeated measurements is equal to two. This setting easily generalizes to a higher dimensional model. However, for sake of clarity, we omit the details.

Throughout the paper, we make the following assumptions:

- (A1) X and ε have square integrable densities f and f_ε w.r.t. Lebesgue measure.
- (A2) $\forall x \in \mathbb{R}, f_\varepsilon^*(x) \neq 0$.
- (A3) f_ε is symmetric around zero.

Let f_Y denote the density of $X + \varepsilon$. By independence of X and ε , $f_Y = f * f_\varepsilon$. Under (A2), we have the equality $f^* = \hat{f}_Y^*/f_\varepsilon^*$. An estimator \hat{f}_Y^* of f_Y^* can be calculated from the data. If f_ε^* is known, an unbiased plug-in-estimator of f^* is then given by $\hat{f}_Y^*/f_\varepsilon^*$. The inverse Fourier transform is then applied to get an estimate of f . However, since neither \hat{f}_Y^* nor $1/f_\varepsilon^*$ are integrable, it is necessary to apply some regularization technique, for example, a spectral cutoff estimator. In this particular case, the estimator of f would be $1/(2\pi) \int_{|u| \leq \pi m} e^{-iu x} \hat{f}_Y^*(u)/f_\varepsilon^*(u) du$. We can notice that this estimator corresponds both to a kernel estimator built with a sinc kernel (Butucea (2004)) or to a projection type estimator as in Comte et al. (2006).

In the present case, the error distribution is assumed to be unknown. To make the problem identifiable, some additional information on the noise is required. In the NS-model, we introduce the empirical characteristic function of ε ,

$$\forall u \in \mathbb{R}, \hat{f}_{\varepsilon, \text{NS}}^*(u) = \frac{1}{M} \sum_{j=1}^M e^{iu\varepsilon_j}.$$

In the same way, f_Y^* is estimated by its empirical version

$$\forall u \in \mathbb{R}, \hat{f}_{Y, \text{NS}}^*(u) = \frac{1}{n} \sum_{j=1}^n e^{iuY_j}.$$

Secondly the identification of f_ε^* is also possible in the model of repeated observations. Under the symmetry assumption **(A3)**, we have the following equalities

$$\begin{aligned} \forall u \in \mathbb{R}, \mathbb{E} \left[e^{iu(Y_{n+j,1} - Y_{n+j,2})} \right] &= \mathbb{E} \left[e^{iu(\varepsilon_{n+j,1} - \varepsilon_{n+j,2})} \right] = \left| \mathbb{E} \left[e^{iu\varepsilon_{n+j,1}} \right] \right|^2 \\ &= \left(\mathbb{E} \left[e^{iu\varepsilon_{n+j,1}} \right] \right)^2 = (f_\varepsilon^*(u))^2. \end{aligned} \quad (5.1)$$

When **(A3)** is violated, the model is much more complicated and requires a completely different approach since f_ε^* is not a real positive function anymore. For further discussion, see [Comte and Kappus \(2015\)](#). Formula (5.1) suggests to define the following unbiased estimator of $(f_\varepsilon^*)^2$:

$$\forall u \in \mathbb{R}, \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u) = \frac{1}{n} \sum_{j=1}^n \cos(u(Y_{j,1} - Y_{j,2})).$$

Moreover, an unbiased estimator of f_Y^* is given by

$$\forall u \in \mathbb{R}, \widehat{f_{Y, \text{RD}}^*}(u) = \frac{1}{2n} \sum_{j=1}^n \left(e^{iuY_{j,1}} + e^{iuY_{j,2}} \right).$$

When considering the empirical characteristic functions, one has to be careful about the fact that the $Y_{j,k}$ are not independent of each other, nor independent of the $\varepsilon_{j,k}$.

Finally, one has to pay attention to the fact that small values of the empirical characteristic function in the denominator lead to unfavorable effects and a bad performance of the estimator. This is an immediate consequence of the fact that a reasonable estimation of the ratio $1/f_\varepsilon^*$ is impossible as soon as the denominator is smaller than the standard deviation. This phenomenon has been investigated in [Neumann \(1997\)](#). This entails the necessity to consider a regularized version of the empirical characteristic function in the denominator. [Delaigle et al. \(2008\)](#) propose a ridge-parameter approach. However, this requires a careful discussion of the choice of the ridge parameter. For this reason, we prefer the completely data driven approach proposed in [Neumann \(1997\)](#), which has also been applied by [Comte and Lacour \(2011\)](#) for the NS-model and by [Comte et al. \(2013\)](#) for the RD-model. This approach corresponds to the following estimators of the ratio:

$$\frac{1}{\widehat{f_{\varepsilon, \text{NS}}^*}(x)} = \frac{\mathbb{1} \left\{ |\widehat{f_{\varepsilon, \text{NS}}^*}(x)| \geq M^{-1/2} \right\}}{\widehat{f_{\varepsilon, \text{NS}}^*}(x)} \quad \text{and} \quad \frac{1}{\widehat{f_{\varepsilon, \text{RD}}^*}(x)} := \frac{\mathbb{1} \left\{ \widehat{f_{\varepsilon, \text{RD}}^{*2}}(x) \geq n^{-1/2} \right\}}{\sqrt{\widehat{f_{\varepsilon, \text{RD}}^{*2}}(x)}}.$$

These definitions lead to defining the empirical versions of f^* as follows:

$$\check{f}_{\text{NS}}^*(u) := \frac{\widehat{f_{Y, \text{NS}}^*}(u)}{\widehat{f_{\varepsilon, \text{NS}}^*}(u)} \quad \text{and} \quad \check{f}_{\text{RD}}^*(u) := \frac{\widehat{f_{Y, \text{RD}}^*}(u)}{\widehat{f_{\varepsilon, \text{RD}}^*}(u)}.$$

Finally, the objects to be estimated are characteristic functions, so the absolute values are bounded by 1. For this reason, one should prefer the regularized versions

$$\widehat{f}_{\text{NS}}^*(u) := \frac{\check{f}_{\text{NS}}^*(u)}{\max\{|\check{f}_{\text{NS}}^*(u)|, 1\}} \quad \text{and} \quad \widehat{f}_{\text{RD}}^*(u) := \frac{\check{f}_{\text{RD}}^*(u)}{\max\{|\check{f}_{\text{RD}}^*(u)|, 1\}}.$$

Given any positive, real valued m , the above considerations lead to defining the spectral cutoff estimators of f as follows:

$$\widehat{f}_{m, \text{NS}}(x) := \frac{1}{2\pi} \int e^{-iux} \widehat{f}_{m, \text{NS}}^*(u) \, du \quad \text{and} \quad \widehat{f}_{m, \text{RD}}(x) := \frac{1}{2\pi} \int e^{-iux} \widehat{f}_{m, \text{RD}}^*(u) \, du, \quad (5.2)$$

with

$$\hat{f}_{m,\text{NS}}^*(u) := \hat{f}_{\text{NS}}^*(u)\mathbb{1}_{[-\pi m, \pi m]}(u) \quad \text{and} \quad \hat{f}_{m,\text{RD}}^*(u) := \hat{f}_{\text{RD}}^*(u)\mathbb{1}_{[-\pi m, \pi m]}(u).$$

Moreover, we use the notation

$$f_m(x) := \frac{1}{2\pi} \int e^{-iux} f_m^*(u) du \quad \text{and} \quad f_m^*(u) := f^*(u)\mathbb{1}_{[-\pi m, \pi m]}(u).$$

5.2.3 Upper risk bounds

The following non-asymptotic risk bounds are valid for the estimators defined in the preceding section.

Proposition 5.2.1.

(i) *In presence of an additional noise sample, under (A1)-(A2), there exists a universal positive constant C such that*

$$\mathbb{E} \left\| f - \hat{f}_{m,\text{NS}} \right\|^2 \leq 2 \|f - f_m\|^2 + C \left(\frac{1}{n} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^2} du + \frac{1}{M} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2} du \right). \quad (5.3)$$

(ii) *In the model of repeated measurements, under (A1) - (A3), there exists a universal positive constant C' such that*

$$\begin{aligned} \mathbb{E} \left\| f - \hat{f}_{m,\text{RD}} \right\|^2 &\leq 2 \|f - f_m\|^2 + \frac{C'}{n} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^2} du \\ &\quad + \frac{C'}{n} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^2}{|f_\varepsilon^*(u)|^2 (|f_\varepsilon^*(u)|^2 \vee n^{-1/2})} du. \end{aligned} \quad (5.4)$$

Remark 7. The first two terms on the right-hand side of Equations (5.3) and (5.4) correspond to the usual terms when the distribution of the errors is known: the squared bias term and a bound on the variance. The last term is due to the estimation of f_ε^* which depends on the considered model. These bounds have already been established in the literature on deconvolution, see Neumann (1997), Delaigle et al. (2008) or Comte and Lacour (2011), Comte et al. (2013). We can hence omit the proof.

In view of the rates of convergence, we observe the following: consider first the NS-model. If $M \geq n$, there is no loss in the rate in comparison to a deconvolution problem with known f_ε . However, a loss in the rate may occur for $M < n$. More precisely, if the ratio M/n is small, in comparison to f^*/f_ε^* , the estimator does not achieve the rates of convergence which are known to be optimal for deconvolution with known error distribution. This is intuitive, since f can only be identified through f_ε and there is no hope to estimate f with high precision when the information on f_ε is not reliable. For a detailed discussion and minimax lower bounds, we refer to Neumann (1997). Next, consider the RD-model. From Equation (5.4), one derives immediately, that there is no loss in the rate, in comparison to deconvolution with known error distribution, if the decay of f^* outbalances the decay of f_ε^* . If this is no longer true, the following holds: the smoother f_ε is, in comparison to f , the worse are the resulting rates of convergence. It can be shown that a loss in the rate is unavoidable in this context (Alexander Meister, personal communication), but to the best of our knowledge, minimax lower bounds have not been published for this particular case.

5.3 Data driven bandwidth selection and oracle bounds

The goal of this section is to provide a strategy for the optimal data driven choice of the smoothing parameter. Given a collection \mathcal{M} of cutoff parameters, which may vary with M and n , the bandwidth \hat{m} should ideally outbalance the bias and variance term displayed in Equations (5.3) and (5.4). This trade-off is easier to realize when the variance term is assumed to be known, which is often the case in the literature on model selection. In a deconvolution problem with perfectly known error distribution \hat{m} should mimic the oracle choice

$$m_{th} = \arg \min_{m \in \mathcal{M}} \left\{ -\|f_m\|^2 + \frac{1}{n} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{(1 - |f_Y^*(u)|^2)}{|f_\varepsilon^*(u)|^2} du \right\}.$$

In the present framework, the considerations are even more involved since the characteristic function in the denominator is unknown and the variance is hence not feasible to actually compute. Following the model selection paradigm, see [Birgé and Massart \(1997\)](#), [Birgé \(1999\)](#) or [Massart \(2003\)](#), we select \hat{m} as the minimizer of a penalized criterion such that

$$\hat{m} = \arg \min_{m \in \mathcal{M}} \left\{ -\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m) \right\}.$$

The penalty term should be chosen large enough to annihilate the fluctuation of \hat{f}_m around its target, for all m in the model collection \mathcal{M} simultaneously, but on the other hand, should ideally be as close as possible to the variance term in order to preserve the non-asymptotic risk bounds. In a model selection problem with known variance, the penalty term is deterministic, which is no longer the case in the present situation.

Before introducing the stochastic penalty terms, we shall need the following definitions: for $\delta > 0$ and $u \in \mathbb{R}$,

$$w(u) := (\log(e + |u|))^{-\frac{1}{2} - \delta}.$$

Moreover

$$k_N(u) := N^{-1/2} (\log N)^{1/2} w(u)^{-1}, \quad N = n, M.$$

The weight function w has been introduced in [Neumann and Reiß \(2009\)](#) and the considerations presented in that paper, combined with ideas given in [Kappus \(2014\)](#) play an important role for our arguments. Since the penalty terms will involve an empirical version of the characteristic function in the denominator, the oracle inequalities depend on a precise control of the deviation of \hat{f}_ε^* from f_ε , simultaneously on the real line. It is shown in [Neumann and Reiß \(2009\)](#) that the distance between both object, weighted by w , is simultaneously small on the real axes. In the penalty terms, there will hence occur a loss of logarithmic order, in comparison to the variance term.

Let us now introduce the stochastic penalty terms. In the NS-model,

$$\begin{aligned} \widehat{\text{pen}}_{\text{NS}}(m) &:= \widehat{\text{pen}}_{1,\text{NS}}(m) + \widehat{\text{pen}}_{2,\text{NS}}(m), \\ &:= \frac{16}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|\tilde{f}_{\varepsilon,\text{NS}}^*(u)|^2} du + \frac{2}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_M^2(u) |\hat{f}_{Y,\text{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\text{NS}}^*(u)|^4} du, \end{aligned} \quad (5.5)$$

with $\tau_Y = \tau_\varepsilon = \sqrt{6}$.

In the RD-model,

$$\begin{aligned} \widehat{\text{pen}}_{\text{RD}}(m) &:= \widehat{\text{pen}}_{1,\text{RD}}(m) + \widehat{\text{pen}}_{2,\text{RD}}(m), \\ &:= \frac{16}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|\tilde{f}_{\varepsilon,\text{RD}}^*(u)|^2} du + \frac{4}{9\pi} \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_n^2(u) |\hat{f}_{Y,\text{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon,\text{RD}}^*(u)|^6} du, \end{aligned} \quad (5.6)$$

with τ_Y as previously and $\tau_\varepsilon = \sqrt{3}$.

The cutoff parameters are selected as the minimizers of the following penalized criteria.

$$\hat{m}_{\text{NS}} = \arg \min_{m \in \mathcal{M}} \left\{ -\|\hat{f}_{m,\text{NS}}\|^2 + \widehat{\text{pen}}_{\text{NS}}(m) \right\} \quad (5.7)$$

$$\hat{m}_{\text{RD}} = \arg \min_{m \in \mathcal{M}} \left\{ -\|\hat{f}_{m,\text{RD}}\|^2 + \widehat{\text{pen}}_{\text{RD}}(m) \right\}. \quad (5.8)$$

Before formulating the oracle bound for the corresponding estimators, let us introduce the deterministic counterparts of the stochastic penalty terms:

$$\begin{aligned} \text{pen}_{\text{NS}}(m) &:= \text{pen}_{1,\text{NS}}(m) + \text{pen}_{2,\text{NS}}(m) \\ &:= \frac{16}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|f_\varepsilon^*(u)|^2} du + \frac{2}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_M^2(u) |f^*(u)|^2}{|f_\varepsilon^*(u)|^4} du. \end{aligned}$$

and

$$\begin{aligned} \text{pen}_{\text{RD}}(m) &:= \text{pen}_{1,\text{RD}}(m) + \text{pen}_{2,\text{RD}}(m) \\ &:= \frac{16}{3\pi} \int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|f_\varepsilon^*(u)|^2} du + \frac{4}{9\pi} \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_n^2(u) |f^*(u)|^2}{|f_\varepsilon^*(u)|^4} du. \end{aligned}$$

We are now ready to formulate the oracle bounds and hence the main result of this paper:

Theorem 5.3.1.

(i) Let \mathcal{M} be a collection of cutoff parameters, with $\max \mathcal{M} \leq \sqrt{M \wedge n}$. Assume that **(A1)** and **(A2)** are satisfied. Let \hat{m}_{NS} be defined by (5.7) and $\hat{f}_{\hat{m}_{\text{NS}},\text{NS}}$ according to (5.2). Then there exists a universal positive constant C^{ad} and a positive constant C depending on the particular choice of τ and δ , but not on any of the underlying distributions, such that

$$\mathbb{E} \|f - \hat{f}_{\hat{m}_{\text{NS}},\text{NS}}\|^2 \leq C^{\text{ad}} \inf_{m \in \mathcal{M}} \left\{ \|f - f_m\|^2 + \text{pen}_{\text{NS}}(m) \right\} + \frac{C}{M \wedge n}.$$

(ii) Let \mathcal{M} be a collection of cutoff parameters with $\max \mathcal{M} \leq \sqrt{n}$. Assume that **(A1)** - **(A3)** are satisfied. Let \hat{m}_{RD} and $\hat{f}_{\hat{m}_{\text{RD}},\text{RD}}$ be defined according to (5.8) and (5.2). Then there is a universal positive constant C^{ad} and a positive constant C depending on the choice of γ and δ , but not on the underlying distributions, such that

$$\mathbb{E} \|f - \hat{f}_{\hat{m}_{\text{RD}},\text{RD}}\|^2 \leq C^{\text{ad}} \inf_{m \in \mathcal{M}} \left\{ \|f - f_m\|^2 + \text{pen}_{\text{RD}}(m) \right\} + \frac{C}{n}.$$

Remark 8. It is remarkable that we are able to establish non-parametric oracle bounds which make sense without specifying any particular semi-parametric model. Related problems are frequently discussed under specific a priori assumptions on the decay behavior of f^* and f_ε^* , see for example Comte et al. (2006) or Comte and Lacour (2011). In the present work, we can completely dispose of any such assumptions, so our approach is as general as possible.

Another interesting point about our considerations is the following: the only assumption imposed on the collection \mathcal{M} of cutoff parameters is some upper bound on the largest index. No further specification is necessary and we may work with an arbitrarily fine grid, allowing good approximation properties. This is a consequence of the fact that our proofs rely on

one sole application of the Talagrand inequality. Additional applications of the Bernstein inequality and sums over \mathcal{M} are not required.

Finally, it is worth emphasizing that, to the best of our knowledge, the non-asymptotic oracle bounds for the RD-model are completely new and the same is true for the bounds in the NS model with $M < n$.

We have considered the problem from a non-asymptotic perspective, but Theorem 5.3.1 entails, from the asymptotic and minimax point of view, the following observation: in those cases where minimax rates of convergence are known, the procedure achieves, up to a logarithmic loss, automatic adaptation over prescribed non-parametric function classes, typically Sobolev-spaces or classes of super-smooth functions.

5.4 Illustrations

5.4.1 Practical estimation procedure

Let us describe first the adaptive procedures as it is implemented for the both models:

- ▷ For $m \in \mathcal{M} = \{m_1, \dots, m_n\}$, compute $-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)$.
- ▷ Choose \hat{m} such as $\hat{m} = \arg \min_{m \in \mathcal{M}} \{-\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m)\}$.
- ▷ Compute $\hat{f}_{\hat{m}}(x) = \frac{1}{2\pi} \int_{-\pi\hat{m}}^{\pi\hat{m}} e^{-ixu} \frac{\hat{f}_Y^*(u)}{\hat{f}_\varepsilon^*(u)} du$.

As often in model selection methods, the values of the constants in the penalty, here denoted by τ_Y and τ_ε , which are obtained from the theory are too large in practice. Therefore a calibration step is required and done: for a small set of densities and different sample sizes the mean integrated squared error (MISE) is computed in order to determine admissible range for the values of the constants (see Baudry et al. (2012) for a description of this step). Finally, the penalties are chosen according to Equations (5.5) and (5.6) with τ_Y replaced by 0.6 and τ_ε by 0.3 for the NS-model and RD-model. Besides we consider the model collection $\mathcal{M} = \{m = k/10, 1 \leq k \leq 25\}$. In practice, one can take k in a much larger set and propose larger values for m ; the first selected value can be followed by another run of the estimation algorithm with a thinner grid of proposals around the selected value. We limited the set here because the proposed values seemed adequate and allowed less time-consuming repeated experiments.

In the sequel we also use the notations \hat{r}^{or} and \hat{r}^{ad} define as follows

$$\hat{r}^{or} = \min_{m \in \mathcal{M}} \hat{\mathbb{E}} \|f - \hat{f}_m\|^2 \quad \text{and} \quad \hat{r}^{ad} = \hat{\mathbb{E}} \|f - \hat{f}_{\hat{m}}\|^2,$$

where $\hat{\mathbb{E}}$ is the approximation of theoretical expectation computed via Monte-Carlo repetitions.

The whole implementation is conducted using R software. The integrated squared error (ISE) $\|f - \hat{f}_{\hat{m}}\|^2$ is computed via a standard approximation and discretization (over 300 points) of the integral on an interval of \mathbb{R} denoted by I . Then the MISE $\mathbb{E}\|f - \hat{f}_{\hat{m}}\|^2$ is computed as the empirical mean of the approximated ISE over 500 simulation samples.

5.4.2 Comparison with Comte and Lacour (2011) and influence of M in the NS-model

We compute different estimators of the signal for different values of M and consider different signal densities and two noises. Following Comte et al. (2006), we study the following densities on the interval I :

- ▷ Laplace distribution, $f(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, $I = [-5, 5]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [-1.5, 26]$.
- ▷ Cauchy distribution : $f(x) = (\pi(1 + x^2))^{-1}$, $I = [-10, 10]$.
- ▷ Standard Gaussian distribution, $I = [-4, 4]$.

All the densities are normalized with unit variance except the Cauchy density.

We consider the two following noise densities with same variance 1/10.

- ▷ **Gaussian noise** : $f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right)$, $f_\varepsilon^*(x) = \exp\left(-\frac{\sigma_\varepsilon^2 x^2}{2}\right)$.
- ▷ **Laplace noise** : $f_\varepsilon(x) = \frac{1}{2\sigma_\varepsilon} \exp\left(-\frac{|x|}{\sigma_\varepsilon}\right)$, $f_\varepsilon^*(x) = \frac{1}{1 + \sigma_\varepsilon^2 x^2}$.

We want to study the influence of the relationship of n and M on the estimation of f in the NS-model. We then consider different values of n and values of $M = \sqrt{n}$ and $M = n$.

Results. The results of the simulations are given in Table 5.1. Table 5.1 illustrates the case where we can recover a preliminary sample of the noise ε (our so-called NS-model). First we see that the risk decreases when the sample size increases. Likewise the risk increases when the variance increases. Secondly the results are very close to those of Comte and Lacour (2011). Nevertheless our procedure is equivalent or better. The main improvement is that our procedure has better performances when the size of the preliminary sample is small. Moreover we show that the procedure does not need big M since we reach the same performances as Comte and Lacour (2011) for $M = n$ when they take $M = n^2$. Indeed if we consider the mixed Gamma distribution for $n = M = 500$, our risk is 0.232 with a Laplace noise, while Comte and Lacour (2011) with $M = n^2$ have 0.382. We can make the same remarks for the other distributions (except for the Laplace distribution) with a sample size 100 or 250 and with the two noise distributions. For the Laplace distribution, the results of Comte and Lacour (2011) are better but they do not outperform ours. Nonetheless, in our model, the risk decreases more rapidly when M and n increase.

Effect of the variance.

We also test how our procedure behaves when the variance is increased. In Table 5.2, we present the results of simulations where the variance takes the values 1/4, 1/2 and 1. We only report the case of a Gaussian error distribution since the results of the Laplace error are very similar. Moreover a Gaussian distribution is a case less favorable. Indeed its Fourier transform decays exponentially and it is known to imply possibly slower asymptotic

f		f_ε	$n = 100$		$n = 250$		$n = 500$	
			Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.
Laplace	$M = \lfloor \sqrt{n} \rfloor$	\hat{r}^{ad}	2.779	2.714	1.790	1.933	1.304	1.403
		\hat{r}^{or}	(2.223)	(1.370)	(1.342)	(0.828)	(0.988)	(0.544)
	$M = n$	\hat{r}^{ad}	2.606	2.620	1.543	1.626	1.136	1.339
		\hat{r}^{or}	(1.848)	(1.357)	(1.107)	(0.786)	(0.796)	(0.539)
Mixed Gamma	$M = \lfloor \sqrt{n} \rfloor$	\hat{r}^{ad}	0.787	0.777	0.377	0.385	0.247	0.247
		\hat{r}^{or}	(0.671)	(0.710)	(0.331)	(0.337)	(0.205)	(0.209)
	$M = n$	\hat{r}^{ad}	0.751	0.725	0.360	0.365	0.232	0.234
		\hat{r}^{or}	(0.642)	(0.682)	(0.325)	(0.319)	(0.194)	(0.197)
Cauchy	$M = \lfloor \sqrt{n} \rfloor$	\hat{r}^{ad}	0.891	0.888	0.433	0.465	0.311	0.325
		\hat{r}^{or}	(0.721)	(0.731)	(0.386)	(0.418)	(0.361)	(0.275)
	$M = n$	\hat{r}^{ad}	0.817	0.806	0.402	0.416	0.285	0.301
		\hat{r}^{or}	(0.687)	(0.687)	(0.361)	(0.379)	(0.232)	(0.254)
Gaussian	$M = \lfloor \sqrt{n} \rfloor$	\hat{r}^{ad}	0.774	0.777	0.450	0.446	0.275	0.303
		\hat{r}^{or}	(0.657)	(0.655)	(0.388)	(0.415)	(0.190)	(0.211)
	$M = n$	\hat{r}^{ad}	0.666	0.644	0.406	0.392	0.247	0.272
		\hat{r}^{or}	(0.512)	(0.540)	(0.345)	(0.325)	(0.127)	(0.126)

Table 5.1: Values of approximated MISE $\mathbb{E} \left(\|f - \hat{f}_{\hat{m}_{NS,NS}}\|^2 \right) \times 100$ averaged over 500 samples with a Laplace and Gaussian noise of variance 1/10, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.

rates. Thus it is more difficult to recover the target density f .

Results. The results are reported in Table 5.2. As before the risk decreases when n and M increase. Similarly when we increase the contamination of the variable of interest by increasing the variance of the error distribution, the risk increases. The procedure performs still well since the adaptive risk is close to the oracle risk around twice bigger.

5.4.3 Illustrations in the RD-model

For this model, we use the same signal and error distributions as described in 5.4.2. We consider different values of n : 200 and 2000, with variance 1/10 and 1/2.

Results. The results are reported in Table 5.3. We note that the values of the MISE are very close for both error distributions. Moreover the adaptive risk is also close to the oracle risk. It is multiplied by approximately 1.5 when the variance of the error distribution is 1/10 and 2 when the variance equals 1/2. Again the risk decreases when n and M increase. And the risk increases when the variance increases.

5.4.4 Comparison with a kernel estimator

Recently, some papers as Comte et al. (2007) have investigated the necessity of inversion in statistical inverse problem. The idea is to compare the performances of a simple kernel estimator directly applied to the data Y_i with the adaptive estimator of the NS-model. This model allows us to choose a non symmetric noise which makes the estimation with a kernel estimator more difficult. That is why we compare our estimator with a kernel in two

			$\sigma_\varepsilon^2 = 1/4$		
f	M	n	100	250	500
Mixed Gamma	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	1.767	0.889	0.471
		\hat{r}^{or}	(0.808)	(0.473)	(0.299)
	n	\hat{r}^{ad}	1.231	0.620	0.387
		\hat{r}^{or}	(0.791)	(0.439)	(0.265)
Gaussian	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	1.663	0.920	0.543
		\hat{r}^{or}	(0.913)	(0.531)	(0.397)
	n	\hat{r}^{ad}	1.245	0.669	0.522
		\hat{r}^{or}	(0.578)	(0.280)	(0.180)

			$\sigma_\varepsilon^2 = 1/2$		
f	M	n	100	250	500
Mixed Gamma	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	2.170	1.229	0.673
		\hat{r}^{or}	(0.940)	(0.652)	(0.456)
Gaussian	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	1.951	1.185	0.867
		\hat{r}^{or}	(1.489)	(0.988)	(0.747)
	n	\hat{r}^{ad}	1.884	1.053	0.690

			$\sigma_\varepsilon^2 = 1$		
f	M	n	100	250	500
Mixed Gamma	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	2.798	1.973	1.232
		\hat{r}^{or}	(1.042)	(0.865)	(0.692)
	n	\hat{r}^{ad}	1.594	1.135	0.951
		\hat{r}^{or}	(0.979)	(0.776)	(0.585)
Gaussian	$\lfloor\sqrt{n}\rfloor$	\hat{r}^{ad}	3.163	2.139	1.898
		\hat{r}^{or}	(2.621)	(1.988)	(1.542)
	n	\hat{r}^{ad}	2.458	1.432	0.982
		\hat{r}^{or}	(1.400)	(0.848)	(0.613)

Table 5.2: Values of approximated MISE $\mathbb{E} \left(\|f - \hat{f}_{\hat{m}_{NS,NS}}\|^2 \right) \times 100$ averaged over 500 samples with a Gaussian noise, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.

contexts: one with a symmetric noise and another one with an asymmetric noise.

First, for the symmetric noise, we take the results of Table 5.2 with variance 1/2 for a Gaussian noise with $M = n$ and compute the kernel estimator for this design.

Secondly we compute the estimator of the signal in the NS-model with $M = \lfloor\sqrt{n}\rfloor$ with an asymmetric noise. We consider two densities defined in the beginning of this section: mixed Gamma and Gaussian. We choose the error distribution as follows

$$\triangleright \textbf{Gamma noise: } f_\varepsilon(x) = x^{\alpha-1} \frac{\beta^\alpha e^{-\beta x}}{\Gamma(\alpha)} \mathbb{1}_{x \geq 0}$$

with parameters $\alpha = 1$ and $\beta = 2$. The variance of the error distribution is then 1/4. For the kernel estimator, we use the function *density* of R with a Gaussian kernel where the bandwidth is selected by cross-validation.

Results. The results are reported in Tables 5.5 and 5.4. In Table 5.4, we see that the results are very close. For the mixed Gamma of the kernel the results are a little bit better

		$\sigma_\varepsilon^2 = 1/10$			
		$n = 200$		$n = 2000$	
f	f_ε	Lap.	Gaus.	Lap.	Gaus.
Laplace	\hat{r}^{ad}	1.706	1.683	0.655	0.725
	\hat{r}^{or}	(1.163)	(1.216)	(0.369)	(0.438)
Mixed Gamma	\hat{r}^{ad}	0.558	0.548	0.075	0.076
	\hat{r}^{or}	(0.381)	(0.391)	(0.061)	(0.060)
Cauchy	\hat{r}^{ad}	0.535	0.595	0.127	0.127
	\hat{r}^{or}	(0.358)	(0.377)	(0.068)	(0.069)
Gaussian	\hat{r}^{ad}	0.338	0.335	0.045	0.050
	\hat{r}^{or}	(0.231)	(0.235)	(0.034)	(0.035)

		$\sigma_\varepsilon^2 = 1/2$			
f	f_ε	Lap.	Gaus.	Lap.	Gaus.
Laplace	\hat{r}^{ad}	3.901	3.539	1.938	2.007
	\hat{r}^{or}	(2.904)	(2.710)	(1.234)	(1.536)
Mixed Gamma	\hat{r}^{ad}	0.857	0.799	0.248	0.226
	\hat{r}^{or}	(0.583)	(0.550)	(0.144)	(0.133)
Cauchy	\hat{r}^{ad}	1.172	1.027	0.312	0.242
	\hat{r}^{or}	(0.599)	(0.617)	(0.181)	(0.174)
Gaussian	\hat{r}^{ad}	0.703	0.586	0.168	0.155
	\hat{r}^{or}	(0.441)	(0.399)	(0.092)	(0.076)

Table 5.3: Values of approximated MISE $\mathbb{E}(\|f - \hat{f}_{\hat{m}_{RD, RD}}\|^2) \times 100$ averaged over 500 samples, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.

but the risk of our estimator decreases more rapidly. For the Gaussian distribution our results are better and the risk decreases also more rapidly when the sample size increases..

For the non symmetric noise the results are reported in Table 5.5. We see that for the mixed Gamma, the kernel estimator performs unexpectedly well. For $n = 100$, the risks are practically the same as those of the adaptive estimator. When n increases the adaptive estimator performs better but the kernel estimator still gives satisfying results.

On the other hand, the results of the Gaussian distribution illustrate well the importance of inversion in statistical inverse problem. Indeed the risk with a kernel estimator is around 3.10^{-2} for the diverse values of n while for the adaptive estimator the risk is divided by 3. More precisely for $n = 100$ the risk of the kernel estimator is multiplied by 2 compared to the adaptive estimator, by 5 for $n = 250$ and by 7 for $n = 500$. So when the error distribution is unknown and the signal to noise ratio (named $s2n$) is not too large, our procedure is worthy of interest. When the $s2n$ gets larger, we only expect the deconvolution procedure not to deteriorate the results compared to direct estimation. It has been checked to be true for known noise density by Comte et al. (2007) see section 4.7 and 4.9. As in practice the value of the $s2n$ is unknown, our procedure is always recommended.

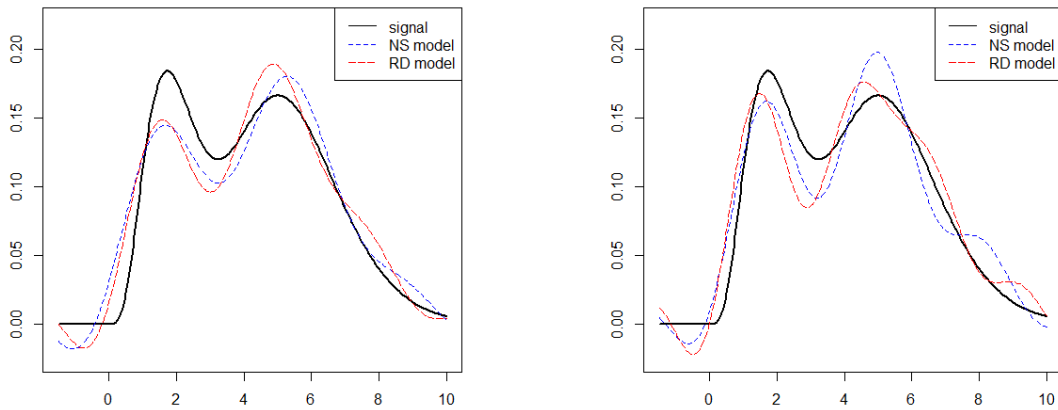
Figures 5.1 and 5.2 illustrate the estimation of a mixed Gamma using both penalized estimators in the cases for $n = M = 200$ and $n = M = 500$. The estimation is made with additional Gaussian and Laplace noises with a variance of $1/10$. The bimodal specificity of the density is well described. Moreover the precision increases with the sample size.

f		$n = 100$	$n = 250$	$n = 500$
Mixed	\hat{r}^{ker}	0.849	0.648	0.526
Gamma	$M = n$ \hat{r}^{ad}	1.483	0.835	0.554
	\hat{r}^{or}	(0.895)	(0.581)	(0.386)
Gaussian	\hat{r}^{ker}	1.673	1.333	1.145
	$M = n$ \hat{r}^{ad}	1.884	1.053	0.690
	\hat{r}^{or}	(0.832)	(0.444)	(0.292)

Table 5.4: Comparison of the method with the results of a Gaussian kernel estimator with a symmetric noise (MISE $\times 100$ averaged over 500 simulations), \hat{r}^{ker} is the risk of the kernel estimator, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.

f		$n = 100$	$n = 250$	$n = 500$
Mixed	\hat{r}^{ker}	1.086	0.881	0.785
Gamma	$M = \lfloor \sqrt{n} \rfloor$ \hat{r}^{ad}	1.063	0.573	0.332
	\hat{r}^{or}	(0.714)	(0.428)	(0.258)
Gaussian	\hat{r}^{ker}	3.181	3.052	2.807
	$M = \lfloor \sqrt{n} \rfloor$ \hat{r}^{ad}	1.364	0.677	0.449
	\hat{r}^{or}	(0.844)	(0.449)	(0.329)

Table 5.5: Comparison of the method with the results of a Gaussian kernel estimator with an asymmetric noise (MISE $\times 100$ averaged over 500 simulations), \hat{r}^{ker} is the risk of the kernel estimator, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.



(a) Laplace noise

(b) Gaussian noise

Figure 5.1: Estimations for $n = M = 200$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.

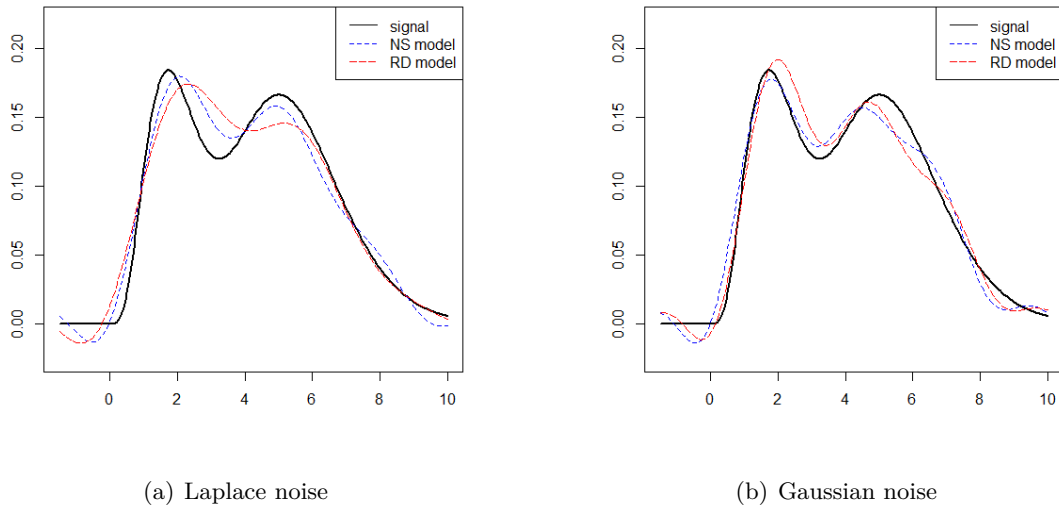


Figure 5.2: Estimations for $n = M = 500$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.

5.5 Concluding remarks

This paper deals with adaptive deconvolution estimation of a density when the noise density is unknown. We have considered two cases: one where a preliminary sample of the noise can be observed and another one where the variable of interest X can be observed repeatedly with independent errors. For both models, we have proposed a theoretical adaptive procedure which automatically makes a data driven bias-variance compromise. Moreover it allows us to not specify rates of convergence since they are mechanically reached. This procedure enables us to treat the problem of adaptive estimation in repeated observation model which is completely new. The estimation procedure relies on the independence of the estimators of the characteristic functions of f_Y^* and f_ε^* . Its advantage is to be very general under weak assumptions. Indeed, that procedure takes into account cases where there can be small number of replications which matches realistic applications as in medicine or economics. Besides of its theoretical properties, our procedure has showed good performances in simulation.

At last we think that our procedure can be extended to density estimation of a random effect in linear mixed-effects model. Indeed we are aware of the work of [Comte and Samson \(2012\)](#) who proposed an adaptive procedure based on deconvolution methods in the unknown-error case which is not optimal and [Dion \(2014\)](#) who used a Lepski's method in the known-error case. In that model, the noise can also be recovered by successive difference similarly to the repeated model but the characteristic function of the noise would be raised to a greater power. We may then propose in the same spirit an adaptive procedure for the random effect in linear mixed-effects model.

