

Mémoire présenté en vue de de l'obtention de l'Habilitation à Diriger des Recherches

Spécialité : Statistiques

Présenté par Marie-Luce Taupin

Déconvolution et Estimation dans les modèles avec erreurs sur les variables.

Rapporteurs:

Béatrice LAURENT
Sara VAN DE GEER
Valentin PATILEA

Soutenance le 09 décembre 2008 devant le jury:

Lucien BIRGE
Valentine GENON-CATALOT
Marc HOFFMANN
Sylvie HUET
Béatrice LAURENT
Pascal MASSART
Valentin PATILEA

Remerciements

Mes premiers mots vont à Pascal Massart et Sylvie Huet qui m'ont initiée à la recherche et aux statistiques. Leurs qualités humaines et mathématiques sont pour moi indiscutables. Je les remercie de leur bienveillance, jamais intrusive, ainsi que de leur disponibilité. Je remercie en particulier Pascal de m'avoir permis d'être accueillie aussi chaleureusement dans le laboratoire de Mathématiques d'Orsay.

Merci à Béatrice Laurent d'avoir accepté d'être rapporteur de mon habilitation. Je la remercie de sa lecture attentive, de ses remarques pertinentes, mais aussi de toutes les discussions fructueuses et amicales que nous avons eues, et ce depuis ma thèse déjà.

Merci à Sara van de Geer et Valentin Patilea d'avoir accepté d'écrire un rapport sur ce travail. J'en suis très flattée.

Merci à Lucien Birgé d'avoir bien voulu faire partie de mon jury. Merci pour sa disponibilité, mais aussi pour les nombreuses discussions très instructives que nous avons eues, notamment sur les minorations et la distance de Hellinger.

Merci à Valentine Genon-Catalot d'avoir bien voulu faire partie du jury, mais aussi pour son aide et ses précieux conseils, bien utiles pour soutenir son habilitation à l'Université Paris Descartes.

Merci à Marc Hoffmann d'avoir bien voulu faire partie de mon jury. J'en suis très flattée.

Un petit mot à Fabienne Comte avec qui j'ai collaboré de nombreuses fois. Merci à Fabienne pour son énergie qui n'a d'égal que son enthousiasme sans cesse renouvelé, sa fraîcheur, ses compétences et son ouverture d'esprit mathématique. Tout ceci en fait une collaboratrice hors pair, avec qui j'ai découvert beaucoup de choses et avec qui c'est un grand plaisir de travailler.

Merci à Catherine Matias pour son enthousiasme, son humour et sa clairvoyance. Catherine fut ma deuxième collaboratrice. C'est grâce à elle que j'ai découvert les chaînes de Markov cachées et leur lien avec les modèles autorégressifs avec erreurs sur les variables.

Merci à Marie-Laure Martin pour sa gentillesse, sa patience et sa sérénité. Marie-Laure m'a fait découvrir les modèles d'analyse de survie. Je l'en remercie encore.

Merci à Cristina Butucea pour son enthousiasme, son sens de l'humour parfois déroutant mais toujours drôle. Elle m'a fait profiter de ses connaissances sur la déconvolution et les minorations.

Merci à Yves Rozenholc pour son énergie hors du commun qui en fait un collaborateur avec qui on ne s'ennuie jamais.

Merci donc à tous mes collaborateurs, sans qui ce mémoire ne serait sans doute pas ce qu'il est.

Je remercie mes anciens collègues de bureau d'Orsay, plus ou moins temporaires : Yannick Heurteaux, Béatrice Laurent, Cécile Durot, Jean-Michel Marin. Merci pour ces moments de partage et d'amitié.

Je remercie également les membres du bâtiment 440 : Cécile, Arnaud, Cathy, Christine, Nathalie, et surtout Gilles et Jean-Michel, qui ont fait de ce couloir au départ excentré, un endroit très chaleureux et plein d'échanges.

Merci à Sana, Vincent, Odile, Liliane, Marie-Anne, pour les discussions mathématiques, mais aussi pour les discussions qui l'étaient moins, notamment informatiques. Merci à tous les autres membres du laboratoire Mathématique d'Orsay qui m'ont accueillie pendant de nombreuses années. J'ai beaucoup apprécié de participer à la vie du laboratoire et d'y travailler en toute liberté. Merci notamment à Elisabeth Gassiat, Jean Bretagnolle, Yves Misiti, Jean Coursol, Isabelle Souriou et Patrick Jakubowicz.

Je remercie tous mes collègues de l'IUT, avec qui j'ai eu le plaisir d'enseigner, de discuter, de partager, notamment Michel Larmande, Jean-Michel Poggi, Dominique Gascon, Michel Levy, Guillaume Bordry, Servane Gey, Marc Métivier, François-Xavier Jollois, Mohamed Mellouk, Philippe Chabault, Caroline Larmande, Clarisse Pantin de la Guere, Bruno Portier, Serge Blumenthal, Thérèse Martelli, François Elazzaoui, Mourad Boughedaoui et Marc Lavielle.

Merci aussi à Florence, Adeline, Elodie, Elisabeth, pour votre aide, votre soutien, pour tous ces moments partagés, et pour les fous rires parfois salutaires.

Je tiens aussi à remercier tous les membres du laboratoire MAP5 de m'avoir accueillie récemment, notamment Christine Graffigne, Anne Estrade, Grégory Nuel et Avner Bar-Hen, dans cette ambiance animée et très sympathique. Un mot pour mes nouveaux collègues de bureaux, certains déjà passés, Laure Coutin, Jean-Christophe Thalabard, Chantal Guihenneuc et Yves Rozenholc. Merci pour tous les moments d'échanges et d'amitié. Merci à Chantal, Adeline et Antoine pour leur soutien, leur aide précieuse pour la découverte de l'Université Paris Descartes, pour leur gentillesse et leur grande disponibilité.

Un mot pour tous mes collègues et amis, certains que j'ai déjà remerciés, d'autres pas encore, Anne-Sophie, Estelle, Marguerite, Frédérique, Gwénaelle, Clara, Pascal, Bertrand, Daniel, Véronique, Stéphanie, Dasha, Nadine, Clémentine, Florence,...

Merci aussi à tous ceux que j'ai oubliés, j'espère qu'ils ne m'en tiendront pas rigueur.

Un dernier mot pour ma famille proche ou moins proche, pour leur soutien, leur amitié et leur aide.

Il est un dernier collaborateur que je n'ai pas encore remercié parce qu'il est bien plus que cela.

Merci Jérôme.

Table des matières

I	Document de synthèse	5
1	Introduction	6
2	Déconvolution	12
2.1	Estimation adaptative de la densité par déconvolution : cadre i.i.d.	12
2.1.1	Introduction	12
2.1.2	Estimateurs non adaptatifs	13
2.1.3	Estimateur adaptatif	14
2.1.4	Borne de risque pour les estimateurs non adaptatifs	15
2.1.5	Estimation adaptative	15
2.1.6	Étude par simulation	16
2.2	Déconvolution adaptative pour une suite faiblement dépendante	19
2.2.1	Mesures de dépendance	20
2.2.2	Exemples de suites faiblement dépendantes	20
2.2.3	Estimateurs	21
2.2.4	Borne de risque pour les estimateurs non adaptatifs	21
2.2.5	Bornes de risque pour l'estimateur adaptatif	21
2.2.6	Estimation adaptative pour f_ε "super smooth".	22
2.2.7	Estimation adaptative pour f_ε "ordinary smooth".	22
2.3	Estimation adaptative dans un modèle ARCH général	23
2.3.1	Introduction	23
2.3.2	Procédure d'estimation	24
2.3.3	Hypothèses de dépendance	24
2.3.4	Borne de risque pour l'estimateur non adaptatif	25
2.3.5	Borne de risque pour l'estimateur adaptatif	25
2.3.6	Estimation de la densité de σ_t^2	26
2.3.7	Exemples et illustration pratique	26
2.4	Estimation de fonctionnelles linéaires dans le modèle de convolution	29
3	Modèles de régression avec erreurs sur les variables	30
3.1	Estimation non paramétrique dans un modèle de régression avec erreurs sur les variables	30
3.1.1	Introduction, modèle et résultats antérieurs	30
3.1.2	Méthode d'estimation	31
3.1.3	Bornes de risque pour les estimateurs sur un modèle	32
3.1.4	Une inégalité de type oracle	32
3.1.5	Illustration par des simulations	33
3.1.6	Perspectives	34
3.2	Estimation semi-paramétrique dans un modèle de régression avec erreurs sur les variables	35
3.2.1	Le modèle	35

3.2.2	Résultats antérieurs	35
3.2.3	Première méthode d'estimation	36
3.2.4	Deuxième méthode d'estimation	37
3.2.5	Propriétés asymptotiques	39
3.2.6	Vitesses de convergence pour des classes de régularité	40
3.2.7	Exemples	40
3.2.8	Perspectives	41
3.3	Estimation dans un modèle autorégressif avec erreurs sur les variables	42
3.3.1	Le modèle, les résultats antérieurs	42
3.3.2	Procédure d'estimation	42
3.3.3	Résultats	43
3.3.4	Perspectives	43
4	Modèles de survie avec erreurs sur les variables	44
4.1	Introduction, résultats antérieurs et modèle considéré	44
4.2	Procédure d'estimation	48
4.3	Propriétés asymptotiques	49
4.4	Choix de la fonction de poids et \sqrt{n} -consistance	49
4.5	Vitesses de convergence pour des classes de régularité classiques	50
4.6	Exemples	50
4.7	Une étude par simulation	51
5	Projets de recherche	55

Première partie
Document de synthèse

Chapitre 1

Introduction

L'objectif de ce mémoire est de présenter mes travaux de recherche et notamment de montrer comment les outils usuels de la déconvolution peuvent être utilisés pour l'estimation dans des modèles à variables mal mesurées, ou mesurées avec erreur. Mes travaux ne sont pas présentés dans l'ordre chronologique, mais plutôt regroupés par thème. Par ailleurs, certains articles sont plus détaillés que d'autres et la présentation des résultats dans ce mémoire peut être légèrement différente des articles originaux, cette nouvelle présentation tenant compte d'améliorations ultérieures.

Je me suis intéressée à des modèles reliant une variable expliquée à une variable explicative mesurée avec une erreur additive. Si Y (ou T) est la variable expliquée et Z la variable explicative, on s'intéresse à la relation qui lie Y (ou T) à Z dans un contexte où la variable Z est mesurée avec une erreur additive. Plus précisément, Z n'est pas observée, mais on observe U liée à Z par la relation

$$U = Z + \varepsilon. \quad (1.0.1)$$

Je renvoie notamment à Fuller (1987), Carroll *et. al.* (1995) ou Schneeweiss et Augustin (2006) pour une présentation d'exemples et de méthodes.

Les modèles que je vais présenter dans ce mémoire sont : le modèle de convolution (1.0.1), des modèles de régression, et des modèles d'analyse de survie. Dans chacun de ces modèles la variable explicative Z (ou covariable) sera non observée. À la place on observe $U = Z + \varepsilon$ et l'on souhaite estimer soit la loi de Z , soit la relation qui lie la variable expliquée à Z , soit les paramètres qui régissent cette relation. La première idée qui vient à l'esprit est de considérer les procédures classiques d'estimation, utilisées quand Z est observée et de remplacer la valeur de Z , non disponible, par l'observation U . Cette idée très souvent appelée idée "naïve" ne fournit pas des estimateurs consistants.

Mes travaux peuvent être regroupés en deux catégories, l'une sur l'estimation non paramétrique et l'autre sur l'estimation semi-paramétrique. La partie sur l'estimation non paramétrique regroupe mes travaux sur la déconvolution et l'estimation dans un modèle de régression non paramétrique avec erreurs sur les variables. La partie sur l'estimation semi-paramétrique regroupe mes travaux sur l'estimation dans les modèles de régression (régression et autorégressif) semi-paramétrique avec erreurs sur les variables et les modèles de survie avec covariable mesurée avec erreur.

Les modèles considérés

Pour tous les modèles étudié, la première étape va consister à retrouver de l'information sur Z en utilisant l'observation U . C'est l'idée de la déconvolution. Je vais donc commencer par présenter mes travaux en déconvolution.

Le modèle de convolution

Cette partie présente les articles Comte *et al.* (2006, 2007, 2008a, 2008b) ainsi que l'article Matias et Taupin (2004). Considérons le modèle

$$U_i = Z_i + \varepsilon_i, \quad i = 1, \dots, n, \quad (1.0.2)$$

avec U_i l'observation, Z_i indépendante de ε_i et ε_i admettant une densité connue f_ε .

Déconvolution On cherche à estimer la densité f_Z des variables Z_1, \dots, Z_n en utilisant les observations U_1, \dots, U_n . L'indépendance de Z_i et ε_i implique que la densité de l'observation U est le produit de convolution de f_Z et f_ε . C'est pourquoi on parle d'estimation de densité par déconvolution ou plus simplement de déconvolution. L'hypothèse de connaissance de f_ε , bien que sujette à critique, est nécessaire pour des raisons d'identifiabilité. Pour des résultats sur l'identifiabilité et l'estimation de la loi de l'erreur ε je renvoie à Neumann (1997), Matias (2002), Meister (2004), Butucea et Matias (2005) ou Butucea, Matias et Pouet (2008).

Dans ce contexte nous avons considéré plusieurs modèles. Le premier est un modèle i.i.d. c'est-à-dire où toutes les variables sont indépendantes et identiquement distribuées. Le second est un modèle dépendant où les ε_i sont i.i.d. et où la suite $(Z_i)_{i \geq 1}$ est strictement stationnaire et satisfait des conditions de dépendance faible récemment introduites par Dedecker et Prieur (2005). Le troisième est un modèle ARCH où $((Y_t, \sigma_t))_{t \in \mathbb{N}}$ est une suite strictement stationnaire de $\mathbb{R} \times \mathbb{R}^+$ satisfaisant la relation

$$Y_t = \sigma_t \eta_t \quad (1.0.3)$$

où $(\eta_t)_{t \in \mathbb{Z}}$ est une suite i.i.d. de variables centrées et de variance finie et pour tout $t \geq 0$, le vecteur aléatoire $(\sigma_i, \eta_{i-1})_{0 \leq i \leq t}$ est indépendant de la suite $(\eta_i)_{i \geq t}$.

Le modèle (1.0.3) est classiquement réécrit *via* une transformation logarithmique sous la forme

$$U_t = Z_t + \varepsilon_t,$$

où $U_t = \ln(Y_t^2)$, $Z_t = \ln(\sigma_t^2)$ et $\varepsilon_t = \ln(\eta_t^2)$. Dans le contexte du modèle (1.0.3), Z_t et ε_t sont indépendantes pour un t fixé, mais les suites $(Z_t)_{t \geq 0}$ et $(\varepsilon_t)_{t \in \mathbb{Z}}$ ne sont pas indépendantes.

Dans ces trois modèles, nous estimons f_Z par sélection de modèles et proposons une nouvelle méthode d'estimation adaptative par déconvolution. Cette méthode d'estimation est nouvelle, tant par la construction du critère d'estimation, que par le choix de la collection de modèles. Ces derniers assurent, comme étape préliminaire, que le meilleur estimateur de la collection atteint la vitesse minimax (dans tous les cas où des minoration ont été établies), que la densité à estimer et la densité de l'erreur soit "ordinary smooth" ou "super smooth". La procédure de sélection de modèles associée à cette collection d'estimateurs fournit un estimateur adaptatif qui atteint les vitesses de convergence minimax dans tous les cas où des minoration ont été établies, et qui atteint la vitesse du meilleur estimateur de la famille à un facteur (logarithmique) négligeable près sinon.