5.6 Proofs

5.6.1 Preliminaries

We start by restating, for the reader's convenience, the following version of Talagrand's inequality:

Lemma 5.6.1. *Let I be some countable index set. For each $i \in I$, let $Z_1^{(i)}, \dots, Z_n^{(i)}$ be independent and identically distributed random variables with values in $[-1, 1]$, defined on the same probability space. Let $v^2 := \sup_{i \in I} \text{Var}[Z_1^{(i)}]$ and $S_n^{(i)} := 1/n \sum_{j=1}^n Z_j^{(i)}$. Then there are universal positive constants c_1 and c_2 such that for any $\kappa > 0$,*

$$\mathbb{P} \left[\left\{ \sup_{i \in I} |S_n^{(i)}| \leq \frac{3}{2} \mathbb{E} \left[\sup_{i \in I} |S_n^{(i)}| \right] + \kappa \right\} \right] \leq 2 \exp \left(- \left(\frac{n\kappa^2}{c_1 v^2 + c_2 \kappa} \right) \right).$$

A proof of this result is given in [Massart \(2003\)](#), see page 170. It follows from the arguments presented therein that for any $r, s > 0$ with $\frac{1}{s} + \frac{1}{r} = 1$, the constants can be chosen $c_1 = 2s^2$ and $c_2 = 6r$.

The following result will be essential for the theoretical justification of the adaptive procedure. The proof was given in [Kappus \(2014\)](#). It uses a fundamental Lemma shown in [Neumann and Reiß \(2009\)](#), combined with Lemma 5.6.1.

Lemma 5.6.2. *Let Z_1, \dots, Z_n be i.i.d. random variables. Assume that $\mathbb{E}|Z_1| \leq m_Z$ for some $m_Z > 0$. Let \hat{f}_Z^* denote the empirical characteristic function. For some $\delta > 0$, let*

$$w(u) = (\log(e + |u|))^{-1/2-\delta}.$$

Assume that $\tau > 2\sqrt{p}$. Then there exists a positive constant C depending on m_Z and the choice of γ and δ such that for arbitrary $n \in \mathbb{N}$:

$$\mathbb{P} \left[\exists u \in \mathbb{R} : |\tilde{f}_Z^*(u) - f_Z^*(u)| > \tau (\log(n))^{1/2} w(u)^{-1} n^{-1/2} \right] \leq C n^{-p} \quad (5.9)$$

Remark 9. In the situation of the preceding Lemma: if the Z_j are assumed to have a symmetric distribution and the empirical characteristic function is defined to be $\hat{f}_Z^*(u) = 1/n \sum_{j=1}^n \cos(uZ_j)$, it is enough to assume $\tau > \sqrt{2p}$ to obtain (5.9).

The next result has been formulated and proved in ([Neumann, 1997](#)).

Lemma 5.6.3. *Let Z_1, \dots, Z_n be i.i.d. random variables. Let f_Z^* be the true and \hat{f}_Z^* the empirical characteristic function. Moreover let $1/\tilde{f}_Z^*(u) = \mathbb{1}\{|\hat{f}_Z^*(u)| \geq n^{-1/2}\}/\hat{f}_Z^*(u)$. Then for arbitrary $p \in \mathbb{N}$, there exists a positive constant C depending only on p such that*

$$\mathbb{E} \left[\left| \frac{1}{\tilde{f}_Z^*(u)} - \frac{1}{f_Z^*(u)} \right|^p \right] \leq C \left(\frac{1}{|f_Z^*(u)|^p} \wedge \frac{n^{-\frac{p}{2}}}{|f_Z^*(u)|^{2p}} \right). \quad (5.10)$$

Remark 10. In the preceding lemma, one should keep in mind that in the RD-model, $Y_{j,1} - Y_{j,2} = \varepsilon_{j,1} - \varepsilon_{j,2}$ plays the role of Z_j .

Lemma 5.6.3 yields the following useful corollary, which allows to compare the stochastic penalty terms to their deterministic counterparts.

Corollary 5.6.4. *Let the penalty terms be defined according to Equations (5.5) and (5.6). Then for some $C > 0$,*

$$\mathbb{E} [\widehat{\text{pen}}_{\text{NS}}(m)] \leq C \text{pen}_{\text{NS}}(m) \quad \text{and} \quad \mathbb{E} [\widehat{\text{pen}}_{\text{RD}}(m)] \leq C \text{pen}_{\text{RD}}(m).$$

Proof. It is enough to consider $\widehat{\text{pen}}_{\text{RD}}$. Equation (5.10) gives immediately for some $C > 0$,

$$\mathbb{E} \left[\int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du \right] \leq C \int_{-\pi m}^{\pi m} \frac{\tau_Y^2 k_n^2(u)}{|f_\varepsilon^*(u)|^2} du.$$

The Cauchy-Schwarz inequality and the estimate $1/|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 \leq n^{1/2}$ imply

$$\begin{aligned} & \mathbb{E} \left[\int_{-\pi m}^{\pi m} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{\tau_\varepsilon^2 k_n^2(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^6} du \right] \\ & \leq 2 \int_{-\pi m}^{\pi m} \mathbb{E}^{1/2} \left[|\hat{f}_{Y, \text{RD}}^*(u) - f_{Y, \text{RD}}^*(u)|^4 \right] \mathbb{E}^{1/2} \left[\frac{\tau_\varepsilon^2 k_n^2(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^{12}} \right] du \\ & \quad + 2 \int_{-\pi m}^{\pi m} \mathbb{E} \left[\frac{\tau_\varepsilon^2 k_n^2(u) |f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^6} \right] du \\ & \leq 2 \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_n^2(u)}{|f_\varepsilon^*(u)|^2} du + 2 \int_{-\pi m}^{\pi m} \frac{\tau_\varepsilon^2 k_n^2(u) |f^*(u)|^2}{|f_\varepsilon^*(u)|^2 (|f_\varepsilon^*(u)|^2 \vee n^{-1/2})} du. \end{aligned}$$

This completes the proof for the RD-model. The arguments for the NS-model are the same, line for line, so we omit the details. The only difference lies in the fact that, in the definition of $\hat{f}_{\varepsilon, \text{NS}}^*$ and $\tilde{f}_{\varepsilon, \text{NS}}^*$, M plays now the role of n as defined in Lemma 5.6.3. \square

5.6.2 A technical auxiliary result

In the sequel, for arbitrary $k > m$, we use the notation

$$\widehat{\text{pen}}_{\text{NS}}(m, k) := \widehat{\text{pen}}_{\text{NS}}(k) - \widehat{\text{pen}}_{\text{NS}}(m) \quad \text{and} \quad \widehat{\text{pen}}_{\text{RD}}(m, k) := \widehat{\text{pen}}_{\text{RD}}(k) - \widehat{\text{pen}}_{\text{RD}}(m),$$

as well as

$$\begin{aligned} \widehat{\text{pen}}_{\ell, \text{NS}}(m, k) &:= \widehat{\text{pen}}_{\ell, \text{NS}}(k) - \widehat{\text{pen}}_{\ell, \text{NS}}(m) \\ \widehat{\text{pen}}_{\ell, \text{RD}}(m, k) &:= \widehat{\text{pen}}_{\ell, \text{RD}}(k) - \widehat{\text{pen}}_{\ell, \text{RD}}(m), \quad \ell = 1, 2. \end{aligned}$$

Moreover,

$$A(m, k) := \{u \in \mathbb{R} : m \leq |u| \leq k\}.$$

The proof of Theorem 3.1 relies on the following auxiliary result which is, in turn, a consequence of Lemma 5.6.2.

Proposition 5.6.5.

(i) *In the model of repeated measurements, there exists an universal positive constant C such that*

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \left\| \hat{f}_{k, \text{RD}} - \hat{f}_{m, \text{RD}} \right\|^2 - 6 \|f_k - f_m\|^2 - \frac{3}{4} \widehat{\text{pen}}_{\text{RD}}(m, k) \right\}_+ \right] \leq \frac{C}{n}.$$

(ii) If an additional sample of the pure noise is available, for some universal positive constant C ,

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \left\| \hat{f}_{k, \text{NS}} - \hat{f}_{m, \text{NS}} \right\|^2 - 6 \|f_k - f_m\|^2 - \frac{3}{4} \widehat{\text{pen}}_{\text{NS}}(m, k) \right\}_+ \right] \leq \frac{C}{n \wedge M}.$$

Proof.

(i) Let us introduce the favorable events

$$\begin{aligned} \mathcal{E}_{Y, \text{RD}} &:= \left\{ \forall u \in \mathbb{R} : |\hat{f}_{Y, \text{RD}}^*(u) - f_Y^*(u)| \leq \tau_Y k_n(u) \right\}. \\ \mathcal{E}_{\varepsilon, \text{RD}} &:= \left\{ \forall u \in \mathbb{R} : |\widehat{f_{\varepsilon, \text{RD}}^{*2}}(u) - f_{\varepsilon}^{*2}(u)| \leq \tau_{\varepsilon} k_n(u) \right\}. \end{aligned}$$

Applying Parseval's equality we can estimate

$$\begin{aligned} \|\hat{f}_{k, \text{RD}} - \hat{f}_{m, \text{RD}}\|^2 &= \frac{1}{2\pi} \|\hat{f}_{k, \text{RD}}^* - \hat{f}_{m, \text{RD}}^*\|^2 = \frac{1}{2\pi} \int_{A(m, k)} \frac{|\hat{f}_{Y, \text{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du \\ &\leq \frac{1}{\pi} \int_{A(m, k)} \frac{|f_{Y, \text{RD}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du + \frac{1}{\pi} \int_{A(m, k)} \frac{|f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du. \end{aligned} \quad (5.11)$$

We start by dealing with the first summand appearing in the last line of (5.11). The definition of $\mathcal{E}_{Y, \text{RD}}$ and the fact that

$$\frac{1}{\pi} \int_{A(m, k)} \frac{\tau_Y^2 k_n^2(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du \leq \frac{3}{8} \widehat{\text{pen}}_{1, \text{RD}}(m, k)$$

immediately imply the following inequality

$$\begin{aligned} &\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{1}{\pi} \int_{A(m, k)} \frac{|f_{Y, \text{RD}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du - \frac{3}{8} \widehat{\text{pen}}_{1, \text{RD}}(m, k) \right\}_+ \mathbf{1}_{\mathcal{E}_{Y, \text{RD}}} \\ &\leq \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{1}{\pi} \int_{A(m, k)} \frac{\tau_Y^2 k_n^2(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du - \frac{3}{8} \widehat{\text{pen}}_{1, \text{RD}}(m, k) \right\}_+ \mathbf{1}_{\mathcal{E}_{Y, \text{RD}}} = 0. \end{aligned}$$

Consider now the second summand in the last line of (5.11). Recall that f_{ε}^* and $\widehat{f_{\varepsilon, \text{RD}}^{*2}}$ are real valued and (formally, with $c/\infty := 0$),

$$|\tilde{f}_{\varepsilon, \text{RD}}(u)|^2 = \begin{cases} \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u), & \text{if } \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u) \geq n^{-1/2} \\ \infty, & \text{else.} \end{cases}$$

This entails the series of inequalities

$$\begin{aligned}
& \frac{1}{\pi} \int_{A(m,k)} \frac{|f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du - 6 \|f_k - f_m\|^2 = \frac{1}{\pi} \int_{A(m,k)} \frac{|f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du - \frac{3}{\pi} \|f_k^* - f_m^*\|^2 \\
& = \frac{1}{\pi} \int_{A(m,k)} |f_Y^*(u)|^2 \left(\frac{1}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} - \frac{3}{|f_\varepsilon^*(u)|^2} \right) du \\
& \leq \frac{1}{\pi} \int_{A(m,k)} |f_Y^*(u)|^2 \left(\frac{1}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} - \frac{3}{|f_\varepsilon^*(u)|^2} \right) \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du \\
& \leq \frac{2}{3\pi} \int_{A(m,k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{f_\varepsilon^{*2}(u) - 3\widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^4} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du \\
& \quad + \frac{2}{\pi} \int_{A(m,k)} |f_Y^*(u) - \hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{f_\varepsilon^{*2}(u) - 3\widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 |f_\varepsilon^*(u)|^2} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du. \quad (5.12)
\end{aligned}$$

For the expression appearing in the last line of formula (5.12), we observe that

$$\begin{aligned}
& \frac{2}{\pi} \int_{A(m,k)} |f_Y^*(u) - \hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{f_\varepsilon^{*2}(u) - 3\widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 |f_\varepsilon^*(u)|^2} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du \\
& \leq \frac{2}{\pi} \int_{A(m,k)} |f_Y^*(u) - \hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 |f_\varepsilon^*(u)|^2} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du \\
& = \frac{2}{\pi} \int_{A(m,k)} \frac{|f_Y^*(u) - \hat{f}_{Y, \text{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du.
\end{aligned}$$

The definition of $\widehat{\text{pen}}_{1, \text{RD}}$ and $\mathcal{E}_{Y, \text{RD}}$ readily imply that

$$\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{2}{\pi} \int_{A(m,k)} \frac{|f_Y^*(u) - \hat{f}_{Y, \text{RD}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2} du - \frac{3}{8} \widehat{\text{pen}}_{1, \text{RD}}(m, k) \right\}_+ \mathbb{1}_{\mathcal{E}_{Y, \text{RD}}} = 0$$

Consider now the second to last line in Equation (5.12). We observe that $3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 \leq |f_\varepsilon^*(u)|^2$ implies $|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2 = \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u) \leq (1/2)|f_\varepsilon^*(u)|^2 - \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)$. From this, we derive that

$$\frac{f_\varepsilon^*(u) - 3\widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^4} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} \leq \frac{1}{2} \frac{|f_\varepsilon^*(u)|^2 - \widehat{f_{\varepsilon, \text{RD}}^{*2}}(u)}{|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^6}.$$

Using this and the definition of $\mathcal{E}_{\varepsilon, \text{RD}}$ and $\widehat{\text{pen}}_{2, \text{RD}}$, we arrive at

$$\begin{aligned}
& \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \int_{A(m,k)} \frac{|\hat{f}_{Y, \text{RD}}^*(u)|^2 |f_\varepsilon^*(u)|^2 - \tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2}{3/2\pi |\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^4} \mathbb{1}_{\{|f_\varepsilon^*(u)|^2 \geq 3|\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2\}} du \right. \\
& \quad \left. - \frac{3}{4} \widehat{\text{pen}}_{2, \text{RD}}(m, k) \right\}_+ \mathbb{1}_{\mathcal{E}_{\varepsilon, \text{RD}}} \\
& \leq \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \int_{A(m,k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u)|^2 - \tilde{f}_{\varepsilon, \text{RD}}^*(u)|^2}{3\pi |\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^6} du - \frac{3}{4} \widehat{\text{pen}}_{2, \text{RD}}(m, k) \right\}_+ \mathbb{1}_{\mathcal{E}_{\varepsilon, \text{RD}}} \\
& \leq \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \int_{A(m,k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{\tau_\varepsilon^2 k_n^2(u)}{3\pi |\tilde{f}_{\varepsilon, \text{RD}}^*(u)|^6} du - \frac{3}{4} \widehat{\text{pen}}_{2, \text{RD}}(m, k) \right\}_+ \mathbb{1}_{\mathcal{E}_{\varepsilon, \text{RD}}} = 0.
\end{aligned}$$

Putting the above together, we have shown

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \|\hat{f}_{k, \text{RD}} - \hat{f}_{m, \text{RD}}\|^2 - 6\|f_k - f_m\|^2 - \frac{3}{4} \widehat{\text{pen}}_{\text{RD}}(m, k) \right\}_+ \mathbf{1}_{\mathcal{E}_{Y, \text{RD}} \cap \mathcal{E}_{\varepsilon, \text{RD}}} \right] = 0.$$

There remains to consider the exceptional set $\mathcal{E}_{Y, \text{RD}}^c \cup \mathcal{E}_{\varepsilon, \text{RD}}^c$. Using the fact that for arbitrary m , the absolute value of $\hat{f}_{m, \text{RD}}^*$ is, by definition, bounded by 1, as well as the fact that $\max \mathcal{M} \leq \sqrt{n}$, we can estimate

$$\begin{aligned} & \mathbb{E} \left[\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \|\hat{f}_{k, \text{RD}}^* - \hat{f}_{m, \text{RD}}^*\|^2 - 6\|f_k^* - f_m^*\|^2 - \frac{3}{4} \widehat{\text{pen}}_{\text{RD}}(m, k) \right\}_+ \mathbf{1}_{(\mathcal{E}_{\varepsilon, \text{RD}} \cap \mathcal{E}_{Y, \text{RD}})^c} \right] \\ & \leq \mathbb{E} \left[\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \|\hat{f}_{k, \text{RD}}^* - \hat{f}_{m, \text{RD}}^*\|^2 \mathbf{1}_{(\mathcal{E}_{\varepsilon, \text{RD}} \cap \mathcal{E}_{Y, \text{RD}})^c} \right] \\ & \leq \sqrt{n} \mathbb{P}((\mathcal{E}_{\varepsilon, \text{RD}} \cap \mathcal{E}_{Y, \text{RD}})^c) \leq \sqrt{n} \left(\mathbb{P}(\mathcal{E}_{Y, \text{RD}}^c) + \mathbb{P}(\mathcal{E}_{\varepsilon, \text{RD}}^c) \right). \end{aligned}$$

Thanks to Lemma 5.6.2, it holds that

$$\sqrt{n} \left(\mathbb{P}(\mathcal{E}_{Y, \text{RD}}^c) + \mathbb{P}(\mathcal{E}_{\varepsilon, \text{RD}}^c) \right) \leq \frac{C}{n}.$$

This completes the proof of part (i).

(ii) The proof of the second part uses essentially the same arguments as the proof of the first part, so we content ourselves with sketching the important steps.

$$\begin{aligned} \mathcal{E}_{Y, \text{NS}} &:= \left\{ \forall u \in \mathbb{R} : |\hat{f}_{Y, \text{NS}}^*(u) - f_Y^*(u)| \leq \tau_Y k_n(u) \right\} \\ \mathcal{E}_{\varepsilon, \text{NS}} &:= \left\{ \forall u \in \mathbb{R} : |\hat{f}_{\varepsilon, \text{NS}}^*(u) - f_\varepsilon^*(u)| \leq \tau_\varepsilon k_M(u) \right\}. \end{aligned}$$

In analogy with (5.11) and (5.12), we find that

$$\begin{aligned} & \|\hat{f}_{k, \text{NS}}^* - \hat{f}_{m, \text{NS}}^*\|^2 - 6\|f_k^* - f_m^*\|^2 \\ & \leq \frac{2}{\pi} \int_{A(m, k)} \frac{|\hat{f}_{Y, \text{NS}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{NS}}^*(u)|^2} du + \frac{1}{3\pi} \int_{A(m, k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u)|^2 - 3|\hat{f}_{\varepsilon, \text{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{NS}}^*(u)|^4} du. \end{aligned}$$

Again, on $\mathcal{E}_{Y, \text{NS}}$,

$$\sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{2}{\pi} \int_{A(m, k)} \frac{|\hat{f}_{Y, \text{NS}}^*(u) - f_Y^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{NS}}^*(u)|^2} du - \frac{3}{4} \widehat{\text{pen}}_{1, \text{NS}}(m, k) \right\}_+ = 0.$$

Using the fact that for $a > 0$, $x, y \in \mathbb{C}$, $|x + y|^2 \leq (1 + a)|x|^2 + (1 + 1/a)|y|^2$ holds, we find that

$$\begin{aligned} & \frac{1}{3\pi} \int_{A(m, k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u)|^2 - 3|\hat{f}_{\varepsilon, \text{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{NS}}^*(u)|^4} du \\ & \leq \frac{1}{2\pi} \int_{A(m, k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u) - \hat{f}_{\varepsilon, \text{NS}}^*(u)|^2}{|\tilde{f}_{\varepsilon, \text{NS}}^*(u)|^4} du. \end{aligned}$$

Consequently, on $\mathcal{E}_{\varepsilon, \text{NS}}$,

$$\begin{aligned} & \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{1}{3\pi} \int_{A(m, k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{|f_\varepsilon^*(u)|^2 - 3|\hat{f}_{\varepsilon, \text{NS}}^*(u)|^2}{|\hat{f}_{\varepsilon, \text{NS}}^*(u)|^4} du - \frac{3}{4} \widehat{\text{pen}}_{2, \text{NS}}(m, k) \right\}_+ \\ & \leq \sup_{\substack{k \geq m \\ k, m \in \mathcal{M}}} \left\{ \frac{1}{2\pi} \int_{A(m, k)} |\hat{f}_{Y, \text{RD}}^*(u)|^2 \frac{\tau_\varepsilon^2 k_M^2(u)}{|\hat{f}_{\varepsilon, \text{NS}}^*(u)|^4} du - \frac{3}{4} \widehat{\text{pen}}_{2, \text{NS}}(m, k) \right\}_+ = 0. \end{aligned}$$

Finally, we find that

$$\mathbb{P} \left[\mathcal{E}_{Y, \text{NS}}^c \cup \mathcal{E}_{\varepsilon, \text{NS}}^c \right] \leq C(n^{-3/2} + M^{-3/2}),$$

leading to the desired bound on the complement set. \square

5.6.3 Proof of the oracle bounds

Proposition 5.6.5 was the essential technical step towards proving the oracle bounds. It shows how the penalty terms control the fluctuation of $\|\hat{f}_m^*\|^2$ around its target, be it in the model of repeated measurements or in the deconvolution model with additional noise sample. Once this result is fixed, the considerations for both models run exactly along the same line. For this reason, we may henceforth drop all subscripts relative to one particular model.

Proof of Theorem 5.3.1

We denote by m^* the oracle cutoff,

$$m^* = \arg \min_{m \in \mathcal{M}} \left\{ -\|\hat{f}_m\|^2 + \text{pen}(m) \right\}.$$

We have

$$\|f - \hat{f}_{\hat{m}}\|^2 \leq 2\|f - \hat{f}_{m^*}\|^2 + 2\|\hat{f}_{m^*} - \hat{f}_{\hat{m}}\|^2. \quad (5.13)$$

Taking expectation and applying Proposition 5.2.1 gives for the first summand on the right hand side

$$\mathbb{E} \left[\|f - \hat{f}_{m^*}\|^2 \right] \leq 2\|f - f_{m^*}\|^2 + \text{pen}(m^*),$$

since the variance term is a fortiori bounded from above by the penalty term. So there remains to consider the second summand on the right hand side of (5.13).

- Consider first the set $\mathcal{G} = \{\hat{m} \leq m^*\}$. Let us notice that on \mathcal{G}

$$\|\hat{f}_{m^*} - \hat{f}_{\hat{m}}\|^2 \mathbf{1}_{\mathcal{G}} = \left(\|\hat{f}_{m^*}\|^2 - \|\hat{f}_{\hat{m}}\|^2 \right) \mathbf{1}_{\mathcal{G}}.$$

Besides according to the definition of \hat{m} , one has the following inequality:

$$-\|\hat{f}_{\hat{m}}\|^2 + \widehat{\text{pen}}(\hat{m}) \leq -\|\hat{f}_{m^*}\|^2 + \widehat{\text{pen}}(m^*), \quad (5.14)$$

which implies

$$-\|\hat{f}_{\hat{m}}\|^2 \leq -\|\hat{f}_{m^*}\|^2 + \widehat{\text{pen}}(m^*).$$

Thus

$$\|\hat{f}_{m^*} - \hat{f}_{\hat{m}}\|^2 \mathbf{1}_{\mathcal{G}} = \left(\|\hat{f}_{m^*}\|^2 - \|\hat{f}_{\hat{m}}\|^2 \right) \mathbf{1}_{\mathcal{G}} \leq \widehat{\text{pen}}(m^*).$$

Taking expectation and applying Corollary 5.6.4 yields for some positive constant C ,

$$\mathbb{E} \left[\|\hat{f} - \hat{f}_{\hat{m}}\|^2 \mathbf{1}_{\mathcal{G}} \right] \leq 2\mathbb{E} \|\hat{f} - \hat{f}_{m^*}\|^2 + 2\mathbb{E} [\widehat{\text{pen}}(m^*)] \leq 2\|f - f_{m^*}\|^2 + 2C\text{pen}(m^*).$$

We have thus proved the desired result on \mathcal{G} :

$$\mathbb{E} \left[\|\hat{f} - \hat{f}_{\hat{m}}\|^2 \mathbf{1}_{\mathcal{G}} \right] \leq C \inf_{m \in \mathcal{M}} \left\{ \|f - f_m\|^2 + \text{pen}(m) \right\}.$$

• Next, we consider the set $\mathcal{G}^c = \{\hat{m} > m^*\}$. It holds that

$$\|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 \mathbf{1}_{\mathcal{G}^c} = 4 \left(\|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 - \frac{3}{4} \|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 \right) \mathbf{1}_{\mathcal{G}^c}.$$

We realize that, by definition of $\hat{f}_{\hat{m}}$, see (5.14), on \mathcal{G}^c ,

$$\begin{aligned} -\frac{3}{4} \|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 &= \frac{3}{4} \left(\|\hat{f}_{m^*}\|^2 - \|\hat{f}_{\hat{m}}\|^2 \right) \\ &\leq \frac{3}{4} \left(\|\hat{f}_{m^*}\|^2 - \|\hat{f}_{m^*}\|^2 + \widehat{\text{pen}}(m^*) - \widehat{\text{pen}}(\hat{m}) \right) = -\frac{3}{4} \widehat{\text{pen}}(m^*, \hat{m}). \end{aligned}$$

It follows from there that

$$\begin{aligned} &\left(\|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 - \frac{3}{4} \|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 \right) \mathbf{1}_{\mathcal{G}^c} \\ &\leq \sup_{\substack{k \geq m^* \\ k \in \mathcal{M}}} \left\{ \|\hat{f}_k - \hat{f}_{m^*}\|^2 - 6\|f_k - f_{m^*}\|^2 - \frac{3}{4} \widehat{\text{pen}}(m^*, k) \right\}_+ + 6 \sup_{k \geq m^*} \|f_k - f_{m^*}\|^2. \end{aligned}$$

Taking expectation and applying Proposition 5.6.5, as well as the monotonicity of the bias term, we conclude that

$$\begin{aligned} \mathbb{E} \left[\|\hat{f} - \hat{f}_{\hat{m}}\|^2 \mathbf{1}_{\mathcal{G}^c} \right] &\leq 2\mathbb{E} \left[\|\hat{f} - \hat{f}_{m^*}\|^2 \right] + 2\mathbb{E} \left[\|\hat{f}_{\hat{m}} - \hat{f}_{m^*}\|^2 \mathbf{1}_{\mathcal{G}^c} \right] \\ &\leq C\mathbb{E} \left[\|\hat{f} - \hat{f}_{m^*}\|^2 \right] + \frac{C}{N}, \end{aligned}$$

with $N = n$ in the RD-model and $N = M \wedge n$ in the NS-model.

This is the desired oracle bound for $\hat{\mathcal{G}}^c$. □

6

Adaptive estimation of marginal random-effects densities in linear mixed-effects model

Ce chapitre est une version modifiée de l'article *Adaptive estimation of marginal random-effects densities in linear mixed-effects model* publié dans *Mathematical Methods of Statistics* **24**(2): 81-109 (2015).

RÉSUMÉ. Dans ce chapitre, nous considérons le problème d'estimation adaptative de densités des effets aléatoires dans les modèles linéaires mixtes. Les modèles linéaires mixtes sont définis par $Y_{k,j} = \alpha_k + \beta_k t_j + \varepsilon_{k,j}$ où $Y_{k,j}$ est la valeur observée pour un individu k au temps t_j pour $k = 1, \dots, N$ et $j = 1, \dots, J$. Les variables aléatoires (α_k, β_k) sont appelées effets aléatoires et représentent les variables aléatoires des individus k . Nous notons leur densité f_α et f_β et supposons que les couples (α_k, β_k) sont indépendants des erreurs de mesure $(\varepsilon_{k,j})$. Nous définissons des estimateurs à noyau de f_α et f_β et présentons une majoration du risque quadratique intégré. Nous donnons aussi des exemples de vitesse de convergence de ces estimateurs. L'accent est mis sur le choix optimal de la fenêtre grâce à de la sélection de modèles par pénalisation conduisant à des inégalités oracles non-asymptotiques dans le cas particulier d'intervalles fixes entre les temps t_j . Enfin nous illustrons la méthode avec des simulations.

ABSTRACT. In this chapter we consider the problem of adaptive estimation of random-effects densities in linear mixed-effects model. The linear mixed-effects model is defined as $Y_{k,j} = \alpha_k + \beta_k t_j + \varepsilon_{k,j}$ where $Y_{k,j}$ is the observed value for individual k at time t_j for $k = 1, \dots, N$ and $j = 1, \dots, J$. Random variables (α_k, β_k) are called random effects and stand for the individual random variables of entity k . We denote their densities f_α and f_β and assume that they are independent of the measurement errors $(\varepsilon_{k,j})$. We introduce kernel estimators of f_α and f_β and present upper risk bounds. We also compute examples of rates of convergence. The focus of this work lies on the near optimal data driven choice of the smoothing parameter using a penalization strategy in the particular case of fixed interval between times t_j . Risk bounds for the adaptive estimators of f_α and f_β are provided. Simulations illustrate the relevance of the methodology.

Contents

6.1	Introduction	143
6.2	Statistical model and estimation procedure	144
6.2.1	Notations	144
6.2.2	Estimation of f_α	145
6.2.3	Estimation of f_β	146
6.3	Model selection	149
6.3.1	Adaptive estimation procedure for f_α	150
6.3.2	Adaptive estimation procedure for f_β	151
6.4	Simulation	152
6.4.1	Implementation of the estimation procedure	152
6.4.2	Simulation setting	152
6.4.3	Simulation results	153
6.5	Concluding remarks	154
6.6	Proofs	156
6.6.1	Sketch of the proof of Proposition 6.2.2	156
6.6.2	Proof of Theorem 6.3.1	156
6.6.3	Proof of Theorem 6.3.2	156
6.6.4	Proof of Lemma 6.6.1	158
6.6.5	Proof of Proposition 6.6.2	161
6.7	Appendix 1	169
6.8	Appendix 2: Supplementary material	170
6.8.1	Proof of Lemma 6.8.2	170
6.8.2	Proof of Lemma 6.8.1	172
6.8.3	Proof of Proposition 6.2.2	172

6.1 Introduction

Mixed models bring together fixed and random effects. They allow analysis of repeated measurements or longitudinal data. In this paper, we concentrate on linear mixed-effects models defined as

$$Y_{k,j} = \alpha_k + \beta_k t_j + \varepsilon_{k,j}, \quad k = 1, \dots, N \quad \text{and} \quad j = 1, \dots, J \quad (6.1)$$

where $Y_{k,j}$ denotes the observed value for individual k at time t_j and (α_k, β_k) represent the individual random variables of entity k . They are known as random effects. The random variables $(\varepsilon_{k,j})$ represent the measurement errors. We denote their densities f_α , f_β and f_ε . We do not assume that α_k and β_k are independent. We make the following assumptions:

- (A1) Times $(t_j)_{0 \leq j \leq J}$ are known and deterministic and $\Delta_j = \Delta$ for all j , $t_j = j\Delta$ and $J \geq 6$.
- (A2) $(\varepsilon_{k,j})_{k,j}$ are *i.i.d.* with distribution f_ε and the Fourier transform of f_ε does not vanish on the real line.
- (A3) (α_k, β_k) are *i.i.d.* with respective distribution f_α and f_β .
- (A4) (α_k, β_k) are independent of $(\varepsilon_{k,j})_{k,j}$.
- (A5) ε is symmetric.

The aim of this paper is to recover the densities f_α and f_β from the data $(Y_{k,j})$ in a nonparametric setting.

Mixed models have been widely studied in a parametric context. For example, [Pinheiro and Bates \(2000\)](#) have considered the problem assuming that both random effects and measurement errors are Gaussian, which enables them to use a maximum likelihood approach. Nonetheless the normality assumption can be too strong in some cases. In this way, [Wu and Zhu \(2010\)](#) relaxed the normality assumption estimating the first four moments of the random-effects density. We can also cite previous works of [Shen and Louis \(1999\)](#) who consider a smoothing method without any assumption on the error distribution f_ε , [Zhang and Davidian \(2001\)](#) and [Vock et al. \(2011\)](#) who propose a semi-nonparametric approach based on the approximation of the random-effects density by an Hermite series assuming that the error distribution is Gaussian. Some approaches are based on normal mixtures, see [Ghidey et al. \(2004\)](#), [Komárek and Lesaffre \(2008\)](#). For nonparametric approach, we can cite [Claeskens and Hart \(2009\)](#) who develop a nonparametric goodness-of-fit test in mixed models providing a nonparametric estimator if the normality hypothesis is rejected. Mixed models are also studied in Bayesian literature, see [Ibrahim and Kleinman \(1998\)](#) who allow the prior to be nonparametric by taking a Dirichlet process.

Here we consider an approach based on deconvolution methods. The convolution model is a classical setting in nonparametric statistics which has been widely studied. There exists a large amount of literature on the subject assuming first that the noise density is known. We can cite [Carroll and Hall \(1988\)](#), [Stefanski \(1990\)](#), [Stefanski and Carroll \(1990\)](#), [Fan \(1991\)](#), [Efromovich \(1997\)](#) and [Delaigle and Gijbels \(2004\)](#) who study rates of convergence and their optimality for kernel estimators. Concerning studies of rate optimality in the minimax sense, we refer to [Butucea \(2004\)](#) and [Butucea and Tsybakov \(2008a,b\)](#). Yet the drawback of these methods is that they all work under the assumption that the error distribution is known. We can also cite, in the known error case, [Dion \(2014\)](#) who study nonparametric estimators based on Lepski's method. However, the main goal of this paper lies in an adaptive choice of a smoothing parameter. For the most part, the adaptive

bandwidth selection in deconvolution models has been addressed with a known error distribution, see for example Pensky and Vidakovic (1999) for wavelet strategy, Comte et al. (2006), Butucea and Comte (2009) for projection strategies, or Meister (2009) and references therein. Adaptive estimation in deconvolution problems with unknown error density has been recently studied in a rigorous way. Several papers focus on that matter as those of Comte and Lacour (2011), Johannes and Schwarz (2013a), Dattner et al. (2016), Kappus (2014) and Kappus and Mabon (2014) (which matches Chapter 5). Rates of convergence have been presented in Neumann (1997) and, more recently, in Johannes (2009), or Meister (2009) under the assumption that a preliminary sample of the noise ε is observed.

More precisely, we follow an approach introduced in deconvolution literature but in presence of repeated measurements. Rates of convergence in a repeated observations model have been presented in Li and Vuong (1998), Neumann (2007), Delaigle et al. (2008) and Comte et al. (2013). More recently, Kappus and Mabon (2014) (see Chapter 5) have achieved a new adaptive procedure in this setting. Their method has the advantage of deriving a nearly optimal data driven choice of the smoothing parameter using a penalization strategy under very weak assumptions: in particular no semi parametric assumption on the shape of the characteristic function of the noise is required. In this paper, we propose to adapt their method in the context of a mixed-effects model.

Model (6.1) has been considered by Comte and Samson (2012). Let us emphasize the novelty of our paper. Comte and Samson (2012) derive theoretical properties of the nonparametric estimators of f_α and f_β only when the error distribution is known. In that particular case they establish oracle inequalities which ensure that their method is adaptive. Since assuming that the noise is known is not realistic, they also define an estimator when it is not. They prove an upper bound for its risk for fixed smoothing parameter. Then they propose an adaptive strategy which is implemented but not studied from a theoretical point of view. Here, we modify the procedure and improve the upper bound. Moreover we prove oracle risk bounds for the adaptive estimators of f_α and f_β in presence of unknown noise. These results are difficult and new. We also derive the rates of convergence for f_β .

This paper is organized as follows. In Section 6.2, we give the notations, specify the statistical model and estimation procedures for f_α and f_β along with upper bounds for both densities and rates of convergence for f_β those of f_α being known. In Section 6.3, we introduce adaptive estimators and propose a new adaptive procedure by penalization in the context of linear mixed-effects model under weak assumptions inspired by the work of Kappus and Mabon (2014) (see Chapter 5). Besides, the theoretical properties of the adaptive estimators are studied. In Section 6.4, we lead a study of the adaptive estimators through simulation experiments. Numerical results are then presented. In the concluding Section 6.5 we give further possible developments or extensions of the method. All the proofs are postponed to Section 6.6.

6.2 Statistical model and estimation procedure

6.2.1 Notations

For two real numbers a and b , we denote $a \vee b = \max(a, b)$, $a \wedge b = \min(a, b)$ and $(a)_+ = \max(a, 0)$. For two functions $\varphi, \psi : \mathbb{R} \rightarrow \mathbb{C}$ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote $\|\varphi\|$ the \mathbb{L}^2 norm of φ defined by $\|\varphi\|^2 = \int_{\mathbb{R}} |\varphi(x)|^2 dx$, $\langle \varphi, \psi \rangle$ the scalar product between φ and ψ defined by $\langle \varphi, \psi \rangle = \int_{\mathbb{R}} \varphi(x) \overline{\psi(x)} dx$. The Fourier transform φ^* is defined by $\varphi^*(x) = \int e^{ixu} \varphi(u) du$. Besides, if φ^* belongs to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, then the function φ is the inverse Fourier transform of φ^* and can be written $\varphi(x) = 1/(2\pi) \int e^{-ixu} \varphi^*(u) du$. Lastly

the convolution product $*$ is defined as $(\varphi * \psi)(x) = \int \varphi(x - u)\psi(u) du$.

6.2.2 Estimation of f_α

Under Assumption **(A1)**, we can write Model (6.1) for $j = 0$ as follows

$$Y_{k,0} = \alpha_k + \varepsilon_{k,0} \quad \text{and} \quad k = 1, \dots, N. \quad (6.2)$$

If observations at $t = 0$ are not available, we simply set $t'_j = t_j - t_1$. Then we would estimate the density of α'_k defined by $\alpha'_k = \alpha_k + \beta_k t_1$ instead. Thus it is interesting that we do not need an independence assumption between α and β . This is a classical deconvolution model in the context of unknown measurement errors (see references above).

The density distribution of $Y_{k,0}$ is noted f_Y . Under Model (6.2) and independence assumptions we have clearly that $f_Y = f_\alpha * f_\varepsilon$ which implies that $f_\alpha^* = f_Y^*/f_\varepsilon^*$. In this case, we have that

$$f_\alpha(x) = \frac{1}{2\pi} \int e^{-iux} f_\alpha^*(u) du = \frac{1}{2\pi} \int e^{-iux} \frac{f_Y^*(u)}{f_\varepsilon^*(u)} du.$$

If f_ε^* were known, we could simply estimate f_α^* with $\hat{f}_Y^*/f_\varepsilon^*$ where \hat{f}_Y^* is an estimator obtained directly from the data with a simple empirical estimator. We should only apply the inverse Fourier transform to get an estimate of f_α . Nevertheless, $1/f_\varepsilon^*$ is not integrable over \mathbb{R} . That is why we cannot compute the inverse Fourier transform over \mathbb{R} . We need to regularize the problem, for example, with a spectral cutoff parameter. For that we introduce the projection of f_α denoted $f_{\alpha,m}$ defined by

$$f_{\alpha,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \frac{f_Y^*(u)}{f_\varepsilon^*(u)} du. \quad (6.3)$$

In this paper the error distribution is assumed to be unknown. To make the problem identifiable, some additional information on the noise is required. In this subsection we derive an estimation procedure for f_ε^* which is specific for the estimation of f_α . When estimating f_β we will not use the same estimator of f_ε^* .

Model (6.2) can be seen as a repeated observation model. Therefore we can recover an estimation of the error distribution from the following data:

$$U_k = Y_{k,4} - Y_{k,3} - (Y_{k,2} - Y_{k,1}) = \varepsilon_{k,4} - \varepsilon_{k,3} - \varepsilon_{k,2} + \varepsilon_{k,1}$$

which imply the following equality under Assumption **(A2)**

$$f_U^*(x) = \mathbb{E} \left[e^{ixU} \right] = |f_\varepsilon^*(x)|^4.$$

Assumption **(A5)** together with **(A2)** imply that f_ε^* is real-valued and positive, so that $f_U^*(x) = (f_\varepsilon^*(x))^4$. As a consequence f_ε^{*4} can be estimated as follows

$$\widehat{f_\varepsilon^{*4}}(x) = \left(\frac{1}{N} \sum_{k=1}^N \cos(xU_k) \right)_+.$$

Nevertheless we need to prevent $\widehat{f_\varepsilon^{*4}}$ to become too small. For that we introduce a regularization of the Fourier transform by truncating the estimator following methods presented in Neumann (1997), Comte and Lacour (2011), Kappus (2014) and in Chapter 5. We define the following threshold

$$k_N(x) = s_N(x)N^{-1/2} \quad (6.4)$$

where $s_N(x) \geq 1$ will be defined later. Now we can introduce another estimator of f_ε^* defined by

$$\check{f}_\varepsilon^*(x) = \begin{cases} (\widehat{f}_\varepsilon^{*4}(x))^{1/4} & \text{if } \widehat{f}_\varepsilon^{*4}(x) \geq k_N(x), \\ (k_N(x))^{1/4} & \text{otherwise.} \end{cases}$$

So using the inverse Fourier transform, we can estimate $f_{\alpha,m}$ as follows

$$\hat{f}_{\alpha,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}_Y^*(u)}{\check{f}_\varepsilon^*(u)} du \quad (6.5)$$

where $\hat{f}_Y^*(u) = 1/N \sum_{k=1}^N e^{iuY_{k,0}}$.

We can state the following upper bound on the \mathbb{L}^2 risk for $\hat{f}_{\alpha,m}$.

Proposition 6.2.1. *Under Assumptions (A1)-(A5), for $k_N(x)$ defined by (6.4) and for $\hat{f}_{\alpha,m}$ defined by (6.5) then there is a positive constant C such that*

$$\mathbb{E} \left\| f_\alpha - \hat{f}_{\alpha,m} \right\|^2 \leq \|f_\alpha - f_{\alpha,m}\|^2 + \frac{C}{N} \left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^2} du + \frac{1}{2\pi} \int_{-\pi m}^{\pi m} s_N^2(u) \frac{|f_\alpha^*(u)|^2}{|f_\varepsilon^*(u)|^8} du \right), \quad (6.6)$$

where $f_{\alpha,m}$ is defined by (6.3) and C is a numerical constant.

The first two terms of the right-hand side of Equation (6.6) correspond to the usual terms when the error distribution is known (see Comte et al. (2006)): a squared bias term ($\|f_\alpha - f_{\alpha,m}\|^2$) and a bound on the variance depending only on f_ε^* . The last term is due to the estimation of f_ε^* and in addition depends on f_α^* . This upper bound, for $s_N \equiv 1$, is smaller than the upper bound in Comte and Samson (2012).

Note that for $s_N \equiv 1$, we get the optimal bound. The rates of convergence, for that kind of bounds, are derived in Comte and Lacour (2011), Comte et al. (2013), Neumann (1997) and Delaigle et al. (2008).

6.2.3 Estimation of f_β

First let us define the estimator of f_β . For the estimation of f_β , we use another approach to see the problem as a deconvolution problem. Without loss of generality we assume that J is even. So for $1 \leq j \leq J/2$, we can transform the data as follows

$$Z_{k,j} = \frac{Y_{k,2j} - Y_{k,2j-1}}{\Delta} = \beta_k + \frac{\varepsilon_{k,2j} - \varepsilon_{k,2j-1}}{\Delta} = \beta_k + \frac{\eta_{k,j}}{\Delta}, \quad \eta_{k,j} = \varepsilon_{k,2j} - \varepsilon_{k,2j-1}.$$

Let us notice that for a fixed j , the $(Z_{k,j})_k$ for $k = 1, \dots, N$ are *i.i.d.* but $Z_{k,j}$ and $Z_{k,l}$ for $j \neq l$ are not independent. It means that we preserve the independence between individuals of the sample.

Since β_k is independent of $\eta_{k,j}$ under Assumption (A4), we can write the following equality $f_{Z_1} = f_\beta * f_{\eta_{k,j}/\Delta}$ which implies $f_{Z_j}^*(x) = f_\beta^*(x) |f_\varepsilon^*(\frac{x}{\Delta})|^2$. Therefore under Assumption (A2) we have $f_\beta^*(x) = f_{Z_1}^*(x) / |f_\varepsilon^*(\frac{x}{\Delta})|^2$. Now using all the observations j we can write that

$$f_\beta^*(x) = \frac{f_{Z_1}^*(x)}{|f_\varepsilon^*(\frac{x}{\Delta})|^2}.$$

Unlike in the estimation of f_α , we do not need to estimate the Fourier transform f_ε^* of the error distribution but only $|f_\varepsilon^*|^2$: that is why we do not assume here that the noise is symmetric. Let us notice the following equality

$$\frac{U_k}{\Delta} = Z_{k,2} - Z_{k,1} = \frac{1}{\Delta} (\varepsilon_{k,4} - \varepsilon_{k,3} - \varepsilon_{k,2} + \varepsilon_{k,1}).$$

Then we have

$$f_{\frac{U}{\Delta}}^*(x) = \mathbb{E} \left[e^{ixU/\Delta} \right] = |f_\varepsilon^*\left(\frac{x}{\Delta}\right)|^4.$$

So $|f_\varepsilon^*|^4$ can be estimated as follows

$$|\widehat{f_\varepsilon^{*4}}|\left(\frac{x}{\Delta}\right) = \left(\frac{1}{N} \sum_{k=1}^N \cos\left(x \frac{U_k}{\Delta}\right) \right)_+.$$

And to prevent the denominator from becoming too small, we regularize the Fourier transform of the error distribution as follows

$$|\tilde{f}_\varepsilon^*\left(\frac{x}{\Delta}\right)|^2 = \begin{cases} |\hat{f}_\varepsilon^*\left(\frac{x}{\Delta}\right)|^2 = \left(|\widehat{f_\varepsilon^{*4}}|\left(\frac{x}{\Delta}\right)\right)^{1/2} & \text{if } |\widehat{f_\varepsilon^{*4}}|\left(\frac{x}{\Delta}\right) \geq k_N\left(\frac{x}{\Delta}\right), \\ (k_N\left(\frac{x}{\Delta}\right))^{1/2} & \text{otherwise.} \end{cases} \quad (6.7)$$

Thus we can estimate f_β^* as follows

$$\hat{f}_\beta^*(x) = \frac{\hat{f}_Z^*(x)}{|\tilde{f}_\varepsilon^*\left(\frac{x}{\Delta}\right)|^2} \quad \text{with} \quad \hat{f}_Z^*(x) = \frac{2}{N(J-4)} \sum_{j=3}^{J/2} \sum_{k=1}^N e^{ixY_{k,j}}, \quad \text{for } j = 3, \dots, J/2.$$

We emphasize that the previous definition uses distinct observations for \tilde{f}_ε^* and \hat{f}_Z^* , so that the numerator and the denominator are independent. This is why Assumption **(A1)** requires $J \geq 6$. We then define $f_{\beta,m}$ as follows

$$f_{\beta,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{f_{Z_1}^*(u)}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} du. \quad (6.8)$$

Applying the inverse Fourier transform, we get an estimate of $f_{\beta,m}$

$$\hat{f}_{\beta,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-ixu} \frac{\hat{f}_Z^*(u)}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} du. \quad (6.9)$$

Thus we can state the following upper bound on the \mathbb{L}^2 risk for $\hat{f}_{\beta,m}$.

Proposition 6.2.2. *Under Assumptions **(A1)**-**(A4)**, for $k_N(x)$ defined by (6.4), then $\hat{f}_{\beta,m}$ defined by (6.9) satisfies*

$$\begin{aligned} \mathbb{E} \left\| f_\beta - \hat{f}_{\beta,m} \right\|^2 &\leq \|f_\beta - f_{\beta,m}\|^2 + \frac{6m}{N} + \frac{12}{N(J-4)} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4} \\ &\quad + \frac{4C_1}{N} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{s_N^2(u) |f_\beta^*(u)|^2}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} du. \end{aligned} \quad (6.10)$$

where $f_{\beta,m}$ is defined by (6.8) and C_1 is defined in Lemma 6.8.2.

The terms of the right-hand side of Equation (6.10) correspond to a squared bias ($\|f_\beta - f_{\beta,m}\|^2$) variance decomposition. Compared to Comte and Samson (2012), this inequality differs from the last term which is smaller than theirs, for $s_N \equiv 1$. The first term of variance with order m/N is the bound we would have if we were in a direct density estimation context. The second term is a classical term appearing in density deconvolution problems when the error distribution is known but the $J - 4$ factor is specific to the repeated measurement framework. The last term of variance is due to the estimation of f_ε^* .

Discussion of the resulting rates

In order to derive the corresponding rates of convergence of the estimator of f_β defined by (6.9), we assume that the density functions f_β and f_ε belong to some nonparametric classes of functions. First, we introduce the following type of smoothness spaces

$$\mathcal{A}(a, r, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}), \int |f^*(u)|^2 e^{2a|u|^r} du \leq L \right\} \quad (6.11)$$

$$\mathcal{S}(\delta, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}), \int |f^*(u)|^2 (1 + u^2)^\delta du \leq L \right\} \quad (6.12)$$

with $r \geq 0$, $a > 0$, $\delta > 1/2$ and $L > 0$. Then if f_β belongs to $\mathcal{A}(a, r, L)$, the squared bias term can be bounded as follows

$$\|f_\beta - f_{\beta,m}\|^2 \leq \frac{L}{2\pi} e^{-2a|\pi m|^r},$$

or if f_β belongs to $\mathcal{S}(\delta, L)$

$$\|f_\beta - f_{\beta,m}\|^2 \leq \frac{L}{2\pi} ((\pi m)^2 + 1)^{-\delta}.$$

To derive the order of the bound on the variance in Equation (6.10), we need more information about the regularity of f_ε . We add the following classical assumption:

There exist positive constants k_0, k'_0, γ, μ , and s such that for any real x

$$k_0(x^2 + 1)^{-\gamma/2} e^{-\mu|x|^s} \leq |f_\varepsilon^*(x)| \leq k'_0(x^2 + 1)^{-\gamma/2} e^{-\mu|x|^s}. \quad (6.13)$$

- We say that $f_\varepsilon^* \in \text{OS}(\gamma)$ (for *ordinary smooth*), if f_ε^* satisfies (6.13) with $\gamma = 0$, $s > 0$ and $\mu > 0$.
- We say that $\gamma = 0$ $f_\varepsilon^* \in \text{SS}(s)$ (for *supersmooth*), if f_ε^* satisfies (6.13) with $s = 0$.

Proposition 6.2.3. *If f_ε^* satisfies (6.13) and $f_\beta \in \mathcal{S}(\delta, L)$ defined by (6.11), then*

$$\mathbb{E} \left\| f_\beta - \hat{f}_{\beta,m} \right\|^2 \leq C m^{-2\delta} + \frac{6m}{N} + C' \frac{m^{4\gamma+1-s} e^{4\mu(\pi m)^s}}{N(J-4)} + C'' \frac{m^{(4\gamma+1-s)\wedge 2(2\gamma-\delta)+4\mu(\pi m)^s}}{N}$$

where C, C' and C'' are positive constants independent of N or J .

Proposition 6.2.4. *Suppose that $f_\varepsilon^* \in \text{OS}(\gamma)$ and $f_\beta \in \mathcal{A}(a, r, L)$ defined by (6.12), then*

$$\mathbb{E} \left\| f_\beta - \hat{f}_{\beta,m} \right\|^2 \leq C e^{-2a(\pi m)^r} + \frac{6m}{N} + C' \frac{m^{4\gamma+1}}{N(J-4)} + \frac{C''}{N}$$

where C, C' and C'' are positive constants independent of N or J .

$f_\beta \in \mathcal{S}(\delta, L, \cdot), f_\varepsilon^* \in \text{OS}(\gamma)$	$f_\beta \in \mathcal{S}(\delta, L), f_\varepsilon^* \in \text{SS}(s)$	$f_\beta \in \mathcal{A}(a, r, L), f_\varepsilon^* \in \text{OS}(\gamma)$
$(NJ)^{\frac{-2\delta}{2\delta+4\gamma+1}} + N^{-\left(\frac{2\delta}{2\delta+1} \wedge \frac{\delta}{2\gamma}\right)}$	$(\log NJ)^{-2\delta/s}$	$\frac{(\log NJ)^{\frac{4\gamma+1}{r}}}{NJ} + \frac{(\log N)^{\frac{1}{r}}}{N}$

Table 6.1: Rates of convergence for the MISE

The rates are reported in Table 6.1. We clearly see that the rates of convergence depend on unknown quantities since they describe the regularity of the density function under estimation as well as the error distribution. When J is considered as a constant, we find the usual rates of convergence corresponding to density deconvolution already presented in the literature. Increasing J will improve the rates.

The idea now for the adaptive estimation is to find a penalty term which have the same order as the bound on the variance. Thus the adaptive estimator will reach automatically the rates of convergence presented in Table 6.1. The following section shows that we can obtain an adaptive procedure under weak assumptions. In particular, we do not assume there that the characteristic function of the error distribution has a particular shape (ordinary smooth or supersmooth). In other words, Assumptions (6.11) or (6.12) and (6.13) are not necessary to derive an adaptive procedure.

6.3 Model selection

In this section, we introduce an adaptive estimator of $|f_\varepsilon^*|^2$ which can be uniformly controlled on the real line. This brings a model selection procedure with very weak assumptions on the error distribution f_ε^* . For that we need to choose a new and adequate $s_N(x)$, larger than previously, which will allow us to apply concentration inequalities of Talagrand type. For $\delta > 0$, let us introduce the weight function w defined as

$$\forall x \in \mathbb{R}, w(x) = (\log(e + |x|))^{-\frac{1}{2}-\delta}$$

which has originally been proposed in Neumann and Reiß (2009). The considerations presented in that paper, combined with ideas given in Kappus (2014) play an important role for the arguments. This function is of high importance since the penalty terms will involve an empirical version of the characteristic function in the denominator ; therefore the oracle inequalities depend on a precise control of the deviation of $\widehat{f_\varepsilon^{*4}}$ from f_ε^{*4} on the real line. It is shown in Neumann and Reiß (2009) that the distance between both objects, weighted by w , is simultaneously small on the real axes. In the penalty terms, there will hence occur a loss of logarithmic order, in comparison to the variance term.

As from now, we set the threshold k_N defined by (6.4) as follows

$$s_N(x) = \kappa (\log N)^{1/2} w(x)^{-1} \tag{6.14}$$

where κ is a positive universal constant.

We want to propose an estimator $\hat{f}_{\hat{m}}$ of f completely data driven. Following the model selection paradigm, see Birgé (1999), Birgé and Massart (1997) or Massart (2003), we select

\hat{m} as the minimizer of a penalized criterion

$$\hat{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|\hat{f}_m\|^2 + \widehat{\text{pen}}(m) \right\}$$

where \mathcal{M}_N describes the model collections. The penalty term should be chosen large enough to counterbalance the fluctuation of \hat{f}_m around \hat{f} , but on the other hand, should ideally not be much larger than the variance terms presented in Equation 6.10. Here the penalty term is stochastic since the variance terms depend on the error distribution which is supposed unknown.

6.3.1 Adaptive estimation procedure for f_α

In this section, we adapt the results of Chapter 5 which proposed a completely data driven procedure in the framework of density estimation in deconvolution problems with unknown error distribution. In the present paper, Model (6.2) can be seen as a repeated observation model, which is studied in their paper. The main difference lies in the fact that in Chapter 5 f_ε^{*2} can be estimated directly from the data with an empirical estimator whereas here only f_ε^{*4} can be estimated. This, in our case, implies that f_ε^* is raised to a greater power in the term specific to the unknown noise density.

We introduce the following notations

$$\begin{aligned} \Theta(m) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|f_\varepsilon^*(u)|^2} du \quad \text{and} \quad \Theta^\alpha(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f_\alpha^*(u)|^2}{|f_\varepsilon^*(u)|^8} du \\ \hat{\Theta}(m) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2}}{|\check{f}_\varepsilon^*(u)|^2} du \quad \text{and} \quad \hat{\Theta}^\alpha(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |\hat{f}_Y^*(u)|^2}{|\check{f}_\varepsilon^*(u)|^{10}} du. \end{aligned}$$

These terms correspond to the deterministic and stochastic bounds on the variance appearing in Equation (6.6). The difference lies in the introduction of the function w essential for the adaptive procedure. Thus, we define an empirical penalty as

$$\widehat{\text{pen}}(m) = 12\lambda^2(m, \hat{\Theta}(m)) \frac{\hat{\Theta}(m)}{n} + 16\kappa^2 \log(Nm) \frac{\hat{\Theta}^\alpha(m)}{N}$$

where κ is the same as in Equation (6.14), and a deterministic penalty

$$\text{pen}(m) = 12\lambda^2(m, \Theta(m)) \frac{\Theta(m)}{N} + 16\kappa^2 \log(Nm) \frac{\Theta^\alpha(m)}{N},$$

with $\lambda(m, D) = \max \left\{ \sqrt{8 \log(1 + Dm^2)}, \frac{16\sqrt{2}}{3\sqrt{N}} \log(1 + Dm^2) \right\}$.

Then, we select the cutoff parameter \hat{m} as a minimizer of the following penalized criterion

$$\hat{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|\hat{f}_{\alpha, m}\|^2 + \widehat{\text{pen}}(m) \right\} \quad (6.15)$$

where $\mathcal{M}_N = \{1, \dots, N\}$. We can now state the following oracle inequality:

Theorem 6.3.1. *Under Assumptions (A1)-(A4), let $\hat{f}_{\alpha, \hat{m}}$ be defined by (6.5) and (6.15). Then there are positive constants C^{ad} and C such that*

$$\mathbb{E} \|f_\alpha - \hat{f}_{\alpha, \hat{m}}\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}_N} \left\{ \|f_\alpha - f_{\alpha, m}\|^2 + \text{pen}(m) \right\} + \frac{C}{N}. \quad (6.16)$$

The latest result is an oracle inequality which means that the squared bias variance compromise is automatically made and completely data driven in a non-asymptotic setting. So rates of convergence are reached without being specified in the framework. This result is of high interest since in deconvolution problems, rates of convergence are classically intricate and depend on the regularity types of the function f under estimation and the error density f_ε^* which is also unknown (see Section 6.2.3).

However the penalty term is not exactly the same as the upper bound terms shown in Equation (6.6). We may wonder if a loss due to adaptation occurs. To answer that question, right-hand side of Equations (6.6) and (6.16) have to be compared. More precisely, since the squared bias term $\|f_\alpha - f_{\alpha,m}\|^2$ is unchanged and N^{-1} is a negligible term, it comes down to compare $\text{pen}(m)$ with

$$\frac{1}{N} \left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^2} du + \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{|f_\alpha^*(u)|^2}{|f_\varepsilon^*(u)|^8} du \right).$$

Clearly the difference lies in the logarithmic terms, and thus, the loss is negligible.

6.3.2 Adaptive estimation procedure for f_β

As in the previous section, we start by defining the bound appearing in the known-error case, then the one appearing in the unknown-error case, with additional w function.

$$\begin{aligned} \Xi(m) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2}}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4} du & \text{and} & \quad \Xi^\beta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |f_\beta^*(u)|^2}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} du. \\ \hat{\Xi}(m) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2}}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4} du & \text{and} & \quad \hat{\Xi}^\beta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du. \end{aligned}$$

We can now define the stochastic penalty associated to the adaptive procedure

$$\begin{aligned} \widehat{\text{qen}}(m) &= \widehat{\text{qen}}_1(m) + \widehat{\text{qen}}_2(m) + \widehat{\text{qen}}_3(m) \\ &= 64 \frac{m}{N} + 16 \frac{\mu^2(m, \hat{\Xi}(m)) \hat{\Xi}(m)}{N(J-4)} + 16\kappa^2 \log(Nm) \frac{\hat{\Xi}^\beta(m)}{N} \end{aligned}$$

and the deterministic penalty

$$\begin{aligned} \text{qen}(m) &= \text{qen}_1(m) + \text{qen}_2(m) + \text{qen}_3(m) \\ &= 64 \frac{m}{N} + 16\mu^2(m, \Xi(m)) \frac{\Xi(m)}{N(J-4)} + 16\kappa^2 \log(Nm) \frac{\Xi^\beta(m)}{N} \end{aligned}$$

with weight $\mu(m, D) = \max \left\{ \sqrt{8 \log(1 + Dm^2)}, \frac{16\sqrt{2}}{3\sqrt{N(J-4)}} \log(1 + Dm^2) \right\}$.

It is worth mentioning that the penalty takes into account the three terms of variance showed in Equation (6.10). So the penalty has the same order as the bounds on the variance. Therefore we select the cutoff parameter \hat{m} as a minimizer of the following penalized criterion

$$\hat{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|\hat{f}_{\beta,m}\|^2 + \widehat{\text{qen}}(m) \right\}. \quad (6.17)$$

Theorem 6.3.2. *Under Assumptions (A1)-(A4), consider $\hat{f}_{\beta, \hat{m}}$ defined by (6.9) and (6.17). Then there are positive constants C^{ad} and C such that*

$$\mathbb{E} \left\| f_{\beta} - \hat{f}_{\beta, \hat{m}} \right\|^2 \leq C^{ad} \inf_{m \in \mathcal{M}_N} \left\{ \|f_{\beta} - f_{\beta, m}\|^2 + \widehat{\text{qen}}(m) \right\} + \frac{C}{N(J-4)} + \frac{C}{N}.$$

The same kind of remarks as after Theorem 6.3.1 hold here. The latest result is an oracle inequality which means that the bias variance compromise is automatically made and completely data driven in an almost non-asymptotic setting. So rates of convergence are reached automatically without being specified in the framework. As far as we know this result is new in the literature.

6.4 Simulation

In this section, we only concentrate on a simulation study of f_{β} . Indeed, the proposed method for the estimation of f_{α} being mainly taken from Chapter 5, we refer to that paper for the performance of the estimator.

The whole implementation is conducted using R software. The integrated squared error $\|f - \hat{f}_{\beta, \hat{m}}\|^2$ is computed via a standard approximation and discretization (over 300 points) of the integral on an interval of \mathbb{R} denoted by I . Then the mean integrated squared error (MISE) $\mathbb{E}\|f - \hat{f}_{\beta, \hat{m}}\|^2$ is computed as the empirical mean of the approximated ISE over 100 simulation samples.

6.4.1 Implementation of the estimation procedure

The adaptive procedure is implemented as follows:

- ▷ For $m \in \mathcal{M}_N = \{m_1, \dots, m_N\}$, compute $-\|f_{\beta, m}\|^2 + \widehat{\text{qen}}(m)$.
- ▷ Choose \hat{m} such as $\hat{m} = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|f_{\beta, m}\|^2 + \widehat{\text{qen}}(m) \right\}$.
- ▷ Compute $\hat{f}_{\beta, \hat{m}}(x) = \int_{-\pi \hat{m}}^{\pi \hat{m}} e^{-ixu} \frac{\hat{f}_Z^*(u)}{|\hat{f}_{\varepsilon}^*(\frac{u}{\Delta})|^2} du$.

Riemann's sums are used to approximate all the integrals. The penalties are chosen according to Theorem 6.3.2 and as in Comte et al. (2007) we consider that m can be fractional by taking the following model collection $\mathcal{M}_N = \{m = k/10, \quad 1 \leq k \leq 25\}$ associated with the following penalty

$$\widehat{\text{qen}}(m) = \kappa_1 \left(\frac{m}{N} + \frac{\log(1 + \hat{\Xi}(m)m^2)\hat{\Xi}(m)}{N(J-4)} \right) + \kappa_2 \log(Nm) \frac{\hat{\Xi}^{\beta}(m)}{N}$$

Moreover the times t_j are chosen as $t_j = j\Delta$ with $\Delta = 2$ and $J = 6$ as in Comte and Samson (2012).

6.4.2 Simulation setting

We consider the four following distributions for β :

- ▷ Standard Gaussian distribution, $I = [-4, 4]$.
- ▷ Cauchy distribution, $f(x) = (\pi(1 + x^2))^{-1}$, $I = [-10, 10]$.
- ▷ Gamma distribution : $5 \cdot \Gamma(25, \frac{1}{25})$, $I = [-1, 13]$.
- ▷ Mixed Gamma distribution : $X = W/\sqrt{5.48}$, with $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$, $I = [-1.5, 26]$.

All the densities are normalized with unit variance except the Cauchy density. In all considered cases, f_α is a standard Gaussian distribution.

We consider the two following noise densities with same variance σ_ε^2 . In the simulation the variance takes the values 1/10 and 1/4. The first one is a Gaussian density (*supersmooth* density) which means $f_\varepsilon^* \in \text{SS}(2)$. The second one is a Laplace density (*ordinary smooth* density) which means $f_\varepsilon^* \in \text{OS}(2)$.

$$\text{Gaussian noise : } f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right), f_\varepsilon^*(x) = \exp\left(-\frac{\sigma_\varepsilon^2 x^2}{2}\right).$$

$$\text{Laplace noise : } f_\varepsilon(x) = \frac{1}{2\sigma_\varepsilon} \exp\left(-\frac{|x|}{\sigma_\varepsilon}\right), f_\varepsilon^*(x) = \frac{1}{1 + \sigma_\varepsilon^2 x^2}.$$

The calibration of the two constants are done with intensive preliminary simulations with a sample size of 500. In the end, we choose $\kappa_1 = \kappa_2 = 1$. We can notice that in Chapter 5 the constant are larger. It seems that the greater the power of f_ε^* is in the denominator, the smaller the constants are.

6.4.3 Simulation results

The results of the simulations are given in Tables 6.2 and 6.3. In both tables, we report the average values of the integrated squared errors multiplied by 100 and their medians (in brackets). A first remark: estimating the Fourier transform of the noise f_ε^* reduces the risk compared to knowing the density of the noise, a fact already pointed in Comte and Lacour (2011). This can be explained by the fact that an additional regularization of the characteristic function of the noise comes in. This regularization is not applied in the procedure when the error distribution is known.

Table 6.2 corresponds to an estimation procedure where the error distribution is a Laplace density while Table 6.3 corresponds to a Gaussian noise. We notice that increasing the sample size improves the estimation and increasing the variances degrades the estimation but in an acceptable way. Concerning the medians of the ISE, they are always lower than the means of the ISE which may indicate that only few estimates degrade global performance.

For all the test densities, the results are very good.

Comparison with existing results. The results of the standard Gaussian distribution, Cauchy distribution and Gamma distribution can be compared to those of Chapter 5. We estimate the same functions but in a more difficult context. In fact we have access to an estimate of $(f_\varepsilon^*)^4$ when they have access to an estimate of f_ε^* or $(f_\varepsilon^*)^2$. That makes the problem tougher since the information about f_ε^* is less precise. Nevertheless, if we compare to Chapter 5 the results are very close which is quite remarkable.

In Table 6.4, we have reported the results of the simulations with Comte and Samson (2012) procedure computed over the same samples used in Tables 6.2 and 6.3. We have

chosen to compare standard Gaussian distribution and Gamma distribution as examples. We can point out that our results are twice better when the error distribution is Gaussian. When the error distribution is a Laplace the results are similar. It is due to the fact that in their procedure their penalty is roughly our penalty q_{en} multiplied by a term depending on the dimension of the model. When the noise is ordinary smooth as a Laplace distribution this term behaves as a constant whereas with a supersmooth noise as a Gaussian it leads to an over penalisation.

So our methodology can handle more cases (supersmooth error) than [Comte and Samson \(2012\)](#) and give very good results of estimation.

		$\sigma_{\varepsilon}^2 = \frac{1}{10}$		$\sigma_{\varepsilon}^2 = \frac{1}{4}$	
		200	2000	200	2000
Gaussian	f_{ε} known	0.344 (0.273)	0.054 (0.045)	0.514 (0.513)	0.129 (0.105)
	f_{ε} unknown	0.331 (0.251)	0.042 (0.033)	0.317 (0.236)	0.057 (0.051)
Cauchy	f_{ε} known	0.625 (0.573)	0.105 (0.097)	0.804 (0.765)	0.216 (0.211)
	f_{ε} unknown	0.507 (0.427)	0.075 (0.071)	0.657 (0.599)	0.090 (0.079)
Gamma	f_{ε} known	0.398 (0.360)	0.069 (0.063)	0.620 (0.517)	0.161 (0.140)
	f_{ε} unknown	0.381 (0.347)	0.051 (0.044)	0.506 (0.430)	0.066 (0.051)
Mixed Gamma	f_{ε} known	0.545 (0.480)	0.095 (0.084)	0.715 (0.620)	0.150 (0.141)
	f_{ε} unknown	0.506 (0.453)	0.082 (0.080)	0.518 (0.495)	0.092 (0.084)

Table 6.2: Results of simulation with Laplace noise. Columns report the average values of $100 \times ISE = 100 \times \|f - \hat{f}_{\beta, \hat{m}}\|^2$ and their medians (in brackets) computed over 100 samples.

6.5 Concluding remarks

This paper deals with the construction of penalized estimators of the random effect densities f_{α} and f_{β} in a linear mixed-effects models. It improves the results of [Comte and Samson \(2012\)](#) to the unknown error case. We have proved oracle risk bounds for the adaptive estimators of f_{α} and f_{β} in presence of unknown noise and derived the rates of convergence for f_{β} . We have also illustrated our procedures on datasets simulations. We have noted that our results for the estimation of f_{β} are twice better when the error is Gaussian since the procedure of [Comte and Samson \(2012\)](#) is over penalized.

Let us emphasize that we do not assume that f_{α} and f_{β} are independent. Hence, a valid problem in this situation may be the estimation of a joint density $\beta f_{\alpha, \beta}$ of α and β . For this purpose, one can build an estimator of $f_{\alpha, \beta}^*$ using the same kind of argument developed

		$\sigma_\varepsilon^2 = \frac{1}{10}$		$\sigma_\varepsilon^2 = \frac{1}{4}$	
		200	2000	200	2000
Gaussian	f_ε known	0.349 (0.296)	0.054 (0.045)	0.650 (0.598)	0.142 (0.128)
	f_ε unknown	0.285 (0.239)	0.038 (0.029)	0.349 (0.273)	0.052 (0.053)
Cauchy	f_ε known	0.588 (0.532)	0.119 (0.112)	0.848 (0.791)	0.272 (0.263)
	f_ε unknown	0.481 (0.449)	0.076 (0.070)	0.680 (0.607)	0.089 (0.084)
Gamma	f_ε known	0.401 (0.332)	0.072 (0.063)	0.956 (0.910)	0.207 (0.190)
	f_ε unknown	0.402 (0.316)	0.049 (0.041)	0.461 (0.418)	0.067 (0.055)
Mixed Gamma	f_ε known	0.504 (0.454)	0.089 (0.084)	0.704 (0.639)	0.163 (0.146)
	f_ε unknown	0.504 (0.446)	0.079 (0.072)	0.552 (0.484)	0.101 (0.093)

Table 6.3: Results of simulation with Gaussian noise. Columns report the average values of $100 \times ISE = 100 \times \|f - \hat{f}_{\beta, \hat{m}}\|^2$ and their medians (in brackets) computed over 100 samples.

		$\sigma_\varepsilon^2 = \frac{1}{10}$		$\sigma_\varepsilon^2 = \frac{1}{4}$	
		200	2000	200	2000
Laplace noise					
	Gaussian	0.434	0.031	0.212	0.051
	Gamma	0.411	0.069	0.517	0.080
Gaussian noise					
	Gaussian	0.972	0.128	4.013	0.101
	Gamma	0.764	0.169	2.572	0.450

Table 6.4: Results of simulation with Laplace and Gaussian noise with [Comte and Samson \(2012\)](#) procedure. Columns report the average values of $100 \times ISE = 100 \times \|f - \hat{f}_{\beta, \hat{m}}\|^2$ computed over 100 samples.

in this paper. Let us notice that

$$f_{\alpha, \beta}^*(u, v) = \frac{\mathbb{E} \left[e^{i(uY_{k,0} + vZ_{k,j})} \right]}{f_\varepsilon^*(u) |f_\varepsilon^*(\frac{v}{\Delta})|^2}$$

so we could estimate $f_{\alpha,\beta}^*$ by replacing $\mathbb{E} \left[e^{i(uY_{k,0} + vZ_{k,j})} \right]$ by an empirical mean. Next as

$$f_{\alpha,\beta}(u, v) = \frac{1}{(2\pi)^2} \iint e^{-i(xu+yv)} f_{\alpha,\beta}^*(u, v) du dv,$$

this implies that a projection estimator would be

$$(\hat{f}_{\alpha,\beta})_{m_1, m_2}(u, v) = \frac{1}{(2\pi)^2} \int_{-\pi m_1}^{\pi m_1} \int_{-\pi m_2}^{\pi m_2} e^{-i(xu+yv)} \hat{f}_{\alpha,\beta}^*(u, v) du dv.$$

This problem requires to choose indepently the cutoffs m_1 and m_2 which would lead to establish an adaptive procedure based on [Goldenshluger and Lepski \(2011\)](#) method. This problem would be very difficult and probably rather technical, this is why it is beyond the scope of the present work.

6.6 Proofs

6.6.1 Sketch of the proof of Proposition 6.2.2

The proof follows the same lines as Proposition 5.2 in [Comte and Samson \(2012\)](#). The difference is that we apply our Lemma 6.8.2 instead.

6.6.2 Proof of Theorem 6.3.1

The proof is similar to the proof of Theorem 6.3.2. See also Chapter 5. Thus the proof is omitted.

6.6.3 Proof of Theorem 6.3.2

Let us introduce some notations: for $k > m$,

$$\hat{\Xi}(m, k) = \hat{\Xi}(k) - \hat{\Xi}(m), \quad \hat{\Xi}^\beta(m, k) = \hat{\Xi}^\beta(k) - \hat{\Xi}^\beta(m).$$

Moreover,

$$\begin{aligned} \widehat{\text{qen}}(m, k) &= \text{qen}_1(m, k) + \text{qen}_2(m, k) + \text{qen}_3(m, k) \\ &= \frac{64(k-m)}{N} + 16 \frac{\hat{\mu}^2(m, k) \hat{\Xi}(m, k)}{N(J-4)} + 16\kappa^2 \log(N(k-m)) \frac{\hat{\Xi}^\beta(m, k)}{N} \end{aligned}$$

with $\hat{\mu}(m, k) = \max \left\{ \sqrt{8 \log(1 + \hat{\Xi}(m, k)m^2)}, \frac{16\sqrt{2}}{3\sqrt{N(J-4)}} \log(1 + \hat{\Xi}(m, k)(k-m)^2) \right\}$.

Now we can start the proof of Theorem 6.3.2. We denote by m^* the oracle cutoff defined by

$$m^* = \arg \min_{m \in \mathcal{M}_N} \left\{ -\|f_{\beta, m}\|^2 + \text{qen}(m) \right\}.$$

We have $\|f_\beta - \hat{f}_{\beta, \hat{m}}\|^2 \leq 2\|f_\beta - \hat{f}_{\beta, m^*}\|^2 + 2\|\hat{f}_{\beta, m^*} - \hat{f}_{\beta, \hat{m}}\|^2$.

• Let us notice on the set $G = \{\hat{m} \leq m^*\}$:

$$\|\hat{f}_{\beta, m^*} - \hat{f}_{\beta, \hat{m}}\|^2 \mathbf{1}_G = \left(\|\hat{f}_{\beta, m^*}\|^2 - \|\hat{f}_{\beta, \hat{m}}\|^2 \right) \mathbf{1}_G.$$

Besides according to the definition of \hat{m} , one has the following inequalities:

$$-\|\hat{f}_{\beta, \hat{m}}\|^2 + \widehat{\text{qen}}(\hat{m}) \leq -\|\hat{f}_{\beta, m^*}\|^2 + \widehat{\text{qen}}(m^*) \quad (6.18)$$

which implies

$$-\|\hat{f}_{\beta, \hat{m}}\|^2 \leq -\|\hat{f}_{\beta, m^*}\|^2 + \widehat{\text{qen}}(m^*).$$

Thus

$$\|\hat{f}_{\beta, m^*} - \hat{f}_{\beta, \hat{m}}\|^2 \mathbf{1}_G = \left(\|\hat{f}_{\beta, m^*}\|^2 - \|\hat{f}_{\beta, \hat{m}}\|^2 \right) \mathbf{1}_G \leq \widehat{\text{qen}}(m^*).$$

Taking expectation, we apply the following Lemma proved in Section 6.6.4.