Nous avons également montré que cette procédure d'estimation adaptative de la densité f_Z fonctionne aussi dans des contextes dépendants variés, avec des conditions de dépendance faible. Plus précisément, dans les deux contextes dépendants nous montrons que sous des conditions de dépendance faibles, l'estimateur adaptatif a la même vitesse que dans le cas i.i.d.. En particulier, les coefficients de dépendance n'interviennent pas dans le terme dominant de la variance, mais uniquement dans un terme additionnel négligeable. De plus la pénalité ne dépend pas des coefficients de dépendance. Ces deux derniers points sont particulièrement nouveaux et intéressants puisqu'ils permettent d'avoir une procédure d'estimation qui ne fait pas intervenir les coefficients de dépendance.

Estimation de fonctionnelles linéaires Cette partie présente l'article Matias et Taupin (2004). Considérons le modèle (1.0.2) avec $\varepsilon \sim \mathcal{N}(0,1)$, dans lequel on souhaite estimer des fonctionnelles linéaires intégrales de f_Z de la forme

$$\Gamma_f(y) = \int f(z)f_\varepsilon(z-x)f_Z(z)dz,$$

pour une fonction f connue. Quand $f \equiv 1$ on retrouve le densité de l'observation U , puisque $f_U = f_Z \star f_\varepsilon = \Gamma_1$. L'intérêt pour ces fonctionnelles linéaires intégrales vient du modèle de régression semi-paramétrique (1.0.4) et plus précisément de l'estimation d'espérances conditionnelles de la forme $\mathbb{E}(f(Z)|U)$ (voir les parties 3.2.3 et 3.3). Dans l'article Matias et Taupin (2004), nous nous sommes concentrées sur les fonctionnelles Γ_f pour des fonctions f polynomiales, exponentielles, ou trigonométriques avec un intérêt tout particulier pour $f \equiv 1$, c'est-à-dire pour l'estimation de f_U .

En utilisant des outils liés à la déconvolution j'ai proposé un estimateur consistant de Γ_f (voir Taupin (2001)). Sa vitesse de convergence dépend de la régularité de $f(x)f_\varepsilon(\cdot-x)$ comme fonction de x comparée à celle de f_ε .

Notre objectif est d'établir des minoration pour des risques minimax associés à des fonctions de perte variées, et de montrer que l'estimateur (2.4.8) atteint ces vitesses optimales quand f est polynomiale, trigonométrique ou exponentielle. Pour chacun des types de fonctions f , nous avons considéré plusieurs risques minimax : le risque quadratique ponctuel, le risque par rapport à la norme \mathbb{L}_p et enfin le risque relié à la norme uniforme.

Un point important à remarquer est que les vitesses (optimales) pour l'estimation de f_U sont très différentes des vitesses pour l'estimation de f_Z . Ceci vient du produit de convolution $f_U = f_Z \star f_\varepsilon$. Comme nous l'avons mentionné, quand f_ε est gaussienne, alors f_U est très régulière, avec notamment une décroissance exponentielle de sa transformée de Fourier. Ceci rend l'estimation de f_U aisée avec des vitesses proches de la vitesse paramétrique alors que l'estimation de f_Z est difficile avec des vitesses de convergence logarithmiques. Par ailleurs, plus f est régulière et plus la vitesse pour l'estimation de Γ_f est rapide. Cependant, bien que les fonctions f que nous avons considérées admettent toutes un prolongement analytique sur le domaine complexe, les vitesses pour l'estimation de Γ_f sont très différentes suivant les cas. Ce travail montre d'ores et déjà que l'utilisation d'outils liés à la déconvolution dans un contexte différent de celui de la déconvolution, permet d'obtenir des vitesses optimales bien plus rapides.

Modèles de régression

Modèle non paramétrique Cette partie correspond à l'article Comte et Taupin (2007) sur le modèle de régression non paramétrique

$$\begin{cases} Y_i &= f(Z_i) + \xi_i, \\ U_i &= Z_i + \varepsilon_i, \end{cases}$$

avec ε_i indépendante de Z_i , et de densité f_ε connue. Dans ce modèle, l'objectif est d'estimer la fonction de régression f en utilisant les observations $(Y_1, U_1), \dots, (Y_n, U_n)$, et sans connaître sa régularité. Le point de départ de notre procédure d'estimation est l'idée classique que la fonction de régression en un point x peut s'écrire sous la forme

$$f(z) = \mathbb{E}(Y|Z = z) = \frac{\int y f_{Z,Y}(z, y) dy}{f_Z(z)} = \frac{(f f_Z)(z)}{f_Z(z)},$$

avec $f_{Z,Y}$ le densité jointe de (Z, Y) . Nous proposons d'estimer la fonction de régression par le ratio de deux estimateurs adaptatifs, c'est-à-dire par $\hat{f} = \tilde{\ell} / \tilde{f}_Z$, où $\tilde{\ell}$ est un estimateur adaptatif de $\ell = f f_Z$ et \tilde{f}_Z est un estimateur adaptatif de f_Z . Ces deux estimateurs vont être obtenus par minimisation de

contrastes pénalisés bien choisis. La méthode d'estimation est en partie basée sur les idées de l'estimation de densité par déconvolution. Nous montrons alors que le risque quadratique intégré de l'estimateur de f est majoré par la somme des risques quadratiques intégrés des estimateurs adaptatifs ℓ et f_Z . Nous avons donc proposé une procédure d'estimation de la fonction de régression, adaptative dans le sens où elle ne requiert pas la connaissance de la régularité de la fonction de régression et avons étudié les vitesses de convergence pour f et f_ε dans deux classes de régularité incluant toutes les deux les fonctions "ordinary smooth" et "super smooth". Cet estimateur atteint les vitesses minimax dans tous les cas où des minorations ont été établies c'est-à-dire dans le cadre d'hypothèses décrit dans Fan et Truong (1993).

Modèle semi-paramétrique Cette partie correspond aux articles Taupin (1998, 2001) et Butucea et Taupin (2008) sur le modèle de régression semi-paramétrique suivant

$$\begin{cases} Y_i &= f_{\theta^0}(Z_i) + \xi_i, \\ U_i &= Z_i + \varepsilon_i, \end{cases} \quad (1.0.4)$$

avec ε_i de densité f_ε connue et indépendante de Z_i et avec ξ_i de loi inconnue. Dans ce modèle, la fonction de régression appartient à une famille paramétrique

$$\mathcal{F} = \{f_\theta, \quad \theta \in \Theta \text{ compact de } \mathbb{R}^d\}.$$

L'objectif est alors d'estimer le paramètre θ^0 , en utilisant les observations $(Y_1, U_1), \dots, (Y_n, U_n)$. J'ai étudié ce problème dans les deux articles Taupin (2001) et Butucea et Taupin (2008). Les deux articles présentent des méthodes d'estimation basées sur une minimisation de deux critères empiriques liés au critère des moindres carrés et qui utilisent des outils liés à la déconvolution.

Dans Taupin (2001), je considère le cas particulier où ε est gaussienne et ξ_i de loi inconnue. Le point de départ de la procédure d'estimation est le critère des moindres carrés $\mathbb{E}[(Y - \mathbb{E}(f_\theta(Z)|U))^2 W(U)]$. L'estimateur construit est consistant et je fournis une majoration de sa vitesse de convergence pour une fonction de régression générale. L'estimateur ainsi construit converge à la vitesse \sqrt{n} dans le cas d'une fonction de régression polynomiale ou exponentielle. Je fournis des majorations de sa vitesse de convergence pour d'autres exemples de fonctions de régression. Cet article est le premier, à ma connaissance, à fournir un estimateur consistant pour un modèle non linéaire général.

Dans Butucea et Taupin (2008), ε_i est de loi connue générale et ξ_i de loi inconnue. Le point de départ de la procédure d'estimation est le critère des moindres carrés $\mathbb{E}[(Y - f_\theta(Z))^2 W(Z)]$. L'estimateur est construit par minimisation d'une version empirique du critère ci-dessus qui utilise des outils liés à la déconvolution. On montre que cet estimateur est consistant, avec une vitesse de convergence qui dépend de la régularité de wf_θ comparée à celle de f_ε . Nous montrons de plus qu'il est \sqrt{n} -consistant et asymptotiquement gaussien pour des modèles de régression classiques sans restriction sur la densité f_ε . Cette méthode d'estimation de θ^0 permet de considérer des modèles non linéaires généraux, des densités f_ε générales et ne requiert pas d'hypothèses a priori sur la loi de Z . Elle présente aussi l'avantage de fournir, sous des hypothèses plus générales, un estimateur plus simple avec des vitesses plus explicites et souvent plus rapides que dans Taupin (2001). En particulier elle permet d'explicitier les vitesses de convergence pour des fonctions de régression et des loi d'erreur ε appartenant à des classes de régularité classiques.

Modèle autorégressif Cette partie correspond à l'article Comte et Taupin (2001) sur le modèle autorégressif avec erreurs sur les variables suivant

$$\begin{cases} Z_i &= f_{\theta^0}(Z_{i-1}) + \xi_i, \\ U_i &= Z_i + \varepsilon_i, \end{cases}$$

avec ε_i de densité f_ε connue et indépendante de Z_i et avec ξ de loi inconnue. Dans ce modèle, la fonction de régression appartient à une famille paramétrique

$$\mathcal{F} = \{f_\theta, \quad \theta \in \Theta \text{ compact de } \mathbb{R}^d\}.$$

L'objectif est alors d'estimer le paramètre θ^0 , en utilisant les observations U_1, \dots, U_n . On suppose que les variables ξ_i et ε_i sont i.i.d. et que les suites $(\xi_i)_{i \in \mathbb{Z}}$ et $(\varepsilon_i)_{i \in \mathbb{Z}}$ sont mutuellement indépendantes. On suppose également que les suites $(Z_i)_{i \in \mathbb{Z}}$ et $(\varepsilon_i)_{i \in \mathbb{Z}}$ sont mutuellement indépendantes. Ce modèle autorégressif avec erreurs sur les variables est un cas particulier de chaîne de Markov cachée, à espace d'état continu et non compact et dont la loi n'est pas complètement paramétrique du fait de la non connaissance de la loi de ξ_i . La procédure d'estimation est basée sur le critère $\mathbb{E}[W(U_{i-1})(U_i - \mathbb{E}(f_\theta(Z_{i-1})|U_{i-1}))^2]$. Nous proposons un estimateur de θ^0 obtenu en minimisant une version empirique de ce critère construite avec des outils liés à la déconvolution. Nous montrons que sous des conditions de stationarité et d'absolue régularité, cet estimateur est consistant pour des fonctions de régression générales et pour une densité f_ε connue générale. Sa vitesse de convergence est la même que celle obtenue dans un cadre i.i.d. dans Taupin (2001).

Modèles d'analyse de données de survie

Cette partie fait l'objet de l'article Martin-Magniette et Taupin (2008). Le modèle considéré est un modèle d'analyse de survie où la relation entre un temps de survie T et une covariable $Z \in \mathbb{R}$ est décrite par le risque instantané conditionnel à Z de la forme

$$R(t|z) = \lim_{\Delta_t \downarrow 0} \frac{1}{\Delta_t} \mathbb{P}(t \leq T < t + \Delta_t | T \geq t, Z = z) \text{ où } R(t|Z) = \eta_{\gamma^0}(t) r_{\beta^0}(Z), \quad (1.0.5)$$

avec η_0 le risque de base et r_{β^0} le risque relatif. Pour $i = 1, \dots, n$ on observe $X_i = \min(T_i, C_i)$ et $D_i = \mathbb{I}_{T_i \leq C_i}$. La covariable Z_i n'est pas observée et à la place on observe

$$U_i = Z_i + \varepsilon_i$$

avec ε_i indépendante de Z_i , et de densité f_ε connue. Dans ce modèle l'objectif est d'estimer le paramètre $\theta^0 = (\gamma^0, \beta^0)^T$ en utilisant les observations i.i.d. (X_i, D_i, U_i) , $i = 1, \dots, n$. Nous proposons d'estimer θ^0 par minimisation d'une version modifiée du critère des moindres carrés qui utilise des outils liés à la déconvolution. On montre alors que cet estimateur est consistant et que sa vitesse de convergence dépend de la régularité de la $W r_\beta$ comparée à celle de f_ε . Nous montrons de plus qu'il est \sqrt{n} -consistant et asymptotiquement gaussien pour les risques instantanés classiques comme ceux du modèle de Cox ou du modèle d'excès de risque. Cette méthode d'estimation permet d'obtenir un estimateur consistant dans des modèles avec des risques relatifs généraux, une densité f_ε connue générale et sans hypothèses à priori sur la loi de Z .

Lien avec la déconvolution

Pour tous ces modèles la première étape consiste à retrouver de l'information sur Z en utilisant l'observation U . C'est l'idée de la déconvolution. Le lien entre déconvolution et estimation non paramétrique dans un modèle avec erreurs sur les variables avait déjà été mis en évidence par Fan *et al.* (1991), Fan et Truong (1993) et Truong (1995). Par contre le lien entre déconvolution et estimation semi-paramétrique dans les modèles (1.0.4) et (1.0.5) est nouveau. Il est maintenant bien connu qu'en déconvolution deux facteurs interviennent pour la vitesse d'estimation : la régularité de la fonction à estimer et la régularité du bruit. Les vitesses les plus lentes sont ainsi obtenues pour les densités f_ε les plus régulières. Par

exemple, si on cherche à estimer f_Z de régularité "classique" quand ε est gaussienne, la vitesse (minimax) est d'ordre logarithmique. Ce résultat est souvent considéré comme décourageant pour l'utilisation de la déconvolution même dans d'autres contextes.

Outre les solutions que mes travaux de recherche apportent dans des modèles réputés délicats, mes travaux montrent que les outils de la déconvolution peuvent être utilisés pour l'estimation de paramètre de \mathbb{R}^d , souvent à la vitesse \sqrt{n} . Quand cette vitesse n'est pas atteinte, les vitesses obtenues sont néanmoins bien plus rapides que les vitesses usuelles de la déconvolution.

L'une des raisons de ce phénomène est la suivante. Toutes les méthodes d'estimation présentées ici sont liées à l'estimation de fonctionnelles linéaires intégrales de f_Z de la forme $\mathbb{E}(\Phi(Z)) = \langle \Phi, f_Z \rangle$. La vitesse d'estimation pour ces fonctionnelles linéaires va être donnée par la régularité de Φ comparée à celle de f_ε . Le point essentiel qui est que si la fonction Φ est plus régulière que f_ε alors ces fonctionnelles s'estiment à la vitesse paramétrique \sqrt{n} et il en est alors de même pour l'estimation du paramètre θ^0 .

Chapitre 2

Déconvolution

2.1 Estimation adaptative de la densité par déconvolution : cadre i.i.d.

Cette partie présente les deux articles écrits en collaboration avec Fabienne Comte et Yves Rozenholc, Comte *et al.* (2006, 2007). Le premier, plus théorique, présente la procédure d'estimation et les propriétés de l'estimateur. Le deuxième est un travail de simulation et d'illustration des propriétés pratiques de l'estimateur construit dans Comte *et al.* (2006). Les programmes et boîtes à outils *Matlab* sont disponibles sur la page web de Yves Rozenholc : <http://www.math-info.univ-paris5.fr/rozen/>.

2.1.1 Introduction

Considérons que l'on cherche à estimer la densité f_Z de Z dans le modèle

$$U = Z + \varepsilon, \quad (2.1.1)$$

où Z et ε sont des variables indépendantes, avec ε admettant une densité connue f_ε . On dispose de n observations U_1, \dots, U_n , i.i.d., copies de U . Dans ce modèle on cherche à estimer la densité inconnue f_Z , sans connaissance a priori sur sa régularité, en utilisant les observations U_1, \dots, U_n et en utilisant la connaissance de f_ε .

Quand on cherche à estimer la densité f_Z par déconvolution, deux facteurs influencent la vitesse de convergence. Le premier facteur est la régularité de la densité à estimer f_Z . Le second facteur est la régularité de la densité de l'erreur f_ε , avec les vitesses de convergence les plus lentes obtenues pour les densités f_ε les plus régulières. Ceci vient de l'indépendance entre Z et ε qui implique que la densité de l'observation U s'écrit comme la convolution $f_U = f_Z * f_\varepsilon$. Or l'opérateur de convolution est régularisant. Ainsi plus f_ε est régulière, plus f_U est régulière, et f_Z devient donc d'autant plus difficile à estimer. Par exemple si ε est gaussienne, alors quelque soit f_Z (même une masse de dirac), f_U admet un prolongement analytique sur tout le domaine complexe. Cela se traduit sur la vitesse d'estimation. Par exemple, si f_ε est la densité gaussienne, et f_Z est de régularité standard, de type Sobolev, avec une transformée de Fourier à décroissance polynomiale, alors la vitesse optimale (au sens minimax) pour l'estimation de f_Z est connue et d'ordre logarithmique (voir Fan (1991)). Dans ce contexte deux types de lois d'erreurs sont classiquement considérées : les densités d'erreur "ordinary smooth", avec une décroissance polynomiale de leur transformée de Fourier et les densités d'erreur dites "super smooth", avec une décroissance exponentielle de leur transformée de Fourier.

Les premiers travaux en déconvolution considèrent l'estimation non adaptative de f_Z par noyau de déconvolution quand f_Z est de régularité "ordinary smooth" connue et que f_ε est "ordinary smooth" ou "supersmooth". On peut citer entre autres Carroll et Hall (1988), Devroye (1992), Fan (1991), Liu et Taylor

(1989), Zhang (1990), et plus récemment Koo (1999), ou Cator (2001). Parmi les articles récents qui ont apporté un éclairage nouveau sur la déconvolution, on peut citer l'article de Pensky et Vidakovic (1999). Ils ont tout d'abord mis en évidence que les vitesses de convergence étaient en fait beaucoup plus rapides quand la densité à estimer est elle même plus régulière, de type "super smooth". Un autre aspect de l'article de Pensky et Vidakovic (1999) est qu'il est le premier à considérer l'estimation adaptative. Leur estimateur construit en utilisant des ondelettes atteint la vitesse minimax dans un certain nombre de cas. Par contre il n'atteint pas la vitesse minimax quand f_Z et f_ε sont toutes les deux de type "super smooth". À ce sujet, je renvoie à Butucea (2004), Butucea et Tsybakov (2007a, 2007b) qui ont étudié les vitesses minimax pour l'estimation de f_Z "super smooth" et proposent également un estimateur adaptatif dans un cas particulier.

D'autres travaux considèrent la déconvolution par noyau avec choix automatique de paramètre de lissage comme ceux de Hesse (1999), Delaigle et Hall (2006), Delaigle et Gijbels (2002, 2004b, 2004a, 2006). Ces derniers n'établissent pas d'inégalité de type oracle permettant d'étudier l'optimalité des estimateurs construits.

Notre objectif est donc triple : proposer un estimateur adaptatif de f_Z , dont la construction ne requiert pas la connaissance de la régularité de f_Z , établir une inégalité oracle qui montre que l'estimateur a la vitesse de convergence optimale au sens minimax et ce aussi bien pour les deux types de lois d'erreurs que pour f_Z dans une large classe de régularité contenant les deux classes de régularité "ordinary" et "super smooth".

Notre estimateur est construit par sélection de modèles, et plus précisément par minimisation d'un contraste pénalisé. Cet estimateur \tilde{f}_Z est adaptatif au sens où sa construction ne nécessite pas de connaître la régularité de f_Z . Nous établissons une majoration non asymptotique du risque quadratique intégré de \tilde{f}_Z , qui assure un compromis automatique entre le biais et une pénalité, qui ne dépend que des observations et de f_ε . Sa vitesse de convergence est la vitesse minimax dans tous les cas où la vitesse minimax optimale est connue, c'est-à-dire dans la plupart des cas. Dans les autres cas, il atteint la vitesse du meilleur estimateur de la collection d'estimateurs, à un facteur logarithmique négligeable près. En particulier notre estimateur possède une meilleure vitesse que celui de Pensky et Vidakovic (1999) ; dans le cas où la densité à estimer f_Z et la densité de l'erreur f_ε sont toutes les deux "super smooth".

2.1.2 Estimateurs non adaptatifs

Le point de départ pour l'estimation de f_Z dans le modèle $U = Z + \varepsilon$ est le critère des moindres carrés en remarquant que

$$f_Z = \arg \min \| t - f_Z \|_2^2 = \arg \min [\| t \|_2^2 - 2 \langle t, f_Z \rangle].$$

Quand les $(Z_i)_{1 \leq i \leq n}$ sont observées, le critère d'estimation est le critère des moindres carrés classique

$$\tilde{\gamma}_n(t) = \| t \|_2^2 - \frac{1}{n} \sum_{i=1}^n t(Z_i), \text{ avec } \mathbb{E}(\tilde{\gamma}_n(t)) = \| t - f_Z \|_2^2 + \| f_Z \|_2^2.$$

Les $(Z_i)_{1 \leq i \leq n}$ n'étant pas observées nous avons proposé un critère empirique, qui estime $\| t \|_2^2 - 2 \langle t, f_Z \rangle$ en utilisant les observations U_1, \dots, U_n . En partant de l'écriture

$$\langle t, f_Z \rangle = \frac{1}{2\pi} \langle t^*, f_Z^* \rangle = \frac{1}{2\pi} \left\langle t^*, \frac{f_U^*}{f_\varepsilon^*} \right\rangle,$$

et en estimant la transformée de Fourier des observations $f_U^*(x)$ par son estimateur empirique naturel $n^{-1} \sum_{k=1}^n e^{ixU_k}$, on propose le contraste suivant :

$$\gamma_n(t) = \frac{1}{n} \sum_{k=1}^n [\| t \|^2 - 2u_t^*(U_k)], \text{ avec } u_t(x) = \frac{1}{2\pi} \left(\frac{t^*(-x)}{f_\varepsilon^*(x)} \right) \text{ et } \mathbb{E}[u_t^*(U)] = \langle t, f_Z \rangle.$$

Associé à ce critère d'estimation, on considère S_m engendré par $\varphi(x) = \sin(\pi x)/(\pi x)$, dont la transformée de Fourier est à support compact. Notons $\varphi_{m,j}(x) = \sqrt{m}\varphi(mx - j)$, $m \in \mathcal{M}_n = \{1, \dots, m_n\}$. Les fonctions $\{\varphi_{m,j}\}_{j \in \mathbb{Z}}$ forment une base orthonormale de l'espace S_m des fonctions de carré intégrable ayant une transformée de Fourier à support inclus dans $[-\pi m, \pi m]$. Ainsi

$$S_m = \left\{ \sum_{j \in \mathbb{Z}} a_{m,j} \varphi_{m,j}, a_{m,j} \in \mathbb{R} \right\} = \{f \in \mathbb{L}_2(\mathbb{R}), \text{ avec } \text{supp}(f^*) \subset [-\pi m, \pi m]\}.$$

Notons $(S_m^{(n)})_{m \in \mathcal{M}_n}$ la collection de modèles avec $S_m^{(n)} = \left\{ \sum_{|j| \leq k_n} a_{m,j} \varphi_{m,j}, a_{m,j} \in \mathbb{R} \right\}$ où k_n est un entier à choisir.

Pour un m arbitraire, un estimateur de f_Z appartenant à $S_m^{(n)}$ est défini par

$$\hat{f}_m^{(n)} = \arg \min_{t \in S_m^{(n)}} \gamma_n(t) = \sum_{|j| \leq K_n} \hat{a}_{m,j} \varphi_{m,j} \quad (2.1.2)$$

où

$$\hat{a}_{m,j} = n^{-1} \sum_{i=1}^n u_{\varphi_{m,j}}^*(U_i) \text{ et } \mathbb{E}(\hat{a}_{m,j}) = \langle f_Z, \varphi_{m,j} \rangle = a_{m,j}.$$

Pour cette construction, on suppose que f_ε satisfait

Il existe $\kappa'_0 \geq \kappa_0 > 0$ et $\gamma \geq 0, \mu \geq 0, \delta \geq 0$ (avec $\gamma > 0$ si $\delta = 0$) tels que

$$\kappa_0(x^2 + 1)^{-\gamma/2} \exp\{-\mu|x|^\delta\} \leq |f_\varepsilon^*(x)| \leq \kappa'_0(x^2 + 1)^{-\gamma/2} \exp\{-\mu|x|^\delta\}. \quad (\mathbf{A}_1^\varepsilon)$$

$$\text{Pour tout } x \in \mathbb{R}, f_\varepsilon^*(x) \neq 0. \quad (\mathbf{A}_2^\varepsilon)$$

Quand $\delta = 0$ dans $(\mathbf{A}_1^\varepsilon)$, les erreurs sont appelées "ordinary smooth". Quand $\delta > 0$ (avec la convention que $\delta > 0$ si et seulement si $\mu > 0$), les erreurs sont dites "super smooth". En effet les densités qui satisfont $(\mathbf{A}_1^\varepsilon)$ avec $\delta > 0$ et $\mu > 0$ sont infiniment dérivables.

Ainsi, nous disposons d'une collection d'estimateurs $(\hat{f}_m^{(n)})_m$ associée à la collection de modèles $(S_m^{(n)})_m$.

2.1.3 Estimateur adaptatif

Notre objectif est de trouver le meilleur modèle \hat{m} dans \mathcal{M}_n , sans connaissance à priori sur la régularité de f_Z , tel que $\hat{f}_{\hat{m}}$ ait un risque du même ordre que $\hat{f}_{\check{m}_{f_Z}^{(n)}}$ avec \check{m}_{f_Z} satisfaisant

$$\check{m}_{f_Z} = \arg \min_{m \in \mathcal{M}_n} \mathbb{E} \|\hat{f}_m^{(n)} - f_Z\|_2^2. \quad (2.1.3)$$

Ce modèle \hat{m} va être construit en minimisant le critère pénalisé

$$\tilde{f}_Z = \hat{f}_{\hat{m}}^{(n)} \text{ avec } \hat{m} = \arg \min_{m \in \mathcal{M}_n} \left[\gamma_n(\hat{f}_m^{(n)}) + \text{pen}(m) \right], \quad (2.1.4)$$

où la pénalité $\text{pen}(m)$ est une fonction des observations qui dépend de f_ε au travers de la quantité

$$\Delta(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|^2} dx. \quad (2.1.5)$$

Plus précisément, pour $\varpi > 1$ et $a > 1$, $\text{pen}(m)$ est définie par

$$\text{pen}(m) = \begin{cases} 4a\varpi \frac{\Delta(m)}{n} & \text{si } 0 \leq \delta < 1/3, \\ 4a\lambda(f_\varepsilon)(\pi m)^{\min((3\delta/2-1/2)+, \delta)} \frac{\Delta(m)}{n} & \text{si } \delta \geq 1/3, \end{cases} \quad (2.1.6)$$

avec $\lambda(f_\varepsilon)$ une constante explicite et connue qui dépend de f_ε . Afin de s'assurer que la pénalité est toujours bornée on ne va considérer que des modèles $S_m^{(n)}$ où $1 \leq m \leq m_n$ avec

$$\pi m_n \leq \begin{cases} n^{1/(2\gamma+1)} & \text{si } \delta = 0 \\ \left[\frac{\ln(n)}{2\mu} + \frac{2\gamma + 1 - \delta}{2\delta\mu} \ln \left(\frac{\ln(n)}{2\mu} \right) \right]^{1/\delta} & \text{si } \delta > 0. \end{cases} \quad (2.1.7)$$

2.1.4 Borne de risque pour les estimateurs non adaptatifs

Nous commençons par donner une majoration du risque quadratique intégré des estimateurs de la collection, sans nous préoccuper pour l'instant de l'adaptation. Cette première étape justifie le choix de la collection de modèles engendré par $\phi(x) = \sin(x)/(\pi x)$, en établissant l'optimalité, au sens minimax, du meilleur estimateur de la collection. Supposons que f_Z satisfait

$$\text{La densité } f_Z \in \mathbb{L}_2(\mathbb{R}) \text{ et il existe } M_2 > 0, \text{ telle que } \int x^2 f_Z^2(x) dx < M_2 < \infty. \quad (\mathbf{A}_3^Z)$$

Proposition 2.1.1 *Sous les hypothèses $(\mathbf{A}_2^\varepsilon)$ et (\mathbf{A}_3^Z) ,*

$$\mathbb{E} \|f_Z - \hat{f}_m^{(n)}\|_2^2 \leq \|f_Z - f_m\|_2^2 + \frac{m^2(M_2 + 1)}{k_n} + \frac{2\Delta(m)}{n},$$

où f_m est la projection orthogonale de f_Z sur S_m .

Précisons la vitesse obtenue pour une classe de régularité classique, définie pour s, r, b des nombres positifs, par

$$\mathcal{S}_{s,r,b}(C_1) = \left\{ \psi \text{ tel que } \int_{-\infty}^{+\infty} |\psi^*(x)|^2 (x^2 + 1)^s \exp\{2b|x|^r\} dx \leq C_1 \right\}. \quad (2.1.8)$$

Par convention $r = 0$ si $b = 0$. Ces classes de régularité sont classiquement considérées en estimation non paramétrique, avec ou sans erreurs. Quand $r = 0$, on retrouve les classes Sobolev. Les densités dans $\mathcal{S}_{s,r,b}(C_1)$ définie par (2.1.8) avec $r > 0$ et $b > 0$ sont infiniment dérivables. Elles admettent une prolongement analytique sur une bande de largeur finie si $r = 1$ et sur tout le domaine complexe si $r = 2$. On déduit alors de la proposition 2.1.1 que sous les conditions $(\mathbf{A}_1^\varepsilon)$, $(\mathbf{A}_2^\varepsilon)$, (\mathbf{A}_3^Z) , et si $k_n \geq n^2$, le risque quadratique intégré est d'ordre

$$(m^2\pi^2 + 1)^{-s} \exp\{-2b\pi^r m^r\} + \frac{(\pi m)^{2\gamma+1-\delta} \exp\{2\mu\pi^\delta m^\delta\}}{n}. \quad (2.1.9)$$

Les choix optimaux de m et les vitesses qui en résultent sont présentés dans le tableau 2.1.4. Remarquons que le choix de k_n ne requiert pas de calcul de compromis et donc ne dépend pas de f_Z .

Le meilleur estimateur de la collection, $\hat{f}_m^{(n)}$ atteint donc la vitesse minimax dans tous les cas où des minoration ont été établies (voir Fan (1991), Butucea (2004), Butucea et Tsybakov (2007a, 2007b) pour les minoration).

2.1.5 Estimation adaptative

Théorème 2.1.1 *Supposons que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$, que f_Z satisfait (\mathbf{A}_3^Z) , et que m_n satisfait (2.1.7). Soit $\hat{f}_m^{(n)}$ défini par (2.1.2) avec $k_n \geq n^2$ et $1 \leq m \leq m_n$. Soit $\text{pen}(m)$ définie par (2.1.6). L'estimateur $\tilde{f}_Z = \hat{f}_m^{(n)}$ défini par (2.1.4) satisfait*

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m(M_2 + 1)}{n} \right] + \frac{\kappa_a}{n},$$

où C_a est une constante qui dépend de a et \overline{C} et κ_a sont des constantes qui dépendent de f_ε , ϖ et a .