Lemma 6.6.1. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_N$*

$$\mathbb{E}[\widehat{\text{qen}}(m)] \leq C \text{qen}(m).$$

It yields for some positive constant C

$$\mathbb{E} \left[\|\hat{f}_{\beta, \hat{m}} - f_{\beta, \hat{m}}\|^2 \mathbf{1}_G \right] \leq 2\mathbb{E} \left[\|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 \right] + 2\mathbb{E}[\widehat{\text{qen}}(m^*)] \leq 2\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 + 2C \text{qen}(m^*).$$

We just proved the desired result on G as $\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 = \|f_{\beta, \hat{m}}\|^2 - \|f_{\beta, m^*}\|^2$ and using the definition of m^*

$$\mathbb{E} \left[\|f_{\beta, \hat{m}} - \hat{f}_{\beta, \hat{m}}\|^2 \mathbf{1}_G \right] \leq C \inf_{m \in \mathcal{M}_N} \left\{ \|f_{\beta, m} - f_{\beta, m^*}\|^2 + \text{qen}(m) \right\}. \quad (6.19)$$

• We now consider the set $G^c = \{\hat{m} > m^*\}$.

$$\begin{aligned} \|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 \mathbf{1}_{G^c} &= \left(\|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 - 4\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m^*, \hat{m}) \right) \mathbf{1}_{G^c} \\ &\quad + \left(4\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 + \frac{1}{2}\widehat{\text{qen}}(m^*, \hat{m}) \right) \mathbf{1}_{G^c} \\ &\leq \sup_{\substack{k \geq m^* \\ k \in \mathcal{M}_N}} \left\{ \|\hat{f}_{\beta, k} - \hat{f}_{\beta, m^*}\|^2 - 4\|f_{\beta, k} - f_{\beta, m^*}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m^*, k) \right\}_+ \\ &\quad + 4\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 + \frac{1}{2} \sum_{\substack{k \geq m^* \\ k \in \mathcal{M}_N}} \widehat{\text{qen}}(m^*, k) \mathbf{1}_{\{\hat{m}=k\}}. \end{aligned} \quad (6.20)$$

Let us first notice the following inequality

$$\forall k > m, \quad \widehat{\text{qen}}(m, k) \leq \widehat{\text{qen}}(k). \quad (6.21)$$

Besides by definition of \hat{m} (see Equation (6.17)), on the set $\{\hat{m} = k\} \cap G^c$ and applying Equation (6.18), one has

$$\begin{aligned} \frac{1}{2}(\widehat{\text{qen}}(k) - \widehat{\text{qen}}(m^*)) &\leq \|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 - \frac{1}{2}\widehat{\text{qen}}(k) + \frac{1}{2}\widehat{\text{qen}}(m^*) \\ \frac{1}{2}\widehat{\text{qen}}(k) &\leq \|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m^*, k) + \frac{1}{2}\widehat{\text{qen}}(m^*) \\ &\leq \left(\|\hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}\|^2 - 4\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 - \frac{1}{2}\widehat{\text{qen}}(m^*, k) \right) + 4\|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 \\ &\quad + \frac{1}{2}\widehat{\text{qen}}(m^*). \end{aligned} \quad (6.22)$$

Now using Equations (6.21) and (6.22)

$$\begin{aligned} \frac{1}{2} \sum_{\substack{k \geq m^* \\ k \in \mathcal{M}_N}} \widehat{\text{qen}}(m^*, k) &\leq \sup_{\substack{k \geq m^* \\ k \in \mathcal{M}_N}} \left\{ \left\| \hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*}^* \right\|^2 - 4 \|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m^*, k) \right\}_+ \\ &\quad + 4 \|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 + \frac{1}{2} \widehat{\text{qen}}(m^*). \end{aligned}$$

From Equation (6.20), we now have

$$\begin{aligned} \left\| \hat{f}_{\beta, \hat{m}} - \hat{f}_{\beta, m^*} \right\|^2 \mathbf{1}_{G^c} &\leq 2 \sup_{\substack{k \geq m^* \\ k \in \mathcal{M}_N}} \left\{ \left\| \hat{f}_{\beta, k} - \hat{f}_{\beta, m^*} \right\|^2 - 4 \|f_{\beta, k} - f_{\beta, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m^*, k) \right\}_+ \\ &\quad + 8 \|f_{\beta, \hat{m}} - f_{\beta, m^*}\|^2 + \frac{1}{2} \widehat{\text{qen}}(m^*). \end{aligned}$$

Taking expectation the first summand is negligible by applying the following Proposition proved in Section 6.6.5.

Proposition 6.6.2. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_N$*

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \left\| \hat{f}_{\beta, k} - \hat{f}_{\beta, m} \right\|^2 - 4 \|f_{\beta, k} - f_{\beta, m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right] \leq \frac{C}{N}.$$

Finally we have $\mathbb{E} \left[\left\| f_{\beta} - \hat{f}_{\beta, \hat{m}} \right\|^2 \mathbf{1}_{G^c} \right] \leq C \left(\|f_{\beta} - f_{\beta, m^*}\|^2 + \text{qen}(m^*) \right) + \frac{C'}{N(J-4)} + \frac{C'}{N}$.
This combining with (6.19) complete the proof. \square

6.6.4 Proof of Lemma 6.6.1

There is nothing to prove for $\widehat{\text{qen}}_1(m)$.

• Consider $\widehat{\text{qen}}_2(m)$. For $q = 1/2$ or 1 , using Cauchy-Schwarz's inequality, we have

$$\mathbb{E} \left[\log^q \left(1 + \hat{\Xi}(m)m^2 \right) \hat{\Xi}(m) \right] \leq \sqrt{\mathbb{E} \left[\log^{2q} \left(1 + \hat{\Xi}(m)m^2 \right) \right] \mathbb{E} \left[\hat{\Xi}^2(m) \right]}.$$

$$\begin{aligned} \hat{\Xi}^2(m) &= \left(\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2}}{\left| \tilde{f}_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^4} du \right)^2 \\ &\leq \frac{1}{4\pi^2} \left(\int_{-\pi m}^{\pi m} w\left(\frac{u}{\Delta}\right)^{-2} \left| \frac{1}{\left| \tilde{f}_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^2} - \frac{1}{\left| f_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^2} + \frac{1}{\left| f_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^2} \right|^2 du \right)^2 \\ &\leq \frac{2}{\pi^2} \left(\int_{-\pi m}^{\pi m} w\left(\frac{u}{\Delta}\right)^{-2} \left| \frac{1}{\left| \tilde{f}_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^2} - \frac{1}{\left| f_{\varepsilon}^*\left(\frac{u}{\Delta}\right) \right|^2} \right|^2 du \right)^2 + 8\Xi^2(m). \end{aligned}$$

Now noticing that we can write the first term of the lastest inequality, we have

$$\begin{aligned} & \mathbb{E} \left(\int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du \right)^2 \\ &= \int_{-\pi m}^{\pi m} \int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} w \left(\frac{v}{\Delta} \right)^{-2} \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} \right|^2 \right] du dv. \end{aligned}$$

Now applying Cauchy-Schwarz's inequality and Lemma 6.8.2 for $p = 2$

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} \right|^2 \right] \\ & \leq \sqrt{\mathbb{E} \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^4 \mathbb{E} \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{v}{\Delta} \right)|^2} \right|^4} \right]} \leq C_2 |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^{-4} |f_\varepsilon^* \left(\frac{v}{\Delta} \right)|^{-4} \end{aligned}$$

So we have

$$\mathbb{E} \left[\hat{\Xi}^2(m) \right] \leq (8C_2 + 8)\Xi^2(m).$$

Besides we have

$$\hat{\Xi}(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w \left(\frac{u}{\Delta} \right)^{-2}}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du + 2\Xi(m).$$

Once again applying Lemma 6.8.2 for $p = 1$

$$\begin{aligned} & \mathbb{E} \left[\int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du \right] \\ & \leq \int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 \right] du \leq C_1 \int_{-\pi m}^{\pi m} w \left(\frac{u}{\Delta} \right)^{-2} |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^{-4} du. \end{aligned}$$

So we have

$$\mathbb{E} \left[\hat{\Xi}(m) \right] \leq (2C_1 + 2)\Xi(m).$$

Now using Jensen's inequality (since log is concave)

$$\begin{aligned} \mathbb{E} \left[\log^{2q} \left(1 + \hat{\Xi}(m)m^2 \right) \right] & \leq \log^{2q} \left(\mathbb{E} \left[1 + \hat{\Xi}(m)m^2 \right] \right) \leq \log^{2q} \left(1 + \mathbb{E} \left[\hat{\Xi}(m) \right] m^2 \right) \\ & \leq \log^{2q} \left(1 + C\Xi(m)m^2 \right) \leq C \log^{2q} \left(1 + \Xi(m)m^2 \right). \end{aligned}$$

Finally $\mathbb{E}[\widehat{\text{qen}}_2(m)] \leq C\text{qen}_2(m)$.

- Consider now $\widehat{\text{qen}}_3(m)$. Another application of Lemma 6.8.2 yields

$$\begin{aligned}
& \frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \leq \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] + \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \leq \frac{C}{N} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} + \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \leq \frac{C}{N} \Xi^\beta(m) + \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right].
\end{aligned}$$

Let us notice that

$$\begin{aligned}
e^{iuZ_{k,j}} - \mathbb{E} \left[e^{iuZ_{k,j}} \right] &= e^{iuZ_{k,j}} - f_\beta^*(u) |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 = e^{iu(\beta_k + \eta_{k,j}/\Delta)} - f_\beta^*(u) |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 \\
&= e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 \right) + \left(e^{iu\beta_k} - f_\beta^*(u) \right) |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2,
\end{aligned}$$

hence

$$\begin{aligned}
\hat{f}_Z^*(u) - f_Z^*(u) &= \frac{1}{N} \sum_{k=1}^N \left(e^{iu\beta_k} - f_\beta^*(u) \right) |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 + \\
&\quad \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} \left(e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 \right) \right). \quad (6.23)
\end{aligned}$$

Then

$$\begin{aligned}
& \frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \leq \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \quad + \frac{2}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} \left| \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} \left(e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 \right) \right) \right|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \quad (6.24)
\end{aligned}$$

Let us consider the first term on the right-hand side of Equation (6.24). We use the fact that $|\tilde{f}_\varepsilon^*(u)|^4 \geq N^{-1/2} (\log N)^{1/2} w(u)^{-1}$ as well as the independence of \hat{f}_β^* and \tilde{f}_ε^* to find

$$\begin{aligned}
& \frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \right] \\
& \leq \frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w\left(\frac{u}{\Delta}\right)^{-2} w(u)^{-2} w(u)^2 |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 N^{-1} (\log N) w\left(\frac{u}{\Delta}\right)^{-2}} du \right] \\
& \leq \mathbb{E} \left[\sup_{u \in \mathbb{R}} |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 w(u)^2 \right] \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4}{|\tilde{f}_\varepsilon^*(u)|^4} du \right].
\end{aligned}$$

Thanks to Theorem 5.1 in Neumann and Reiß (2009), for some positive constant C ,

$$\mathbb{E} \left[\sup_{u \in \mathbb{R}} |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 w(u)^2 \right] \leq \frac{C}{N}.$$

Applying Lemma 6.8.2 for $p = 1$, we get

$$\mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \right] \leq \frac{C}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2} |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \leq Cm.$$

which means that there exists C such that

$$\frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w \left(\frac{u}{\Delta} \right)^{-2} |\hat{f}_\beta^*(u) - f_\beta^*(u)|^2 |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^{12}} du \right] \leq C \frac{m}{N}.$$

Now let us consider the second term on the right-hand side of Equation (6.24) and let us notice that

$$\begin{aligned} & \mathbb{E} \left| \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} \left(e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right) \right|^2 \\ &= \frac{4}{N^2(J-4)^2} \mathbb{E} \left[\sum_{k,k'=1}^N \sum_{j,j'=3}^{J/2} e^{iu(\beta_k - \beta_{k'})} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \left(e^{iu\eta_{k',j'}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right] \\ &= \frac{4}{N^2(J-4)^2} \sum_{k,k'=1}^N \sum_{j,j'=3}^{J/2} \mathbb{E} \left[e^{iu(\beta_k - \beta_{k'})} \right] \mathbb{E} \left[\left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \left(e^{iu\eta_{k',j'}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right] \end{aligned}$$

The term $e^{iu(\beta_k - \beta_{k'})}$ is equal to 1 if $k = k'$ otherwise $\mathbb{E}[e^{iu(\beta_k - \beta_{k'})}] = |f_\beta^*(u)|^2 \leq 1$. Moreover if $k \neq k'$ and $j \neq j'$, $\mathbb{E} \left[\left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \left(e^{iu\eta_{k',j'}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right] = 0$, then

$$\begin{aligned} & \mathbb{E} \left| \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} \left(e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right) \right|^2 \\ & \leq \frac{4}{N(J-4)^2} \sum_{j=3}^{J/2} \mathbb{E} \left[\left(e^{iu\eta_{1,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \left(e^{iu\eta_{1,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right] \\ & \leq \frac{4}{N(J-4)} (1 - |f_\varepsilon^*(u)|^4) \leq \frac{4}{N(J-4)}. \end{aligned}$$

Noticing the independence between the numerator and the denominator, using that $|\tilde{f}_\varepsilon^*(u)|^4 \geq N^{-1/2} (\log N)^{1/2} w(u)^{-1}$ and applying Lemma 6.8.2 for $p = 1$, we have

$$\begin{aligned} & \frac{1}{N} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w \left(\frac{u}{\Delta} \right)^{-2} \left| \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} \left(e^{iu\beta_k} \left(e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 \right) \right) \right|^2}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^{12}} du \right] \\ & \leq \frac{4}{N(J-4)} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w \left(\frac{u}{\Delta} \right)^{-2}}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \right] \leq \frac{C}{N(J-4)} \Xi(m). \end{aligned}$$

This completes the proof. \square

6.6.5 Proof of Proposition 6.6.2

Before proving Proposition 6.6.2, we first to need prove two auxiliary lemmas. In the sequel, C will always denote some universal positive constant, but the value may vary from line to line. For $k > m$, let us introduce the following notation: $A(m, k) := \{u \in \mathbb{R}, |u| \in [\pi m, \pi k]\}$.

Lemma 6.6.3. *For an estimator of \tilde{f}_ε^* defined by (6.7), assume $\kappa > \sqrt{c_1 p}$. Let $\tau \geq 2\kappa$ and $x \geq 1$. Then for some positive constant C*

$$\mathbb{P} \left[\exists u \in \mathbb{R} : |(\tilde{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u)| > \tau (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2} \right] \leq Cx^{-p} N^{-p}.$$

Proof.

$$\begin{aligned} \left| (\tilde{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u) \right| &\leq \left| (\tilde{f}_\varepsilon^*)^4(u) - (\hat{f}_\varepsilon^*)^4(u) \right| + \left| (\hat{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u) \right| \\ &\leq 2k_N(u) + \left| (\hat{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u) \right|. \end{aligned}$$

By Lemma 6.7.3, we have

$$\begin{aligned} &\mathbb{P} \left[\exists u \in \mathbb{R} : |(\tilde{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u)| > \tau (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2} \right] \\ &\leq \mathbb{P} \left[\exists u \in \mathbb{R} : \left| (\hat{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u) \right| + 2k_N(u) > \tau (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2} \right] \\ &\leq \mathbb{P} \left[\exists u \in \mathbb{R} : |(\hat{f}_\varepsilon^*)^4(u) - (f_\varepsilon^*)^4(u)| > (\tau - 2\kappa) (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2} \right] \\ &\leq Cx^{-p} N^{-p}. \end{aligned}$$

□

Lemma 6.6.4. *In the situation of the preceding Lemma*

$$\mathbb{P} \left[\exists u \in \mathbb{R} : \left| |\tilde{f}_\varepsilon^*(u)|^2 - |f_\varepsilon^*(u)|^2 \right| \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| < |f_\varepsilon^*(u)|\}} > \frac{\tau (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2}}{|\tilde{f}_\varepsilon^*(u)|^2} \right] \leq Cx^{-p} N^{-p}.$$

Proof. This is a direct consequence of Lemma 6.6.3 using the fact that for $x, y \geq 0$, $|\sqrt{x} - \sqrt{y}| \leq \frac{|x-y|}{2\sqrt{x \wedge y}}$ holds. □

Lemma 6.6.5. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_N$*

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{24} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \leq \frac{C}{N}.$$

Proof.

$$\begin{aligned} &\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{24} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\ &\leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \frac{1}{2\pi} \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{24} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\ &\leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \sup_{t \in S(m,k)} \left| \frac{1}{2\pi} \int_{A(m,k)} t^*(u) \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) du \right|^2 - \frac{\pi(k-m)}{N} \right\}_+ \right]. \end{aligned}$$

We then study the following empirical process

$$\nu_N(t) = \frac{1}{N} \sum_{k=1}^N \left(\frac{1}{2\pi} \int_{A(m,k)} \overline{t^*(u)} e^{iu\beta_k} du - \mathbb{E} \left[\frac{1}{2\pi} \int_{A(m,k)} \overline{t^*(u)} e^{iu\beta_k} du \right] \right)$$

and define the following space: $S(m, k) = \{\text{Supp}(t) \subset A(m, k), \|t\| = 1\}$. Then we can write

$$\begin{aligned} \sup_{t \in S(m, k)} |\nu_N(t)|^2 &= \left| \int_{A(m, k)} \frac{1}{2\pi} \overline{t^*(u)} \left(\frac{1}{N} \sum_{k=1}^N e^{iu\beta_k} - f_\beta^*(u) \right) du \right|^2 \\ &\leq \sup_{t \in S(m, k)} \frac{1}{4\pi^2} \int_{A(m, k)} |t^*(u)|^2 du \int_{A(m, k)} \left| \frac{1}{N} \sum_{k=1}^N e^{iu\beta_k} - f_\beta^*(u) \right|^2 du \\ &\leq \frac{1}{2\pi} \int_{A(m, k)} \left| \frac{1}{N} \sum_{k=1}^N e^{iu\beta_k} - f_\beta^*(u) \right|^2 du, \end{aligned}$$

hence

$$\mathbb{E} \left[\sup_{t \in S(m, k)} |\nu_N(t)|^2 \right] \leq \frac{1}{2\pi} \int_{A(m, k)} \mathbb{E} \left[\left| \frac{1}{N} \sum_{k=1}^N e^{iu\beta_k} - f_\beta^*(u) \right|^2 \right] du \leq \frac{(k-m)}{N} := H^2.$$

For the variance term, we have

$$\text{Var} \left[\left| \frac{1}{2\pi} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \right| \right] \leq \mathbb{E} \left[\left| \frac{1}{2\pi} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \right|^2 \right].$$

Let us notice that the expectation can be rewritten as follows

$$\begin{aligned} \mathbb{E} \left[\left| \frac{1}{2\pi} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \right|^2 \right] &= \mathbb{E} \left[\frac{1}{4\pi^2} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \int_{A(m, k)} \overline{t^*(v)} e^{-iv\beta_1} dv \right] \\ &= \frac{1}{4\pi^2} \iiint_{\mathbb{R} \times A(m, k) \times A(m, k)} f_\beta(x) e^{i(u-v)x} \overline{t^*(u)} t^*(v) du dv dx \\ &= \frac{1}{4\pi^2} \iint_{A(m, k) \times A(m, k)} f_\beta^*(u-v) \overline{t^*(u)} t^*(v) du dv. \end{aligned}$$

Now applying Cauchy-Schwarz's inequality

$$\begin{aligned} \text{Var} \left[\left| \frac{1}{2\pi} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \right| \right] &\leq \frac{1}{4\pi^2} \iint_{A(m, k) \times A(m, k)} |f_\beta^*(u-v) \overline{t^*(u)} t^*(v)| du dv \\ &\leq \frac{1}{4\pi^2} \sqrt{\iint_{A(m, k) \times A(m, k)} |f_\beta^*(u-v)| |\overline{t^*(u)} t^*(v)|^2 du dv} \sqrt{\iint_{A(m, k) \times A(m, k)} |f_\beta^*(u-v)| |t^*(v)|^2 du dv} \\ &\leq \frac{1}{4\pi^2} \sqrt{\iint_{A(m, k) \times \mathbb{R}} |f_\beta^*(w)| |\overline{t^*(u)} t^*(v)|^2 du dv} \sqrt{\iint_{\mathbb{R} \times A(m, k)} |f_\beta^*(w)| |t^*(v)|^2 dw dv} \\ &\leq \frac{1}{4\pi^2} \iint_{A(m, k) \times A(m, k)} |f_\beta^*(w)| |\overline{t^*(u)} t^*(v)|^2 du dv \leq \frac{\|f_\beta^*\|_1}{2\pi}. \end{aligned}$$

So we choose ν as follows

$$\sup_{t \in S(m, k)} \text{Var} \left[\left| \frac{1}{2\pi} \int_{A(m, k)} \overline{t^*(u)} e^{iu\beta_1} du \right| \right] \leq \frac{\|f_\beta^*\|_1}{2\pi} := \nu.$$

And noticing that $\left| \frac{1}{2\pi} \int_{A(m,k)} \overline{t^*(u)} e^{iux} du \right|^2 \leq k - m$, we choose M_1 as

$$\sup_{x \in \mathbb{R}} \left| \frac{1}{2\pi} \int_{A(m,k)} \overline{t^*(u)} e^{iux} du \right| \leq \sqrt{k - m} := M_1.$$

We can now apply Talagrand's inequality

$$\begin{aligned} & \mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{24} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\ & \leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} K_1 \left(\frac{\|f_\beta^*\|_1}{N} e^{-K_2 \frac{k-m}{\|f_\beta^*\|_1}} + K_2 \frac{k-m}{N^2} e^{-K_3 \frac{\sqrt{N(k-m)}}{\sqrt{k-m}}} \right). \end{aligned}$$

Finally

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{24} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \leq \frac{C}{N}.$$

□

Lemma 6.6.6. *There is a positive constant C such that for any arbitrary $m \in \mathcal{M}_N$*

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du - \frac{1}{16} \widehat{\text{qen}}_2(m, k) \right\}_+ \right] \leq \frac{C}{N(J-4)}$$

with $S_{NJ}(u) = \frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} e^{iu\beta_k} (e^{iu\eta_{k,j}/\Delta} - |f_\varepsilon^*(\frac{u}{\Delta})|^2)$.

Proof. We introduce the notation $\mathbb{E}[X | \tilde{f}_\varepsilon^*, \beta]$ which corresponds to the conditional expectation of a random variable X given β_1, \dots, β_N and $\varepsilon_{k,j}$ for $j = 1, 2$ and $k = 1, \dots, N$.

$$\begin{aligned} & \mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du - \frac{1}{16} \widehat{\text{qen}}_2(m, k) \right\}_+ \right] \\ & \leq \mathbb{E} \left[\sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du - \frac{\hat{\mu}^2(m, k) \hat{\Xi}(m, k)}{N(J-4)} \right\}_+ \right] \\ & \leq \mathbb{E} \left[\sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \frac{1}{2\pi} \int_{A(m,k)} \mathbb{E} \left[\left\{ \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} - \frac{\hat{\mu}^2(m, k)}{N(J-4) |\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right\}_+ \middle| \tilde{f}_\varepsilon^*, \beta \right] du \right] \end{aligned}$$

Now $\left(\frac{2}{N(J-4)} \sum_{k=1}^N \sum_{j=3}^{J/2} e^{iu\beta_k} e^{iu\eta_{k,j}/\Delta} \right) / |\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^2$ (conditional on $\tilde{f}_\varepsilon^*(\frac{u}{\Delta})$ and β_1, \dots, β_N) is the sum of $N(J-4)$ independent and identically distributed random variables with

variance $v^2 \leq 1/|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4$ which are surely bounded by $1/|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^2$. Thus Lemma 6.7.1 gives

$$\begin{aligned} & \mathbb{E} \left[\left\{ \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} - \frac{\hat{\mu}^2(m, k)}{N(J-4)|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right\}_+ \middle| \tilde{f}_\varepsilon^*, \beta \right] \\ & \leq \frac{32}{N(J-4)|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \exp\left(-\frac{\hat{\mu}^2(m, k)}{8}\right) \\ & \quad + \frac{128\sqrt{2}}{N^2(J-4)^2|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \exp\left(-\frac{3}{16\sqrt{2}}\sqrt{N(J-4)}\hat{\mu}(m, k)\right) \\ & \leq \frac{32}{N(J-4)|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} (k-m)^{-2}\hat{\Xi}(m, k)^{-1} + \frac{128\sqrt{2}}{N^2(J-4)^2|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} (k-m)^{-2}\hat{\Xi}(m, k)^{-1} \end{aligned}$$

where we used the fact that

$$\hat{\mu}(m, k) \leq \max \left\{ \sqrt{8 \log \left(1 + \hat{\Xi}(m, k)(k-m)^2 \right)}, \frac{16\sqrt{2}}{3\sqrt{N(J-4)}} \log \left(1 + \hat{\Xi}(m, k)(k-m)^2 \right) \right\}.$$

We have thus shown for a universal positive constant C that for any $m, k \in \mathcal{M}_N$

$$\begin{aligned} & \int_{A(m, k)} \mathbb{E} \left[\left\{ \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} - \frac{\hat{\mu}^2(m, k)}{N(J-4)|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right\}_+ \middle| \tilde{f}_\varepsilon^*, \beta \right] du \\ & \leq \frac{C}{N(J-4)} (k-m)^{-2}\hat{\Xi}(m, k)^{-1} \int_{A(m, k)} \frac{du}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \leq \frac{C}{N(J-4)} (k-m)^{-2}. \end{aligned}$$

Finally,

$$\mathbb{E} \left[\sup_{\substack{k > m \\ k \in \mathcal{M}_N}} \left\{ \frac{1}{2\pi} \int_{A(m, k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du - \frac{1}{16} \widehat{\text{qen}}_2(m, k) \right\}_+ \right] \leq \frac{C}{N(J-4)}.$$

□

Proof of Proposition 6.6.2. Using Plancherel's formula, we get

$$\begin{aligned} & \|\hat{f}_{\beta, k} - \hat{f}_{\beta, m}\|^2 \\ & = \frac{1}{2\pi} \int_{A(m, k)} \frac{|\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbb{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| > |f_\varepsilon^*(\frac{u}{\Delta})|\}} du + \frac{1}{2\pi} \int_{A(m, k)} \frac{|\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbb{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\ & \leq \frac{1}{\pi} \int_{A(m, k)} \left(\frac{|\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} + \frac{|f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right) \mathbb{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| > |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\ & \quad + \frac{1}{\pi} \int_{A(m, k)} \left(|\hat{f}_Z^*(u)|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^2} - \frac{1}{|f_\varepsilon^*(\frac{u}{\Delta})|^2} \right|^2 + \frac{|\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right) \mathbb{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du. \end{aligned}$$

Then it follows that

$$\begin{aligned}
\|\hat{f}_{\beta,k} - \hat{f}_{\beta,m}\|^2 &\leq \frac{1}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| > |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&+ \frac{1}{\pi} \int_{A(m,k)} \frac{|f_Z^*(u)|^2}{|f_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| > |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&+ \frac{1}{\pi} \int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})^2 - |f_\varepsilon^*(\frac{u}{\Delta})|^2|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4 |f_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&+ \frac{2}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|f_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&+ \frac{2}{\pi} \int_{A(m,k)} \frac{|f_Z^*(u)|^2}{|f_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du.
\end{aligned}$$

Therefore we can write

$$\begin{aligned}
\|\hat{f}_{\beta,k} - \hat{f}_{\beta,m}\|^2 &\leq \frac{2}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du + \frac{2}{\pi} \int_{A(m,k)} \frac{|f_Z^*(u)|^2}{|f_\varepsilon^*(\frac{u}{\Delta})|^4} du \\
&+ \frac{1}{\pi} \int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})^2 - |f_\varepsilon^*(\frac{u}{\Delta})|^2|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4 |f_\varepsilon^*(\frac{u}{\Delta})|^4} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&\leq \frac{2}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Z^*(u) - f_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} du + 4 \|f_{\beta,k} - f_{\beta,m}\|^2 \\
&+ \frac{1}{\pi} \int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^2 - |f_\varepsilon^*(\frac{u}{\Delta})|^2|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^8} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&:= I_1(m, k) + 4 \|f_{\beta,k} - f_{\beta,m}\|^2 + I_2(m, k). \tag{6.25}
\end{aligned}$$

To bound $I_2(m, k)$, we introduce the following set

$$\begin{aligned}
C(m, k) &= \left\{ \forall u \in \mathbb{R} : \left| |\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^2 - |f_\varepsilon^*(\frac{u}{\Delta})|^2 \right|^2 \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} \leq \frac{4\kappa^2 \log(N(k-m)) w(\frac{u}{\Delta})^{-2} N^{-1}}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^4} \right\}
\end{aligned}$$

On $C(m, k)$, the following inequalities can be deduced

$$\begin{aligned}
I_2(m, k) &\leq 8\kappa^2 \log(N(k-m)) N^{-1} \frac{1}{2\pi} \int_{A(m,k)} \frac{w(\frac{u}{\Delta})^{-2} |\hat{f}_Z^*(u)|^2}{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})|^{12}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(\frac{u}{\Delta})| \leq |f_\varepsilon^*(\frac{u}{\Delta})|\}} du \\
&\leq 8\kappa^2 \log(N(k-m)) N^{-1} \hat{\Xi}^\beta(m, k) := \frac{1}{2} \widehat{\text{qen}}_3(m, k). \tag{6.26}
\end{aligned}$$

To bound $I_1(m, k)$, we use Equation (6.23) and the notation S_{NJ} defined in Lemma 6.6.6.

Thus we have

$$\begin{aligned}
I_1(m, k) &\leq \frac{4}{\pi} \int_{A(m, k)} \frac{\left| \frac{1}{N} \sum_{k=1}^N |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 (e^{iu\beta_k} - f_\beta^*(u)) \right|^2}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du + \frac{4}{\pi} \int_{A(m, k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \\
&\leq \frac{8}{\pi} \int_{A(m, k)} \frac{\left| \frac{1}{N} \sum_{k=1}^N |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 (e^{iu\beta_k} - f_\beta^*(u)) \right|^2}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du + \frac{4}{\pi} \int_{A(m, k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \\
&\quad + \frac{8}{\pi} \int_{A(m, k)} \left| \frac{1}{N} \sum_{k=1}^N |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du \\
&\leq \frac{8}{\pi} \int_{A(m, k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du + \frac{4}{\pi} \int_{A(m, k)} \frac{|S_{NJ}(u)|^2}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \\
&\quad + \frac{8}{\pi} \int_{A(m, k)} |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4 \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du.
\end{aligned} \tag{6.27}$$

We can now write the following inequalities, using Equations (6.25), (6.26) and (6.27)

$$\begin{aligned}
&\left\| \hat{f}_{\beta, k} - \hat{f}_{\beta, m} \right\|^2 - 4 \|f_{\beta, k} - f_{\beta, m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \\
&\leq \frac{8}{\pi} \int_{A(m, k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{4} \widehat{\text{qen}}_1(m, k) + \frac{4}{\pi} \int_{A(m, k)} \frac{|S_{NJ}(u)|^2}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4} du \\
&\quad - \frac{1}{2} \widehat{\text{qen}}_2(m, k) \\
&\quad + \frac{8}{\pi} \int_{A(m, k)} |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^4 \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} - \frac{1}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2} \right|^2 du \\
&\quad - \frac{1}{4} \widehat{\text{qen}}_1(m, k) + \frac{1}{2} \widehat{\text{qen}}_3(m, k) - \frac{1}{2} \widehat{\text{qen}}_3(m, k) \\
&\quad + \frac{1}{\pi} \int_{A(m, k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2 - |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^2}{|f_\varepsilon^* \left(\frac{u}{\Delta} \right)|^8} \mathbb{1}_{\{|\tilde{f}_\varepsilon^* \left(\frac{u}{\Delta} \right)| \leq |f_\varepsilon^* \left(\frac{u}{\Delta} \right)|\}} du \mathbb{1}_{\{C(m, k)^c\}}.
\end{aligned} \tag{6.28}$$

Then we take the expaction and get the following upper bound

$$\begin{aligned}
\mathfrak{E} &:= \mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \left\| \hat{f}_{\beta, k} - \hat{f}_{\beta, m} \right\|^2 - 4 \|f_{\beta, k} - f_{\beta, m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right] \\
&\leq \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \left\| \hat{f}_{\beta, k} - \hat{f}_{\beta, m} \right\|^2 - 4 \|f_{\beta, k} - f_{\beta, m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right].
\end{aligned}$$

Besides injecting Equation (6.28) into the previous inequality gives

$$\begin{aligned}
\mathfrak{E} &\leq 16 \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \frac{1}{2\pi} \int_{A(m,k)} \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 du - \frac{1}{64} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\
&\quad + \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \frac{8}{\pi} \int_{A(m,k)} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} - \frac{1}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} \right|^2 du \right. \right. \\
&\quad \left. \left. - \frac{1}{4} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\
&\quad + 8 \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|S_{NJ}(u)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4} du - \frac{1}{16} \widehat{\text{qen}}_2(m, k) \right\}_+ \right] \\
&\quad + \frac{1}{\pi} \sum_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 - |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)| \leq |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|\}} du \mathbf{1}_{\{C(m,k)^c\}} \right].
\end{aligned}$$

Now noticing that $\left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2$ and $\left| \frac{1}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} - \frac{1}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} \right|^2$ are independent and applying Lemma 6.8.2, we have

$$\begin{aligned}
&\frac{8}{\pi} \mathbb{E} \left[\int_{A(m,k)} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \left| \frac{1}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} - \frac{1}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} \right|^2 du \right] \\
&= \frac{8}{\pi} \int_{A(m,k)} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \mathbb{E} \left[\left| \frac{1}{N} \sum_{k=1}^N (e^{iu\beta_k} - f_\beta^*(u)) \right|^2 \right] \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} - \frac{1}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2} \right|^2 \right] du \\
&\leq \frac{8}{\pi} \int_{A(m,k)} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \frac{1}{N} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} du \leq \frac{16(k-m)}{N} := \frac{1}{4} \widehat{\text{qen}}_1(m, k).
\end{aligned}$$

Lemma 6.6.3 implies that $\mathbb{P}[C(m, k)^c] \leq N^{-3}(k-m)^{-3}$, we then get

$$\begin{aligned}
&\mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 - |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2}{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)| \leq |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|\}} du \mathbf{1}_{\{C(m,k)^c\}} \right] \\
&\leq 4\mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Z^*(u)|^2 \frac{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4}{k_N^4 \left(\frac{u}{\Delta}\right)} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)| \leq |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|\}} du \mathbf{1}_{\{C(m,k)^c\}} \right] \\
&\leq 4\mathbb{E} \left[\int_{A(m,k)} \kappa^{-4} (\log N)^{-2} w\left(\frac{u}{\Delta}\right)^4 N^2 du \mathbf{1}_{\{C(m,k)^c\}} \right] \leq 4\kappa^{-4} (\log N)^{-2} N^2 (k-m) \mathbb{P}[C(m, k)^c] \\
&\leq 4\kappa^{-4} (\log N)^{-2} N^2 (k-m) N^{-3} (k-m)^{-3} \leq 4\kappa^{-4} N^{-1} (k-m)^{-2}.
\end{aligned}$$

Finally applying Lemma 6.6.5 and 6.6.6, we have

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \mathcal{M}_N}} \left\{ \|\hat{f}_{\beta,k} - \hat{f}_{\beta,m}\|^2 - 4\|f_{\beta,k} - f_{\beta,m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right] \leq C \left(\frac{1}{N} + \frac{2}{N(J-4)} \right).$$

This completes the proof of Proposition 6.6.2. \square

6.7 Appendix 1

We remind, for the readers convenience, some useful results.

Lemma 6.7.1. *Let X_1, \dots, X_n be i.i.d. random variables with $\text{Var}[X_1] \leq v^2$ and suppose that almost surely $\|X_1\|_\infty \leq B$. Let $S_n = 1/n \sum_{j=1}^n (X_j - \mathbb{E}[X_1])$. Let $\mathbb{E}|S_n| \leq H$. Then*

$$\mathbb{E} \left[\left\{ |S_n|^2 - H^2 \right\}_+ \right] \leq 32 \frac{v^2}{n} \exp \left(-n \frac{H^2}{8v^2} \right) + 128\sqrt{2} \frac{B^2}{n^2} \exp \left(-n \frac{H}{\frac{16\sqrt{2}}{3}B} \right).$$

Lemma 6.7.2. *(Talagrand's inequality). Let I be some countable index set. For each $i \in I$, let $X_1^{(i)}, \dots, X_n^{(i)}$ be centered i.i.d. random variables, defined on the same probability space, with $\|X_1^{(i)}\| \leq B$ for some $B < \infty$. Let $v^2 := \sup_{i \in I} \text{Var} X_1^{(i)}$. Then for arbitrary $\epsilon > 0$, there are positive constants c_1 and $c_2 = c_2(\epsilon)$ depending only on ϵ such that for any $\kappa > 0$:*

$$\mathbb{P} \left[\left\{ \sup_{i \in I} |S_n^{(i)}| \geq (1 + \epsilon) \mathbb{E} \left[\sup_{i \in I} |S_n^{(i)}| \right] + \kappa \right\} \right] \leq 2 \exp \left(-n \left(\frac{\kappa^2}{c_1 v^2} \wedge \frac{\kappa}{c_2 B} \right) \right).$$

A proof can be found, for example, on page 170 in Massart (2003).

Next we give some technical results which will be essential for the proofs.

Lemma 6.7.3. *In the definition of \tilde{f}_ϵ^* , assume $\kappa > \sqrt{c_1 p}$. Let $\tau \geq 2\kappa$ and $x \geq 1$. Then for some positive constant C*

$$\mathbb{P} \left[\exists u \in \mathbb{R} : |\hat{f}_\epsilon^*(u)^4 - f_\epsilon^*(u)^4| > \tau (\log(Nx))^{1/2} w(u)^{-1} N^{-1/2} \right] \leq Cx^{-p} N^{-p}.$$

See Lemma 5.5 in Kappus (2014) for the proof.

Lemma 6.7.4 (Lemma 2 p.35 (Butucea and Tsybakov (2008a))). *Let γ, μ , and s be positive constants then for any $m > 0$*

$$\int_0^m (x^2 + 1)^\gamma e^{2\mu x^s} dx \approx m^{2\gamma+1-s} e^{2\mu m^s}.$$

We introduce the notation $g(x) \lesssim h(x)$ if there exists a positive constant C such that for all x , $g(x) \leq Ch(x)$ and the notation $g(x) \asymp h(x)$ if $g(x) \lesssim h(x)$ and $h(x) \lesssim g(x)$.