TAB. 2.1 – Choix \check{m} et vitesses correspondantes sous $(\mathbf{A}_1^\varepsilon)$ - $(\mathbf{A}_2^\varepsilon)$ et (2.1.8).

		f_ε	
		$\delta = 0$ ordinary smooth	$\delta > 0$ supersmooth
$r = 0$ Sobolev(s)	$\pi\check{m} = O(n^{1/(2s+2\gamma+1)})$ rate = $O(n^{-2s/(2s+2\gamma+1)})$ <i>vitesse minimax</i>		$\pi\check{m} = [\ln(n)/(2\mu + 1)]^{1/\delta}$ rate = $O((\ln(n))^{-2s/\delta})$ <i>vitesse minimax</i>
f_Z $r > 0$ \mathcal{C}^∞	$\pi\check{m} = [\ln(n)/2b]^{1/r}$ rate = $O\left(\frac{\ln(n)^{(2\gamma+1)/r}}{n}\right)$ <i>minimax rate</i>		\check{m} solution de $\check{m}^{2s+2\gamma+1-r} \exp\{2\mu(\pi\check{m})^\delta + 2b\pi^r \check{m}^r\}$ = $O(n)$ <i>vitesse minimax si $r < \delta$ et $s = 0$</i> <i>minorations non établies dans les autres cas</i>

Deux cas peuvent se produire suivant l'ordre de grandeur de la pénalité $\text{pen}(m)$. Soit cette pénalité est du même ordre de grandeur que la variance $\Delta(m)/n$ et dans ce cas \tilde{f}_Z atteint la vitesse optimale donnée par l'inégalité (2.1.3). Soit la pénalité $\text{pen}(m)$ est légèrement supérieure à la variance $\Delta(m)/n$. Dans ce cas la vitesse de convergence de l'estimateur adaptatif est donnée par le compromis entre le biais et la pénalité. Si dans ce compromis, c'est le biais qui reste le terme dominant, alors la vitesse de convergence de \tilde{f}_Z reste la vitesse optimale. Si dans ce compromis, c'est la pénalité qui domine alors le risque quadratique intégré est d'ordre légèrement supérieur au risque quadratique du meilleur estimateur donné par l'inégalité oracle (2.1.3). Dans ce cas la perte due à l'adaptation est d'ordre logarithmique. Ceci ne se produit que quand $r \geq \delta > 1/3$.

La vitesse de convergence de \tilde{f}_Z est la vitesse minimax dans les cas où les minorations ont été établies. Sa vitesse de convergence est en particulier plus rapide que celle de l'estimateur adaptatif de Pensky et Vidakovic (1999) quand $(r > 0, \delta > 0)$. En effet ce dernier n'atteint la vitesse minimax que dans les cas $(\delta = 0, r = 0)$, $(\delta = 0, r > 0)$, et $(\delta > 0, r = 0)$, mais pas quand $(r > 0, \delta > 0)$.

2.1.6 Étude par simulation

Je présente ici une partie de l'étude par simulation que nous avons effectuée. Cette partie se trouve détaillée dans l'article Comte *et al.* (2006). Je renvoie à l'article Comte *et al.* (2007) pour une étude par simulation beaucoup plus détaillée et approfondie. Les programmes *Matlab* correspondants sont disponibles sur la page web de Yves Rozenholc à l'adresse : <http://www.math-info.univ-paris5.fr/rozen/>.

L'erreur intégrée $\text{ISE}(\hat{f}_{\hat{m}}^{(n)}) = \|\hat{f}_{\hat{m}}^{(n)} - f\|_2^2$ est calculée en utilisant des approximations d'intégrales. Le risque quadratique intégré noté MISE (Mean Integrated Squared Error), avec

$$\text{MISE}(\hat{f}_{\hat{m}}^{(n)}) = \mathbb{E}\|\hat{f}_{\hat{m}}^{(n)} - f\|_2^2$$

est calculé comme la moyenne sur 500 simulations des $\|\hat{f}_m^{(n)} - f\|_2^2$. Nous illustrons notre méthode sur plusieurs exemples de densité à estimer et pour deux types de densité d'erreur f_ε .

Lois d'erreurs et pénalités associées

• **Cas 1 : ε Laplace (ou Double exponentielle).** Dans ce cas, $f_\varepsilon(x) = e^{-\sqrt{2}|x|}/\sqrt{2}$, et $f_\varepsilon^*(x) = (1 + x^2/2)^{-1}$. La pénalité proposée est la suivante

$$\text{pen}(m) = \frac{6\pi m}{n} \left(1 + \frac{(\ln(m))^{2.5}}{m} + \frac{\pi^2}{3}\sigma^2 m^2 + \frac{\pi^4}{20}\sigma^4 m^4 \right).$$

• **Cas 2 : ε Gaussienne.** Dans ce cas $f_\varepsilon(x) = 1/\sqrt{2\pi}e^{-x^2/2}$, et $f_\varepsilon^*(x) = e^{-x^2/2}$ et la pénalité est

$$\text{pen}(m) = \frac{6\pi m}{n} \left(1 + \frac{(\ln(m))^{2.5}}{m} + \frac{\pi^2\sigma^2 m^2}{3} \right) \left(\int_0^\pi \exp(\sigma^2 m^2 x^2) dx / \pi \right),$$

où l'intégrale est approchée numériquement.

Densités f_Z à estimer

On considère les densités f_Z suivantes :

(a) Chi2(3)-type distribution, $Z = 1/\sqrt{6}U$, $U \sim \chi^2(3)$ où $U \sim \Gamma(\frac{3}{2}, \frac{1}{2})$.

(b) Densité de la loi de Laplace.

(c) Mélange de lois Gamma, $Z = 1/\sqrt{5.48}W$ avec $W \sim 0.4\Gamma(5, 1) + 0.6\Gamma(13, 1)$.

(d) Loi de Cauchy, $f_Z(x) = (1/\pi)(1/(1+x^2))$, $f_Z^*(x) = e^{-|x|}$.

(e) Gaussienne, $Z \sim \mathcal{N}(0, \sigma^2)$ avec $\sigma = 1$.

(f) Mélange de lois gaussiennes : $Z \sim \sqrt{2}V$ avec $V \sim 0.5\mathcal{N}(-3, 1) + 0.5\mathcal{N}(2, 1)$.

On a choisi $k_n = 2^8$. À part pour la densité de la loi de Cauchy qui a une variance infinie, les variables Z sont de variance 1. De telle sorte que $s2n = 1/\sigma^2$ représente le ratio de la variance de Z sur celle du bruit ε et que une grande valeur de $s2n$ revient à considérer qu'il y a très peu de bruit.

Résultats

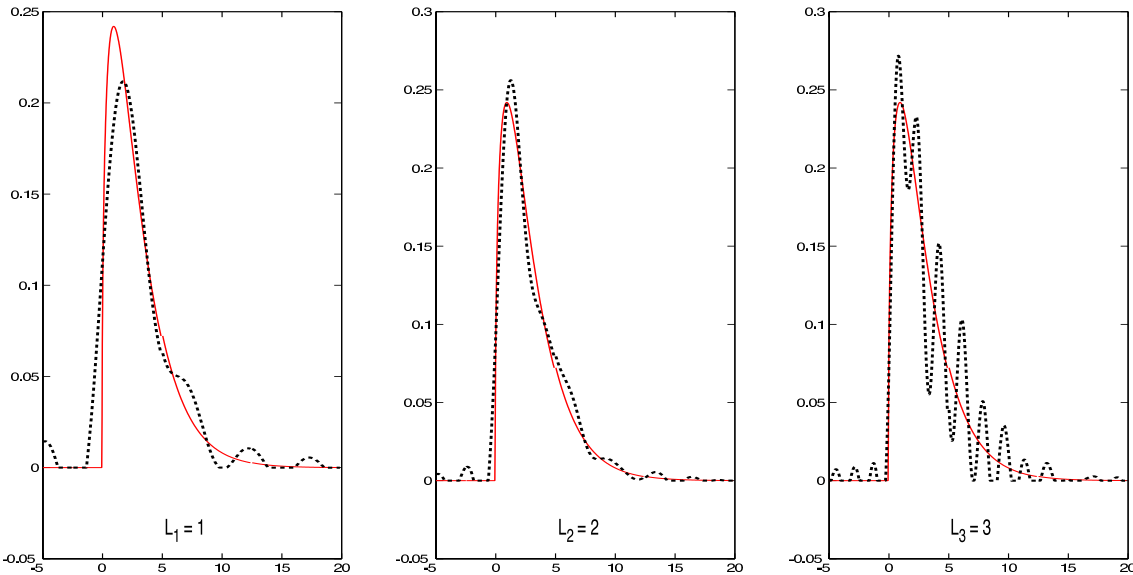


FIG. 2.1 – Graphe de l'estimateur (pointillé) et de la vraie densité $\chi^2(3)$ (a) (trait plein) - ε Laplace - $n = 750$, $s2n=10$, quand $m = 1$ (gauche), $m = 2$ (milieu), $m = 3$ (droite). L'algorithme choisit $\hat{m} = 2$.

Cette figure 2.1 compare les estimateurs $\hat{f}_m^{(n)}$ obtenus avec $m = 1, 2$ et 3 , et justifie le choix $\hat{m} = 2$ de l'algorithme. Le tableau 2.2 presente les MISE pour les deux types d'erreurs, pour les différentes densités à estimer, pour différents $s2n$ et pour différentes tailles d'échantillons. Clairement le MISE est d'autant plus faible que le niveau de bruit est faible.

TAB. 2.2 – Moyenne des MISE $\times 100$ obtenue avec $N = 500$ échantillons, $n = 100, 250, 500, 1000, 2500$ et $s2n = 2, 4, 10, 100, 1000$. Densités (a) : Chi2(3), (b) : Laplace, (c) : Mélange de Gamma, (d) : Cauchy, (e) Gaussienne, (f) : Mélange de gaussiennes.

$\times 10^{-2}$		$n = 100$		$n = 250$		$n = 500$		$n = 1000$		$n = 2500$	
f_Z	$s2n$	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.	Lap.	Gaus.
(a)	2	2.02	4.15	1.39	2.37	1.18	1.72	1.06	1.36	1.03	1.12
	4	1.52	1.79	1.21	1.27	1.07	1.13	1.04	1.04	0.654	0.996
	10	1.31	1.31	1.13	1.11	1.01	1.03	0.505	0.995	0.345	0.974
	10^2	1.22	1.23	0.72	0.884	0.409	0.411	0.327	0.335	0.179	0.232
	10^3	1.22	1.21	0.651	0.638	0.391	0.382	0.293	0.298	0.157	0.157
(b)	2	3.7	10.6	2.17	5.2	1.61	3.03	1.41	2.07	1.2	1.48
	4	2.5	2.99	1.66	1.93	1.33	1.46	1.26	1.25	0.817	1.12
	10	1.9	1.97	1.43	1.42	1.35	1.22	0.723	1.12	0.441	1.06
	10^2	1.69	1.64	0.883	1.06	0.607	0.538	0.453	0.385	0.343	0.211
	10^3	1.68	1.65	0.814	0.79	0.593	0.561	0.411	0.379	0.284	0.24
(c)	2	1.32	3.96	0.547	1.88	0.292	1.01	0.148	0.533	0.06	0.224
	4	0.79	1.05	0.316	0.453	0.151	0.224	0.0815	0.116	0.0361	0.0497
	10	0.495	0.524	0.194	0.215	0.103	0.11	0.0543	0.0565	0.024	0.0246
	10^2	0.369	0.384	0.152	0.149	0.0789	0.0785	0.0409	0.0412	0.0194	0.0186
	10^3	0.364	0.353	0.149	0.15	0.0762	0.0767	0.0404	0.0406	0.0184	0.0185
(d)	2	2.72	9.09	1.22	4.26	0.645	2.3	0.353	1.25	0.158	0.513
	4	1.66	2.27	0.716	0.967	0.364	0.514	0.205	0.28	0.138	0.127
	10	1.15	1.13	0.437	0.46	0.249	0.257	0.215	0.142	0.219	0.0764
	10^2	0.815	0.783	0.373	0.351	0.351	0.271	0.206	0.201	0.147	0.0962
	10^3	0.783	0.78	0.366	0.355	0.34	0.331	0.189	0.189	0.121	0.118
(e)	2	2.74	9.21	1.1	4.08	0.605	2.14	0.296	1.06	0.143	0.446
	4	1.59	2.23	0.591	0.878	0.362	0.457	0.229	0.227	0.463	0.0894
	10	0.885	1.02	0.397	0.42	0.372	0.21	0.515	0.112	0.229	0.046
	10^2	0.711	0.713	0.565	0.432	0.396	0.394	0.279	0.195	0.171	0.15
	10^3	0.739	0.705	0.606	0.592	0.352	0.355	0.259	0.246	0.167	0.145
(f)	2	2.97	9.98	1.26	4.45	0.693	2.31	0.328	1.26	0.132	0.509
	4	1.73	2.37	0.709	1.02	0.375	0.478	0.185	0.257	0.0751	0.105
	10	1.14	1.21	0.463	0.466	0.237	0.242	0.118	0.122	0.0468	0.0515
	10^2	0.851	0.817	0.359	0.352	0.166	0.167	0.0866	0.0867	0.034	0.0351
	10^3	0.823	0.828	0.344	0.327	0.169	0.163	0.0845	0.0839	0.0334	0.0336

Le tableau 2.3 compare les performances de l'estimateur adaptatif \tilde{f}_Z avec celui présenté dans Delaigle et Gijbels (2004b). Cette comparaison est faite pour les densités (a), (c), (e) et (f) correspondantes aux densités #2, #6, #1 et #3 dans l'article de Delaigle et Gijbels (2004b).

TAB. 2.3 – Médiane des ISE obtenues par Delaigle et Gijbels (2004) avec un noyau de déconvolution et avec 4 stratégies différentes de choix de paramètre de lissage. Comparaison avec \tilde{f}_Z , estimateur par projection pénalisé. On donne la médiane et la moyenne des risques. .

density f_Z	method	$n = 100$		$n = 250$	
		ε Lap.	ε Gaus.	ε Lap.	ε Gaus.
(a) or #2 $\chi^2(3)$ ($s2n=4$)	DG, lower median	0.015	0.018	—	—
	DG, higher median	0.018	0.022	—	—
	Proj. : median	0.014	0.016	—	—
	Proj. : mean	0.015	0.018	—	—
(c) or #6 Mix.Gamma ($s2n=10$)	DG, lower median	—	—	0.0021	0.0023
	DG, higher median	—	—	0.0024	0.0026
	Proj. : median	—	—	0.0017	0.0020
	Proj., mean	—	—	0.0019	0.0021
(e) or #1 Gauss ($s2n=4$)	DG, lower median	0.0071	0.0080	0.0041	0.0051
	DG, higher median	0.011	0.012	0.0059	0.0072
	Proj. : median	0.012	0.017	0.0049	0.0066
	Proj. : mean	0.016	0.022	0.0059	0.0088
(f) or #3 Mix.Gauss ($s2n=4$)	DG, lower median	0.018	0.027	0.011	0.020
	DG, higher median	0.031	0.034	0.023	0.028
	Proj. : median	0.016	0.022	0.0063	0.0088
	Proj. : mean	0.017	0.024	0.0071	0.010

2.2 Déconvolution adaptative pour une suite faiblement dépendante

Cette partie fait l’objet de l’article Comte *et al.* (2008a) écrit en collaboration avec Fabienne Comte et Jérôme Dedecker.

On considère le modèle 2.1.1, avec les mêmes notations que dans la partie 2.1, mais sous des conditions de dépendance. Plus précisément, on souhaite estimer la densité f_Z de Z dans le modèle 2.1.1, quand les variables aléatoires Z_1, \dots, Z_n sont identiquement distribuées et satisfont les conditions de dépendance faible introduites par Dedecker et Prieur (2005). Les erreurs $\varepsilon_1, \dots, \varepsilon_n$ sont indépendantes et identiquement distribuées et on suppose également que la suite $(\varepsilon_i)_{1 \leq i \leq n}$ est indépendante de la suite $(Z_i)_{1 \leq i \leq n}$. Des résultats d’estimation de densité par déconvolution ont déjà été obtenus sous des conditions plus restrictives de mélange fort au sens de Rosenblatt (1956) (voir notamment Masry (1991, 1993c, 1993a, 1994, 2003)).