Lemma 6.7.5. *If f_ϵ^* satisfies assumption (6.13) then*

$$\begin{aligned} \int_{-\pi m}^{\pi m} \frac{du}{|f_\epsilon^*(\frac{u}{\Delta})|^4} &\asymp (\pi m)^{4\gamma+1-s} e^{4\mu(\pi m)^s}, \\ \int_{-\pi m}^{\pi m} \frac{|f_\beta^*(u)|^2}{|f_\epsilon^*(\frac{u}{\Delta})|^8} du &\lesssim (\pi m)^{(4\gamma+1-s) \wedge 2(2\gamma-\delta)_+} e^{4\mu(\pi m)^s} \mathbf{1}_{\{s>r\}} \\ &\quad + (\pi m)^{2(2\gamma-\delta)_+} e^{2(2\mu-a)(\pi m)^s} \mathbf{1}_{\{r=s, 2\mu \geq a\}} + \mathbf{1}_{\{r>s\} \cup \{2\mu \leq a\}}. \end{aligned}$$

A proof can be found in Comte and Lacour (2011).

6.8 Appendix 2: Supplementary material

To prove Propositions 6.2.1 and 6.2.2, we need the two following technical lemma.

Lemma 6.8.1. *Let $q \geq 1$, under Assumption (A2) there exists a constant C_q such that*

$$\mathbb{E} \left[\left| \frac{1}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^{2q} \right] \leq C_q \left(\frac{1}{|f_\varepsilon^*(x)|^{2q}} \wedge \frac{k_N^{2q}(x)}{|f_\varepsilon^*(x)|^{10q}} \right).$$

Lemma 6.8.2. *Let $p \geq 1$, under assumption (A2) there exists a constant C_p such that*

$$\mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \leq C_p \left(\frac{1}{|f_\varepsilon^*(x)|^{4p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}} \right).$$

6.8.1 Proof of Lemma 6.8.2

We start by proving Lemma 6.8.2 since Lemma 6.8.1 is obtained as a consequence of it. Let $p \geq 1$ be. Using that $1/|\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2 \leq 1/|\tilde{f}_\varepsilon^*(x)|^4$, we have

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &= \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 < k_N(x)\}} \left| \frac{1}{\sqrt{k_N(x)}} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] + \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &\leq |f_\varepsilon^*(x)|^{-4p} \mathbb{P} \left[|\tilde{f}_\varepsilon^*(x)|^4 < k_N(x) \right] \frac{\left| |f_\varepsilon^*(x)|^2 - \sqrt{k_N(x)} \right|^{2p}}{k_N(x)^p} \\ &\quad + \mathbb{E} \left[\frac{\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| |\tilde{f}_\varepsilon^*(x)|^4 - |f_\varepsilon^*(x)|^4 \right|^{2p}}{|f_\varepsilon^*(x)|^{4p} |\tilde{f}_\varepsilon^*(x)|^{4p} \left| |\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2 \right|^{2p}} \right] \\ &\leq |f_\varepsilon^*(x)|^{-4p} \frac{\left| |f_\varepsilon^*(x)|^2 - \sqrt{k_N(x)} \right|^{2p}}{k_N(x)^p} + \mathbb{E} \left[\frac{\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| |\tilde{f}_\varepsilon^*(x)|^4 - |f_\varepsilon^*(x)|^4 \right|^{2p}}{|f_\varepsilon^*(x)|^{4p} |\tilde{f}_\varepsilon^*(x)|^{8p}} \right] \quad (6.29) \end{aligned}$$

• 1st case : $|f_\varepsilon^*(x)|^4 < 2k_N(x)$. In this case we have $\frac{1}{|f_\varepsilon^*(x)|^{4p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}} = |f_\varepsilon^*(x)|^{-4p}$.

Then starting from (6.29), we get

$$\begin{aligned} & \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \leq 9^p |f_\varepsilon^*(x)|^{-4p} + |f_\varepsilon^*(x)|^{-4p} \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \frac{\left| |\tilde{f}_\varepsilon^*(x)|^4 - |f_\varepsilon^*(x)|^4 \right|^{2p}}{|f_\varepsilon^*(x)|^{8p}} \right] \\ &\leq 9^p |f_\varepsilon^*(x)|^{-4p} + |f_\varepsilon^*(x)|^{-4p} k_N(x)^{-2p} \mathbb{E} \left[\left| \widehat{|f_\varepsilon^*|^4}(x) - |f_\varepsilon^*(x)|^4 \right|^{2p} \right] \\ &\leq 9^p |f_\varepsilon^*(x)|^{-4p} + |f_\varepsilon^*(x)|^{-4p} k_N(x)^{-2p} N^{-p} \\ &\leq 9^p |f_\varepsilon^*(x)|^{-4p} + |f_\varepsilon^*(x)|^{-4p} (\log N)^{-p} w(x)^{2p} N^p N^{-p} \\ &\leq O(|f_\varepsilon^*(x)|^{-4p}). \end{aligned}$$

• 2nd case : $|f_\varepsilon^*(x)|^4 \geq 2k_N(x)$. In this case we have $\frac{1}{|f_\varepsilon^*(x)|^{4p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}} = \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}}$.
Now using the Markov and Rosenthal inequalities.

$$\begin{aligned} \mathbb{P} \left[\left| \tilde{f}_\varepsilon^*(x) \right|^4 \leq k_N(x) \right] &= \mathbb{P} \left[\widehat{|f_\varepsilon^{*4}|}(x) \leq k_N(x) \right] \leq \mathbb{P} \left[\left| \widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4 \right| > |f_\varepsilon^*(x)|^4 - k_N(x) \right] \\ &\leq \mathbb{P} \left[\left| \widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4 \right| > |f_\varepsilon^*(x)|^4 / 2 \right] \\ &\leq \frac{\mathbb{E} \left| \widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4 \right|^{2p}}{\left(|f_\varepsilon^*(x)|^4 / 2 \right)^{2p}} \leq \frac{c_p N^{-p}}{|f_\varepsilon^*(x)|^{8p}}. \end{aligned}$$

Then we can bound the first term of Equation (6.29) as follows

$$\begin{aligned} |f_\varepsilon^*(x)|^{-4p} \mathbb{P} \left[|\tilde{f}_\varepsilon^*(x)|^4 < k_N(x) \right] &\frac{\left| |f_\varepsilon^*(x)|^2 - \sqrt{k_N(x)} \right|^{2p}}{k_N(x)^p} \\ &\leq |f_\varepsilon^*(x)|^{-4p} \mathbb{P} \left[|\tilde{f}_\varepsilon^*(x)|^4 < k_N(x) \right] \frac{\left(1 + \sqrt{k_N(x)} \right)^{2p}}{k_N(x)^p} \\ &\leq |f_\varepsilon^*(x)|^{-4p} \mathbb{P} \left[|\tilde{f}_\varepsilon^*(x)|^4 < k_N(x) \right] \frac{C k_N(x)^p}{k_N(x)^p} \\ &\leq O(N^{-p} |f_\varepsilon^*(x)|^{-12p}) \leq O(k_N(x)^{2p} |f_\varepsilon^*(x)|^{-12p}). \end{aligned} \quad (6.30)$$

Moreover using this time that $1/|\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2 \leq 1/|f_\varepsilon^*(x)|^4$, we can bound the second term of Equation (6.29) as follows

$$\begin{aligned} &\mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &= \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} \right|^{2p} \left| \frac{|\widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4}{|\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &= |f_\varepsilon^*(x)|^{-4p} \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} + \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \left| \frac{|\widehat{|f_\varepsilon^{*4}|}(x) |f_\varepsilon^*(x)|^4}{|\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2} \right|^{2p} \right] \end{aligned}$$

We then deduce the following bounds

$$\begin{aligned} &\mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &\leq 2^{2p-1} |f_\varepsilon^*(x)|^{-8p} \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{|\widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4}{|\tilde{f}_\varepsilon^*(x)|^2 + |f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\ &\quad + 2 |f_\varepsilon^*(x)|^{-8p} \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \frac{\left| \widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4 \right|^{4p}}{|\tilde{f}_\varepsilon^*(x)|^{4p}} \right] \\ &\leq 2^{2p-1} |f_\varepsilon^*(x)|^{-12p} \mathbb{E} \left[\left| \widehat{|f_\varepsilon^{*4}|}(x) - |f_\varepsilon^*(x)|^4 \right|^{2p} \right] \\ &\quad + 2 |f_\varepsilon^*(x)|^{-12p} \mathbb{E} \left[\mathbb{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \frac{\left| \widehat{|f_\varepsilon^{*4}|}(x) \right|^4 - |f_\varepsilon^*(x)|^4}{|\tilde{f}_\varepsilon^*(x)|^{4p}} \right] \end{aligned}$$

which implies

$$\begin{aligned}
& \mathbb{E} \left[\mathbf{1}_{\{|\tilde{f}_\varepsilon^*(x)|^4 \geq k_N(x)\}} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \right] \\
& \leq 2^{2p-1} |f_\varepsilon^*(x)|^{-12p} N^{-p} + 2 |f_\varepsilon^*(x)|^{-12p} k_N(x)^{-p} N^{-2p} \\
& \leq 2^{2p-1} |f_\varepsilon^*(x)|^{-12p} N^{-p} + 2 |f_\varepsilon^*(x)|^{-12p} N^p N^{-2p} \\
& \leq O(|f_\varepsilon^*(x)|^{-12p} N^{-p}) \leq O(k_N(x)^{2p} |f_\varepsilon^*(x)|^{-12p}). \tag{6.31}
\end{aligned}$$

Then gathering Equations (6.30) and (6.31), we just proved that if $|f_\varepsilon^*(x)|^4 \geq 2k_N(x)$ then

$$\mathbb{E} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \leq O(k_N(x)^{2p} |f_\varepsilon^*(x)|^{-12p}).$$

In the end: $\mathbb{E} \left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \leq C_p \left(\frac{1}{|f_\varepsilon^*(x)|^{4p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}} \right)$. □

6.8.2 Proof of Lemma 6.8.1

Under (A5) and applying Lemma 6.8.2, we have

$$\begin{aligned}
& \mathbb{E} \left[\left| \frac{1}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^{2p} \right] \\
& = \mathbb{E} \left[\left| \frac{1}{|\tilde{f}_\varepsilon^*(x)|^2} - \frac{1}{|f_\varepsilon^*(x)|^2} \right|^{2p} \middle/ \left| \frac{1}{\tilde{f}_\varepsilon^*(x)} + \frac{1}{f_\varepsilon^*(x)} \right|^{2p} \right] \\
& = |f_\varepsilon^*(x)|^{2p} \mathbb{E} \left[\left| \frac{1}{(\tilde{f}_\varepsilon^*(x))^2} - \frac{1}{(f_\varepsilon^*(x))^2} \right|^{2p} \right] \\
& = |f_\varepsilon^*(x)|^{2p} C_p \left(\frac{1}{|f_\varepsilon^*(x)|^{4p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{12p}} \right) = C_p \left(\frac{1}{|f_\varepsilon^*(x)|^{2p}} \wedge \frac{k_N^{2p}(x)}{|f_\varepsilon^*(x)|^{10p}} \right).
\end{aligned}$$

□

6.8.3 Proof of Proposition 6.2.2

As aforementioned $\hat{f}_{\beta,m}$ can be seen as a projection estimator. We can then write the following equality using Pythagoras' theorem

$$\|f_\beta - \hat{f}_{\beta,m}\|^2 = \|f_\beta - f_{\beta,m}\|^2 + \|f_{\beta,m} - \hat{f}_{\beta,m}\|^2.$$

Now using Plancherel's formula, we can write

$$\begin{aligned}
\|f_{\beta,m} - \hat{f}_{\beta,m}\|^2 & = \frac{1}{2\pi} \int |f_{\beta,m}^*(u) - \hat{f}_{\beta,m}^*(u)|^2 du = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Z^*(u)}{\tilde{f}_\varepsilon^*(\frac{u}{\Delta})^2} - \frac{f_Z^*(u)}{f_\varepsilon^*(\frac{u}{\Delta})^2} \right|^2 du \\
& \leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} |\hat{f}_Z^*(u) R(\frac{u}{\Delta})|^2 du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} \left| \frac{\hat{f}_Z^*(u) - f_Z^*(u)}{f_\varepsilon^*(\frac{u}{\Delta})^2} \right|^2 du \tag{6.32}
\end{aligned}$$

with $R\left(\frac{u}{\Delta}\right) = 1/\tilde{f}_\varepsilon^*\left(\frac{u}{\Delta}\right)^2 - 1/f_\varepsilon^*\left(\frac{u}{\Delta}\right)^2$. Taking the expectation, we get

$$\mathbb{E} \left\| f_{\beta,m} - \hat{f}_{\beta,m} \right\|^2 \leq \frac{1}{\pi} \int_{-\pi m}^{\pi m} \mathbb{E} \left[\left| \hat{f}_Z^*(u) R\left(\frac{u}{\Delta}\right) \right|^2 \right] du + \frac{1}{\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} \mathbb{E} \left[\left| \hat{f}_Z^*(u) - f_Z^*(u) \right|^2 \right] du.$$

Yet we can write

$$\mathbb{E} \left[\left| \hat{f}_Z^*(u) - f_Z^*(u) \right|^2 \right] = \frac{4}{N(J-4)^2} \sum_{j=3}^{J/2} \text{Var} \left(e^{iuZ_{1,j}} \right) + \frac{4}{N(J-4)^2} \sum_{\substack{3 \leq j, j' \leq J/2 \\ j \neq j'}} \text{Cov} \left(e^{iuZ_{1,j}}, e^{iuZ_{1,j'}} \right),$$

which implies that

$$\begin{aligned} \mathbb{E} \left| \hat{f}_Z^*(u) - f_Z^*(u) \right|^2 &\leq \frac{4}{N(J-4)^2} \left(\frac{J-4}{2} + \frac{(J-4)^2}{4} (1 - |f_\beta^*(u)|^2) |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \right) \\ &\leq \frac{2}{N(J-4)} \left(1 + \frac{J-4}{2} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4 \right), \end{aligned}$$

hence

$$\frac{1}{\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} \mathbb{E} \left[\left| \hat{f}_Z^*(u) - f_Z^*(u) \right|^2 \right] du \leq \frac{1}{\pi} \frac{2}{N(J-4)} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} du + \frac{2m}{N}. \quad (6.33)$$

Now noticing that $|\hat{f}_Z^*(u) - f_Z^*(u)|$ and $|R\left(\frac{u}{\Delta}\right)|$ are independent and applying Lemma 6.8.2 for $p = 1$, we get

$$\begin{aligned} &\int_{-\pi m}^{\pi m} \mathbb{E} \left[\left| \hat{f}_Z^*(u) R\left(\frac{u}{\Delta}\right) \right|^2 \right] du \\ &= \int_{-\pi m}^{\pi m} \mathbb{E} \left[\left| \hat{f}_Z^*(u) - f_Z^*(u) + f_Z^*(u) \right|^2 |R\left(\frac{u}{\Delta}\right)|^2 \right] du \\ &\leq 2 \int_{-\pi m}^{\pi m} |f_Z^*(u)|^2 \mathbb{E} |R\left(\frac{u}{\Delta}\right)|^2 du + 2 \int_{-\pi m}^{\pi m} \mathbb{E} \left| \hat{f}_Z^*(u) - f_Z^*(u) \right|^2 \mathbb{E} |R\left(\frac{u}{\Delta}\right)|^2 du \\ &\leq 2 \int_{-\pi m}^{\pi m} |f_Z^*(u)|^2 \frac{k_N^2(u)}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du + 2 \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} \frac{2}{N(J-4)} \left(1 + \frac{J-4}{2} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^2 \right) du \\ &\leq 2C_1 \int_{-\pi m}^{\pi m} |f_Z^*(u)|^2 \frac{k_N^2(u)}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du + \frac{4m\pi}{N} + \frac{4}{N(J-4)} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} du \quad (6.34) \end{aligned}$$

Plugging (6.33) and (6.34) into Equation (6.32) yields

$$\begin{aligned} \mathbb{E} \left\| f_{\beta,m} - \hat{f}_{\beta,m} \right\|^2 &\leq \frac{6m}{N} + \frac{6}{\pi N(J-4)} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} du + \frac{2C_1}{\pi} \int_{-\pi m}^{\pi m} |f_Z^*(u)|^2 \frac{k_N^2(u)}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{12}} du \\ &\leq \frac{6m}{N} + \frac{12}{N(J-4)} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^{-4} du + \frac{2C_1}{\pi} \frac{1}{N} \int_{-\pi m}^{\pi m} s_N^2(u) \frac{|f_\beta^*(u)|^2}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} du, \end{aligned}$$

as $|f_Z^*(u)|^2 = |f_\beta^*(u)|^2 |f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4$. In the end

$$\begin{aligned} \mathbb{E} \left\| f_\beta - \hat{f}_{\beta,m} \right\|^2 &\leq \|f_\beta - f_{\beta,m}\|^2 + \frac{6m}{N} + \frac{12}{N(J-4)} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^4} \\ &\quad + \frac{4C_1}{N} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} s_N^2(u) \frac{|f_\beta^*(u)|^2}{|f_\varepsilon^*\left(\frac{u}{\Delta}\right)|^8} du. \end{aligned}$$

□

7

Estimation of convolution in the model with noise

Ce chapitre est une version modifiée de l'article *Estimation of convolution in the model with noise* publié avec F. Comte, C. Chesneau et F. Navarro dans *Journal of Nonparametric Statistics* **27**(3): 286-315 (2015).

RÉSUMÉ. Nous examinons l'estimation de la ℓ -ième convolée d'une densité d'une variable aléatoire X non observée à partir de n i.i.d observations du modèle de convolution $Y = X + \varepsilon$. Premièrement, nous supposons que la densité du bruit ε est connu et définissons des estimateurs non adaptatifs pour lesquels nous donnons une majoration du risque quadratique intégré. En particulier, sous hypothèses de régularité des densités de X et ε , nous montrons que la vitesse de convergence paramétrique $1/n$ peut être atteinte. Ensuite nous construisons un estimateur adaptatif par méthode de pénalisation ayant des performances proches de l'estimateur non adaptatif. Le prix d'adaptation est une perte logarithmique dans la vitesse. Les résultats sont étendus au cas où la densité du bruit est inconnue sous condition de disposer d'un échantillon préliminaire du bruit. Enfin, nous effectuons des simulations pour illustrer nos résultats théoriques.

ABSTRACT. We investigate the estimation of the ℓ -fold convolution of the density of an unobserved variable X from n i.i.d. observations of the convolution model $Y = X + \varepsilon$. We first assume that the density of the noise ε is known and define non adaptive estimators, for which we provide bounds for the mean integrated squared error (MISE). In particular, under some smoothness assumptions on the densities of X and ε , we prove that the parametric rate of convergence $1/n$ can be attained. Then we construct an adaptive estimator using a penalization approach having similar performances to the non adaptive one. The price for its adaptivity is a logarithmic term. The results are extended to the case of unknown noise density, under the condition that an independent noise sample is available. Lastly, we report a simulation study to support our theoretical findings.

Contents

7.1	Motivations	177
7.2	Estimation with known noise density	178
7.2.1	Notations and assumption	178
7.2.2	Estimators when the noise density is known	178
7.2.3	Risk bound for $\hat{g}_{\ell,m}$	179
7.2.4	Rates of convergence for $\hat{g}_{\ell,m}$	180
7.2.5	Risk bound on the adaptive estimator $\hat{g}_{\ell,\hat{m}}$	181
7.3	Estimation with unknown noise density	182
7.3.1	Risk bound for fixed m	182
7.3.2	Adaptive procedure	184
7.4	Illustration	185
7.5	Proofs	188
7.5.1	Proof of Proposition 7.2.1.	188
7.5.2	Proof of Corollary 7.2.2	189
7.5.3	Proof of Theorem 7.2.3.	191
7.5.4	Proof of Proposition 7.5.1	193
7.5.5	Proof of the rates in Table 7.2	195
7.5.6	Proof of Theorem 7.3.4	195
7.5.7	Proof of Lemma 7.5.3	197
7.5.8	Proof of Proposition 7.5.4	199
7.6	Appendix	203

7.1 Motivations

In actuarial or financial sciences, quantities of interest often involve sums of random variables. For example, in the individual risk model, the total amount of claims on a portfolio of insurance contracts is described by the sum of all claims on the individual policies. Consequently, it is natural to aim at estimating the density of sums of random variables, and estimating these directly from observations of the variables. A solution to this question is proposed in Chesneau et al. (2013). But the measures associated to the variables are also often not so precise, and clearly, measurement errors can occur. Thus, we consider the model

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \quad (7.1)$$

where the Y_j 's are i.i.d. with density f_Y , X_j 's are i.i.d. with density f and ε_j 's i.i.d. with density f_ε . We assume that X_j and ε_j are independent, and that f is unknown whereas, in a first step, f_ε is known. All random variables considered in this work are real valued. Note that our assumptions imply that $f_Y(u) = (f \star f_\varepsilon)(u)$, where $(h \star k)(t) = \int h(t-x)k(x)dx$ denotes the convolution product. Our aim is to estimate the ℓ -fold convolution of f for any integer $\ell \geq 1$, i.e. to estimate g_ℓ defined by

$$g_\ell(u) = \underbrace{(f \star \dots \star f)}_{\ell \text{ times}}(u). \quad (7.2)$$

A large and growing body of literature has investigated this problem without noise ($\varepsilon_j = 0$). In this case, various methods and results can be found in e.g. Frees (1994), Saavedra and Cao (2000), Ahmad and Fan (2001), Ahmad and Mugdadi (2003), Prakasa Rao (2004), Schick and Wefelmeyer (2004, 2007), Du and Schick (2007), Giné and Mason (2007), Nickl (2007), Nickl (2009) and Chesneau et al. (2013). However, the modeling of some complex phenomena implies that error disrupts the observation of X_j . Thus, Model (7.1) with $\varepsilon_j \neq 0$ (the so-called "convolution model") has been widely considered and the estimation of f in this context is a well-known problem ("deconvolution estimator" corresponding to the case $\ell = 1$). See e.g. Carroll and Hall (1988), Devroye (1989), Fan (1991), Pensky and Vidakovic (1999), Fan and Koo (2002), Butucea (2004), Butucea and Tsybakov (2008b), Comte et al. (2006), Delaigle and Gijbels (2006), Johannes (2009) and Meister (2009). The consideration of $\varepsilon_j \neq 0$ and $\ell \geq 1$ is a more general challenge with potential applications in many fields. A first approach has been investigated in Chesneau and Navarro (2014) via wavelet methods, with convincing numerical illustrations but sub-optimal theoretical results.

In this paper, we consider Model (7.1) with a special emphasis on the influences of ℓ and ε_j on the estimation of g_ℓ . In a first part, we introduce non-adaptive estimators combining kernel methods and Fourier analysis. Then we determine an upper bound of their mean integrated squared error (MISE) and exhibit fast rates of convergence according to different kinds of smoothness of f and f_ε (and suitable values of a tuning parameter). In particular, as proved in Saavedra and Cao (2000) and Chesneau et al. (2013) for the case $\varepsilon_j = 0$, we show that the parametric rate of convergence $1/n$ can be attained under some assumptions on ℓ , f and f_ε . In a second part, we develop an adaptive version of this estimator using a penalization approach. It has similar performance to the non-adaptive one, up to a logarithmic factor. This result significantly improves (Chesneau and Navarro, 2014, Proposition 5.1) in terms of rates of convergence and assumptions on the model.

In a second time, we assume that the distribution of the noise is unknown. Then, for identifiability purpose, some additional information on the errors has to be available. As in Neumann (1997), Comte and Lacour (2011) and Kappus and Mabon (2014) (see

Chapter 5), we consider the setting where an independent noise sample of size M is available. Other contexts may bring relevant information for estimation, as for example a repeated measurement setting, see [Delaigle et al. \(2008\)](#), [Bonhomme and Robin \(2010\)](#), [Neumann \(2007\)](#), [Kappus and Mabon \(2014\)](#) (see Chapter 5); these settings are not considered here. We shall take advantage of the tools developed in [Kappus \(2014\)](#) and [Kappus and Mabon \(2014\)](#) (see Chapter 5) to provide an adaptive strategy in this framework, which is more realistic but more difficult than the known noise density case.

The paper is structured as follows. We first study in Section 7.2 the case where the noise density is known. Our estimators are introduced in Section 7.2.2. Their performances in terms of MISE and rates of convergence are presented in Sections 7.2.3 and 7.2.4. The adaptive strategy is proved to reach an automatic bias-variance compromise up to negligible terms in Section 7.2.5. Section 7.3 is devoted to the case of unknown noise density: instead, to preserve identifiability of the problem, we assume that an independent noise sample is available. New estimates are proposed and studied: an adaptive procedure is proposed and proved to reach the best risk bound as possible, up to logarithmic terms. Simulation results are presented in Section 7.4. All proofs are relegated to Section 7.5.

7.2 Estimation with known noise density

7.2.1 Notations and assumption

For two real numbers a and b , we denote by $a \wedge b$ the minimum of a and b . For two functions φ and ψ belonging to $\mathbb{L}^1(\mathbb{R}) \cap \mathbb{L}^2(\mathbb{R})$, we denote by φ^* the Fourier transform defined by $\varphi^*(x) = \int \varphi(t)e^{-ixt} dt$, by $\|\varphi\|$ the \mathbb{L}^2 -norm of φ , $\|\varphi\|^2 = \int |\varphi(u)|^2 du$, by $\langle \varphi, \psi \rangle$ the scalar product $\langle \varphi, \psi \rangle = \int \varphi(u)\bar{\psi}(u)du$. We recall the inverse Fourier formula $\varphi(x) = (2\pi)^{-1} \int e^{iux} \varphi^*(u) du$, and that, for the convolution product $(\varphi \star \psi)(t) = \int \varphi(t-x)\psi(x)dx$, we have by Fourier transform $(\varphi \star \psi)^*(u) = \varphi^*(u)\psi^*(u)$. Lastly $\langle \varphi, \psi \rangle = (2\pi)^{-1} \langle \varphi^*, \psi^* \rangle$.

Moreover, we work all along the paper under the assumption

(A1) $f_\varepsilon^*(u) \neq 0, \forall u \in \mathbb{R}$.

7.2.2 Estimators when the noise density is known

We define now our estimators, considering that the observations are drawn from Model (7.1) and assuming that f_ε is known and satisfies **(A1)**.

The idea for this construction is based on the following relation. First under independence of X_j and ε_j in Model (7.1), we have $f_Y = f \star f_\varepsilon$, which implies $f_Y^* = f^* f_\varepsilon^*$. Moreover, $g_\ell^*(u) = (f^*(u))^\ell$, thus we have, using **(A1)**,

$$g_\ell(x) = \frac{1}{2\pi} \int g_\ell^*(u) e^{iux} du = \frac{1}{2\pi} \int (f^*(u))^\ell e^{iux} du = \frac{1}{2\pi} \int \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell e^{iux} du.$$

Then f_Y^* can easily be estimated by its empirical counterpart

$$\hat{f}_Y^*(u) = \frac{1}{n} \sum_{j=1}^n e^{-iuY_j}.$$

But a pure plug-in approach leads to integrability problems, this is why a cutoff has to be introduced, which can be seen as a truncated version of the empirical estimator: $\hat{f}_Y^*(u) \mathbf{1}_{[-\pi m, \pi m]}(u)$. Thus, we propose the estimator, $\forall m \in \mathbb{N}$,

$$\hat{g}_{\ell,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left(\frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell e^{iux} du. \quad (7.3)$$

The cutoff can also be seen as a particular kernel (the so-called "sinc" kernel) strategy, and the estimator (7.3) corresponds for $\ell = 1$ to the one of Fan (1991), Butucea (2004), Butucea and Tsybakov (2008b), Comte and Lacour (2011), see also Meister (2009) and references therein.

Naturally, the performance of $\hat{g}_{\ell,m}$ depends on the choice of m . We make $\hat{g}_{\ell,m}$ adaptive by considering the estimator: $\hat{g}_{\ell,\hat{m}}$, where \hat{m} is the integer defined by

$$\hat{m} = \arg \min_{m \in \{1, \dots, n\}} \left(-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_\ell(m) \right). \quad (7.4)$$

where, for numerical constant ϑ specified later,

$$\text{pen}_\ell(m) = \vartheta 2^{2\ell-1} \lambda_\ell^{2\ell}(m, \Delta_\ell(m)) \frac{\Delta_\ell(m)}{n^\ell}, \quad \Delta_\ell(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(u)|^{2\ell}} du \quad (7.5)$$

and

$$\lambda_\ell(m, D) = \max \left\{ 2^{-1/\ell+1} \left(\log(1 + m^2 D) \right)^{1/2}, \frac{2^{2-1/(2\ell)}}{\sqrt{n}} \log(1 + m^2 D) \right\}. \quad (7.6)$$

7.2.3 Risk bound for $\hat{g}_{\ell,m}$

Now we can prove the following risk bound for the MISE, a decomposition which explains our adaptive strategy and will allow to discuss rates.

Proposition 7.2.1. *Consider data from Model (7.1), g_ℓ given by (7.2) for $\ell \geq 1$, and assume that (A1) is fulfilled and f is square integrable. Let $\hat{g}_{\ell,m}$ be defined by (7.3). Then for any nonzero integer m , we have*

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{2\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^{\ell} \binom{\ell}{k} \frac{C_k}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du, \quad (7.7)$$

with $C_k = (72k)^{2k} (2k/(2k-1))^k$.

Inequality (7.7) deserves two comments. First, we note that this bound is non asymptotic, and so is the adaptive strategy inspired from it. Secondly, we can see the usual antagonism between the bias term $(2\pi)^{-1} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du$ which is smaller for large m and the variance term which gets larger when m increases.

Now, we can see how Inequality (7.7) allows us to explain the definition of the model selection criterion given in (7.4). The bias term can also be written $\|g_\ell\|^2 - \|g_{\ell,m}\|^2$ with $g_{\ell,m}^*(u) = g_\ell^*(u) \mathbf{1}_{[-\pi m, \pi m]}(u)$; as $\|g_\ell\|^2$ is a constant, only $-\|g_{\ell,m}\|^2$ is estimated, by $-\|\hat{g}_{\ell,m}\|^2$. Usually, the variance is replaced by its explicit bound, given here by the second right-hand-side term of (7.7). It is noteworthy that here, the variance bound is the sum of ℓ terms and only the last term is involved in the penalty. This is rather unexpected and specific to our particular setting. These considerations explain the definition of (7.4) and the selection procedure. Note that the main difference between the term corresponding to $k = \ell$ in the variance bound (7.7) and the penalty given by (7.5)-(7.6) lies in additional logarithmic terms contained in the definition of $\lambda_\ell(m, \Delta_\ell(m))$: they represent the price of adaptation, but it is worth stressing that they do not require any parametric assumption on the rate of decrease of f^* or f_ε^* .

7.2.4 Rates of convergence for $\hat{g}_{\ell,m}$

Now, from an asymptotic point of view, we can investigate the rates of convergence of our estimators according to the smoothness assumptions on f and f_ε . The following balls of function spaces are considered, for positive α, β ,

$$\mathcal{A}(\beta, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}); \int |f^*(u)|^2 \exp(2|u|^\beta) du \leq L, \sup_{u \in \mathbb{R}} (|f^*(u)|^2 \exp(2|u|^\beta)) \leq L \right\}$$

and

$$\mathcal{S}(\alpha, L) = \left\{ f \in \mathbb{L}^1 \cap \mathbb{L}^2(\mathbb{R}); \int |f^*(u)|^2 (1+u^2)^\alpha du \leq L, \sup_{u \in \mathbb{R}} (|f^*(u)|^2 (1+u^2)^\alpha) \leq L \right\}.$$

As the noise density is assumed to be known, regularity assumptions on f_ε^* are standardly formulated in the following way, for positive a, b :

[OS(a)] f_ε^* is OS(a) if $\forall u \in \mathbb{R}$,

$$\frac{c_\varepsilon}{(1+u^2)^a} \leq |f_\varepsilon^*(u)|^2 \leq \frac{C_\varepsilon}{(1+u^2)^a}.$$

[SS(b)] f_ε^* is SS(b) if $\forall u \in \mathbb{R}$,

$$c_\varepsilon e^{-2|u|^b} \leq |f_\varepsilon^*(u)|^2 \leq C_\varepsilon e^{-2|u|^b}.$$

Note that the lower bound for f_ε^* is the one used to obtain upper bound of the risk. These classical sets allow to compute rates, depending on the types of f and f_ε and to get the following consequence of Proposition 7.2.1.

Corollary 7.2.2. *Under the assumptions of Proposition 7.2.1, the following table exhibits the values of m and φ_n such that φ_n is the rate of convergence satisfying “there exists a constant $C_\ell^* > 0$ such that $\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq C_\ell^* \varphi_n$ ”, according some smoothness assumptions on f and f_ε .*

	$f \in \mathcal{A}(\beta, L), f_\varepsilon$ is OS(a)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is SS(b)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is OS(a)
m_{opt}	$\frac{1}{\pi} \left(\frac{\log(n)}{2\ell} \right)^{1/\beta}$	$\frac{1}{\pi} \left(\frac{\log(n)}{4\ell} \right)^{1/b}$	$n^{k_0/(2ak_0+2\alpha k_0+1)}$
φ_n	n^{-1} for $\ell \geq 2$ $\frac{[\log(n)]^{(2a+1)/\beta}}{n}$ for $\ell = 1$	$(\log(n))^{-2\ell\alpha/b}$	$n^{-2\alpha\ell/(2a+2\alpha+1)} \vee n^{-1}$

where k_0 is the integer such that $a_\ell(k_0 - 1) < 2\alpha < a_\ell(k_0)$ where $a_\ell(0) = 0$, $a_\ell(k) = (2ak+1)/(\ell-k)$, $k = 1, \dots, \ell-1$, with convention $a_1(1) = +\infty$. In the case: $f \in \mathcal{S}(\alpha, L)$, f_ε is OS(a), for $\ell \geq 2$, if $2\alpha = a_\ell(1)$, the rate is $\log(n)/n$, and if $2\alpha > a_\ell(1)$, the rate is $1/n$.

The parametric rate of convergence $1/n$ is attained for the case $f \in \mathcal{A}(\beta, L)$, f_ε is OS(a) for $\ell \geq 2$, and in the case $f \in \mathcal{S}(\alpha, L)$, f_ε is OS(a) when $\ell > 1 + (a + 1/2)/\alpha$. To the best of our knowledge, Theorem 7.2.1 is the first result showing rates of convergence for an estimator in the context of Model (7.1) for different types of smoothness on f and f_ε .

The results correspond to classical deconvolution rates for $\ell = 1$ (see Comte et al. (2006)). Moreover, this improves the rate of convergence determined in Proposition 5.1 of Chesneau and Navarro (2014).

The case where both f and f_ε are supersmooth is not detailed. For $\ell = 1$, the rates corresponding to this case, are implicit solutions of optimization equations (see Butucea and Tsybakov (2008b)) and recursive formula are given in Lacour (2006). Explicit polynomial rates can be obtained, and also rates faster than logarithmic but slower than polynomial. We do not investigate them here for sake of conciseness. We can emphasize that the function f being unknown, its regularity is also unknown, and the asymptotic optimal choice of m cannot be done in practice. This is what makes an automatic choice of m mandatory to obtain a complete estimation procedure. The automatic resulting trade-off avoids tedious computations of asymptotic rates and perform at best for all sample sizes.

7.2.5 Risk bound on the adaptive estimator $\hat{g}_{\ell, \hat{m}}$

Theorem 7.2.3 determines risk bound for the MISE of the estimator $\hat{g}_{\ell, \hat{m}}$.

Theorem 7.2.3. *Consider Model (7.1) under Assumption (A1). Let $g_{\ell, m}$ such that $g_{\ell, m}^* = g_\ell^* \mathbf{1}_{[-\pi m, \pi m]}$ and let $\hat{g}_{\ell, \hat{m}}$ be defined by (7.3)-(7.4) and define*

$$\mathcal{V}_\ell(m) := \frac{2^\ell - 1}{\pi} \sum_{k=1}^{\ell} \binom{\ell}{k} \frac{C_k}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du. \quad (7.8)$$

Then, there exists ϑ_0 such that for any $\vartheta \geq \vartheta_0$, we have

$$\mathbb{E}(\|\hat{g}_{\ell, \hat{m}} - g_\ell\|^2) \leq 2^{2\ell+1} \min_{m \in \{1, \dots, n\}} \left(\|g_\ell - g_{\ell, m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}_\ell(m) \right) + \frac{c_\ell}{n^\ell}$$

where c_ℓ is a constant independent of n . Note that any $\vartheta_0 \geq 2$ suits.

The result in Theorem 7.2.3 shows that the adaptive estimator $\hat{g}_{\ell, \hat{m}}$ performs an automatic trade-off between the bias and variance terms described in Proposition 7.2.1. Moreover, the bound on the risk given in Theorem 7.2.3 is non asymptotic, and holds for any sample size. We note that $\vartheta_0 = 2$ suits, and the larger ϑ_0 , the larger $\text{pen}_\ell(m)$ and thus the bound. In practice, this constant is fixed by preliminary simulation experiments.

As a consequence, we can derive asymptotic results. Table 7.1 exhibits the rates of convergence $\varphi_n^{(a)}$ satisfying “there exists a constant $C_\ell^{**} > 0$ such that $\mathbb{E}(\|\hat{g}_{\ell, \hat{m}} - g_\ell\|^2) \leq C_\ell^{**} \varphi_n^{(a)}$ ”, depending on smoothness assumptions on f and f_ε .

	$f \in \mathcal{A}(\beta, L), f_\varepsilon$ is OS(a)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is SS(b)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is OS(a)
$\varphi_n^{(a)}$	$\log(n)/n, \ell \geq 2$	$(\log(n))^{-2\ell\alpha/b}$	$(n/\log(n))^{-2\alpha\ell/(2a+2\alpha+1)} \vee (\log(n)/n)$

Table 7.1: Examples of rates of convergence $\varphi_n^{(a)}$ of the estimator $\hat{g}_{\ell, \hat{m}}$

Theorem 7.2.3 implies thus that $\hat{g}_{\ell, \hat{m}}$ attains similar rates of convergence to $\hat{g}_{\ell, m}$ with automatic “optimal” choices for m . The only difference is possibly a logarithmic term.

7.3 Estimation with unknown noise density

We consider now that f_ε is unknown, but still satisfying **(A1)**. Then, the estimation problem is no longer identifiable, unless additional information is available. To that aim, we assume that we have at hand a sample $\varepsilon'_1, \dots, \varepsilon'_M$ of i.i.d. replications of ε_1 , independent of the first sample.