Quand on s’intéresse à l’estimation de la densité f_Z dans un contexte classique sans erreur ($(Z_i)_{1 \leq i \leq n}$ observées) et dépendant (voir par exemple Viennet (1997) ou Dedecker et Prieur (2005)), la vitesse d’estimation est la même que dans le cadre i.i.d. et les coefficients de mélange apparaissent dans la constante devant le terme de variance. La vitesse de convergence n’est certes pas changée, mais la constante peut être très différente.

De même, quand on s’intéresse à l’estimation adaptative de la densité f_Z dans un contexte classique sans erreur ($(Z_i)_{1 \leq i \leq n}$ observées) et dépendant, les coefficients de mélange apparaissent lors de la procédure d’estimation adaptative. Par exemple, le niveau de seuillage proposé dans Tribouley et Viennet (1998) aussi bien que la pénalité dans Comte et Merlevède (2002) dépendent tous les deux de la somme des coefficients de β -mélange (forts) des $(Z_i)_{i \geq 1}$. Or cette somme des coefficients de mélange est connue pour être difficilement estimable.

Dans ce contexte, notre objectif est d’estimer par déconvolution la densité f_Z de façon adaptative

quand les $(Z_i)_{1 \leq i \leq n}$ sont faiblement dépendantes. L'apport de notre travail est donc multiple. Tout d'abord nous envisageons des conditions de dépendance faible dans un contexte de déconvolution adaptative ce qui est complètement nouveau, à la fois du point de vue de la déconvolution, mais aussi du point de vue de l'adaptation. Ceci permet d'envisager des exemples de modèles plus généraux que les modèles mélangeants au sens du β -mélange fort. Ensuite, par une procédure d'estimation adaptative de f_Z analogue à celle décrite dans la partie 2.1, et avec une pénalité qui ne dépend pas des coefficients de dépendance, nous établissons une inégalité de type oracle qui montre que le risque quadratique intégré de notre estimateur adaptatif est le même que celui obtenu dans le cadre i.i.d. (avec les propriétés d'optimalité mentionnées dans la partie 2.1) plus un terme additionnel négligeable. Ainsi, contrairement au cas classique sans erreur, les coefficients de dépendance n'interviennent pas dans la procédure adaptative d'estimation et n'interviennent dans la vitesse qu'au travers d'un terme supplémentaire négligeable.

2.2.1 Mesures de dépendance

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace de probabilité. Soit Y une variable aléatoire à valeur dans un espace de Banach $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$. Notons $\Lambda_{\kappa}(\mathbb{B})$ l'ensemble des fonctions κ -Lipschitziennes de $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$ dans \mathbb{R} , c'est-à-dire telles que $|f(x) - f(y)| \leq \kappa \|x - y\|_{\mathbb{B}}$. Soit \mathcal{M} une σ -algèbre de \mathcal{A} . Soit $\mathbb{P}_{Y|\mathcal{M}}$ la distribution conditionnelle de Y sachant \mathcal{M} , \mathbb{P}_Y la distribution de Y , et soit $\mathcal{B}(\mathbb{B})$ la σ -algèbre borélienne de $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$.

Définissons

$$\beta(\mathcal{M}, \sigma(Y)) = \mathbb{E} \left(\sup_{A \in \mathcal{B}(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(A) - \mathbb{P}_Y(A)| \right),$$

et si $\mathbb{E}(\|Y\|) < \infty$, $\tau(\mathcal{M}, Y) = \mathbb{E} \left(\sup_{f \in \Lambda_1(\mathbb{B})} |\mathbb{P}_{Y|\mathcal{M}}(f) - \mathbb{P}_Y(f)| \right).$

Les coefficients $\beta(\mathcal{M}, \sigma(Y))$ sont les coefficients de mélange classique introduits par Rozanov et Volkonskii (1960). Le coefficient $\tau(\mathcal{M}, Y)$ a été introduit par Dedecker et Prieur (2005).

Soit $\mathbf{X} = (X_i)_{i \geq 1}$ une suite strictement stationnaire de variables aléatoires réelles. Pour tout $k \geq 0$, les coefficients $\beta_{\mathbf{X},1}(k)$ et $\tau_{\mathbf{X},1}(k)$ sont définis par

$$\beta_{\mathbf{X},1}(k) = \beta(\sigma(X_1), \sigma(X_{1+k})), \quad (2.2.1)$$

$$\text{et si } \mathbb{E}(|X_1|) < \infty, \quad \tau_{\mathbf{X},1}(k) = \tau(\sigma(X_1), X_{1+k}). \quad (2.2.2)$$

Sur \mathbb{R}^l , on considère la norme $\|x - y\|_{\mathbb{R}^l} = l^{-1}(|x_1 - y_1| + \dots + |x_l - y_l|)$. Soit $\mathcal{M}_i = \sigma(X_k, 1 \leq k \leq i)$. Les coefficients $\beta_{\mathbf{X},\infty}(k)$ et $\tau_{\mathbf{X},\infty}(k)$ sont définis par

$$\beta_{\mathbf{X},\infty}(k) = \sup_{i \geq 1, l \geq 1} \sup \{ \beta(\mathcal{M}_i, \sigma(X_{i_1}, \dots, X_{i_l})), i + k \leq i_1 < \dots < i_l \},$$

et si $\mathbb{E}(|X_1|) < \infty$, $\tau_{\mathbf{X},\infty}(k) = \sup_{i \geq 1, l \geq 1} \sup \{ \tau(\mathcal{M}_i, (X_{i_1}, \dots, X_{i_l})), i + k \leq i_1 < \dots < i_l \}.$

2.2.2 Exemples de suites faiblement dépendantes

Les exemples de suites β -mélangeants sont maintenant bien connus (voir les livres de Doukhan (1994) et Bradley (2007)). Néanmoins de nombreux processus, même markoviens, ne sont pas β -mélangeants. Par exemple considérons $(\epsilon_i)_{i \geq 1}$ une suite i.i.d. telle que $\mathbb{P}(\epsilon_1 = 1) = \mathbb{P}(\epsilon_1 = 0) = 1/2$. Alors la solution stationnaire $(Z_i)_{i \geq 0}$ de l'équation $Z_n = 0.5(Z_{n-1} + \epsilon_n)$ où Z_0 est indépendante de $(\epsilon_i)_{i \geq 1}$ n'est pas β -mélangeante. Par contre dans cet exemple particulier $\tau_{\mathbf{X},\infty}(k) \leq 2^{-k}$. Je renvoie à Comte *et al* (2008a) et surtout à Dedecker et Prieur (2005) pour une présentation de nombreux exemples de suites τ -dépendantes mais non β -mélangeantes.

2.2.3 Estimateurs

Les estimateurs sont les mêmes que ceux présentés dans la partie 2.1 consacrée au cadre i.i.d.. Plus précisément, la collection d'estimateurs (non adaptatifs) est celle définie par (2.1.2) et l'estimateur adaptatif est celui défini par (2.1.4) avec une pénalité qui sera précisée.

2.2.4 Borne de risque pour les estimateurs non adaptatifs

La proposition suivante montre que le risque quadratique intégré d'un estimateur de la collection est du même ordre que dans le cas i.i.d. avec un terme supplémentaire dû à la dépendance.

Proposition 2.2.1 *Sous les conditions $(\mathbf{A}_2^\varepsilon)$ et (\mathbf{A}_3^Z) ,*

$$\mathbb{E}\|f_Z - \hat{f}_m^{(n)}\|_2^2 \leq \|f_Z - f_m\|_2^2 + \frac{m^2(M_2 + 1)}{k_n} + \frac{2\Delta(m)}{n} + \frac{2R_{n,m}}{n},$$

où $\Delta(m)$ est défini par (2.1.5) et

$$R_{n,m} = \frac{1}{\pi} \sum_{k=2}^n \int_{-\pi m}^{\pi m} |\text{Cov}(e^{ixZ_1}, e^{ixZ_k})| dx. \quad (2.2.3)$$

De plus, $R_{n,m} \leq \min(R_{n,m,\beta}, R_{n,m,\tau})$, où $R_{n,m,\beta} = 4m \sum_{k=1}^{n-1} \beta_{\mathbf{Z},1}(k)$ et $R_{n,m,\tau} = \pi m^2 \sum_{k=1}^{n-1} \tau_{\mathbf{Z},1}(k)$.

Si $\sum_{k>0} \beta_{\mathbf{Z},1}(k) < \infty$, et $k_n > n^2$, le risque est du même ordre que dans le cas i.i.d., borné par l'inégalité (2.1.9). Si $\sum_{k>0} \tau_{\mathbf{Z},1}(k) < \infty$, $k_n > n^2$ et si $\gamma > 1/2$ quand $\delta = 0$ dans $(\mathbf{A}_1^\varepsilon)$, le risque est du même ordre que dans le cas i.i.d., borné par l'inégalité (2.1.9).

En conclusion, sous des conditions de dépendance très faibles, l'estimateur non pénalisé a le même risque que l'estimateur non pénalisé dans le cas i.i.d. plus un terme supplémentaire négligeable. Ceci est à relier aux résultats connus pour l'estimation d'une densité dans un cadre classique dépendant (voir par exemple Viennet (1997) ou Dedecker et Prieur (2005)). En effet, quand les $(Z_i)_{1 \leq i \leq n}$ sont observées (c'est-à-dire $\varepsilon_i = 0$), les coefficients de mélanges des $(Z_i)_{i \geq 1}$ interviennent devant le terme de variance dominant. Ici ils n'interviennent que dans un terme additionnel négligeable.

2.2.5 Bornes de risque pour l'estimateur adaptatif

Nous établissons une première majoration du risque de l'estimateur adaptatif \tilde{f}_Z valable sous des conditions de dépendance faibles.

Théorème 2.2.1 *Supposons que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$, que f_Z satisfait (\mathbf{A}_3^Z) , et que m_n satisfait (2.1.7). Soit $\hat{f}_m^{(n)}$ défini par (2.1.2) avec $k_n \geq n^2$ et $1 \leq m \leq m_n$. Soit $\text{pen}(m)$ définie par (2.1.6). L'estimateur $\tilde{f}_Z = \hat{f}_m^{(n)}$ défini par (2.1.4) satisfait*

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m(M_2 + 1)}{n} \right] + \frac{\overline{C}(R_{n,m_n} + m_n)}{n},$$

où $R_{n,m}$ est défini par (2.2.3), C_a est une constante qui dépend de a , et \overline{C} est une constante qui dépend de f_ε , ϖ et a .

Comparons le risque \tilde{f}_Z obtenu ici avec celui obtenu dans le théorème 2.1.1 dans le cadre i.i.d.. Le terme $\inf_{m \in \{1, \dots, m_n\}} [\|f_Z - f_m\|_2^2 + \text{pen}(m) + m(M_2 + 1)/n]$ est le risque de \tilde{f}_Z quand toutes les variables sont i.i.d.. La structure de dépendance des $(Z_i)_{i \geq 1}$ amène donc un terme supplémentaire $n^{-1}(R_{n, m_n} + m_n)$ qui peut être relié aux coefficients de dépendance faible $\beta_{\mathbf{Z}, 1}(k)$ et $\tau_{\mathbf{Z}, 1}(k)$ définis dans (2.2.1)-(2.2.2) via la fin de la proposition 2.2.1.

Si ces coefficients de dépendance sont sommables et si la densité de l'erreur f_ε est "super smooth" alors le terme $n^{-1}(R_{n, m_n} + m_n)$ est négligeable et le risque de \tilde{f}_Z est celui du contexte indépendant.

Par contre si f_ε est "ordinary smooth", le terme m_n/n peut ne pas être négligeable. Dans ce cas le théorème 2.2.1 ne permet pas de retrouver la vitesse du cadre i.i.d. Pour la retrouver il est nécessaire de considérer des conditions de dépendance plus restrictives.

2.2.6 Estimation adaptative pour f_ε "super smooth".

Comme conséquence directe du théorème 2.2.1, nous avons le corollaire suivant.

Corollaire 2.2.1 *Supposons que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$ avec $\delta > 0$, que f_Z satisfait (\mathbf{A}_3^Z) , et que m_n satisfait (2.1.7). Soit $\text{pen}(m)$ définie par (2.1.6). Soit $\hat{f}_m^{(n)}$ défini par (2.1.2) avec $k_n \geq n^2$ et $1 \leq m \leq m_n$.*

1. *Si $\sum_{k>0} \beta_{\mathbf{Z}, 1}(k) < \infty$, alors $\tilde{f}_Z = \hat{f}_{\hat{m}}^{(n)}$ défini par (2.1.4) satisfait*

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m(M_2 + 1)}{n} \right] + \frac{\overline{C}(\ln(n))^{1/\delta}}{n},$$

où C_a est une constante qui dépend de a et \overline{C} est une constante qui dépend de f_ε , ϖ , a et de $\sum_{k>0} \beta_{\mathbf{Z}, 1}(k)$.

2. *Si $\sum_{k>0} \tau_{\mathbf{Z}, 1}(k) < \infty$, alors $\tilde{f}_Z = \hat{f}_{\hat{m}}^{(n)}$ définie par (2.1.4) satisfait*

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m(M_2 + 1)}{n} \right] + \frac{\overline{C}(\ln(n))^{2/\delta}}{n},$$

où C_a est une constante qui dépend de a et \overline{C} est une constante qui dépend de f_ε , a et de $\sum_{k>0} \tau_{\mathbf{Z}, 1}(k)$.

Les termes qui dépendent de des puissances de $\ln(n)$ sont négligeables par rapport aux autres. Par conséquent le risque de l'estimateur \tilde{f}_Z construit avec la même pénalité que dans le cas i.i.d. est d'ordre $\inf_{m \in \{1, \dots, m_n\}} [\|f_Z - f_m\|_2^2 + \text{pen}(m)]$, c'est-à-dire du meilleur ordre de grandeur possible, comme dans le cas i.i.d. Notons que ce théorème est établi sous des conditions de dépendance très faibles. Ce résultat est à relier aux résultats connus en estimation de densité dans un cadre classique mélangeant (sans erreurs). En effet quand les $(Z_i)_{1 \leq i \leq n}$ sont observées ($\varepsilon_i = 0$), le niveau de seuillage proposé dans Tribouley et Viennet (1998) aussi bien que la pénalité dans Comte et Merlevède (2002) dépendent tous les deux de la somme des coefficients de β -mélange (forts) des $(Z_i)_{i \geq 1}$, connue pour être difficilement estimable.

2.2.7 Estimation adaptative pour f_ε "ordinary smooth".

Nous établissons maintenant un résultat d'adaptation quand f_ε est "ordinary smooth", sous des conditions de dépendance plus restrictives que dans le théorème 2.2.1, mais permettant d'avoir une majoration du risque de \tilde{f}_Z comme dans le cas i.i.d., et toujours avec une pénalité qui ne dépend pas des coefficients de dépendance. Pour $a > 1$ et $\varpi > 1$, on définit $\text{pen}(m)$ par

$$\text{pen}(m) = 4a\varpi \frac{\Delta(m)}{n}. \quad (2.2.4)$$

Théorème 2.2.2 *Supposons que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$ avec $\delta = 0$, que f_Z satisfait (\mathbf{A}_3^Z) , et que m_n satisfait (2.1.7). Soient $\text{pen}(m)$ définie (2.2.4) et $\hat{f}_m^{(n)}$ défini par (2.1.2) avec $k_n \geq n^2$ et $1 \leq m \leq m_n$.*

1. *Si $\beta_{\mathbf{Z},\infty}(k) = O(k^{-(1+\theta)})$ pour un $\theta > (2\gamma + 3)/(2\gamma + 1)$, alors $\tilde{f}_Z = \hat{f}_m^{(n)}$ défini par (2.1.4) satisfait*

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m(M_2 + 1)}{n} \right] + \frac{\bar{C}}{n}, \quad (2.2.5)$$

où C_a est une constante qui dépend de a et \bar{C} est une constante qui dépend de f_ε , a , ϖ et $\sum_{k>0} \beta_{\mathbf{Z},\infty}(k)$.