7.3.1 Risk bound for fixed m

We have to replace in (7.3) the noise characteristic function f_ε^* by an estimator, and to prevent it from getting too small in the denominator. Thus, we propose the estimator, $\forall m \in \mathbb{N}$,

$$\tilde{g}_{\ell,m}(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} \right)^\ell e^{iux} du,$$

where still $\hat{f}_Y^*(u) = n^{-1} \sum_{j=1}^n e^{-iuY_j}$ and now

$$\tilde{f}_\varepsilon^*(u) = \begin{cases} \hat{f}_\varepsilon^*(u) & \text{if } |\hat{f}_\varepsilon^*(u)| \geq k_M(u) \\ k_M(u) & \text{otherwise} \end{cases} \quad \text{where } \hat{f}_\varepsilon^*(u) = \frac{1}{M} \sum_{j=1}^M e^{-iu\varepsilon'_j}, \quad (7.10)$$

and

$$k_M(u) = \kappa s_M(u) M^{-1/2}. \quad (7.11)$$

Here κ is a numerical constant, which will be calibrated in the simulations.

The definition of s_M is simply 1 when considering one estimator in the collection, and more elaborated to deal with the adaptive case. To study the risk bound of the estimator, we recall the following key Lemma (see Neumann (1997), Comte and Lacour (2011) and Kappus (2014)).

Lemma 7.3.1. *Assume that **(A1)** holds. Let $s_M(u) \equiv 1$ in ((7.11) for \tilde{f}_ε defined by (7.10), then for any $p \geq 1$, there exists a constant τ_p such that, $\forall u \in \mathbb{R}$,*

$$\mathbb{E} \left(\left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right|^{2p} \right) \leq \tau_p \left(\frac{1}{|f_\varepsilon^*(u)|^{2p}} \wedge \frac{(k_M(u))^{2p}}{|f_\varepsilon^*(u)|^{4p}} \right).$$

To study the MISE of the $\tilde{g}_{\ell,m}$, we write $\|\tilde{g}_{\ell,m} - g_\ell\|^2 = \|g_\ell - g_{\ell,m}\|^2 + \|\tilde{g}_{\ell,m} - g_{\ell,m}\|^2$, which starts the bias-variance decomposition. Then we can split the integrated variance term of $\tilde{g}_{\ell,m}$ into three terms

$$\begin{aligned} \|\tilde{g}_{\ell,m} - g_{\ell,m}\|^2 &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} + \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell - \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^\ell \right|^2 du \\ &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \left| \sum_{k=1}^{\ell} \binom{\ell}{k} \left(\frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^{\ell-k} \left(\frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right)^k \right|^2 du \\ &\leq \frac{2^\ell - 1}{2\pi} \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| \frac{\hat{f}_Y^*(u)}{\tilde{f}_\varepsilon^*(u)} - \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2k} du \\ &\leq \frac{2^\ell}{2\pi} (T_1 + T_2 + T_3) \end{aligned}$$

with

$$\begin{aligned}
T_1 &= \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| \frac{\hat{f}_Y^*(u) - f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2k} du \\
T_2 &= \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| (\hat{f}_Y^*(u) - f_Y^*(u)) \left(\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right) \right|^{2k} du \\
T_3 &= \int_{-\pi m}^{\pi m} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \left| \frac{f_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2(\ell-k)} \left| f_Y^*(u) \left(\frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right) \right|^{2k} du.
\end{aligned}$$

Clearly, T_1 is the term corresponding to known f_ε^* and we have $\mathbb{E}(T_1) \leq 3^{2\ell-1} \mathcal{V}_\ell$. Using the first bound in Lemma 7.3.1 and the independence of the samples, we have also $\mathbb{E}(T_2) \leq 3^{2\ell-1} (\max_{1 \leq k \leq \ell} \tau_k) \mathcal{V}_\ell$. The novelty comes from T_3 , where we use the second bound given in Lemma 7.3.1:

$$\begin{aligned}
\mathbb{E}(T_3) &\leq \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \int_{-\pi m}^{\pi m} \frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2(\ell-k)}} \tau_k \frac{M^{-k}}{|f_\varepsilon^*(u)|^{4k}} du \\
&\leq \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \tau_k \frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2(\ell+k)}} du \\
&= \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \tau_k \frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du := \mathcal{W}_\ell.
\end{aligned}$$

Therefore, the following result holds.

Proposition 7.3.2. *Consider Model (7.1) under (A1). Let $\tilde{g}_{\ell,m}$ be defined by (7.9)-(7.11) with $s_M(u) \equiv 1$. Then there exists constants $C_\ell^{**}, C_\ell^{***} > 0$ (independent of n) such that*

$$\mathbb{E}(\|\tilde{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + C_\ell^{**} \mathcal{V}_\ell + C_\ell^{***} \mathcal{W}_\ell,$$

where \mathcal{V}_ℓ is defined by (7.8) and \mathcal{W}_ℓ by (7.12).

We can see that if $M \geq n$, we have

$$\frac{1}{M^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du \leq \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du$$

and thus $\mathcal{W}_\ell \leq C_\ell^{(4)} \mathcal{V}_\ell$. Thus, for $M \geq n$, the risk of $\tilde{g}_{\ell,m}$ has the same order as the risk of $\hat{g}_{\ell,m}$ and the estimation of the noise characteristic function does not modify the rate of the estimator. The results given in Corollary 7.2.2 are thus still valid. Obviously, if the noise sample size M is smaller than n , it may imply deterioration of the rates. Moreover, the above result is a generalization of Neumann (1997) and Comte and Lacour (2011), which both correspond to the case $\ell = 1$.

More generally, we can derive asymptotic results, following the discussion in Johannes and Schwarz (2013b), with the analogy for noise levels $(\varepsilon, \sigma) = (1/n, 1/M)$.

Corollary 7.3.3. *Under the assumptions of Proposition 7.3.2 and for $\ell \geq 2$, we have: “there exists a constant $K_\ell > 0$ and a choice $m^* = \min(m_{opt}, m_{opt}^{(2)})$ such that $\mathbb{E}(\|\tilde{g}_{\ell,m^*} - g_\ell\|^2) \leq K_\ell \max(\varphi_n, \psi_M)$ ”, depending on smoothness assumptions on f and f_ε , with φ_n and m_{opt} as in Corollary 7.2.2 and ψ_M and $m_{opt}^{(2)}$, given in Table 7.2.*

	$f \in \mathcal{A}(\beta, L), f_\varepsilon$ is OS(a)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is SS(b)	$f \in \mathcal{S}(\alpha, L), f_\varepsilon$ is OS(a)
$m_{opt}^{(2)}$	$\frac{1}{\pi} \left(\frac{\log M}{2\ell} \right)^{1/\beta}$	$\frac{1}{\pi} \left(\frac{\log M}{4\ell} \right)^{1/b}$	$M^{\frac{k_0}{2ak_0+1}}$
ψ_M	$1/M$	$(\log M)^{-2\ell\alpha/b}$	if $\alpha < a + 1/2$, $M^{-2\ell\alpha/(2a+1)}$ if $\alpha = a + 1/2$, $\log(M)/M$ if $\alpha > a + 1/2$, $1/M$

Table 7.2: Examples of values of $m_{opt}^{(2)}$ and ψ_M , for $\ell \geq 2$, k_0 is an integer such that $b_\ell(k_0 - 1) < 2\alpha \leq b_\ell(k_0)$, for $b_\ell(k) = (2ak + 1)/\ell$, $k = 1, \dots, \ell$ and $b_\ell(0) = 0$.

7.3.2 Adaptive procedure

Adaptive estimation here is a difficult task. We apply the methodology recently proposed by Kappus and Mabon (2014) (see Chapter 5) in the context $\ell = 1$, improving the results of Comte and Lacour (2011). In particular the estimator \tilde{f}_ε^* is now taken with s_M in (7.11) defined as follows. Let for a given $\delta > 0$,

$$\forall u \in \mathbb{R}, \quad w(u) = (\log(e + |u|))^{-\frac{1}{2}-\delta},$$

originally introduced in Neumann and Reiß (2009), and \tilde{f}_ε^* given by (7.10)-(7.11) and by

$$s_M(u) = (\log M)^{1/2} w^{-1}(u). \quad (7.13)$$

It follows from Chapter 6 that Lemma 7.3.1 still holds with this choice of $s_M(u)$.

With this threshold, the bound given in Proposition 7.3.2 becomes

$$\mathbb{E}(\|\tilde{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \mathbb{V}_\ell$$

with

$$\mathbb{V}_\ell = C_\ell^{**} \mathcal{V}_\ell + C_\ell^{****} \sum_{k=1}^{\ell} \binom{\ell}{k} 3^{2k-1} \tau_k \frac{\log^k(M)}{M^k} \int_{-\pi m}^{\pi m} \frac{w^{-2k}(u) |f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2k}} du. \quad (7.14)$$

Clearly, the definition of s_M here implies logarithmic loss if we compare the last term above with \mathcal{W}_ℓ given by (7.12). Then using k_M defined by (7.11) and (7.13) and considering $\tilde{g}_{\ell,m}$ defined by (7.9)-(7.10), we propose the following adaptive strategy. We set first

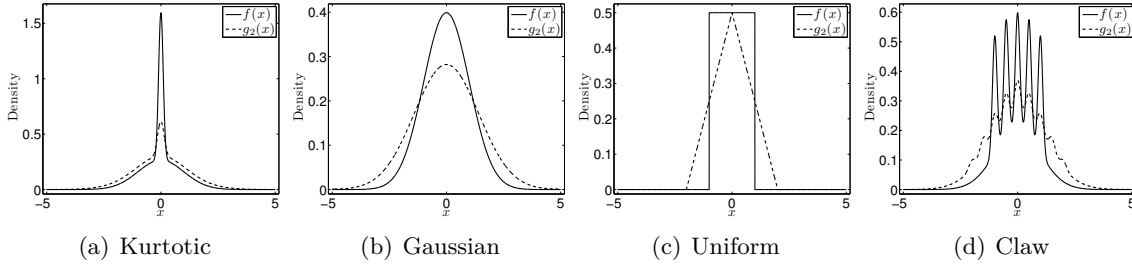
$$\hat{\Delta}_\ell(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du \quad \text{and} \quad \hat{\Delta}_\ell^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du,$$

as estimators of $\Delta_\ell(m)$ given by (7.5) and of $(1/(2\pi)) \int_{-\pi m}^{\pi m} |f^*(u)|^{2\ell} / |f_\varepsilon^*(u)|^{2\ell} du$. Due to the additional $w(u)$ factors, these terms are replaced by

$$\Delta_\ell^*(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du, \quad \Delta_\ell^f(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |f^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du.$$

We can now define the stochastic penalty associated with the adaptive procedure

$$\begin{aligned} \widehat{\text{qen}}_\ell(m) &= \widehat{\text{qen}}_{\ell,1}(m) + \widehat{\text{qen}}_{\ell,2}(m) \\ &= 2^{4\ell} \lambda_\ell^{2\ell}(m, \hat{\Delta}_\ell(m)) \frac{\hat{\Delta}_\ell(m)}{n^\ell} + 2^{2\ell} \kappa^{2\ell} \log^\ell(Mm) \frac{\hat{\Delta}_\ell^f(m)}{M^\ell} \end{aligned}$$

Figure 7.1: Test densities f (solid) and g_2 (dashed).

which estimates the deterministic quantities

$$\begin{aligned} \text{qen}_\ell(m) &= \text{qen}_{\ell,1}(m) + \text{qen}_{\ell,2}(m) \\ &= 2^{2\ell+1} \lambda_\ell^{2\ell}(m, \Delta_\ell^*(m)) \frac{\Delta_\ell^f(m)}{n^\ell} + 2^{2\ell+1} \kappa^2 \log^\ell(Mm) \frac{\Delta_\ell^f(m)}{M^\ell} \end{aligned}$$

where the weights $\lambda_\ell(m, D)$ are defined by (7.6). Then, we select the cutoff parameter \tilde{m} as a minimizer of the following penalized criterion

$$\tilde{m} = \arg \min_{m \in \{1, \dots, n\}} \left\{ -\|\tilde{g}_{\ell, m}\|^2 + \widehat{\text{qen}}_\ell(m) \right\}. \quad (7.15)$$

Theorem 7.3.4. Consider Model (7.1) under (A1), associated with an independent noise sample $\varepsilon'_1, \dots, \varepsilon'_M$. Let the estimator $\tilde{g}_{\ell, \tilde{m}}$ be defined by (7.9) and (7.15). Then there are positive constants C^{ad} and C such that

$$\mathbb{E} \left(\|g_\ell - \tilde{g}_{\ell, \tilde{m}}\|^2 \right) \leq C^{ad} \inf_{m \in \{1, \dots, n\}} \left\{ \|g_\ell - g_{\ell, m}\|^2 + \mathbb{V}_\ell(m) + \text{qen}_\ell(m) \right\} + \frac{C}{n^\ell} + \frac{C}{M^\ell},$$

where \mathbb{V}_ℓ is defined by (7.14).

This result is a generalization of Kappus and Mabon (2014) (see Chapter 5), which corresponds to the case $\ell = 1$. As in the known density case, we do not use the whole variance for penalization, but only the last terms. Moreover, the final estimator automatically reaches the bias-variance compromise, up to logarithmic terms.

7.4 Illustration

We now illustrate the theoretical results by a simulation study within the context described in Section 7.2 and Section 7.3. That is, we consider the problem of estimating the density g_ℓ , emphasizing the case $\ell = 2$, for both known and unknown types of noises. All simulations have been implemented under Matlab.

The performance of the proposed method is studied for four sets of test distributions for X_j representing different degrees of smoothness (see Figure 7.1),

- ▷ Kurtotic distribution: $\frac{2}{3}\mathcal{N}(0, 1) + \frac{1}{3}\mathcal{N}(0, (1/10)^2)$,
- ▷ Standard normal distribution: $\mathcal{N}(0, 1)$,
- ▷ Uniform distribution: $\mathcal{U}(-1, 1)$,
- ▷ Claw distribution: $\frac{1}{2}\mathcal{N}(0, 1) + \sum_{l=0}^4 \mathcal{N}(l/2 - 1, (1/2)^2)$.

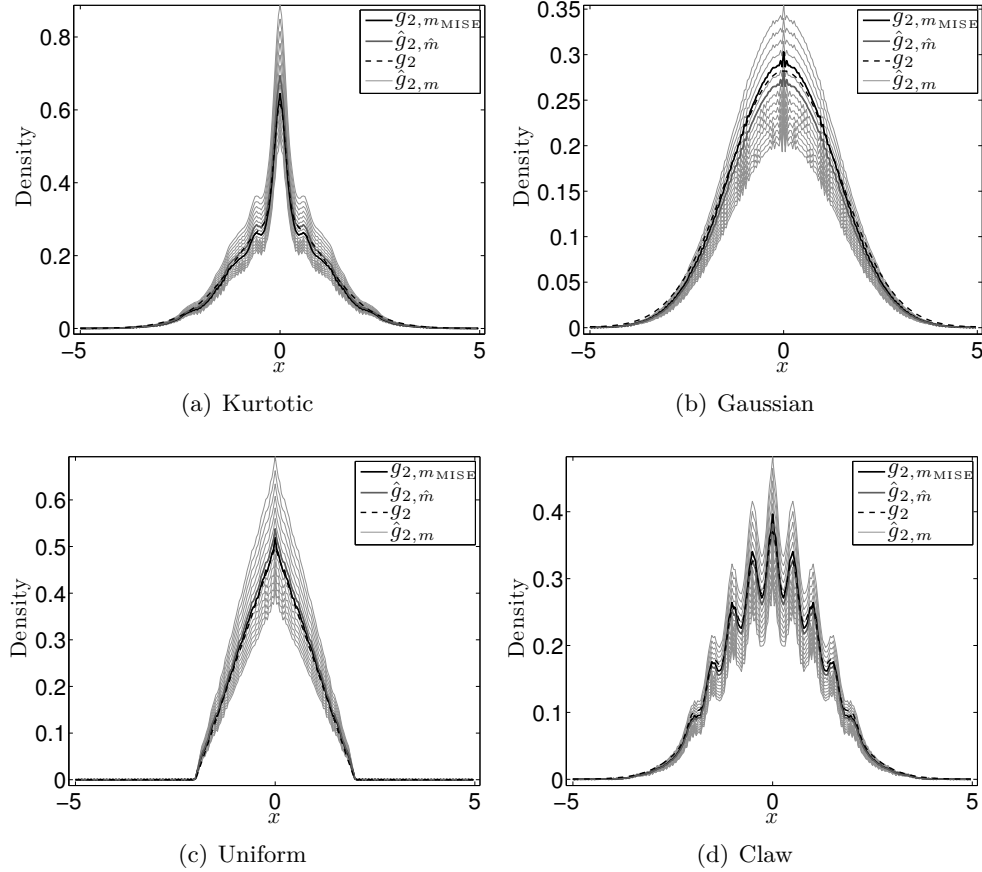


Figure 7.2: Laplace noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.

Note that the two Gaussian mixture densities are taken from [Marron and Wand \(1992\)](#).

We consider two types of error density f_ε , with the same variance $\sigma_\varepsilon^2 = 0.25$. The first one is Laplace distribution, with f_ε^* OS(2), and the second one is the Gaussian distribution with f_ε^* SS(2).

$$(a) \text{ Laplace error: } f_\varepsilon(x) = \frac{1}{2\sigma_\varepsilon} \exp\left(-\frac{|x|}{\sigma_\varepsilon}\right), \quad f_\varepsilon^*(x) = \frac{1}{1+\sigma_\varepsilon^2 x^2}.$$

$$(b) \text{ Gaussian error: } f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right), \quad f_\varepsilon^*(x) = \exp\left(-\frac{\sigma_\varepsilon^2 x^2}{2}\right),$$

In the case of an entirely known noise distribution, the constant ϑ in (7.5) is set to 1.8 for all the tests and in the unknown case the penalties are chosen according to Theorem 7.3.4. We propose the following penalty:

$$\widehat{\text{qen}}_\ell(m) = \kappa_1 \lambda_\ell^{2\ell}(m, \hat{\Delta}_\ell(m)) \frac{\hat{\Delta}_\ell(m)}{n^\ell} + \kappa_2 \kappa^{2\ell} \log^\ell(Mm) \frac{\hat{\Delta}_\ell^f(m)}{M^\ell}$$

with the associated constants penalties $\kappa_1 = \kappa_2 = 1$ and $\kappa = 1.8$. The choice of the constants $\vartheta, \kappa_1, \kappa_2$ and κ is assessed by preliminary intensive simulation experiments (see explanation about slope heuristics and practical calibration methods in [Baudry et al. \(2012\)](#)). As in [Chesneau et al. \(2013\)](#), the estimators proposed in this paper are well suited for FFT-based method. Thus, the resulting estimators $\hat{g}_{\ell, \hat{m}}$ and $\tilde{g}_{\ell, \tilde{m}}$ are simple to implement and fast

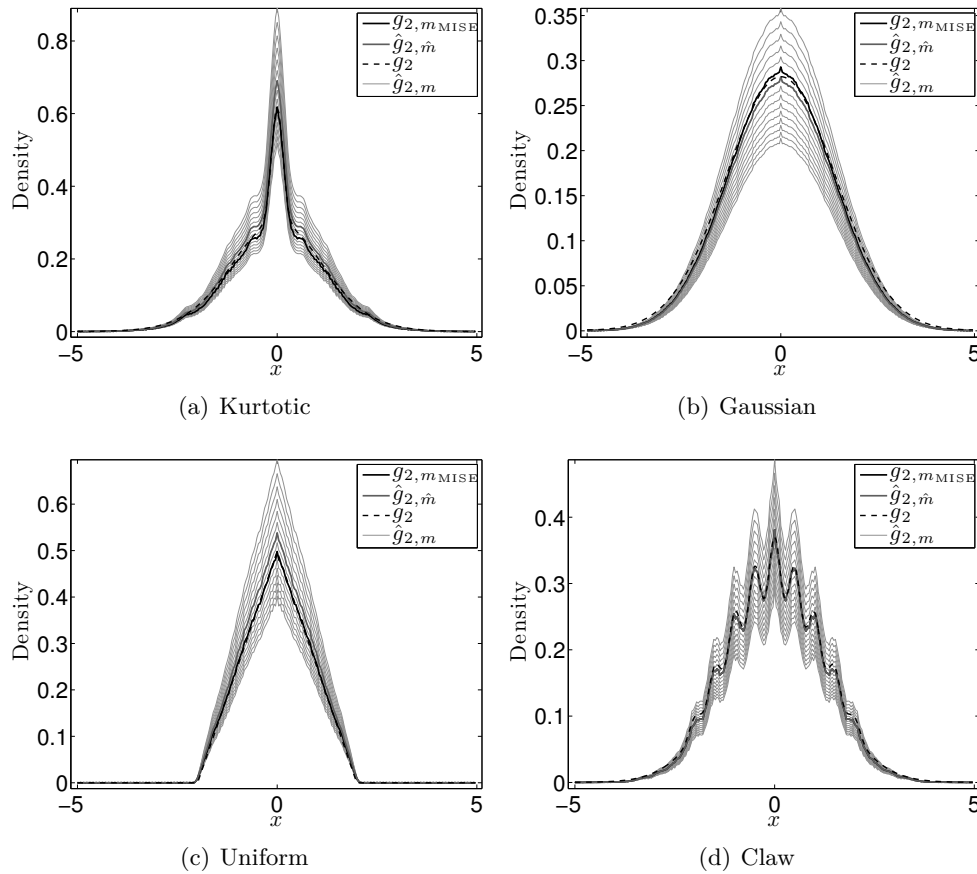


Figure 7.3: Gaussian noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.

which allows us to perform the penalized criterion in a reasonable time. For numerical implementation, we consider an interval $[a, b]$ that covers the range of the data and the density estimates were evaluated at $M = 2^r$ equally spaced points $t_i = a + i(b - a)/M$, $i = 0, 1, \dots, M - 1$, between a and b , with $r = 8$, $b = -a = 5$ and M is the number of discretization points. In each case, the grid of m values that we have considered consisted of 50 values from $0.5\hat{m}_0$ to $1.1\hat{m}_0$ where $\hat{m}_0 = n^{-\ell/5}$ denotes a pilot bandwidth (see e.g. Silverman (1986)).

In order to illustrate Theorem 7.2.3, we study the influence of the noise type on the numerical performances of the proposed estimator. For each density, samples of size $n = 1000$ were generated, with aim to estimate g_ℓ from X_1, \dots, X_n drawn from any of the test densities. The results are depicted in Figure 7.2 and Figure 7.3 for Laplace and Gaussian noise respectively. Figures 7.2-7.3 show the results of the numerical simulation of our adaptive estimator $\hat{g}_{\ell,m}$. Figure 7.4(a) contains a plot of the penalized criterion function versus the kernel bandwidth m and Figure 7.4(b) the estimated MISE as a function of m . We give this plot for kurtotic distribution and Laplace noise only, but the behaviour is the same in all the other cases. It is clear from Figure 7.4(a), that the value of \hat{m} is the unambiguous minimizer of $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_\ell(m)$. We also see that \hat{m} provides a result close to m_{MISE} : in the Laplace case, for the Kurtotic density, the bandwidth which minimizes $MISE(m)$ in this case is $m_{MISE} = 0.1306$ and $\hat{m} = 0.1257$. This holds true for all test densities. In practice, the minimum of the MISE is close to the minimum of the penalized

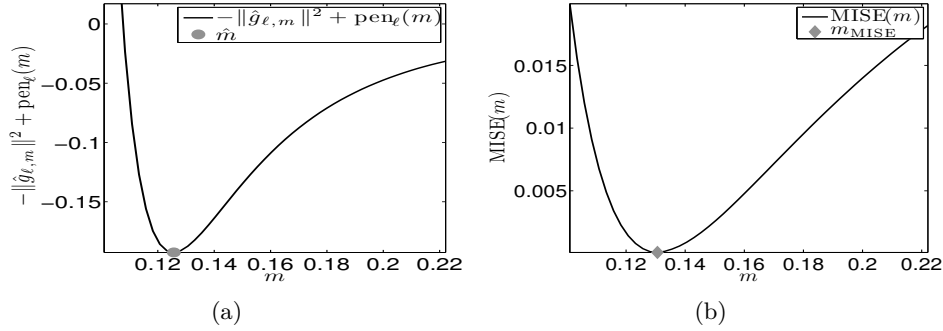


Figure 7.4: Laplace noise, Kurtotic density. (a): Graph of the function $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_{\ell}(m)$ against the smoothing parameter m . (b): $\text{MISE}(m)$. The gray diamond represents the global minimizer of $\text{MISE}(m)$ and the gray circle represents the global minimizer of $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_{\ell}(m)$.

	Noise			
	Gaussian		Laplace	
	known	unknown	known	unknown
Kurtotic	0.9200	0.4842	0.6678	0.4919
Gaussian	0.0285	0.0397	0.0350	0.0389
Uniform	1.4902	1.0301	1.1795	1.0300
Claw	0.1130	0.0494	0.0556	0.0485

Table 7.3: $1000 \times \text{MISE}$ values from 100 replications of sample sizes $n = 1000$

loss function, thus supporting the choice dictated by our theoretical procedure. Therefore, the proposed estimator provides effective results for the four test densities.

We also compare the performance of $\hat{g}_{\ell,\hat{m}}$ with that of $\tilde{g}_{\ell,\tilde{m}}$. For each density, samples with size $n = 1000$ were generated and the MISE was approximated as an average of the Integrated Squared Error over 100 replications. Table 7.3 presents the MISE from 100 repetitions for the two types of errors. As noted in Comte and Lacour (2011), estimating the characteristic function of the noise reduces the risk compared to knowing it. Indeed, for virtually all cases, $\tilde{g}_{\ell,\tilde{m}}$ consistently showed lower risk than $\hat{g}_{\ell,\hat{m}}$, with the exception of the Gaussian density for which $\hat{g}_{\ell,\hat{m}}$ performs slightly better.

7.5 Proofs

7.5.1 Proof of Proposition 7.2.1.

Upper bound on the MISE of $\hat{g}_{\ell,m}$.

Let us introduce the function $g_{\ell,m}$ defined by

$$g_{\ell,m}^*(u) = g_{\ell}^*(u) \mathbf{1}_{[-\pi m, \pi m]}(u) = (f^*(u))^{\ell} \mathbf{1}_{[-\pi m, \pi m]}(u) = \left(\frac{f_Y^*(u)}{f_{\varepsilon}^*(u)} \right)^{\ell} \mathbf{1}_{[-\pi m, \pi m]}(u).$$

The support of $(g_{\ell} - g_{\ell,m})^*$ being $[-\pi m, \pi m]^c$, disjoint from the support of $(g_{\ell,m} - \hat{g}_{\ell,m})^*$, we get

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_{\ell}\|^2) = \|g_{\ell} - g_{\ell,m}\|^2 + \mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2).$$

The Parseval identity yields

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

The same argument gives

$$\mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{\mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right)}{|f_\varepsilon^*(u)|^{2\ell}} du.$$

Using the binomial theorem and the inequality $|\sum_{i=1}^m b_i a_i|^2 \leq (\sum_{i=1}^m |b_i|) \sum_{i=1}^m |b_i| |a_i|^2$, we get

$$\begin{aligned} \mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) &= \mathbb{E}\left(|(\hat{f}_Y^*(u) - f_Y^*(u) + f_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) \\ &= \mathbb{E}\left(\left|\sum_{k=1}^{\ell} \binom{\ell}{k} (\hat{f}_Y^*(u) - f_Y^*(u))^k (f_Y^*(u))^{\ell-k}\right|^2\right) \\ &\leq (2^\ell - 1) \sum_{k=1}^{\ell} \binom{\ell}{k} \mathbb{E}\left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2k}\right) |f_Y^*(u)|^{2(\ell-k)}. \end{aligned}$$

The Marcinkiewicz & Zygmund Inequality applied to the n i.i.d. random variables U_1, \dots, U_n with $U_j = e^{-iuY_j} - \mathbb{E}(e^{-iuY_1})$, $|U_j| \leq 2$ and the exponent $p = 2k$ (see Appendix) yields

$$\mathbb{E}\left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2k}\right) \leq C_k \frac{1}{n^k},$$

with $C_k = (72k)^{2k} (2k/(2k-1))^k$. Therefore

$$\mathbb{E}\left(|(\hat{f}_Y^*(u))^\ell - (f_Y^*(u))^\ell|^2\right) \leq (2^\ell - 1) \sum_{k=1}^{\ell} \binom{\ell}{k} C_k \frac{1}{n^k} |f_Y^*(u)|^{2(\ell-k)}.$$

Since $f_Y^*(u) = f^*(u)f_\varepsilon^*(u)$, we obtain

$$\mathbb{E}(\|g_{\ell,m} - \hat{g}_{\ell,m}\|^2) \leq \frac{2^\ell - 1}{2\pi} \sum_{k=1}^{\ell} \binom{\ell}{k} C_k \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du.$$

Putting (7.16), (7.17) and (7.18) together, we get the desired result:

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \frac{1}{2\pi} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^{\ell} \binom{\ell}{k} C_k \frac{1}{n^k} \int_{-\pi m}^{\pi m} \frac{|f^*(u)|^{2(\ell-k)}}{|f_\varepsilon^*(u)|^{2k}} du. \quad \square$$

7.5.2 Proof of Corollary 7.2.2

Case $f \in \mathcal{A}(\beta, L)$ and f_ε is OS(a). The bias term is bounded by $L^\ell e^{-2\ell(\pi m)^\beta}$ since

$$\begin{aligned} \int_{|u| \geq \pi m} |f^*(u)|^{2\ell} du &= \int_{|u| \geq \pi m} |f^*(u)|^2 e^{2|u|^\beta} |f^*(u)|^{2(\ell-1)} e^{2(\ell-1)|u|^\beta} e^{-2\ell|u|^\beta} du \\ &\leq L^\ell e^{-2\ell(\pi m)^\beta}. \end{aligned}$$

All the terms containing powers of $|f^*|$ in the numerator are integrable as f_ε is ordinary smooth. Therefore they are of order less than $1/n$. Then we have the general bound

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq L^\ell e^{-2\ell(\pi m)^\beta} + \frac{2^\ell - 1}{2\pi} \frac{C_\ell}{n^\ell} \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^{2\ell}} + \frac{C}{n}$$

where $\int_{-\pi m}^{\pi m} du/|f_\varepsilon^*(u)|^{2\ell} = O(m^{2\ell a+1})$. Choosing $m = \pi^{-1}(\log(n)/(2\ell))^{1/\beta}$ gives the rate $1/n$ if $\ell \geq 2$ and $(\log n)^{(2a+1)/\beta}/n$ if $\ell = 1$.

Case $f \in \mathcal{S}(\alpha, L)$ and f_ε is SS(b). Applying Lemma 7.6.2, we obtain

$$\begin{aligned} \mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) &\leq L^\ell (\pi m)^{-2\ell\alpha} + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^{\ell} \binom{\ell}{k} C_k \frac{L^{\ell-k}}{c_\varepsilon n^k} \int_{-\pi m}^{\pi m} (1+u^2)^{-\alpha(\ell-k)} e^{2k|u|^b} du \\ &\leq L^\ell (\pi m)^{-2\ell\alpha} + \frac{2^\ell - 1}{2\pi} \sum_{k=1}^{\ell-1} \binom{\ell}{k} C_k \frac{L^{\ell-k}}{c_\varepsilon n^k} e^{2k(\pi m)^b} \int (1+u^2)^{-\alpha} du + C \frac{1}{c_\varepsilon n^\ell} m^{1-b} e^{2\ell(\pi m)^b}. \end{aligned}$$

The choice $m = \pi^{-1}(\log(n)/(4\ell))^{1/b}$ makes all variance terms of order at most $O(1/\sqrt{n})$, so that it implies a rate $(\log(n))^{-2\ell\alpha/b}$, which is logarithmic, but faster when ℓ increases.

Case $f \in \mathcal{S}(\alpha, L)$, f_ε is OS(a).

In this case, there exists a constant $K_\ell > 0$ such that

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq K_\ell \left(m^{-2\alpha\ell} + \sum_{k=1}^{\ell-1} \frac{m^{(2ak-2\alpha(\ell-k)+1)} \log(m)^{\mathbf{1}_{2ak-2\alpha(\ell-k)+1=0}}}{n^k} + \frac{m^{2\alpha\ell+1}}{n^\ell} \right).$$

Then α must be compared to the increasing sequence $a_\ell(k) = (2ak+1)/(\ell-k)$.

First case: $2\alpha < a_\ell(1)$. Then the trade-off between $m^{-2\alpha\ell}$ and $m^{2a-2\alpha(\ell-1)+1}/n$ implies the choice $m_{opt} = n^{1/(2a+2\alpha+1)}$ and the rate $n^{-2\alpha\ell/(2a+2\alpha+1)}$, where $2\alpha\ell/(2a+2\alpha+1) \in (0, 1)$.

In this situation, the terms following have orders

$$\sum_{k=2}^{\ell-1} \frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} + \frac{m_{opt}^{2\alpha\ell+1}}{n^\ell}$$

and we notice that

$$\frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} = n^{\frac{2ak-2\alpha(\ell-k)+1}{2a+2\alpha+1}-k} = n^{\frac{-2\alpha\ell-k+1}{2a+2\alpha+1}} \leq n^{\frac{-2\alpha\ell}{2a+2\alpha+1}}$$

since $k \geq 2$. Therefore the rate is of order $n^{-2\alpha\ell/(2a+2\alpha+1)}$. Notice that making compromise using another variance term can be checked to give larger choice of m and thus worse rate, when inserted in the first variance term. Note also that the resulting rate is the usual rate of deconvolution to the power ℓ .

Second case: $2\alpha > a_\ell(1)$. The term corresponding to $k=1$ is of order $O(1/n)$. Let $k_0 \geq 2$ be such that $a_\ell(k_0-1) < 2\alpha < a_\ell(k_0)$. Then the risk bound is

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq C \left(m^{-2\alpha\ell} + \frac{1}{n} + \dots + \frac{1}{n^{k_0-1}} + \sum_{k=k_0}^{\ell-1} \frac{m^{(2ak-2\alpha(\ell-k)+1)}}{n^k} + \frac{m^{2\alpha\ell+1}}{n^\ell} \right).$$

Then the compromise is made between $m^{-2\alpha\ell}$ and $m^{2ak_0-2\alpha(\ell-k_0)+1}/n^{k_0}$ and implies the choice $m_{opt} = n^{k_0/(2ak_0+2\alpha k_0+1)}$ and the rate $n^{-2\alpha\ell k_0/(2ak_0+2\alpha k_0+1)}$. But clearly this term is

smaller than n^{-1} . Moreover, we can check that plugging the above m_{opt} in the other terms let them less than $1/n$. Namely

$$\frac{m_{opt}^{2\alpha\ell+1}}{n^\ell} = n^{-\frac{2\alpha\ell k_0 + \ell - k_0}{2\alpha k_0 + 2\alpha k_0 + 1}} \leq n^{-1}, \quad \frac{m_{opt}^{2ak-2\alpha(\ell-k)+1}}{n^k} = n^{-\frac{2\alpha k_0 \ell + k - k_0}{2\alpha k_0 + 2\alpha k_0 + 1}} \leq n^{-1}, \quad \text{for } k = k_0 + j, j \geq 1.$$

Consequently, we have, if $2\alpha > a_\ell(1)$ that is

$$\text{if } 2\alpha > \frac{2a}{\ell-1} + \frac{1}{\ell-1}, \text{ then } \mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq Cn^{-1}.$$

If $2\alpha = a_\ell(k_0)$ for $k_0 \geq 2$, then the rate is still of order $1/n$ since the additional term $\log(n)/n^{k_0}$ does not change the rate, and if $2\alpha = a_\ell(1)$, then the rate is of order $\log(n)/n$. Gathering both cases, we get that, if $f \in \mathcal{S}(\alpha, L)$ and f_ε is OS(a), the rate is $n^{-2\alpha\ell/(2a+2\alpha+1)\vee n^{-1}}$, except if $2\alpha = a_\ell(1)$ where the rate is of order $\log(n)/n$. \square

7.5.3 Proof of Theorem 7.2.3.

The result of Proposition 7.2.1 can be written

$$\mathbb{E}(\|\hat{g}_{\ell,m} - g_\ell\|^2) \leq \|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m), \quad (7.19)$$

where $\mathcal{V}_\ell(m)$ is defined by (7.8). Let us define the oracle

$$m^* = \arg \min_{m \in \{1, \dots, n\}} \left\{ -\|g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}_\ell(m) \right\}$$

and note that, since $\|g_\ell - g_{\ell,m}\|^2 = \|g_\ell\|^2 - \|g_{\ell,m}\|^2$, we also have

$$m^* = \arg \min_{m \in \{1, \dots, n\}} \left\{ \|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}_\ell(m) \right\}.$$

Now, we start by writing that

$$\|g_\ell - \hat{g}_{\ell,\hat{m}}\|^2 \leq 2\|g_\ell - \hat{g}_{\ell,m^*}\|^2 + 2\|\hat{g}_{\ell,m^*} - \hat{g}_{\ell,\hat{m}}\|^2. \quad (7.20)$$

i) We consider the random set $\Omega = \{\hat{m} \leq m^*\}$, on which it holds that

$$\|\hat{g}_{\ell,m^*} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{1}_\Omega = (\|\hat{g}_{\ell,m^*}\|^2 - \|\hat{g}_{\ell,\hat{m}}\|^2) \mathbf{1}_\Omega. \quad (7.21)$$

Now, the definition of \hat{m} implies that

$$-\|\hat{g}_{\ell,\hat{m}}\|^2 + \text{pen}_\ell(\hat{m}) \leq -\|\hat{g}_{\ell,m^*}\|^2 + \text{pen}_\ell(m^*), \quad (7.22)$$

and thus

$$\begin{aligned} \|\hat{g}_{\ell,m^*}\|^2 - \|\hat{g}_{\ell,\hat{m}}\|^2 &\leq \text{pen}_\ell(m^*) - \text{pen}_\ell(\hat{m}) \\ &\leq \text{pen}_\ell(m^*). \end{aligned}$$

Then, (7.21) and (7.23) imply $\|\hat{g}_{\ell,m^*} - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{1}_\Omega \leq \text{pen}_\ell(m^*)$. Plugging this in (7.20) yields

$$\begin{aligned} \mathbb{E}(\|g_\ell - \hat{g}_{\ell,\hat{m}}\|^2 \mathbf{1}_\Omega) &\leq 2\mathbb{E}(\|g_\ell - \hat{g}_{\ell,m^*}\|^2) + 2\text{pen}_\ell(m^*) \\ &\leq 2(\|g_\ell - g_{\ell,m^*}\|^2 + \mathcal{V}_\ell(m^*)) + 2\text{pen}_\ell(m^*) \text{ with (7.19),} \\ &= 2 \min_{m \in \{1, \dots, n\}} \left\{ \|g_\ell - g_{\ell,m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}_\ell(m) \right\} \text{ by definition of } m^*. \end{aligned}$$

Thus we have

$$\mathbb{E} \left(\|g_\ell - \hat{g}_{\ell, \hat{m}}\|^2 \mathbf{1}_\Omega \right) \leq 2 \min_{m \in \{1, \dots, n\}} \left\{ \|g_\ell - g_{\ell, m}\|^2 + \mathcal{V}_\ell(m) + \text{pen}_\ell(m) \right\},$$

which ends step **i**).

ii) Now, we study the bound on Ω^c . Let us define, for $k > m$

$$\begin{aligned} \text{pen}_\ell(m, k) &= \vartheta 2^{2\ell-1} \frac{\Lambda_\ell^{2\ell}(m, k)}{n^\ell} (\Delta_\ell(k) - \Delta_\ell(m)), \\ \Lambda_\ell^{2\ell}(m, k) &= \max \left\{ 2^{2\ell-1} \left(\log(1 + (k-m)^2 (\Delta_\ell(k) - \Delta_\ell(m))) \right)^\ell, \right. \\ &\quad \left. \frac{2^{4\ell-1}}{n^\ell} \left(\log(1 + (k-m)^2 (\Delta_\ell(k) - \Delta_\ell(m))) \right)^{2\ell} \right\}. \end{aligned}$$

First we write

$$\begin{aligned} \|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, \hat{m}}\|^2 \mathbf{1}_{\Omega^c} &= \left(\|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, \hat{m}}\|^2 - 2^{2\ell-1} \|g_{\ell, \hat{m}} - g_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, \hat{m}) \right) \mathbf{1}_{\Omega^c} \\ &\quad + \left(2^{2\ell-1} \|g_{\ell, \hat{m}} - g_{\ell, m^*}\|^2 + \frac{1}{2} \text{pen}_\ell(m^*, \hat{m}) \right) \mathbf{1}_{\Omega^c} \\ &\leq \sup_{k > m^*} \left\{ \|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, k}\|^2 - 2^{2\ell-1} \|g_{\ell, k} - g_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \mathbf{1}_{\Omega^c} \\ &\quad + 2^{2\ell-1} \|g_\ell - g_{\ell, m^*}\|^2 + \frac{1}{2} \sum_{k > m^*} \text{pen}_\ell(m^*, k) \mathbf{1}_{k=\hat{m}}. \end{aligned}$$

Now, Inequality (7.22) can be written

$$\frac{1}{2} \text{pen}_\ell(\hat{m}) \leq \|\hat{g}_{\ell, \hat{m}}\|^2 - \|\hat{g}_{\ell, m^*}\|^2 + \text{pen}_\ell(m^*) - \frac{1}{2} \text{pen}_\ell(\hat{m})$$

and we notice that, for $k = \hat{m} > m^*$, then $\|\hat{g}_{\ell, \hat{m}}\|^2 - \|\hat{g}_{\ell, m^*}\|^2 = \|\hat{g}_{\ell, \hat{m}} - \hat{g}_{\ell, m^*}\|^2 = \|\hat{g}_{\ell, k} - \hat{g}_{\ell, m^*}\|^2$ and $\text{pen}_\ell(m^*, k) \leq \text{pen}_\ell(k)$, so that

$$\begin{aligned} \frac{1}{2} \text{pen}_\ell(m^*, k) &\leq \|\hat{g}_{\ell, k} - \hat{g}_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) + \text{pen}_\ell(m^*) + \frac{1}{2} \text{pen}_\ell(m^*, k) - \frac{1}{2} \text{pen}_\ell(k) \\ &\leq \left(\|\hat{g}_{\ell, k} - \hat{g}_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) - 2^{2\ell-1} \|g_{\ell, m^*} - g_{\ell, k}\|^2 \right) \\ &\quad + \text{pen}_\ell(m^*) + 2^{2\ell-1} \|g_{\ell, m^*} - g_{\ell, k}\|^2 \\ &\leq \sup_{k > m^*} \left(\|\hat{g}_{\ell, k} - \hat{g}_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) - 2^{2\ell-1} \|g_{\ell, m^*} - g_{\ell, k}\|^2 \right) \\ &\quad + \text{pen}_\ell(m^*) + 2^{2\ell-1} \|g_{\ell, m^*} - g_\ell\|^2. \end{aligned}$$

It follows that, on Ω^c ,

$$\begin{aligned} \frac{1}{2} \sum_{k > m^*} \text{pen}_\ell(m^*, k) \mathbf{1}_{\hat{m}=k} &\leq \sup_{k > m^*} \left(\|\hat{g}_{\ell, k} - \hat{g}_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) - 2^{2\ell-1} \|g_{\ell, m^*} - g_{\ell, k}\|^2 \right) \\ &\quad + \text{pen}_\ell(m^*) + 2^{2\ell-1} \|g_{\ell, m^*} - g_\ell\|^2. \end{aligned}$$

Inserting this in (7.24) implies

$$\begin{aligned} \|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, \hat{m}}\|^2 \mathbf{1}_{\Omega^c} &\leq 2 \sup_{k > m^*} \left\{ \|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, k}\|^2 - 2^{2\ell-1} \|g_{\ell, k} - g_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \mathbf{1}_{\Omega^c} \\ &\quad + 2^{2\ell} \|g_\ell - g_{\ell, m^*}\|^2 + \text{pen}_\ell(m^*). \end{aligned}$$

We can prove the following proposition, see Section 7.5.4.

Proposition 7.5.1. *Under the assumptions of Theorem 7.2.3, there exists a constant K_ℓ such that*

$$\mathbb{E} \left(\sup_{k > m^*} \left\{ \|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, k}\|^2 - 2^{2\ell-1} \|g_{\ell, k} - g_{\ell, m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \right) \leq \frac{K_\ell}{n^\ell}.$$

Then Inequality (7.25) and Proposition 7.5.1 imply

$$\begin{aligned} \mathbb{E}(\|\hat{g}_{\ell, m^*} - \hat{g}_{\ell, \hat{m}}\|^2 \mathbf{1}_{\Omega^c}) &\leq 2^{2\ell} \|g_\ell - g_{\ell, m^*}\|^2 + \text{pen}_\ell(m^*) + \frac{2K_\ell}{n^\ell} \\ &\leq 2^{2\ell} [\|g_\ell - g_{\ell, m^*}\|^2 + \mathcal{V}_\ell(m^*) + \text{pen}_\ell(m^*)] + \frac{2K_\ell}{n^\ell} \end{aligned}$$

which gives the result on Ω^c , by using again the definition of m^* and plugging this in (7.20). Gathering the two steps **i)** and **ii)** gives the result of Theorem 7.2.3. \square

The rates of convergence follow from those of Theorem 7.2.1 and Corollary 7.2.2.

7.5.4 Proof of Proposition 7.5.1

The proof of Proposition 7.5.1 relies on the following Lemma:

Lemma 7.5.2. *Let X_1, \dots, X_n be centered independent and identically distributed random variables, with $|X_i| \leq b$, $\forall i$ a.s. and $\text{Var}(X_i) \leq v^2$. Then, for $S_n = X_1 + \dots + X_n$, and $H > 0$,*

$$\mathbb{E} \left(\left| \frac{S_n}{n} \right|^p - H^p \right)_+ \leq \frac{2^p \Gamma(p/2 + 1)}{n^{p/2}} v^p \exp(-n2^{2/p-2} H^2/v^2) + \frac{2^{2p} \Gamma(p+1)}{n^p} b^p \exp(-n2^{1/p-2} H/b).$$

Proof of Lemma 7.5.2. The proof relies on Bernstein Inequality, which gives

$$\mathbb{P} \left(\left| \frac{S_n}{n} \right| \geq x \right) \leq 2 \exp \left(-\frac{nx^2/2}{v^2 + bx} \right) \leq 2 \left\{ \exp \left(-\frac{nx^2}{2v^2} \right) + \exp \left(-\frac{nx}{2b} \right) \right\}.$$

Thus

$$\begin{aligned} \mathbb{E} \left(\left| \frac{S_n}{n} \right|^p - H^p \right)_+ &\leq \int_0^{+\infty} \mathbb{P} \left(\left| \frac{S_n}{n} \right|^p - H^p \geq z \right) dz \\ &\leq \int_0^{+\infty} \mathbb{P} \left(\left| \frac{S_n}{n} \right| \geq [H^p + z]^{1/p} \right) dz \\ &\leq 2 \int_0^{+\infty} \left\{ \exp \left(-\frac{n[H^p + z]^{2/p}}{2v^2} \right) + \exp \left(-\frac{n[H^p + z]^{1/p}}{2b} \right) \right\} dz \end{aligned}$$

Now we use that by concavity inequality, we have that, for $q \geq 1$ and $x, y \geq 0$, $(x+y)^{1/q} \geq 2^{1/q-1}(x^{1/q} + y^{1/q})$, which implies for $p \geq 2$,

$$\begin{aligned} \mathbb{E} \left(\left| \frac{S_n}{n} \right|^p - H^p \right)_+ &\leq 2 \exp(-n2^{2/p-2} H^2/v^2) \int_0^{+\infty} \exp \left(-n \frac{2^{2/p-1} z^{2/p}}{2v^2} \right) dz \\ &\quad + 2 \exp(-n2^{1/p-2} H/b) \int_0^{+\infty} \exp \left(-n \frac{2^{1/p-1} z^{1/p}}{2b} \right) dz \\ &= \frac{2^p \Gamma(p/2 + 1)}{n^{p/2}} (v^2)^{p/2} \exp(-n2^{2/p-2} H^2/v^2) \\ &\quad + \frac{2^{2p} \Gamma(p+1)}{n^p} b^p \exp(-n2^{1/p-2} H/b), \end{aligned}$$

which is the announced result. \square

Proof of Proposition 7.5.1. For $k > m^*$, we have

$$\begin{aligned}
\|\hat{g}_{\ell,k} - \hat{g}_{\ell,m^*}\|^2 &= \frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \left| \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2\ell} du \\
&\leq \frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du \\
&\quad + \frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1} |f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du \\
&= \frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du + 2^{2\ell-1} \|g_{\ell,k} - g_{\ell,m^*}\|^2.
\end{aligned}$$

Therefore

$$\begin{aligned}
&\mathbb{E} \left(\sup_{k > m^*} \left\{ \|\hat{g}_{\ell,m^*} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1} \|g_{\ell,k} - g_{\ell,m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \right) \\
&\leq \mathbb{E} \left(\sup_{k > m^*} \left\{ \frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \right) \\
&\leq \sum_{k > m^*} \mathbb{E} \left(\frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2} \text{pen}_\ell(m^*, k) \right)_+ \\
&\leq \sum_{k > m^*} \mathbb{E} \left(\frac{1}{2\pi} \int_{\pi m^* \leq |u| \leq \pi k} \frac{2^{2\ell-1}}{|f_\varepsilon^*(u)|^{2\ell}} \left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell} - \frac{\Lambda_\ell^{2\ell}(m^*, k)}{n^\ell} \right)_+ du \right),
\end{aligned}$$

for $\vartheta \geq 2$. We apply now Lemma 7.5.2 with $b = v = 1$, $p = 2\ell$, and

$$\frac{\Lambda_\ell^{2\ell}(m, k)}{n^\ell} = H^{2\ell} \geq \frac{2^{2\ell-1}}{n^\ell} \left(\log(1 + (k - m^*)^2 (\Delta_\ell(k) - \Delta_\ell(m^*))) \right)^\ell$$

in the first exponential term of the bound, and

$$\frac{\Lambda_\ell^{2\ell}(m, k)}{n^\ell} = H^{2\ell} \geq \frac{2^{4\ell-1}}{n^{2\ell}} \left(\log(1 + (k - m^*)^2 (\Delta_\ell(k) - \Delta_\ell(m^*))) \right)^{2\ell}$$

in the second term. We obtain

$$\begin{aligned}
\mathbb{E} \left(|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell} - \frac{\Lambda_\ell^{2\ell}(m^*, k)}{n^\ell} \right)_+ &\leq \frac{2^{2\ell} \ell!}{n^\ell} \frac{1}{(k - m^*)^2 (\Delta_\ell(k) - \Delta_\ell(m^*))} \\
&\quad + \frac{2^{4\ell} (2\ell)!}{n^{2\ell}} \frac{1}{(k - m^*)^2 (\Delta_\ell(k) - \Delta_\ell(m^*))}.
\end{aligned}$$

Plugging this into (7.26) yields

$$\begin{aligned}
&\mathbb{E} \left(\sup_{k > m^*} \left\{ \|\hat{g}_{\ell,m^*} - \hat{g}_{\ell,k}\|^2 - 2^{2\ell-1} \|g_{\ell,k} - g_{\ell,m^*}\|^2 - \frac{1}{2} \text{pen}_\ell(m^*, k) \right\}_+ \right) \\
&\leq 2^{2\ell-1} \left(\frac{2^{2\ell} \ell!}{n^\ell} + \frac{2^{4\ell} (2\ell)!}{n^{2\ell}} \right) \sum_{k > m^*} \frac{1}{(k - m^*)^2} \leq \frac{\pi^2 2^{6\ell} (2\ell)!}{6 n^\ell}.
\end{aligned}$$

This ends the proof of Proposition 7.5.1. \square

7.5.5 Proof of the rates in Table 7.2

We consider here the compromise between the bias term and the terms depending on M . In the case $f \in \mathcal{A}(\beta, L)$, f_ε is OS(a), the terms are

$$L^\ell e^{-2\ell(\pi m)^\beta} + \frac{C}{M}$$

so that the choice $2\ell(\pi m_{opt}^{(2)})^\beta = \log(M)$ gives a term of order $1/M$.

In the case $f \in \mathcal{S}(\alpha, L)$, f_ε is SS(b), the terms are

$$L^\ell (\pi m)^{-2\alpha\ell} + C \sum_{k=1}^{\ell} \frac{e^{2k(\pi m)^b}}{M^k}$$

so that the choice $\pi m_{opt}^{(2)} = [(\log M)/(4\ell)]^{1/b}$ gives a term of order $(\log M)^{-2\alpha\ell/b}$.

In the case $f \in \mathcal{S}(\alpha, L)$, f_ε is OS(a), the terms are

$$C \left(m^{-2\alpha\ell} + \sum_{k=1}^{\ell} \frac{m^{(2ak-2\alpha\ell+1)+[\log(m)]^{1_{2ak-2\alpha\ell+1=0}}}}{M^k} \right).$$

The study follows a scheme analogous to the study of Corollary 7.2.2 with terms $a_\ell(k)$ replaced by $b_\ell(k) = (2ak + 1)/\ell$ which is also an increasing sequence in k . If $2\alpha < b_\ell(1)$, the compromise between the bias and the term for $k = 1$ leads to choose $m_{opt}^{(2)} = M^{1/(2a+1)}$ for a trade-off of order $M^{-2\alpha\ell/(2a+1)}$. All terms corresponding to $k \geq 2$ are then negligible with respect to this order. If $2\alpha = b_\ell(1)$, the rate is $\log M/M$. If $2\alpha > b_\ell(1)$, let $k_0 \geq 2$ be such that $b_\ell(k_0 - 1) < 2\alpha \leq b_\ell(k_0)$, the terms can be written

$$C \left(m^{-2\alpha\ell} + \frac{1}{M} + \cdots + \frac{1}{M^{k_0-1}} + \frac{m^{2ak_0+2\alpha\ell+1}[\log(m)]^{1_{2\alpha=b_\ell(k_0)}}}{M^{k_0}} + \sum_{k=k_0+1}^{\ell} \frac{m^{(2ak-2\alpha\ell+1)}}{M^k} \right).$$

Therefore, choosing $m_{opt}^{(2)} = M^{k_0/(2ak_0+1)}$ gives terms of index $k \geq k_0$ and a bias negligible w.r.t. M^{-1} . Therefore the rate is of order M^{-1} .

7.5.6 Proof of Theorem 7.3.4

Let us first introduce some notations. For $k > m$, set

$$\hat{\Delta}_\ell(m, k) = \hat{\Delta}_\ell(k) - \hat{\Delta}_\ell(m) \quad \text{and} \quad \hat{\Delta}_\ell^f(m, k) = \hat{\Delta}_\ell^f(k) - \hat{\Delta}_\ell^f(m),$$

$$\begin{aligned} \widehat{\text{qen}}(m, k) &:= 2^{4\ell} \hat{\lambda}_1^{2\ell}(m, k) \frac{\hat{\Delta}_\ell(m, k)}{n^\ell} + 2^{2\ell} \kappa^{2\ell} \log^\ell(M(k-m)) \frac{\hat{\Delta}_\ell^f(m, k)}{M^\ell} \\ &:= \widehat{\text{qen}}_1(m, k) + \widehat{\text{qen}}_2(m, k) \end{aligned}$$

with

$$\hat{\lambda}_1^{2\ell}(m, k) = \max \left\{ 2^{2\ell-1} \log^\ell \left(1 + \hat{\Delta}_\ell(m, k)(k-m)^2 \right), \frac{2^{4\ell-1}}{n^\ell} \log^{2\ell} \left(1 + \hat{\Delta}_\ell(m, k)(k-m)^2 \right) \right\}.$$

Now we can start the proof of Theorem 7.3.4. We denote by m^* the oracle cutoff defined by

$$m^* = \arg \min_{m \in \{1, \dots, n\}} \left\{ -\|g_{\ell, m}\|^2 + \mathbb{V}_\ell(m) + \text{qen}_\ell(m) \right\},$$

where $\mathbb{V}_\ell(m)$ is defined by (7.14). We have

$$\|g_\ell - \tilde{g}_{\ell, \tilde{m}}\|^2 \leq 2 \|g_\ell - \tilde{g}_{\ell, m^*}\|^2 + 2 \|\tilde{g}_{\ell, m^*} - \tilde{g}_{\ell, \tilde{m}}\|^2 \quad (7.27)$$

• Let us notice on the set $G = \{\tilde{m} \leq m^*\}$:

$$\|\tilde{g}_{\ell, m^*} - \tilde{g}_{\ell, \tilde{m}}\|^2 \mathbf{1}_G = \left(\|\tilde{g}_{\ell, m^*}\|^2 - \|\tilde{g}_{\ell, \tilde{m}}\|^2 \right) \mathbf{1}_G.$$

Besides according to the definition of \tilde{m} , one has the following inequality:

$$-\|\tilde{g}_{\ell, \tilde{m}}\|^2 + \widehat{\text{qen}}_\ell(\tilde{m}) \leq -\|\tilde{g}_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*), \quad (7.28)$$

which implies $-\|\tilde{g}_{\ell, \tilde{m}}\|^2 \leq -\|\tilde{g}_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*)$. Thus

$$\|\tilde{g}_{\ell, m^*} - \tilde{g}_{\ell, \tilde{m}}\|^2 \mathbf{1}_G = \left(\|\tilde{g}_{\ell, m^*}\|^2 - \|\tilde{g}_{\ell, \tilde{m}}\|^2 \right) \mathbf{1}_G \leq \widehat{\text{qen}}_\ell(m^*).$$

Taking expectation, we apply the following Lemma

Lemma 7.5.3. *There exists a positive constant C such that for any arbitrary $m \in \{1, \dots, n\}$*

$$\mathbb{E} [\widehat{\text{qen}}_\ell(m)] \leq C \text{qen}_\ell(m).$$

It yields for some positive constant C

$$\begin{aligned} \mathbb{E} \left[\|g_\ell - \tilde{g}_{\ell, \tilde{m}}\|^2 \mathbf{1}_G \right] &\leq 2\mathbb{E} \left[\|g_\ell - \tilde{g}_{\ell, m^*}\|^2 \right] + 2\mathbb{E} [\widehat{\text{qen}}_\ell(m^*)] \\ &\leq 2 \|g_\ell - g_{\ell, m^*}\|^2 + 2\mathbb{V}_\ell(m^*) + 2C \text{qen}_\ell(m^*). \end{aligned}$$

We just proved the result on G

$$\mathbb{E} \left[\|g_\ell - \tilde{g}_{\ell, \tilde{m}}\|^2 \mathbf{1}_G \right] \leq C \inf_{m \in \{1, \dots, n\}} \left\{ \|g_\ell - g_{\ell, m}\|^2 + \mathbb{V}_\ell(m) + \text{qen}_\ell(m) \right\}. \quad (7.29)$$

• We now consider the set $G^c = \{\tilde{m} > m^*\}$.

$$\begin{aligned} \|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 \mathbf{1}_{G^c} &= \left(\|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m^*, \tilde{m}) \right) \mathbf{1}_{G^c} \\ &\quad + \left(2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 + \frac{1}{2} \widehat{\text{qen}}(m^*, \tilde{m}) \right) \mathbf{1}_{G^c} \\ &\leq \sup_{\substack{k \geq m^* \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell, k} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, k} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m^*, k) \right\}_+ \\ &\quad + 2^{4\ell-1} \|g_\ell - g_{\ell, m^*}\|^2 + \frac{1}{2} \sum_{\substack{k \geq m^* \\ k \in \{1, \dots, n\}}} \widehat{\text{qen}}(m^*, k) \mathbf{1}_{\{\tilde{m} = k\}}. \end{aligned} \quad (7.30)$$

Let us first notice the following inequality

$$\forall k > m, \quad \widehat{\text{qen}}(m, k) \leq \widehat{\text{qen}}_\ell(k). \quad (7.31)$$

Besides by definition of \tilde{m} (see (7.15)), on the set $\{\tilde{m} = k\} \cap G^c$ and applying (7.28), we have

$$\frac{1}{2} (\widehat{\text{qen}}_\ell(k) - \widehat{\text{qen}}_\ell(m^*)) \leq \|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 - \frac{1}{2} (\widehat{\text{qen}}_\ell(k) - \widehat{\text{qen}}_\ell(m^*))$$

so that

$$\begin{aligned} \frac{1}{2} \widehat{\text{qen}}_\ell(k) &\leq \|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, k) + \widehat{\text{qen}}_\ell(m^*) \\ &\leq \left(\|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, k) \right) \\ &\quad + 2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*) \\ &\leq \left(\|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, k) \right) \\ &\quad + 2^{4\ell-1} \|g_\ell - g_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*). \end{aligned}$$

Now using Inequalities (7.31) and (7.32) on G^c gives

$$\begin{aligned} \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, \tilde{m}) &= \frac{1}{2} \sum_{\substack{k \geq m^* \\ k \in \{1, \dots, n\}}} \widehat{\text{qen}}_\ell(m^*, k) \mathbf{1}_{\{\tilde{m} = k\}} \\ &\leq \sup_{\substack{k \geq m^* \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell, k} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, \tilde{m}} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, k) \right\}_+ \\ &\quad + 2^{4\ell-1} \|g_\ell - g_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*). \end{aligned}$$

From Inequality (7.30), we now have

$$\begin{aligned} \|\tilde{g}_{\ell, \tilde{m}} - \tilde{g}_{\ell, m^*}\|^2 \mathbf{1}_{G^c} &\leq 2 \sup_{\substack{k \geq m^* \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell, k} - \tilde{g}_{\ell, m^*}\|^2 - 2^{4\ell-1} \|g_{\ell, k} - g_{\ell, m^*}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m^*, k) \right\}_+ \\ &\quad + 2^{4\ell} \|g_\ell - g_{\ell, m^*}\|^2 + \widehat{\text{qen}}_\ell(m^*). \end{aligned}$$

Taking expectation the first summand is negligible by applying the following Proposition.

Proposition 7.5.4. *Under (A1), there is a positive constant C such that for any arbitrary $m \in \{1, \dots, n\}$*

$$\mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell, k} - \tilde{g}_{\ell, m}\|^2 - 2^{4\ell-1} \|g_{\ell, k} - g_{\ell, m}\|^2 - \frac{1}{2} \widehat{\text{qen}}_\ell(m, k) \right\}_+ \right] \leq \frac{C}{n^\ell} + \frac{C}{M^\ell}.$$

Finally gathering the last result with (7.27) and using Lemma 7.5.3, we have

$$\mathbb{E} \left[\|g_\ell - \tilde{g}_{\ell, \tilde{m}}\|^2 \mathbf{1}_{G^c} \right] \leq C \left(\|g_\ell - g_{\ell, m^*}\|^2 + \mathbb{V}_\ell(m^*) + \text{qen}(m^*) \right) + C' \left(\frac{1}{n^\ell} + \frac{1}{M^\ell} \right).$$

This, combined with (7.29), completes the proof. \square

7.5.7 Proof of Lemma 7.5.3

Before proving Lemma 7.5.3, we first need to prove an auxiliary lemma. In the sequel, C will always denote some universal positive constant, but the value may vary from line to line.

Lemma 7.5.5. *For an estimator of f_ε^* defined by (7.10), assume $\kappa > \sqrt{c_1 p}$. Let $\tau \geq 2\kappa$ and $x \geq 1$. Then for some positive constant C*

$$\mathbb{P} \left[\exists u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > \tau (\log(Mx))^{1/2} w(u)^{-1} M^{-1/2} \right] \leq Cx^{-p} M^{-p}.$$

Proof of Lemma 7.5.5. We write the decomposition

$$|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| \leq |\tilde{f}_\varepsilon^*(u) - \hat{f}_\varepsilon^*(u)| + |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| \leq 2k_M(u) + |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|.$$

Using the previous inequality and applying Lemma 5.5 in Kappus (2014), we have

$$\begin{aligned} & \mathbb{P} \left[\exists u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > \tau (\log(Mx))^{1/2} w(u)^{-1} M^{-1/2} \right] \\ & \leq \mathbb{P} \left[\exists u \in \mathbb{R} : |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| + 2k_M(u) > \tau (\log(Mx))^{1/2} w(u)^{-1} M^{-1/2} \right] \\ & \leq \mathbb{P} \left[\exists u \in \mathbb{R} : |\hat{f}_\varepsilon^*(u) - f_\varepsilon^*(u)| > (\tau - 2\kappa) (\log(Mx))^{1/2} w(u)^{-1} M^{-1/2} \right] \\ & \leq Cx^{-p} M^{-p}. \quad \square \end{aligned}$$

Proof of Lemma 7.5.3. For $q = 1/2$ or 1 , using Cauchy-Schwarz's inequality, we have

$$\mathbb{E} \left[\log^q \left(1 + \hat{\Delta}_\ell(m) m^2 \right) \hat{\Delta}_\ell(m) \right] \leq \sqrt{\mathbb{E} \left[\log^{2q} \left(1 + \hat{\Delta}_\ell(m) m^2 \right) \right]} \mathbb{E} \left[\hat{\Delta}_\ell^2(m) \right]$$

Let $A_p(x) = \left| \frac{1}{\tilde{f}_\varepsilon^*(x)} - \frac{1}{f_\varepsilon^*(x)} \right|^{2p} / \left(\frac{1}{|f_\varepsilon^*(x)|^{2p}} \wedge \frac{k_M^{2p}(x)}{|f_\varepsilon^*(x)|^{4p}} \right)$. We have

$$\begin{aligned} \hat{\Delta}_\ell(m) &= \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} w(u)^{-2\ell} \left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} + \frac{1}{f_\varepsilon^*(u)} \right|^{2\ell} du \\ &\leq 2^{2\ell-1} \Delta_\ell^*(m) + \frac{2^{2\ell-2}}{\pi} \int_{-\pi m}^{\pi m} w(u)^{-2\ell} \left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right|^{2\ell} du \\ &\leq 2^{2\ell-1} \Delta_\ell^*(m) + 2^{2\ell-1} \Delta_\ell^*(m) \sup_{u \in \mathbb{R}} A_\ell(u) \leq 2^{2\ell-1} \Delta_\ell^*(m) (1 + \sup_{u \in \mathbb{R}} A_\ell(u)) \end{aligned}$$

and applying Lemma 5.6 and Corollary 5.7 in Kappus (2014) stating that the expectation of $\sup_{u \in \mathbb{R}} A_\ell(u)$ is bounded, we get $\mathbb{E} \left[\hat{\Delta}_\ell^2(m) \right] \leq C(\Delta_\ell^*)^2(m)$. By Jensen inequality (since log is concave)

$$\begin{aligned} \mathbb{E} \left[\log^{2q} \left(1 + \hat{\Delta}_\ell(m) m^2 \right) \right] &\leq \log^{2q} \left(\mathbb{E} \left[1 + \hat{\Delta}_\ell(m) m^2 \right] \right) \leq \log^{2q} \left(1 + \mathbb{E} \left[\hat{\Delta}_\ell(m) \right] m^2 \right) \\ &\leq \log^{2q} (1 + \sqrt{C} \Delta_\ell^*(m) m^2) \leq C' \log^{2q} \left(1 + \Delta_\ell^*(m) m^2 \right). \end{aligned}$$

So

$$\mathbb{E} \left[\log^q \left(1 + \hat{\Delta}_\ell(m) m^2 \right) \hat{\Delta}_\ell(m) \right] \leq C \log^q \left(1 + \Delta_\ell^*(m) m^2 \right) \Delta_\ell^*(m)$$

which means $\mathbb{E}[\widehat{\text{qen}}_1(m)] \leq C \text{qen}_1(m)$. Consider now $\widehat{\text{qen}}_2(m)$. As we mentioned above that it follows from Kappus (2014) that $\mathbb{E}(\sup_{u \in \mathbb{R}} A_\ell(u))$ is bounded by a constant, this

yields

$$\begin{aligned}
& \frac{1}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] \\
& \leq \frac{2^{2\ell-1}}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] + \frac{2^{2\ell-1}}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] \\
& \leq \frac{2^{4\ell-2}}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] \mathbb{E} \left[1 + \sup_{u \in \mathbb{R}} A_\ell(u) \right] \\
& \quad + \frac{2^{2\ell-1}}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] \\
& \leq \frac{2^{4\ell-2}}{M^\ell} \Delta_\ell^f(m) \mathbb{E} \left[1 + \sup_{u \in \mathbb{R}} A_\ell(u) \right] + \frac{2^{2\ell-1}}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right].
\end{aligned}$$

We use the fact that $\tilde{f}_\varepsilon^*(u) \geq M^{-1/2}(\log M)^{1/2} w(u)^{-1}$, $w(u) \leq 1$ as well as the independence of \hat{f}_Y^* and \tilde{f}_ε^* to find

$$\begin{aligned}
& \frac{1}{M^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} du \right] \\
& = \frac{1}{M^\ell} \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \mathbb{E} \left[|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell} \right] \mathbb{E} \left[\frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \right] du \leq \frac{c_\ell}{n^\ell} \mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du \right],
\end{aligned}$$

where c_ℓ is the constant of the [Rosenthal \(1970\)](#) inequality . Applying the same arguments as for the bounding of $\mathbb{E} [\widehat{\text{qen}}_1(m)]$, we get $\mathbb{E} \left[\frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{w(u)^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du \right] \leq C \Delta_\ell^*(m)$. This completes the proof. \square

7.5.8 Proof of Proposition 7.5.4

For $k > m$, let us introduce the following notation

$$A(m, k) := \{u \in \mathbb{R}, |u| \in [\pi m, \pi k]\}.$$

We need to prove auxiliary lemmas before proving Proposition 7.5.4.

Lemma 7.5.6. *There is a positive constant C such that for any arbitrary $m \in \{1, \dots, n\}$*

$$\mathbb{E} \left[\sup_{k \geq m} \left\{ \frac{1}{2\pi} \int_{A(m, k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{4\ell}} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \leq \frac{C}{n^\ell}.$$

Proof of Lemma 7.5.6.

$$\begin{aligned}
& \mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \{1, \dots, n\}}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{4\ell} \widehat{\text{qen}}_1(m, k)} \right\}_+ \right] \\
& \leq \mathbb{E} \left[\sum_{\substack{k \geq m \\ k \in \{1, \dots, n\}}} \mathbb{E} \left[\left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{4\ell} \widehat{\text{qen}}_1(m, k)} \right\}_+ \middle| \tilde{f}_\varepsilon^* \right] \right] \\
& \leq \mathbb{E} \left[\sum_{\substack{k \geq m \\ k \in \{1, \dots, n\}}} \frac{1}{2\pi} \int_{A(m,k)} \mathbb{E} \left[\left\{ \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{\hat{\lambda}_1^{2\ell}(m, k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}} \right\}_+ \middle| \tilde{f}_\varepsilon^* \right] du \right].
\end{aligned}$$

Now $\hat{f}_Y^*(u)/\tilde{f}_\varepsilon^*(u)$ (conditional on $\tilde{f}_\varepsilon^*(u)$) is the sum of n i.i.d. random variables with variance $v^2 \leq 1/|\tilde{f}_\varepsilon^*(u)|^2$ which are bounded by $2/\tilde{f}_\varepsilon^*(u)$. Thus Lemma 7.5.2 gives

$$\begin{aligned}
& \mathbb{E} \left[\left\{ \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{\hat{\lambda}_1^{2\ell}(m, k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}} \right\}_+ \middle| \tilde{f}_\varepsilon^* \right] \\
& \leq 2^{2\ell} \Gamma(\ell + 1) \frac{n^{-\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \exp\left(-2^{1/\ell-2} \hat{\lambda}_1^2(m, k)\right) + 2^{4\ell} \Gamma(2\ell + 1) \frac{n^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \exp\left(-n^{1/2} 2^{1/(2\ell)-2} \hat{\lambda}_1(m, k)\right) \\
& \leq 2^{2\ell} \Gamma(\ell + 1) \frac{n^{-\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} (k - m)^{-2} \hat{\Delta}_\ell(m, k)^{-1} + 2^{4\ell} \Gamma(2\ell + 1) \frac{n^{-2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} (k - m)^{-2} \hat{\Delta}_\ell(m, k)^{-1}
\end{aligned}$$

where we used the fact that

$$\hat{\lambda}_1(m, k) \leq \max \left\{ \sqrt{2^{-1/\ell+2} \log\left(1 + \hat{\Delta}_\ell(m, k)(k - m)^2\right)}, \frac{2^{-1/2\ell+2}}{\sqrt{n}} \log\left(1 + \hat{\Delta}_\ell(m, k)(k - m)^2\right) \right\}.$$

We have thus shown for a universal positive constant C that, for any $m, k \in \{1, \dots, n\}$, $k > m$,

$$\begin{aligned}
& \int_{A(m,k)} \mathbb{E} \left[\left\{ \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} - \frac{\hat{\lambda}_1^{2\ell}(m, k)}{n^\ell |\tilde{f}_\varepsilon^*(u)|^{2\ell}} \right\}_+ \middle| \tilde{f}_\varepsilon^* \right] du \\
& \leq \frac{C}{n^\ell} (k - m)^{-2} \hat{\Delta}_\ell(m, k)^{-1} \int_{A(m,k)} \frac{du}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \leq \frac{C}{n^\ell} (k - m)^{-2}.
\end{aligned}$$

Finally $\mathbb{E} \left[\sup_{k > m} \left\{ (2\pi)^{-1} \int_{A(m,k)} (|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}) / |\tilde{f}_\varepsilon^*(u)|^{2\ell} du - 2^{-4\ell} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \leq C/n^\ell$, which ends the proof. \square

Proof of Proposition 7.5.4. Applying Plancherel's formula we get

$$\begin{aligned}
\|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 &= \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du \\
&= \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| > |f_\varepsilon^*(u)|\}} du + \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\
&:= S_1 + S_2.
\end{aligned}$$

Let us consider S_1 .

$$S_1 \leq \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + \underbrace{\frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} \frac{|f_Y^*(u)|^{2\ell}}{|f_\varepsilon^*(u)|^{2\ell}} du}_{2^{2\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2}$$

now S_2 ,

$$\begin{aligned} S_2 &= \frac{1}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} + \frac{1}{f_\varepsilon^*(u)} \right|^{2\ell} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\ &\leq \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \left| \frac{1}{\tilde{f}_\varepsilon^*(u)} - \frac{1}{f_\varepsilon^*(u)} \right|^{2\ell} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\ &\quad + \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} \left| \frac{\hat{f}_Y^*(u)}{f_\varepsilon^*(u)} \right|^{2\ell} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du := S_{2,1} + S_{2,2}. \end{aligned}$$

Yet for the second term of (7.34), we can notice

$$S_{2,2} \leq \frac{2^{4\ell-2}}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + 2^{4\ell-2} \|g_{\ell,k} - g_{\ell,m}\|^2. \quad (7.35)$$

For the first term of (7.34) we can write

$$\begin{aligned} S_{2,1} &= \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell} |f_\varepsilon^*(u)|^{2\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\ &\leq \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du. \end{aligned} \quad (7.36)$$

Let us introduce the set following set

$$C(m, k) = \left\{ \forall u \in \mathbb{R} : |\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^2 \leq \kappa^2 \log(M(k-m)) w(u)^{-2} M^{-1} \right\} \quad (7.37)$$

On $C(m, k)$, the following inequalities can be deduced

$$\begin{aligned} &\frac{1}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\ &\leq \kappa^{2\ell} \log^\ell(M(k-m)) M^{-\ell} \frac{1}{2\pi} \int_{A(m,k)} \frac{w(u)^{-2\ell} |\hat{f}_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \\ &\leq \kappa^{2\ell} \log^\ell(M(k-m)) M^{-\ell} \hat{\Delta}_\ell^f(m, k) \leq \frac{1}{2^{2\ell}} \widehat{\text{qen}}_2(m, k). \end{aligned}$$

Gathering Equations (7.33), (7.35) and (7.36), we have

$$\begin{aligned} \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 &\leq \frac{2^{4\ell-2}}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 \\ &\quad + \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du [\mathbf{1}_{C(m,k)} + \mathbf{1}_{C(m,k)^c}] \\ &\leq \frac{2^{4\ell-2}}{\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du + 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 + \frac{1}{2} \widehat{\text{qen}}_2(m, k) \\ &\quad + \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c}. \end{aligned} \quad (7.38)$$

Starting from (7.38), we can now write the following inequalities

$$\begin{aligned}
& \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \leq \frac{2^{4\ell-2}}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du \\
& - \frac{1}{2} \widehat{\text{qen}}_1(m, k) + \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c} \\
& \leq 2^{4\ell-1} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{4\ell}} \widehat{\text{qen}}_1(m, k) \right\}_+ \\
& + \frac{2^{2\ell-1}}{2\pi} \int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c}.
\end{aligned}$$

Taking expectation we get

$$\begin{aligned}
& \mathbb{E} \left[\sup_{\substack{k \geq m \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right] \\
& \leq \mathbb{E} \left[2^{4\ell-1} \sup_{\substack{k > m \\ k \in \{1, \dots, n\}}} \left\{ \frac{1}{2\pi} \int_{A(m,k)} \frac{|\hat{f}_Y^*(u) - f_Y^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{2\ell}} du - \frac{1}{2^{4\ell}} \widehat{\text{qen}}_1(m, k) \right\}_+ \right] \\
& + \frac{2^{2\ell-1}}{2\pi} \sum_{\substack{k > m \\ k \in \{1, \dots, n\}}} \mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c} \right].
\end{aligned}$$

The first term is bounded by applying Lemma 7.5.6. For the second, we can notice that on $C(m, k)^c$ defined by (7.37) following Lemma 7.5.5 for $x = k - m$ and $p = 3\ell$, we have $\mathbb{P}[C(m, k)^c] \leq M^{-3\ell} (k - m)^{-3\ell}$ and we get

$$\begin{aligned}
& \sum_{\substack{k > m \\ k \in \{1, \dots, n\}}} \mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{|\tilde{f}_\varepsilon^*(u) - f_\varepsilon^*(u)|^{2\ell}}{|\tilde{f}_\varepsilon^*(u)|^{4\ell}} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c} \right] \\
& \leq \sum_{\substack{k > m \\ k \in \{1, \dots, n\}}} \mathbb{E} \left[\int_{A(m,k)} |\hat{f}_Y^*(u)|^{2\ell} \frac{2^{2\ell} |f_\varepsilon^*(u)|^{2\ell}}{k_M^{4\ell}(u)} \mathbf{1}_{\{|\tilde{f}_\varepsilon^*(u)| \leq |f_\varepsilon^*(u)|\}} du \mathbf{1}_{C(m,k)^c} \right] \\
& \leq \sum_{\substack{k > m \\ k \in \{1, \dots, n\}}} \mathbb{E} \left[\int_{A(m,k)} 2^{2\ell} \kappa^{-4\ell} (\log M)^{-2\ell} w(u)^{4\ell} M^{2\ell} du \mathbf{1}_{C(m,k)^c} \right] \\
& \leq \sum_{\substack{k > m \\ k \in \{1, \dots, n\}}} 2^{2\ell} \kappa^{-4\ell} (\log M)^{-2\ell} M^{2\ell} (k - m) \mathbb{P}[C(m, k)^c] \leq \frac{C}{M^\ell},
\end{aligned}$$

and applying Lemma 7.5.6, we have

$$\mathbb{E} \left[\sup_{\substack{k > m \\ k \in \{1, \dots, n\}}} \left\{ \|\tilde{g}_{\ell,k} - \tilde{g}_{\ell,m}\|^2 - 2^{4\ell-1} \|g_{\ell,k} - g_{\ell,m}\|^2 - \frac{1}{2} \widehat{\text{qen}}(m, k) \right\}_+ \right] \leq \frac{C}{n^\ell} + \frac{C}{M^\ell}.$$

This completes the proof. \square

7.6 Appendix

We recall the following result, which can be found in [Shiryaev and Wilson \(1995\)](#) (inequality (26) p.498).