2. *Si $\gamma > 1/2$ dans $(\mathbf{A}_1^\varepsilon)$ et $\tau_{\mathbf{Z},\infty}(k) = O(k^{-(1+\theta)})$ pour $\theta > (2\gamma + 5)/(2\gamma + 1)$, alors $\tilde{f}_Z = \hat{f}_m^{(n)}$ défini par (2.1.4) satisfait (2.2.5), où \bar{C} est une constante qui dépend de f_ε , a , ϖ et $\sum_{k>0} \tau_{\mathbf{Z},\infty}(k)$.*

Pour des erreurs "ordinary smooth", l'estimateur \tilde{f}_Z construit dans le cadre i.i.d. a le même risque dans le cadre i.i.d. que dans le contexte dépendant. Mais ceci n'est vrai que sous des conditions de dépendance légèrement plus restrictives que celles du théorème 2.2.1. En effet nous faisons maintenant intervenir les coefficients $\beta_{\mathbf{Z},\infty}(k)$ et $\tau_{\mathbf{Z},\infty}(k)$, avec des conditions sur leur décroissance.

2.3 Estimation adaptative dans un modèle ARCH général

Cette partie présente l'article Comte *et al.* (2008b), écrit en collaboration avec Fabienne Comte et Jérôme Dedecker.

2.3.1 Introduction

Considérons $((Y_t, \sigma_t))_{t \in \mathbb{N}}$ une suite strictement stationnaire de $\mathbb{R} \times \mathbb{R}^+$ satisfaisant la relation

$$Y_t = \sigma_t \eta_t \quad (2.3.6)$$

où $(\eta_t)_{t \in \mathbb{Z}}$ est une suite i.i.d. de variables centrées et de variance finie et pour tout $t \geq 0$, le vecteur aléatoire $(\sigma_i, \eta_{i-1})_{0 \leq i \leq t}$ est indépendant de la suite $(\eta_i)_{i \geq t}$.

De tels modèles sont usuellement utilisés en finance, où $(\sigma_t^2)_{t \in \mathbb{Z}}$, est la volatilité, que l'on cherche à étudier et qui n'est pas observée. Une grande variété de modèles paramétriques ont été proposés depuis le premier modèle ARCH(1) de Engle (1982). On peut notamment citer les modèles GARCH(p, q) de Bollerslev (1986), et leurs extensions décrites dans Duan (1997). Il n'est en général pas possible de calculer la densité stationnaire de σ_t , même pour ces modèles paramétriques. C'est l'une des raisons pour lesquelles il est intéressant de l'estimer, sans à priori sur sa régularité.

Nous proposons une méthode d'estimation adaptative de la densité $\ln(\sigma_t^2)$ à partir des observations $(Y_t)_{1 \leq t \leq n}$. Puis nous décrivons une méthode permettant de remonter à l'estimation adaptative de la densité de σ_t^2 .

Le modèle (2.3.6) est classiquement réécrit *via* une transformation logarithmique sous la forme

$$U_t = Z_t + \varepsilon_t, \quad (2.3.7)$$

où $U_t = \ln(Y_t^2)$, $Z_t = \ln(\sigma_t^2)$ et $\varepsilon_t = \ln(\eta_t^2)$. Dans le contexte du modèle (2.3.6), les variables Z_t et ε_t sont indépendantes pour un t fixé, mais les suites $(Z_t)_{k \geq 0}$ et $(\varepsilon_k)_{k \in \mathbb{Z}}$ ne sont pas indépendantes.

L'objectif est alors d'estimer la densité de $Z_t = \ln(\sigma_t^2)$, quand la densité f_ε de l'erreur $\varepsilon_t = \ln(\eta_t^2)$ est connue. On va en particulier chercher un estimateur adaptatif de f_Z construit sans connaissance à priori de sa régularité, construit à partir des observations $U_t = \ln(Y_t^2)$ et de la connaissance de f_ε . Puisque

Z_t et ε_t sont indépendantes pour chaque t , la densité commune f_U de l'observation U_t est le produit de convolution $f_U = f_Z * f_\varepsilon$.

Ce problème est donc lié au problème de la déconvolution que nous avons décrit dans les parties (2.1) et 2.2. Dans le cadre i.i.d. décrit dans la partie 2.1, les variables $(Z_t)_{t \geq 0}$ et $(\varepsilon_t)_{t \geq 0}$ sont i.i.d. et par conséquent les suites $(Z_t)_{t \geq 0}$ et $(\varepsilon_t)_{t \geq 0}$ sont indépendantes. Dans la partie 2.2, la suite $(Z_t)_{t \geq 0}$ n'était pas i.i.d., mais la suite $(\varepsilon_t)_{t \geq 0}$ l'était. Quand on considère le modèle (2.3.6), l'indépendance entre $(Z_t)_{t \geq 0}$ et $(\varepsilon_t)_{t \in \mathbb{Z}}$ n'est plus vérifiée, même si à t fixé on garde l'indépendance entre Z_t et ε_t .

Il existe d'autres travaux qui considèrent ce modèle. On peut citer par exemple l'article de van Es *et al.* (2005) qui étudie ce problème mais sous des conditions moins générales. Ils considèrent le même modèle avec η_t est gaussienne, la densité f_Z de Z_t est deux fois dérivable et le processus (U_t, Z_t) est α -mélangeant.

Notre travail s'inscrit dans un contexte plus général que ceux des travaux antérieurs. Tout d'abord, nous considérons le problème de l'estimation adaptative de la densité de $Z_t = \ln(\sigma_t^2)$ sans connaissance a priori de sa régularité. Ensuite, nos hypothèses de dépendance sont des hypothèses de dépendance faible nettement moins restrictives que celles qui ont été considérées auparavant. Elles permettent notamment d'étudier des modèles ARCH généraux, comme les modèles ARCH(∞) décrits dans Giraitis *et al.* (2000). Enfin, nous ne nous restreignons pas à une loi d'erreur donnée, mais considérons des lois d'erreur générales.

Avec une procédure d'estimation adaptative de f_Z analogue à celle décrite dans la partie 2.1, et avec une pénalité qui ne dépend pas des coefficients de dépendance, nous établissons une inégalité de type oracle qui montre que le risque quadratique intégré de notre estimateur adaptatif est le même que celui obtenu dans le cadre i.i.d. (avec les propriétés d'optimalité mentionnées dans la partie 2.1) plus un terme additionnel négligeable. Ainsi, comme dans le cas dépendant et contrairement au cas classique sans erreur, les coefficients de dépendance n'interviennent pas dans la procédure adaptative d'estimation et n'interviennent dans la vitesse qu'au travers d'un terme supplémentaire, négligeable sous certaines conditions ou au pire du même ordre que le terme de variance principal. La procédure d'estimation adaptative est la même dans le cas i.i.d, dans le cas dépendant et dans le cas ARCH, ce qui la rend particulièrement robuste et donc attractive. Le modèle ARCH est un modèle qui du fait de sa structure fait appel à des hypothèses de dépendance particulières qui nécessitent des outils spécifiques notamment pour les majorations des risques des estimateurs. Nous concluons notre travail par une présentation d'exemples de modèles ARCH et des conditions de dépendance qu'ils satisfont.

2.3.2 Procédure d'estimation

La procédure d'estimation est celle décrite dans la partie 2.1. On considère donc $(\hat{f}_m^{(n)})_{1 \leq m \leq m_n}$ la collection d'estimateurs définis par (2.1.2), obtenu par minimisation de contraste sur un modèle $S_m^{(n)}$. De la même façon on définit l'estimateur adaptatif de f_Z par (2.1.4) avec une pénalité du type (2.1.6).

Avant d'établir les bornes de risque, je précise les conditions de dépendance sous lesquelles les résultats sont valides, en utilisant les notations et les définitions de la partie 2.2.1.

2.3.3 Hypothèses de dépendance

Après la transformation logarithmique ($U_t = \ln(Y_t^2)$, $Z_t = \ln(\sigma_t^2)$), le processus d'intérêt est

$$(W_t)_{t \in \mathbb{Z}} = ((U_t, Z_t))_{t \in \mathbb{Z}}. \quad (2.3.8)$$

Précisons les définitions de la partie 2.2.1 pour un processus dans \mathbb{R}^2 . Soit $(W_t)_{t \geq 0}$ une suite strictement stationnaire à valeurs dans \mathbb{R}^2 . Considérons la norme $\|x - y\|_{\mathbb{R}^2} = |x_1 - y_1| + |x_2 - y_2|$ sur \mathbb{R}^2 . Pour tout $k \geq 0$, on définit les coefficients

$$\text{si } \mathbb{E}(\|W_0\|_{\mathbb{R}^2}) < \infty, \quad \tau_1(k) = \tau(\sigma(W_0), W_k). \quad (2.3.9)$$

On considère également la norme $\|x - y\|_{(\mathbb{R}^2)^l} = l^{-1}(\|x_1 - y_1\|_{\mathbb{R}^2} + \dots + \|x_l - y_l\|_{\mathbb{R}^2})$ sur $(\mathbb{R}^2)^l$. Soit $\mathcal{M}_i = \sigma(W_k, 0 \leq k \leq i)$. Si $\mathbb{E}(\|W_1\|_{\mathbb{R}^2}) < \infty$, les coefficients $\tau_\infty(k)$ sont définis par

$$\tau_\infty(k) = \sup_{i \geq 0} \sup_{l \geq 1} \{\tau(\mathcal{M}_i, (W_{i_1}, \dots, W_{i_l})), i + k \leq i_1 < \dots < i_l\}. \quad (2.3.10)$$

On dit que le processus $(W_t)_{t \geq 0}$ est τ -dépendant si les coefficients $\tau_\infty(k)$ tendent vers zéro quand k tend vers l'infini. On dit qu'il est géométriquement τ -dépendant si il existe $a < 1$ et $C > 0$ tels que $\tau_\infty(k) \leq Ca^k$ pour tout $k \geq 1$. Considérons dorénavant les coefficients de dépendance définis comme dans (2.3.9) et (2.3.10) avec $(W_t)_{t \in \mathbb{Z}} = ((U_t, Z_t))_{t \in \mathbb{Z}}$.

Les hypothèses de dépendance pour le modèle (2.3.7) sont résumées par :

$$\left\{ \begin{array}{l} - \text{les variables } \varepsilon_i \text{ sont i.i.d.}, \\ - \text{Le vecteur aléatoire } (Z_i, \varepsilon_{i-1})_{0 \leq i \leq t} \text{ est indépendant de la suite } (\varepsilon_i)_{i \geq t}, \\ - \text{Le processus } (W_t)_{t \in \mathbb{Z}} \text{ est strictement stationnaire et } \tau\text{-dépendant.} \end{array} \right. \quad (2.3.11)$$

2.3.4 Borne de risque pour l'estimateur non adaptatif

Proposition 2.3.1 *Soit $\Delta(m)$ défini par (2.1.5). Si f_ε satisfait $(\mathbf{A}_1^\varepsilon)$ et si f_Z satisfait (\mathbf{A}_3^Z) , alors $\hat{f}_m^{(n)}$ défini par (2.1.2) satisfait*

$$\mathbb{E}\|f_Z - \hat{f}_m^{(n)}\|_2^2 \leq \|f_Z - f_m\|_2^2 + \frac{m^2(M_2 + 1)}{n} + \frac{2\Delta(m)}{n} + \frac{2R_m}{n},$$

$$\text{où } R_m = \frac{1}{\pi} \sum_{k=2}^n \int_{-\pi m}^{\pi m} \left| \frac{\text{Cov}(e^{ixU_1}, e^{ixZ_k})}{f_\varepsilon^*(-x)} \right| dx.$$

De plus, $R_m \leq R_{m,\tau}$, où

$$R_{m,\tau} = \pi m \Delta_{1/2}(m) \sum_{k=1}^{n-1} \tau_1(k),$$

$$\text{avec } \tau_1 \text{ défini par (2.3.9), et où } \Delta_{1/2}(m) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} \frac{1}{|f_\varepsilon^*(x)|} dx.$$

Comme dans le cas dépendant décrit dans la partie 2.2, le risque quadratique intégré de $\hat{f}_m^{(n)}$ est le risque du cas i.i.d. plus un terme supplémentaire $2R_m/n$. Ce terme additionnel fait intervenir les coefficients de dépendance et la quantité $\Delta_{1/2}(m)$ spécifique au modèle ARCH. Sous certaines conditions, le terme $\Delta_{1/2}(m)/n$ est négligeable ou au pire du même ordre que $\Delta(m)/n$. Plus précisément, si $\delta = 0$, et $\gamma > 1$ ou si $\delta > 0$ dans $(\mathbf{A}_1^\varepsilon)$, et si $\sum_{k \geq 1} \tau_1(k) < +\infty$, alors $\hat{f}_m^{(n)}$ défini par (2.1.2) satisfait

$$\mathbb{E}\|f_Z - \hat{f}_m^{(n)}\|_2^2 \leq \frac{C_1}{2\pi} (m^2 \pi^2 + 1)^{-s} \exp\{-2b\pi^r m^r\} + \frac{2\lambda_1(f_\varepsilon, \kappa_0)\Gamma(m)}{n} + \frac{C_2}{n} \Gamma(m) o_m(1), \quad (2.3.12)$$

avec C_1 et C_2 des constantes finies. La constante C_2 dépend de $\sum_{k \geq 1} \tau_1(k)$. L'estimateur $\hat{f}_m^{(n)}$ a donc un risque quadratique du même ordre que dans le cas i.i.d..

2.3.5 Borne de risque pour l'estimateur adaptatif

Pour $a > 1$, considérons la pénalité $\text{pen}(m)$ définie par

$$\text{pen}(m) = \begin{cases} 48a \frac{\Delta(m)}{n} & \text{if } 0 \leq \delta < 1/3, \\ 16a\lambda_3 \frac{\Delta(m) (\pi m)^{\min((3\delta/2-1/2)_+, \delta)}}{n} & \text{si } \delta \geq 1/3, \end{cases} \quad (2.3.13)$$

où $\Delta(m)$ est défini par (2.1.5) et $\lambda_3(f_\varepsilon)$ est une constante qui dépend de f_ε . La pénalité est donc du même ordre que celle du modèle i.i.d. à des constantes près.

Théorème 2.3.1 *Supposons (2.3.8)-(2.3.11), que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$, que f_Z satisfait (\mathbf{A}_3^Z) et que m_n satisfait (2.1.7). Soit $\hat{f}_m^{(n)}$ défini par (2.1.2) avec $1 \leq m \leq m_n$ et $\text{pen}(m)$ définie par (2.3.13). Supposons de plus que l'une des deux conditions suivantes a lieu :*

1. $\delta = 0$, $\gamma \geq 3/2$ dans $(\mathbf{A}_1^\varepsilon)$ et $\tau_\infty(k) = O(k^{-(1+\theta)})$ pour un $\theta > 3 + 2/(1 + 2\gamma)$
2. $\delta > 0$ dans $(\mathbf{A}_1^\varepsilon)$ et $\tau_\infty(k) = O(k^{-(1+\theta)})$ pour un $\theta > 3$,

où τ_∞ est défini par (2.3.10). Alors, l'estimateur $\tilde{f}_Z = \hat{f}_m^{(n)}$ défini par (2.1.4) satisfait

$$\mathbb{E}(\|f_Z - \tilde{f}_Z\|_2^2) \leq C_a \inf_{m \in \{1, \dots, m_n\}} \left[\|f_Z - f_m\|_2^2 + \text{pen}(m) + \frac{m^2(M_2 + 1)}{n} \right] + \frac{\bar{C}}{n}, \quad (2.3.14)$$

où C_a est une constante qui dépend de a et \bar{C} est une constante qui dépend de f_ε , a , et $\sum_{k \geq 1} \tau_\infty(k)$.

L'estimateur \tilde{f}_Z est construit avec une pénalité qui ne dépend pas de la somme des coefficients de dépendance et qui est la même que dans le cas i.i.d.. Ceci est d'un intérêt tout particulier car la somme des coefficients de dépendance est connue pour être difficilement estimable.

L'estimateur \tilde{f}_Z est donc adaptatif dans le sens où sa construction ne requiert pas de connaître la régularité de f_Z et que son risque quadratique est du même ordre de grandeur que le meilleur des risques des estimateurs de la collection. Il n'y a pas à ma connaissance de résultats d'optimalité (au sens minimax) dans un tel modèle. Nous ne pouvons donc pas affirmer qu'il atteint ou non la vitesse minimax, mais nous établissons qu'il a exactement la même vitesse que dans le cas i.i.d., où des minoration ont été établies.

2.3.6 Estimation de la densité de σ_t^2 .

La procédure proposée permet d'estimer la densité de $Z_t = \ln(\sigma_t^2)$. Je décris maintenant comment on obtient un estimateur de la densité de σ_t^2 ainsi qu'une majoration de son risque quadratique intégré.

Pour $u > 0$, si f_σ est la densité de σ_t^2 , on a la relation $f_\sigma(u) = f_Z(\ln(u))/u$. On estime donc f_σ par

$$\hat{f}_\sigma(u) = \frac{\tilde{f}(\ln(u))}{u}, \quad \text{pour } u > 0.$$

Le risque quadratique intégré de \hat{f}_σ sur $[a, \infty[$, avec $a > 0$, est du même ordre que $\mathbb{E}(\|\tilde{f}_Z - f_Z\|_2^2)$. En effet nous avons les relations

$$\mathbb{E}\left(\int_0^\infty |\hat{f}_\sigma(t) - f_\sigma(t)|^2 t dt\right) = \mathbb{E}(\|\tilde{f}_Z - f_Z\|_2^2) \quad \text{et} \quad \mathbb{E}\left(\int_a^\infty |\hat{f}_\sigma(t) - f_\sigma(t)|^2 dt\right) \leq \frac{1}{a} \mathbb{E}(\|\tilde{f}_Z - f_Z\|_2^2).$$

2.3.7 Exemples et illustration pratique

Exemples de modèles ARCH satisfaisant les conditions de dépendance

Dans cette partie nous présentons des exemples de modèles ARCH (2.3.6) pour lesquels le théorème 2.3.1 s'applique. Parmi ces exemples de modèles, on peut citer notamment deux grandes familles.

Une première famille de modèles est composée des modèles (2.3.1) décrits par

$$Y_t = \sigma_t \eta_t, \quad \text{avec } \sigma_t = f(\eta_{t-1}, \eta_{t-2}, \dots). \quad (2.3.1)$$

où f est une fonction mesurable. Une deuxième famille est composée des modèles décrits par

$$Y_t = \sigma_t \eta_t, \quad \text{avec } \sigma_t = f(\sigma_{t-1}, \eta_{t-1}) \quad \text{et} \quad \sigma_0 \text{ indépendant de } (\eta_t)_{t \geq 0}. \quad (2.3.2)$$

Dans ce cas σ_t est une chaîne de Markov stationnaire et on supposera pour simplifier que $\mathbb{E}(\eta_0^2) = 1$.

Le modèle ARCH a été introduit par Engle (1982) et généralisé par Bollerslev (1986) avec la classe des modèles GARCH(p, q) définis par $Y_t = \sigma_t \eta_t$ et

$$\sigma_t^2 = a + \sum_{i=1}^p a_i Y_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2 \quad (2.3.3)$$

où les coefficients $a, a_i, i = 1, \dots, p$ et $b_j, j = 1, \dots, q$ sont des réels positifs. Sous la condition $\sum_{i=1}^p a_i + \sum_{j=1}^q b_j < 1$, le modèle admet une unique solution stationnaire de la forme (2.3.1) (voir Bougerol et Picard (1992a, 1992b) et Ango Nzé (1992)).

Un exemple de modèle linéaire général ARCH(∞) décrit par Giraitis *et al.* (2000) est :

$$\sigma_t^2 = a + \sum_{j=1}^{\infty} a_j Y_{t-j}^2, \quad (2.3.4)$$

où $a \geq 0$ et $a_j \geq 0$. Si $\sum_{j \geq 1} a_j < 1$, alors il existe une unique solution strictement stationnaire au modèle (2.3.4), qui est de la forme (2.3.1).

Pour les modèles satisfaisant (2.3.2), on peut citer les modèles dits GARCH(1, 1) "augmentés" introduits par Duan (1997) :

$$\Lambda(\sigma_t^2) = c(\eta_{t-1})\Lambda(\sigma_{t-1}^2) + h(\eta_{t-1}), \quad (2.3.5)$$

où Λ est une fonction continue et coissante de \mathbb{R}^+ . D'autres modèles satisfaisant (2.3.2) sont les modèles ARCH non linéaires (voir Doukhan (1994), p. 106-107), pour lesquels :

$$\sigma_t = f(\sigma_{t-1} \eta_{t-1}). \quad (2.3.6)$$

Il existe une solution stationnaire au modèle (2.3.6) dès que la densité de η_0 est strictement positive au voisinage de 0 et que $\limsup_{|x| \rightarrow \infty} |f(x)/x| < 1$.

La proposition suivantes fournit des conditions générales suffisantes assurant que les conditions de dépendance du théorème 2.3.1 sont vérifiées.

Proposition 2.3.2 *Soit Y_t et σ_t satisfaisant (2.3.1) ou (2.3.2). Supposons que σ_0^2 et η_0^2 admettent des densités bornées et que $\mathbb{E}(\eta_0^2) = 1$.*

- 1) *Pour le modèle ARCH(∞) (2.3.4), nous avons*
 - *si $a_j = 0$, pour $j \geq J$, alors $((Z_t, U_t))_{t \geq 0}$ est géométriquement τ -dépendant.*
 - *Si $a_j = O(b^j)$ pour un $b < 1$ alors $\tau_\infty(n) = O(\kappa^{\sqrt{n}})$ pour un $\kappa < 1$.*
 - *Si $a_j = O(j^{-b})$ pour un $b > 1$ alors $\tau_\infty(n) = O(n^{-b/2}(\ln(n))^{b+2})$.*
- 2) *Pour le modèle (2.3.2), si il existe $\kappa < 1$ tel que*

$$\mathbb{E}(|(f(x, \eta_0))^2 - (f(y, \eta_0))^2|) \leq \kappa |x^2 - y^2|, \quad (2.3.7)$$

alors $((Z_t, U_t))_{t > 0}$ est géométriquement τ -dépendant avec $\tau_\infty(n) = O(n(\sqrt{\kappa})^n)$.

Un exemple de chaîne de Markov satisfaisant (2.3.7) est le modèle autorégressif $\sigma_t^2 = h(\sigma_{t-1}^2) + r(\eta_{t-1})$ pour une fonction h κ -lipschitzienne.

Illustration pratique

Les programmes *Matlab* sont disponibles sur la page de Yves Rozenholc à l'adresse : <http://www.math-info.univ-paris5.fr/~rozen/>.

Soit $\varepsilon_t \rightsquigarrow \ln[\mathcal{N}(0, 1)]^2$ et la pénalité

$$\text{pen}(m) = \left(1 + \frac{\ln(\pi m)^{2.5}}{(1 + \sigma_\varepsilon^2)\pi m}\right) \int_{-\pi m}^{\pi m} \frac{dx}{|f_\varepsilon^*(x)|^2}$$

Considérons les trois modèles dépendants suivants

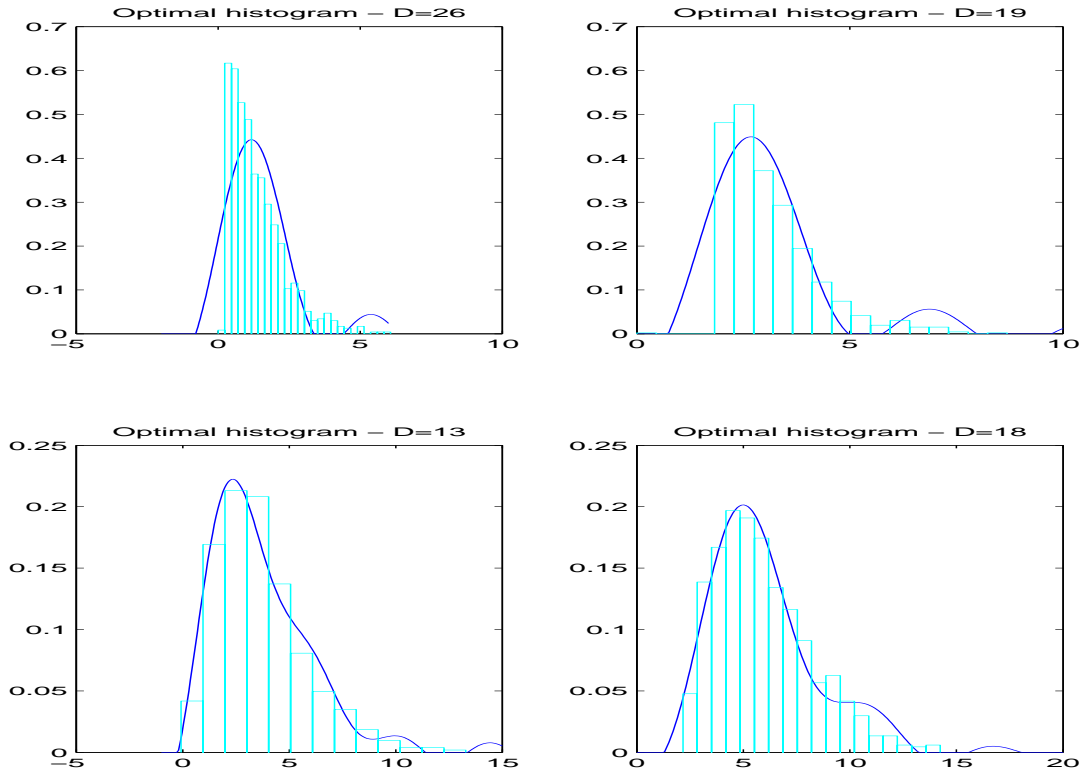


FIG. 2.2 – Deconvolution de deux modèles ARCH (haut) et de deux modèles de de Vries (bas). Courbes estimées et histogramme optimal sur l’observation directe de $\ln(\sigma_t^2)$ pour $n = 1000$ données simulées.

M1 modèle GARCH(1,1) (i.e. $p = q = 1$ dans (2.3.3)) avec paramètres $a = 1, a_1 = 0.7, b_1 = 0.2$ (comme dans van Es *et al.* (2005)),

M2 modèle GARCH(1,1) ($p = q = 1$ dans (2.3.3)) avec paramètres $a = 5, a_1 = 0.79$ et $b_1 = 0.2$.

M3 $Y_t = \sigma_t \eta_t$ et $\sigma_t^2 = \tau^2 \sigma_{t-1}^2 + 1/\eta_{t-1}^2$, avec des variables η_t , i.i.d. $\mathcal{N}(0, 1)$ et $\tau^2 = 0.5$ ou $\tau^2 = 0.8$ (voir l’exemple de de Vries (1991), équation (10)).

Dans ces modèles, la vraie densité stationnaire à estimer n’est pas connue. Nous allons donc comparer l’estimateur construit par déconvolution avec l’estimateur adaptatif obtenu par histogramme optimal et construit avec les observations directes $\ln(\sigma_t^2)$, décrit dans Birgé et Rozenholc (2006). On utilise cet histogramme comme référence pour l’allure de la vraie densité stationnaire. Cette démarche n’est évidemment valable que dans un contexte de simulation où les Z_t sont aussi disponibles. Le niveau de bruit apparaît au travers de $s2n = \widehat{\text{var}}(U)/\sigma_\varepsilon^2 - 1$, où $\widehat{\text{var}}(U)$ est la variance empirique des observations.

Pour le modèle M1, on obtient de bons résultats. La partie en haut à gauche de la Figure 2.2 montre l’estimateur basé sur l’observation des $(U_t)_{1 \leq t \leq n}$ avec $U_t = \ln(Y_t^2)$. On peut remarquer que le pic est légèrement coupé. Le résultat est meilleur dans la partie en haut à droite de la Figure 2.2 (modèle M2), quand $s2n$ augmente (il arrive à 0.35 au lieu de 0.13 dans le cas précédent). Pour le modèle M3, on choisit $\tau^2 = 0.5$ (Figure 2.2-en bas à gauche) et $\tau^2 = 0.8$ (Figure 2.2-en bas à droite). On a alors $s2n=0.6$ (gauche) et 1.34 (droite) dans le bas de la Figure 2.2. Ici la procédure donne de bons résultats, avec une pénalité bien calibrée.

2.4 Estimation de fonctionnelles linéaires dans le modèle de convolution

Cette partie décrit brièvement l'article Matias et Taupin (2004). J'ai choisi de ne pas détailler plus cette présentation afin de ne pas rendre ce mémoire trop technique. Considérons le modèle

$$U_i = Z_i + \varepsilon_i, \quad i = 1, \dots, n,$$

avec U_i l'observation, Z_i indépendante de ε_i et $\varepsilon_i \sim \mathcal{N}(0, 1)$. On souhaite estimer des fonctionnelles linéaires intégrales de f_Z de la forme

$$\Gamma_f(x) = \int f(z) f_\varepsilon(z - x) f_Z(z) dz,$$

pour une fonction f connue. Quand $f \equiv 1$ on retrouve la densité de l'observation U , puisque $f_U = f_Z \star f_\varepsilon \equiv \Gamma_1$. En utilisant des outils liés à la déconvolution j'ai proposé dans Taupin (2001) d'estimer Γ_f par

$$\widehat{\Gamma}_{f,n}(y_0) = \int f(x) f_\varepsilon(y_0 - x) \widehat{f}_Z(x) dx, \quad \text{avec} \quad \widehat{f}_Z(x) = \frac{1}{n} \sum_{i=1}^n K_{n,C_n}(x - U_i), \quad (2.4.8)$$

C_n étant une suite qui tend vers l'infini, K_{n,C_n} étant un noyau de déconvolution défini par sa transformée de Fourier via

$$K_{n,C_n}^*(t) = \frac{K_n^*(t)}{f_\varepsilon^*(t)} = \frac{K^*(t/C_n)}{f_\varepsilon^*(t)}, \quad (2.4.9)$$

avec K un noyau à choisir, ayant une transformée de Fourier à support compact. Cet estimateur est consistant (voir Taupin (2001)), avec une vitesse qui dépend de la régularité de $f(x) f_\varepsilon(\cdot - x)$ comme fonction de x comparée à celle de f_ε .

Dans l'article Matias et Taupin, nous établissons des minoration pour des risques minimax associés à des fonctions de perte variées. Pour chacun des types de fonctions f , nous avons considéré plusieurs risques : le risque quadratique ponctuel, le risque par rapport à la norme \mathbb{L}_p et enfin le risque relié à la norme uniforme. Les minoration des risques minimax associés montrent que l'estimateur (2.4.8) atteint la vitesse optimale quand f est polynomiale, trigonométrique ou exponentielle.

Chapitre 3

Modèles de régression avec erreurs sur les variables

3.1 Estimation non paramétrique dans un modèle de régression avec erreurs sur les variables

Cette partie présente les résultats de l'article Comte et Taupin (2007).

3.1.1 Introduction, modèle et résultats antérieurs

On observe $(Y_i, U_i)_{1 \leq i \leq n}$ un n -échantillon de variables aléatoires i.i.d. satisfaisant

$$\begin{cases} Y_i = f(Z_i) + \xi_i, & \mathbb{E}(\xi_i) = 0, \\ U_i = Z_i + \varepsilon_i, \end{cases} \quad (3.1.1)$$

où les $Z_i, \xi_i, \varepsilon_i$ sont indépendantes. L'objectif est d'estimer la fonction de régression f à partir des observations $(Y_i, U_i)_{i=1, \dots, n}$. Les variables aléatoires $(Z_i)_{i=1, \dots, n}$ ne sont pas observées et ont une densité commune f_Z inconnue. Dans ce modèle, les variables ξ_i sont centrées de densité inconnue, tandis que les erreurs $(\varepsilon_i)_{i=1, \dots, n}$ ont une densité commune connue f_ε .