Lemma 7.6.1 (Marcinkiewicz & Zygmund Inequality). *Let n be a positive integer, $p > 1$ and U_1, \dots, U_n be n zero mean independent random variables such that $\sup_{j \in \{1, \dots, n\}} \mathbb{E}(|U_j|^p) < \infty$. Then, for $B_p = (18p)^p(p/(p-1))^{p/2}$,*

$$\mathbb{E}(|\sum_{j=1}^n U_j|^p) \leq B_p \mathbb{E}(|\sum_{j=1}^n |U_j|^2|^{p/2}),$$

We also give the following useful lemma, see Lemma 2 p. 35 in [Butucea and Tsybakov \(2008b\)](#). For two functions u, v , we denote $u(x) \lesssim v(x)$ if there exists a positive constant C not depending on x such that $u(x) \leq Cv(x)$ and $u(x) \approx v(x)$ if $u(x) \lesssim v(x)$ and $v(x) \lesssim u(x)$.

Lemma 7.6.2. ([Butucea and Tsybakov \(2008b\)](#)) *Consider c, s nonnegative real numbers, and γ a real such that $2\gamma > -1$ if $c = 0$ or $s = 0$. Then, for any $m > 0$,*

- $\int_{-m}^m (x^2 + 1)^\gamma \exp(c|x|^s) dx \approx m^{2\gamma+1-s} e^{cm^s},$

and if in addition $2\gamma > 1$ if $c = 0$ or $s = 0$,

- $\int_m^\infty (x^2 + 1)^{-\gamma} \exp(-c|x|^s) dx \approx m^{-2\gamma+1-s} e^{-cm^s}.$

Bibliographie

- Abramovich, F., Pensky, M., and Rozenholc, Y. (2013). Laplace deconvolution with noisy observations. *Electron. J. Stat.*, 7 :1094–1128.
- Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York, ninth dover printing, tenth gpo printing edition.
- Ahmad, I. and Fan, Y. (2001). Optimal bandwidth for kernel density estimators of functions of observations. *Statist. Probab. Lett.*, 51 :245–251.
- Ahmad, I. and Mugdadi, A. (2003). Analysis of kernel density estimation of functions of random variables. *J. Nonparametric Stat.*, 15 :579–605.
- Barron, A., Birgé, L., and Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields*, 113(3) :301–413.
- Baudry, J.-P., Maugis, C., and Michel, B. (2012). Slope heuristics : overview and implementation. *Statistics and Computing*, 22(2) :455–470.
- Belomestny, D., Comte, F., and Genon-Catalot, V. (2016). Laguerre estimation for k -monotone densities observed with noise. preprint hal-01252143.
- Birgé, L. (1999). An alternative point of view on Lepski’s method. In *State of the art in probability and statistics : Festschrift for Willem R. van Zwet*, pages 113–133. IMS Lecture Notes-Monograph Series.
- Birgé, L. and Massart, P. (1997). From model selection to adaptive estimation. In *Festschrift for Lucien Le Cam*, pages 55–87. Springer.
- Birgé, L. and Massart, P. (2007). Minimal penalties for gaussian model selection. *Probab. Theory Relat. Fields*, 138(1-2) :33–73.
- Bongioanni, B. and Torrea, J. L. (2009). What is a Sobolev space for the Laguerre function systems? *Studia Math.*, 192(2) :147–172.
- Bonhomme, S. and Robin, J.-M. (2010). Generalized non-parametric deconvolution with an application to earnings dynamics. *Review of Economic Studies*, 77(2) :491–533.
- Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Can. J. Stat.*, 32(2) :181–192.
- Butucea, C. and Comte, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli*, 15(1) :69–98.
- Butucea, C. and Matias, C. (2005). Minimax estimation of the noise level and of the deconvolution density in a semiparametric convolution model. *Bernoulli*, 11(2) :309–340.

- Butucea, C., Matias, C., and Pouet, C. (2008). Adaptivity in convolution models with partially known noise distribution. *Electron. J. Stat.*, 2 :897–915.
- Butucea, C. and Tsybakov, A. (2008a). Sharp optimality in density deconvolution with dominating bias I. *Theory Proba. Appl.*, 52(1) :24–39.
- Butucea, C. and Tsybakov, A. (2008b). Sharp optimality in density deconvolution with dominating bias II. *Theory Proba. Appl.*, 52(2) :237–249.
- Cai, T. T. and Low, M. G. (2003). A note on nonparametric estimation of linear functionals. *Ann. Statist.*, 31(4) :1140–1153.
- Cai, T. T. and Low, M. G. (2005). Adaptive estimation of linear functionals under different performance measures. *Bernoulli*, 11(2) :341–358.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Am. Stat. Assoc.*, 83(404) :1184–1186.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2009). Improving point and interval estimators of monotone functions by rearrangement. *Biometrika*, 96(3) :559–575.
- Chesneau, C., Comte, F., and Navarro, F. (2013). Fast nonparametric estimation for convolutions of densities. *Can. J. Stat.*, 41 :617–636.
- Chesneau, C. and Navarro, F. (2014). On a plug-in wavelet estimator for convolutions of densities. *J. Stat. Theory Pract.*, 8 :653–673.
- Claeskens, G. and Hart, J. D. (2009). Goodness-of-fit tests in mixed models. *TEST*, 18(2) :213–239.
- Comte, F. (2004). Kernel deconvolution of stochastic volatility models. *J. Time Ser. Anal.*, 25(4) :563–582.
- Comte, F. (2015). *Estimation non-paramétrique*. Spartacus.
- Comte, F., Cuenod, C.-A., Pensky, M., and Rozenholc, Y. (2015a). Laplace deconvolution and its application to dynamic contrast enhanced imaging. *J. R. Stat. Soc., Ser. B*.
- Comte, F., Dedecker, J., and Taupin, M. (2008). Adaptive density deconvolution with dependent inputs. *Math. Methods Stat.*, 17(2) :87–112.
- Comte, F. and Genon-Catalot, V. (2015). Adaptive Laguerre density estimation for mixed Poisson models. *Electron. J. Stat.*, 9 :1112–1148.
- Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *J. Multivariate Anal.*, 140 :31–46.
- Comte, F. and Lacour, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *J. R. Stat. Soc., Ser. B*, 73 :601–627.
- Comte, F. and Lacour, C. (2013). Anisotropic adaptive kernel deconvolution. *Ann. Inst. H. Poincaré Probab. Statist.*, 49(2) :569–609.
- Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Can. J. Stat.*, 3(34) :431–452.

- Comte, F., Rozenholc, Y., and Taupin, M.-L. (2007). Finite sample penalization in adaptive density deconvolution. *J. Stat. Comput. Simulation*, 77(11) :977–1000.
- Comte, F. and Samson, A. (2012). Nonparametric estimation of random-effects densities in linear mixed-effects model. *J. Nonparametric Stat.*, 24(4) :951–975.
- Comte, F., Samson, A., and Stirnemann, J. (2013). Deconvolution estimation of onset of pregnancy with replicate observations. *Scand. J. Stat.*, 41(2) :325–345.
- Comte, F., Samson, A., and Stirnemann, J. J. (2015b). Hazard estimation for censored data contaminated with additive measurement error : Application to length of pregnancy. *MAP5 2015-18*.
- Dattner, I., Goldenshluger, A., and Juditsky, A. (2011). On deconvolution of distribution functions. *Ann. Statist.*, 39(5) :2477–2501.
- Dattner, I. and Reiser, B. (2013). Estimation of distribution functions in measurement error models. *J. Stat. Plann. Inference*, 143(3) :479 – 493.
- Dattner, I., Reiß, M., and Trabs, M. (2016). Adaptive quantile estimation in deconvolution with unknown error distribution. *Bernoulli*, 22(1) :143–192.
- De Canditiis, D. and Pensky, M. (2016). Estimation of delta-contaminated density of the random intensity of poisson data. *arXiv preprint arXiv :1509.00500, to appear in Electronic Journal of Statistics*.
- Delaigle, A. and Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Stat. Math.*, 56(1) :19–47.
- Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Stat. Sin.*, 16 :773 –788.
- Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.*, 36(2) :665–685.
- Devroye, L. (1989). Consistent deconvolution in density estimation. *Can. J. Stat.*, 17 :235–239.
- Diggle, P. J. and Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *J. R. Stat. Soc., Ser. B*, 55(2) :523–531.
- Dion, C. (2014). New adaptive strategies for nonparametric estimation in linear mixed models. *J. Stat. Plann. Inference*, 150 :30 – 48.
- Du, J. and Schick, A. (2007). Root-n consistency and functional central limit theorems. *Internat. J. Statist. Management Systems*, 2 :67–87.
- Duval, C. (2015). A note on a fixed point method for deconvolution. preprint MAP5 2015-27.
- Efromovich, S. (1997). Density estimation for the case of supersmooth measurement errors. *J. Am. Stat. Assoc.*, 92 :526–535.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.*, 19(3) :1257–1272.

- Fan, J. (1993). Adaptively local one-dimensional subproblems with application to a deconvolution problem. *Ann. Statist.*, 21(2) :600–610.
- Fan, J. and Koo, J. (2002). Wavelet deconvolution. *IEEE transactions on information theory*, 48 :734–747.
- Frees, E. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, 89 :17–525.
- Ghidey, W., Lesaffre, E., and Eilers, P. (2004). Smooth random effects distribution in a linear mixed model. *Biometrics*, 60(4) :945–953.
- Giné, E. and Mason, D. (2007). On local U-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, 35 :1105–1145.
- Goldenshluger, A. (1999). On pointwise adaptive nonparametric deconvolution. *Bernoulli*, 5(5) :907–925.
- Goldenshluger, A. and Lepski, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli*, 14 :1150–1190.
- Goldenshluger, A. and Lepski, O. (2009). Structural adaptation via lp-norm oracle inequalities. *Probab. Theory Relat. Fields*, 143 :41–71.
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation : Oracle inequalities and adaptive minimax optimality. *Ann. Statist.*, 39(3) :1608–1632.
- Goldenshluger, A. and Lepski, O. (2013). General Selection Rule from a Family of Linear Estimators. *Theory of Probability & Its Applications*, 57(2) :209–226.
- Goldenshluger, A. and Lepski, O. (2014). On adaptive minimax density estimation on a \mathbb{R}^d . *Probab. Theory Relat. Fields*, 159(3-4) :479–543.
- Groeneboom, P. and Jongbloed, G. (2003). Density estimation in the uniform deconvolution model. *Stat. Neerl.*, 57(1) :136–157.
- Groeneboom, P. and Wellner, J. A. (1992). *Information bounds and nonparametric maximum likelihood estimation*, volume 19 of *DMV Seminar*. Birkhäuser Verlag, Basel.
- Hall, P. and Lahiri, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.*, 36(5) :2110–2134.
- Hebiri, M., Loubes, J.-M., and Rochet, P. (2014). Aggregation for linear inverse problems. preprint hal-00994717.
- Horn, R. and Johnson, C. (1990). *Matrix Analysis*. Cambridge University Press.
- Hu, Y. and Ridder, G. (2004). Estimation of Nonlinear Models with Measurement Error Using Marginal Information. Econometric Society 2004 North American Summer Meetings 21, Econometric Society.
- Ibrahim, J. G. and Kleinman, K. P. (1998). Semiparametric Bayesian methods for random effects models. In *Practical nonparametric and semiparametric Bayesian statistics*, volume 133 of *Lecture Notes in Statist.*, pages 89–114. Springer, New York.
- Jirak, M., Meister, A., and Reiß, M. (2014). Adaptive function estimation in nonparametric regression with one-sided errors. *Ann. Statist.*, 42(5) :1970–2002.

- Johannes, J. (2009). Deconvolution with unknown error distribution. *Ann. Statist.*, 37(5a) :2301–2323.
- Johannes, J. and Schwarz, M. (2013a). Adaptive circular deconvolution by model selection under unknown error distribution. *Bernoulli*, 19(5A) :1576–1611.
- Johannes, J. and Schwarz, M. (2013b). Adaptive gaussian inverse regression with partially unknown operator. *Communications in Statistics - Theory and Methods*, 42 :1343–1362.
- Jongbloed, G. (1998). Exponential deconvolution : two asymptotically equivalent estimators. *Stat. Neerl.*, 52(1) :6–17.
- Kappus, J. (2014). Adaptive nonparametric estimation for Lévy processes observed at low frequency. *Stochastic Processes Appl.*, 124 :730–758.
- Kappus, J. and Mabon, G. (2014). Adaptive density estimation in deconvolution problems with unknown error distribution. *Electron. J. Stat.*, 8(2) :2879–2904.
- Kerkycharian, G., Lepski, O., and Picard, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Relat. Fields*, 121(2) :137–170.
- Klein, T. and Rio, E. (2005). Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3) :1060–1077.
- Komárek, A. and Lesaffre, E. (2008). Generalized linear mixed model with a penalized Gaussian mixture as a random effects distribution. *Comput. Statist. Data Anal.*, 52(7) :3441–3458.
- Lacour, C. (2006). Rates of convergence for nonparametric deconvolution. *C. R. Acad. Sci. Paris*, Ser. I 342(11) :877–882.
- Lacour, C. and Massart, P. (2015). Minimal penalty for Goldenshluger-Lepski method. *ArXiv e-prints 1503.00946*.
- Laurent, B., Ludeña, C., and Prieur, C. (2008). Adaptive estimation of linear functionals by model selection. *Electron. J. Stat.*, 2 :993–1020.
- Lepski, O. and Willer, T. (2015). Lower bounds in the convolution structure density model. *To appear in Bernoulli*.
- Lepskii, O. V. (1991). On a Problem of Adaptive Estimation in Gaussian White Noise. *Theory of Probability & Its Applications*, 35(3) :454–466.
- Li, T. and Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *J. Multivariate Anal.*, 65 :139–165.
- Lounici, K. (2007). Generalized mirror averaging and D-convex aggregation. *Math. Methods Stat.*, 16(3) :246–259.
- Mabon, G. (2014). Adaptive deconvolution on the nonnegative real line. preprint MAP5 2014-33.
- Mabon, G. (2016). Adaptive deconvolution of linear functionals on the nonnegative real line. *accepted for publication in J. Stat. Plann. Inference*, preprint MAP5 2015-24.
- Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *Ann. Statist.*, 20(2) :712–736.

- Masry, E. (1993). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes Appl.*, 47(1) :53–74.
- Masry, E. (2002). Multivariate probability density estimation for associated processes : Strong consistency and rates. *Stat. Probab. Lett.*, 58(2) :205–219.
- Massart, P. (2003). *Concentration Inequalities and Model selection*, volume 1896 of *Lecture Notes in Mathematics*. Springer, Berlin. Lectures from the 33rd Summer School on Probability Theory held in Saint-Flour, July 6–23, 2003.
- Matias, C. (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM : Probab. Stat.*, 6 :271–292.
- Meister, A. (2004). On the effect of misspecifying the error density in a deconvolution problem. *Canad. J. Statist.*, 32(4) :439–449.
- Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Stat. Sin.*, 16 :195–211.
- Meister, A. (2009). *Deconvolution Problems in Nonparametric Statistics*, volume 193 of *Lecture Notes in Statistics*. Springer-Verlag, Berlin.
- Meister, A. and Neumann, M. H. (2010). Deconvolution from non-standard error densities under replicated measurements. *Stat. Sin.*, 20(4) :1609–1636.
- Nemirovski, A. (2000). Topics in non-parametric statistics. *Lectures on Probability Theory and Statistics : Ecole d’Ete de Probabilites de Saint-Flour XXVIII-1998*, 28 :85.
- Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametric Stat.*, 7(4) :307–330.
- Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *J. Multivariate Anal.*, 98(10) :1955–1968.
- Neumann, M. H. and Reiß, M. (2009). Nonparametric estimation for Lévy processes from low frequency observations. *Bernoulli*, 15(1) :223–248.
- Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Relat. Fields*, 138 :411–449.
- Nickl, R. (2009). On convergence and convolutions of random signed measures. *J. Theoret. Probab.*, 22 :38–56.
- Pensky, M. (2002). Density deconvolution based on wavelets with bounded supports. *Stat. Probab. Lett.*, 56(3) :261–269.
- Pensky, M. (2014). Minimax theory of estimation of linear functionals of the deconvolution density with or without sparsity. *ArXiv e-prints 1411.1660*.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.*, 27(6) :2033–2053.
- Pinheiro, J. and Bates, D. (2000). *Mixed-effect models in S and Splus*. Statistics and Computing Series. Springer-Verlag, New York, NY.
- Prakasa Rao, B. (2004). Moment inequalities for supremum of empirical processes of u-statistic structure and application to density estimation. *J.Iran. Statist. Soc.*, 3 :59–68.

- Rebafka, T. and Roueff, F. (2015). Nonparametric estimation of the mixing density using polynomials. *Math. Methods Stat.*, 24(3) :200–224.
- Rebelles, G. (2015). Pointwise adaptive estimation of a multivariate density under independence hypothesis. *Bernoulli*, 21(4) :1984–2023.
- Rebelles, G. (2015). Structural adaptive deconvolution under l_p -losses. eprint arXiv :1504.06246.
- Reiß, M. and Selk, L. (2013). Estimating nonparametric functionals efficiently under one-sided errors. Preprint, arXiv :1407.4229.
- Rigollet, P. and Tsybakov, A. (2007). Linear and convex aggregation of density estimators. *Math. Methods Stat.*, 16(3) :260–280.
- Rosenthal, H. (1970). On the subspaces of \mathbb{L}_p ($p \geq 2$) spanned by sequences of independent random variables. *Isr. J. Math.*, pages 273–303.
- Roueff, F. and Rydén, T. (2005). Nonparametric estimation of mixing densities for discrete distributions. *Ann. Statist.*, 33(5) :2066–2108.
- Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, 28 :799–815.
- Schick, A. and Wefelmeyer, W. (2004). Root n consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, 16 :925–935.
- Schick, A. and Wefelmeyer, W. (2007). Root n consistent density estimators of convolutions in weighted \mathbb{L}_1 -norms. *J. Statist. Plann. Inference*, 137 :1765–1774.
- Shen, W. and Louis, T. (1999). Empirical bayes estimation via the smoothing by roughening approach. *J. Comput. Graph. Stat.*, 8(4) :800–823.
- Shiryaev, A. and Wilson, S. (1995). *Probability*. Graduate Texts in Mathematics. Springer New York.
- Silverman, B. W. (1986). *Density estimation : for statistics and data analysis*. Chapman & Hall, London.
- Söhl, J. and Trabs, M. (2012). A uniform central limit theorem and efficiency for deconvolution estimators. *Electron. J. Stat.*, 6 :2486–2518.
- Stefanski, L. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Stat. Probab. Lett.*, 9 :229–235.
- Stefanski, S. and Carroll, R. (1990). Deconvoluting kernel density estimators. *Statistics*, 21 :169–184.
- Tsybakov, A. (2000). On the best rate of adaptive estimation in some inverse problems. *C. R. Acad. Sci., Paris, Sér. I, Math.*, 330(9) :835–840.
- Tsybakov, A. (2003). Optimal rates of aggregation. *Proceedings of COLT-2003, Lecture Notes in Artificial Intelligence*, 2777 :303–313.
- Tsybakov, A. (2013). Aggregation and high-dimensional statistics. Lecture notes Saint Flour.

- Tsybakov, A. B. (2009). *Introduction to nonparametric estimation*. Springer Series in Statistics. Springer, New York. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.
- van Es, B. (2011). Combining kernel estimators in the uniform deconvolution problem. *Stat. Neerl.*, 65(3) :275–296.
- van Es, B., Jongbloed, G., and van Zuijlen, M. (1998). Isotonic inverse estimators for nonparametric deconvolution. *Ann. Statist.*, 26(6) :2395–2406.
- Vareschi, T. (2015). Noisy Laplace deconvolution with error in the operator. *J. Stat. Plann. Inference*, 157–158 :16–35.
- Vock, D., Davidian, M., Tsiatis, A., and Muir, A. (2011). Mixed model analysis of censored longitudinal data with flexible random-effects density. *Biostatistics*, 13 :61–73.
- Wu, P. and Zhu, L. (2010). An orthogonality-based estimation of moments for linear mixed models. *Scand. J. Stat.*, 37 :253–263.
- Zhang, C.-H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.*, 18(2) :806–831.
- Zhang, D. and Davidian, M. (2001). Linear mixed models with flexible distributions of random effects for longitudinal data. *Biometrics*, 57(3) :795–802.

Table des figures

2.1	Estimation of the mixed Gamma density with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (bottom left for $n = 200$ and bottom right $n = 2000$), with $\sigma^2 = 1/10$	51
2.2	Estimation of the Gamma density $\frac{2}{\sqrt{20}} \cdot \Gamma(20, \frac{1}{2})$ with Laguerre method (top left for $n = 200$ and top right $n = 2000$) and with Fourier method (bottom left for $n = 200$ and bottom right $n = 2000$), with $\sigma^2 = 1/10$	52
3.1	\tilde{m} in function of $\kappa \in [10^{-4}, 10]$ (15 equidistant calibration constants, some are superposed) plus $\kappa = 0$ (top) and $\kappa = 20$ (bottom), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0, 5]$ for the β and $[0, 10]$ for the χ^2	84
3.2	\tilde{m} in function of $\kappa \in [10^{-4}, 10^{-3}]$ (15 equidistant calibration constants, some are superposed), $n = 5000$ and $\sigma^2 = 1/10$ and 15 equidistant x_0 in the interval $[0, 5]$ for the β and $[0, 10]$ for the χ^2	84
3.3	Estimation of the Gamma density: left for $n = 200$ and right $n = 2000$; $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.	86
3.4	Estimation of the Mixed Gamma density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.	87
3.5	Estimation of the log Normal density: left for $n = 200$ and right $n = 2000$, $\sigma^2 = 1/10$. In bold black line the true density, in orange data-driven estimators computed on various samples and in green dashed line the oracle (the best in the collection) estimator.	87
5.1	Estimations for $n = M = 200$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.	131
5.2	Estimations for $n = M = 500$ of a mixed Gamma (bold black line), in blue dashed line for the NS estimator and in red longdashed for the RD estimator.	132
7.1	Test densities f (solid) and g_2 (dashed).	185
7.2	Laplace noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.	186
7.3	Gaussian noise. True density (dotted), density estimates (gray) and sample of 20 estimates (thin gray) out of 50 proposed to the selection algorithm obtained with a sample of $n = 1000$ data.	187

- 7.4 Laplace noise, Kurtotic density. (a): Graph of the function $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_{\ell}(m)$ against the smoothing parameter m . (b): $\text{MISE}(m)$. The gray diamond represents the global minimizer of $\text{MISE}(m)$ and the gray circle represents the global minimizer of $-\|\hat{g}_{\ell,m}\|^2 + \text{pen}_{\ell}(m)$ 188

Liste des tableaux

1.1	Vitesse de convergence de l'estimateur \hat{f}_m pour le risque quadratique intégré.	15
1.2	Vitesse de convergence de l'estimateur \tilde{f}_m pour le risque quadratique intégré.	23
2.1	Rates of convergence for the MISE	43
2.2	Results of simulation: MISE $\mathbb{E}(\ f - \hat{f}_m\ ^2) \times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$	49
2.3	Results of simulation: MISE $\mathbb{E}(\ S_X - \hat{S}_{X, \tilde{m}_2}\ ^2) \times 100$ averaged over 200 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$	50
3.1	Results of simulation for the global estimation based on Chapter 2. MISE $\mathbb{E}(\ f - \hat{f}_m\ ^2) \times 100$ averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$	85
3.2	Results of simulation for the pointwise estimation. Approximation of the integrated risk multiplied by 100 averaged over 500 samples. σ^2 denotes the level of variance of the noise. $\sigma^2 = 0$ corresponds to the model without noise ($Y = 0$). The noise is $\Gamma(2, \frac{1}{\sqrt{20}})$ for $\sigma^2 = \frac{1}{10}$ and $\Gamma(2, \frac{1}{\sqrt{8}})$ for $\sigma^2 = \frac{1}{4}$. <i>oracle</i> = $\text{length}(I)/300 \sum_{k=1}^{300} \min_{m \in \mathcal{M}} \mathbb{E} f(x_0) - f_m(x_0) ^2$	86
4.1	Results of simulations. Columns report the average values $100 \times \text{ISE} = 100 \times \ f_w^{\mathbb{S}} - f\ ^2$ (aggregate) computed over 200 samples.	111
5.1	Values of approximated MISE $\mathbb{E}(\ f - \hat{f}_{\hat{m}_{\text{NS}, \text{NS}}}\ ^2) \times 100$ averaged over 500 samples with a Laplace and Gaussian noise of variance $1/10$, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.	128
5.2	Values of approximated MISE $\mathbb{E}(\ f - \hat{f}_{\hat{m}_{\text{NS}, \text{NS}}}\ ^2) \times 100$ averaged over 500 samples with a Gaussian noise, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.	129
5.3	Values of approximated MISE $\mathbb{E}(\ f - \hat{f}_{\hat{m}_{\text{RD}, \text{RD}}}\ ^2) \times 100$ averaged over 500 samples, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.	130

5.4	Comparison of the method with the results of a Gaussian kernel estimator with a symmetric noise (MISE $\times 100$ averaged over 500 simulations), \hat{r}^{ker} is the risk of the kernel estimator, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.	131
5.5	Comparison of the method with the results of a Gaussian kernel estimator with an asymmetric noise (MISE $\times 100$ averaged over 500 simulations), \hat{r}^{ker} is the risk of the kernel estimator, \hat{r}^{ad} is the risk of the adaptive estimator, \hat{r}^{or} is the risk of the oracle estimator, M is the size of the noise and n is the size of the Y_i 's.	131
6.1	Rates of convergence for the MISE	149
6.2	Results of simulation with Laplace noise. Columns report the average values of $100 \times ISE = 100 \times \ f - \hat{f}_{\beta, \hat{m}}\ ^2$ and their medians (in brackets) computed over 100 samples.	154
6.3	Results of simulation with Gaussian noise. Columns report the average values of $100 \times ISE = 100 \times \ f - \hat{f}_{\beta, \hat{m}}\ ^2$ and their medians (in brackets) computed over 100 samples.	155
6.4	Results of simulation with Laplace and Gaussian noise with Comte and Samson (2012) procedure. Columns report the average values of $100 \times ISE = 100 \times \ f - \hat{f}_{\beta, \hat{m}}\ ^2$ computed over 100 samples.	155
7.1	Examples of rates of convergence $\varphi_n^{(a)}$ of the estimator $\hat{g}_{\ell, \hat{m}}$	181
7.2	Examples of values of $m_{opt}^{(2)}$ and ψ_M , for $\ell \geq 2$, k_0 is an integer such that $b_\ell(k_0 - 1) < 2\alpha \leq b_\ell(k_0)$, for $b_\ell(k) = (2ak + 1)/\ell$, $k = 1, \dots, \ell$ and $b_\ell(0) = 0$	184
7.3	1000 \times MISE values from 100 replications of sample sizes $n = 1000$	188

Table des matières

Résumé	vii
Abstract	ix
Remerciements	xi
1 Introduction	1
1.1 Estimation non-paramétrique adaptative	2
1.1.1 Estimation par projection	3
1.1.2 Étude du risque et choix non adaptatif de la dimension	4
1.1.3 Sélection de modèles et inégalités oracles	5
1.1.4 Inégalités de concentration	8
1.2 Modèle de convolution	10
1.2.1 Problème et modèle statistique	10
1.2.2 Méthodes de déconvolution à bruit connu	12
1.2.3 Méthodes de déconvolution à bruit inconnu	21
1.3 Perspectives	27
I Estimation adaptative dans le modèle de convolution sur \mathbb{R}^+	31
2 Adaptive deconvolution on the nonnegative real line	33
2.1 Introduction	35
2.1.1 Model	35
2.1.2 Bibliographical context	35
2.1.3 General strategy	36
2.1.4 Plan of the paper	37
2.2 Statistical model and estimation procedure	38
2.2.1 Notations	38
2.2.2 Laguerre basis	38
2.2.3 Projection estimator of the density function	38
2.2.4 Projection estimator of the survival function	39
2.3 Bounds on the \mathbb{L}^2 risk	40
2.3.1 Upper bounds	40
2.3.2 Rates of convergence	41
2.3.3 In what context does the Laguerre procedure improve the Fourier approach ?	43
2.4 Model selection	45
2.4.1 Adaptive density estimation	45
2.4.2 Adaptive survival function estimation	46

2.5	Illustrations	47
2.5.1	Simulation setting	47
2.5.2	Practical estimation procedure	48
2.5.3	Simulation results	49
2.6	Concluding remarks	50
2.7	Proofs	52
2.7.1	Proof of Section 2.3	52
2.7.2	Proof of Theorem 2.4.1	54
2.7.3	Proof of Theorem 2.4.2 and Corollary 2.4.3	60
2.8	Appendix	65
3	Adaptive deconvolution of linear functionals on the nonnegative real line	67
3.1	Introduction	69
3.2	Statistical model and estimation procedure	71
3.2.1	Notations	71
3.2.2	Laguerre basis	71
3.2.3	Projection estimator of the linear functional	71
3.2.4	Upper bound on the mean squared error and rates of convergence	74
3.3	Model selection and adaptive estimation	78
3.4	Particular case of pointwise estimation	80
3.4.1	Theoretical results	80
3.4.2	Illustrations	82
3.5	Proofs	88
3.5.1	Proofs of Section 3.2	88
3.5.2	Proofs of Section 3.3	90
3.6	Appendix 1	97
3.7	Appendix 2	97
4	Density aggregation on the nonnegative real line in the convolution model	99
4.1	Introduction	101
4.1.1	Motivations	101
4.1.2	Aim and model	102
4.1.3	Notations	103
4.2	Classical aggregation problem	104
4.3	Linear aggregation of Laguerre estimators	105
4.3.1	Estimation of the weights of aggregation	106
4.3.2	Oracle inequality for the aggregate	108
4.3.3	Discussion	108
4.4	Illustration	110
4.5	Proofs	112
4.5.1	Proof of Lemma 4.3.5	112
4.5.2	Proof of Theorem 4.3.3	112
4.5.3	Proof of Corollary 4.3.4	113
II	Estimation adaptative de densité dans le modèle de convolution avec bruit inconnu	115
5	Adaptive density estimation in deconvolution problems with unknown error distribution	117

5.1	Introduction	119
5.2	Statistical model, estimation procedure and risk bounds	120
5.2.1	Notations	121
5.2.2	Statistical model and estimators	121
5.2.3	Upper risk bounds	123
5.3	Data driven bandwidth selection and oracle bounds	124
5.4	Illustrations	126
5.4.1	Practical estimation procedure	126
5.4.2	Comparison with Comte and Lacour (2011) and influence of M in the NS-model	127
5.4.3	Illustrations in the RD-model	128
5.4.4	Comparison with a kernel estimator	128
5.5	Concluding remarks	132
5.6	Proofs	133
5.6.1	Preliminaries	133
5.6.2	A technical auxiliary result	134
5.6.3	Proof of the oracle bounds	138
6	Adaptive estimation of marginal random-effects densities in linear mixed-effects model	141
6.1	Introduction	143
6.2	Statistical model and estimation procedure	144
6.2.1	Notations	144
6.2.2	Estimation of f_α	145
6.2.3	Estimation of f_β	146
6.3	Model selection	149
6.3.1	Adaptive estimation procedure for f_α	150
6.3.2	Adaptive estimation procedure for f_β	151
6.4	Simulation	152
6.4.1	Implementation of the estimation procedure	152
6.4.2	Simulation setting	152
6.4.3	Simulation results	153
6.5	Concluding remarks	154
6.6	Proofs	156
6.6.1	Sketch of the proof of Proposition 6.2.2	156
6.6.2	Proof of Theorem 6.3.1	156
6.6.3	Proof of Theorem 6.3.2	156
6.6.4	Proof of Lemma 6.6.1	158
6.6.5	Proof of Proposition 6.6.2	161
6.7	Appendix 1	169
6.8	Appendix 2: Supplementary material	170
6.8.1	Proof of Lemma 6.8.2	170
6.8.2	Proof of Lemma 6.8.1	172
6.8.3	Proof of Proposition 6.2.2	172
7	Estimation of convolution in the model with noise	175
7.1	Motivations	177
7.2	Estimation with known noise density	178
7.2.1	Notations and assumption	178
7.2.2	Estimators when the noise density is known	178

7.2.3	Risk bound for $\hat{g}_{\ell,m}$	179
7.2.4	Rates of convergence for $\hat{g}_{\ell,m}$	180
7.2.5	Risk bound on the adaptive estimator $\hat{g}_{\ell,\hat{m}}$	181
7.3	Estimation with unknown noise density	182
7.3.1	Risk bound for fixed m	182
7.3.2	Adaptive procedure	184
7.4	Illustration	185
7.5	Proofs	188
7.5.1	Proof of Proposition 7.2.1.	188
7.5.2	Proof of Corollary 7.2.2	189
7.5.3	Proof of Theorem 7.2.3.	191
7.5.4	Proof of Proposition 7.5.1	193
7.5.5	Proof of the rates in Table 7.2	195
7.5.6	Proof of Theorem 7.3.4	195
7.5.7	Proof of Lemma 7.5.3	197
7.5.8	Proof of Proposition 7.5.4	199
7.6	Appendix	203
Bibliographie		205
Table des figures		214
Liste des tableaux		216

Estimation non-paramétrique adaptative pour des modèles bruités

RÉSUMÉ. Dans cette thèse, nous nous intéressons au problème d'estimation de densité dans le modèle de convolution. Ce cadre correspond aux modèles avec erreurs de mesures additives, c'est-à-dire que nous observons une version bruitée de la variable d'intérêt. Pour mener notre étude, nous adoptons le point de vue de l'estimation non-paramétrique adaptative qui repose sur des procédures de sélection de modèle développées par Birgé & Massart ou sur les méthodes de Lepski. Cette thèse se divise en deux parties. La première développe des méthodes spécifiques d'estimation adaptative quand les variables d'intérêt et les erreurs sont des variables aléatoires positives. Ainsi nous proposons des estimateurs adaptatifs de la densité ou encore de la fonction de survie dans ce modèle, puis de fonctionnelles linéaires de la densité cible. Enfin nous suggérons une procédure d'agrégation linéaire. La deuxième partie traite de l'estimation adaptative de densité dans le modèle de convolution lorsque la loi des erreurs est inconnue. Dans ce cadre il est supposé qu'un échantillon préliminaire du bruit est disponible ou que les observations sont disponibles sous forme de données répétées. Les résultats obtenus pour des données répétées dans le modèle de convolution permettent d'élargir cette méthodologie au cadre des modèles linéaires mixtes. Enfin cette méthode est encore appliquée à l'estimation de la densité de somme de variables aléatoires observées avec du bruit.

MOTS-CLÉS. Modèles de convolution. Modèles de durées. Modèles mixtes. Estimation non-paramétrique. Estimation adaptative. Estimation par projection. Sélection de modèles. Méthodes de Goldenshluger et Lepski. Agrégation. Vitesses optimales minimax.

Nonparametric adaptive estimation in measurement error models

ABSTRACT. In this thesis, we are interested in nonparametric adaptive estimation problems of density in the convolution model. This framework matches additive measurement error models, which means we observe a noisy version of the random variable of interest. To carry out our study, we follow the paradigm of model selection developed by Birgé & Massart or criterion based on Lepski's method. The thesis is divided into two parts. In the first one, the main goal is to build adaptive estimators in the convolution model when both random variables of interest and errors are distributed on the nonnegative real line. Thus we propose adaptive estimators of the density along with the survival function, then of linear functionals of the target density. This part ends with a linear density aggregation procedure. The second part of the thesis deals with adaptive estimation of density in the convolution model when the distribution is unknown and distributed on the real line. To make this problem identifiable, we assume we have at hand either a preliminary sample of the noise or we observe repeated data. So, we can derive adaptive estimation with mild assumptions on the noise distribution. This methodology is then applied to linear mixed models and to the problem of density estimation of the sum of random variables when the latter are observed with an additive noise.

KEYWORDS. Convolution models. Duration models. Mixed models. Nonparametric estimation. Adaptive estimation. Projection estimators. Model selection. Goldenshluger and Lepski method. Aggregation. Minimax optimal rates.