Comme en déconvolution, deux facteurs interviennent sur la vitesse d'estimation de la fonction de régression dans un modèle avec erreurs sur les variables : la régularité de la fonction de régression mais aussi la régularité de la densité du bruit f_ε . Et comme en déconvolution, plus f_ε est régulière, plus il est difficile de retrouver de l'information sur Z et donc d'estimer la fonction de régression. Les vitesses de convergence les plus lentes sont donc obtenues pour les f_ε les plus régulières (ε gaussienne par exemple).

Les premiers résultats d'estimation non paramétrique dans le modèle (3.1.1) ont été obtenus par Fan *et al.* (1991), Fan et Truong (1993). Ils ont proposé un estimateur du type Nadaraya-Watson, construit comme le ratio de deux estimateurs à noyaux par déconvolution. Parmi les auteurs qui ont étudié ce problème, on peut citer Fan et Masry (1992), Masry (1993b), Ioannides et Alevizos (1997). Ces articles considèrent différentes approches (risques \mathbb{L}_p , normalité asymptotique...) sous des hypothèses variées (mélange, régularité,...). En particulier quand la fonction de régression et la densité f_Z admettent toutes les deux des dérivées jusqu'à l'ordre k , Fan et Truong (1993) étudient l'optimalité pour l'estimation de f . Quand la densité f_ε est soit "ordinary" soit "super smooth", ils établissent des majorations et des minoration pour le risque quadratique ponctuel et pour le risque \mathbb{L}_p restreint à un ensemble compact.

En utilisant une approche différente, Koo et Lee (1998) ont proposé une méthode d'estimation basée sur les B -splines. Leurs résultats ne s'appliquent que pour une densité d'erreur f_ε "ordinary smooth".

Dans les travaux mentionnés ci-dessus, la fonction de régression et la densité de Z appartiennent à la même classe de régularité connue de type "ordinary smooth". Nous proposons une méthode d'estimation

adaptative de la fonction de régression, dans le sens où sa construction ne requiert pas la connaissance à priori de sa régularité. De plus nous établissons une inégalité de type oracle pour le risque quadratique intégré sur un compact de l'estimateur adaptatif et nous étudions les vitesses de convergence pour f et f_ε dans deux classes de régularité incluant toutes les deux les fonctions "ordinary smooth" et "super smooth". Notre estimateur atteint les vitesses minimax dans tous les cas où des minoration ont été établies.

3.1.2 Méthode d'estimation

Nous proposons une méthode d'estimation de la fonction de régression f , qui ne requiert aucune connaissance à priori sur sa régularité, ni sur celle de la densité f_Z .

Notre méthode d'estimation est basée sur l'idée classique que la fonction de régression peut s'écrire

$$f(z) = \mathbb{E}(Y|Z = z) = \frac{\int y f_{Z,Y}(z, y) dy}{f_Z(z)} = \frac{(f f_Z)(z)}{f_Z(z)},$$

avec $f_{Z,Y}$ le densité jointe de (Z, Y) . Par conséquent la fonction de régression peut être estimée par le ratio de deux estimateurs, c'est-à-dire par $\tilde{f} = \tilde{\ell} / \tilde{f}_Z$, où $\tilde{\ell}$ est un estimateur adaptatif de $\ell = f f_Z$ et \tilde{f}_Z est un estimateur adaptatif de f_Z . Ces deux estimateurs vont être obtenus par minimisation de contrastes pénalisés bien choisis. La méthode d'estimation reprend les idées et les notations de la partie 2.1 sur l'estimation de la densité par déconvolution dans un cadre i.i.d. .

Estimateurs non adaptatifs

Comme pour l'estimation de densité par déconvolution, on part du fait que

$$\ell = \arg \min_t \|t - \ell\|_2^2 = \arg \min_t \|t\|_2^2 - 2\langle \ell, t \rangle = \arg \min_t [\|t - \ell\|_2^2 - \|\ell\|_2^2].$$

On remarque ensuite que $\langle \ell, t \rangle = \mathbb{E}(f(Z)t(Z)) = \mathbb{E}(Yt(Z))$. Nous avons donc besoin d'estimer $\mathbb{E}(Yt(Z))$ en utilisant les observations U_1, \dots, U_n . Pour un m arbitraire, et avec les notations de la partie 2.1, on considère un estimateur de ℓ sur $S_m^{(n)}$ défini par

$$\hat{\ell}_m^{(n)} = \arg \min_{t \in S_m^{(n)}} \gamma_{n,\ell}(t), \quad (3.1.2)$$

où, pour $t \in S_m^{(n)}$,

$$\gamma_{n,\ell}(t) = \|t\|_2^2 - 2n^{-1} \sum_{i=1}^n (Y_i u_t^*(U_i)) \text{ avec } u_t(x) = (2\pi)^{-1} t^*(-x) / f_\varepsilon^*(-x) \text{ et } \mathbb{E}(Y_1 u_t^*(U_1)) = \langle t, \ell \rangle.$$

La collection d'estimateurs non adaptatifs $\hat{f}_{m_{f_Z}}^{(n)}$ de f_Z est définie par (2.1.2) dans la partie (2.1).

Les estimateurs non adaptatifs de f sont définis par $\hat{f}_{m_\ell, m_{f_Z}} = \hat{\ell}_{m_\ell} / \hat{f}_{m_{f_Z}}$ où m_ℓ et m_{f_Z} sont donnés par les compromis entre carré du biais et variance de chacun des estimateurs. La construction et l'étude de ces estimateurs requièrent de se restreindre à un compact A et de considérer les hypothèses suivantes.

$$f \in \mathcal{F}_G = \{\phi \text{ telle que } \sup_{x \in G} |\phi(x)| \leq \kappa_{\infty, G} < \infty\}, \text{ où } G \text{ est le support de la densité } f_Z. \quad (\mathbf{A}_1)$$

$$\text{Il existe des constantes positives } g_0, g_1 \text{ telles que pour tout } x \in A, g_0 \leq g(x) \leq g_1. \quad (\mathbf{A}_2)$$

Estimateurs adaptatifs

L'estimateur adaptatif de f_Z est \tilde{f}_Z défini dans la partie (2.1). L'estimateur adaptatif de ℓ est défini de façon analogue par

$$\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell}^{(n)} \text{ avec } \hat{m}_\ell^{(n)} = \arg \min_{m \in \mathcal{M}_{n,\ell}} \left[\gamma_{n,\ell}(\hat{\ell}_m^{(n)}) + \text{pen}_\ell(m) \right], \quad (3.1.3)$$

où pour κ' une constante numérique,

$$\text{pen}_\ell(m) = \kappa'[1 + \hat{m}_2(Y)]\Delta(m)/n \text{ et } \hat{m}_2(Y) = \frac{1}{n} \sum_{i=1}^n Y_i^2.$$

Pour $r \in \mathbb{R}$ et $d > 0$, on note $r^{(d)} = \text{sign}(r) \min(|r|, d)$. Pour a_n bien choisie, l'estimateur adaptatif de f est \tilde{f} de f défini par

$$\tilde{f} = (\tilde{\ell}/\tilde{f}_Z)^{(a_n)}. \quad (3.1.4)$$

3.1.3 Bornes de risque pour les estimateurs sur un modèle

Proposition 3.1.1 *Notons ℓ_m la projection orthogonale de ℓ sur S_m . Soit $\hat{\ell}_m$ défini par (3.1.2). Si ℓ et f_Z satisfont (\mathbf{A}_3^Z) , alors*

$$\mathbb{E}(\|\hat{\ell}_m^{(n)} - \ell\|_2^2) \leq \|\ell - \ell_m\|_2^2 + 2\mathbb{E}(Y_1^2)\Delta(m)/n + (\kappa_{\mathcal{L}} + \|\ell\|_1^2)(\pi m)^2/k_n$$

En combinant cette borne pour l'estimation de ℓ et la proposition 2.1.1 pour l'estimation de f_Z on obtient la borne de risque suivante pour l'estimateur de la fonction de régression.

Proposition 3.1.2 *Soit f_ε satisfaisant $(\mathbf{A}_1^\varepsilon)$, f_Z satisfaisant (\mathbf{A}_2) et f_Z appartenant à $\mathcal{S}_{s_{f_Z}, r_{f_Z}, b_{f_Z}}(\kappa_{f_Z})$ avec $s_{f_Z} > 1/2$ si $r_{f_Z} = 0$. Notons $\hat{f}_{m_\ell, m_{f_Z}} = \hat{\ell}_{m_\ell}/\hat{f}_{Z, m_{f_Z}}$ l'estimateur de f non adaptatif défini par $\hat{f}_{m_\ell, m_{f_Z}} = \hat{\ell}_{m_\ell}/\hat{f}_{Z, m_{f_Z}}$ avec m_ℓ et m_{f_Z} qui réalisent le meilleur compromis biais-variance pour chacun des deux estimateurs de ℓ et f_Z . Si $a_n = n^k$ avec $k > 0$ et $k_n \geq n^{3/2}$, alors pour n assez grand*

$$\mathbb{E}(\|\hat{f}_{m_\ell, m_{f_Z}} - f\|_{\mathbb{A}}^2) \leq C_0[\mathbb{E}(\|\ell - \hat{\ell}_{m_\ell}\|_2^2) + \mathbb{E}(\|f_Z - \hat{f}_{Z, m_{f_Z}}\|_2^2)] + o(n^{-1}),$$

où C_0 est une constante qui dépend de f_Z .

On constate donc que le risque quadratique intégré de $\hat{f}_{m_\ell, m_{f_Z}}$ est majoré par la somme des risques quadratiques intégrés $\hat{\ell}_{m_\ell}$ et $\hat{f}_{m_{f_Z}}$. Ceci vient très probablement de la méthode d'estimation. En effet, dans le contexte de l'estimation classique d'une fonction de régression, l'estimateur de Nadaraya-Watson présente déjà la particularité de faire intervenir la régularité de f_Z pour l'estimation de la fonction de régression. Dans le cadre classique, sans erreurs, il existe d'autres méthodes qui ne présentent pas cet inconvénient, comme par exemple les méthodes d'estimation par polynômes locaux ou par projection.

3.1.4 Une inégalité de type oracle

Théorème 3.1.1 *Supposons que f_ε satisfait $(\mathbf{A}_1^\varepsilon)$, f_Z et ℓ satisfont (\mathbf{A}_3^Z) . Soient $\hat{\ell}_m$ et \hat{f}_m les estimateurs définis par (3.1.2) et (2.1.2) avec $k_n > n$, $m_{n,\ell} \leq m_n$, $m_{n,g} \leq m_n$ et m_n comme dans (2.1.7).*

Estimation adaptative de ℓ . *Sous (\mathbf{A}_1) , si $\mathbb{E}|\xi_1|^8 < \infty$, $\tilde{\ell} = \hat{\ell}_{\hat{m}_\ell}$ défini par (3.1.3) satisfait*

$$\mathbb{E}(\|\ell - \tilde{\ell}\|_2^2) \leq K' \inf_{m \in \mathcal{M}_{n,\ell}} [\|\ell - \ell_m\|_2^2 + (\pi m)^2(\kappa_{\mathcal{L}} + \|\ell\|_1^2)/n + \mathbb{E}(\text{pen}_\ell(m))] + c'/n,$$

où K' est une constante numérique et où c' est une autre constante qui dépend de f_ε et de ℓ .

Estimation adaptative de f . Notons $f_{Z,m}$ la projection orthogonale de f_Z sur S_m . Soit f_Z satisfaisant **(A₂)** et appartenant à $\mathcal{S}_{s_{f_Z}, r_{f_Z}, b_{f_Z}}(\kappa_{f_Z})$ défini par (2.1.8) avec $s_{f_Z} > 1/2$ si $r_{f_Z} = 0$. Supposons que $\mathbb{E}|\xi_1|^8 < \infty$. Soit \tilde{f} définie par (3.1.4). Si $k_n \geq n^{3/2}$, $a_n = n^k$ avec $k > 0$, alors pour n assez grand

$$\begin{aligned} \mathbb{E}(\|(f - \tilde{f})\mathbb{I}_A\|_2^2) &\leq C_0 \inf_{m \in \mathcal{M}_{n,\ell}} [\|\ell - \ell_m\|_2^2 + (\pi m)^2(\kappa_{\mathcal{L}} + \|\ell\|_1^2)/n + \mathbb{E}(\text{pen}_\ell(m))] \\ &+ C_1 \inf_{m \in \mathcal{M}_{n,f_Z}} [\|f_Z - f_{Z,m}\|_2^2 + (\pi m)^2(\kappa_{\mathcal{G}} + 1)/n + \text{pen}_{f_Z}(m)] + c/n \end{aligned}$$

où C_0, C_1 sont des constantes qui dépendent de ℓ et f_Z respectivement et c est une constante qui dépend de f_ε, f et f_Z .

On constate que le risque quadratique intégré de l'estimateur adaptatif de f est majoré par la somme des risques des estimateurs adaptatifs de ℓ et de f_Z . Si la densité f_Z est plus régulière que la fonction de régression, alors l'estimateur a pour risque le plus petit risque possible pour estimer f .

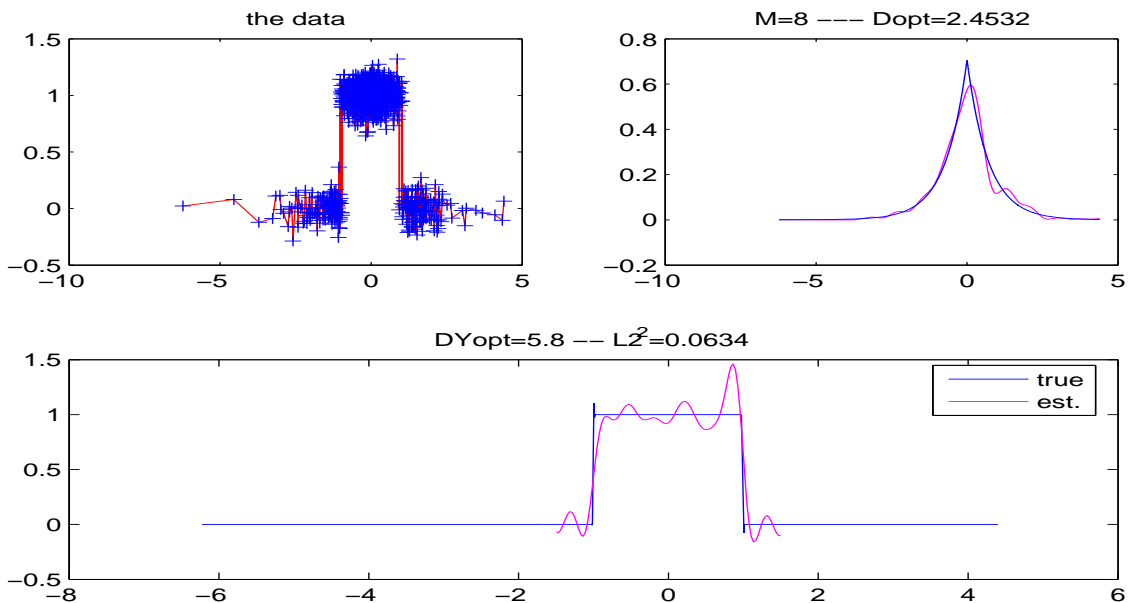
On peut montrer que si f et f_Z admettent toutes les deux des dérivées jusqu'à l'ordre k , alors l'estimateur adaptatif \tilde{f} atteint la vitesse minimax établie par Fan (1991) et Fan et Truong (1993).

3.1.5 Illustration par des simulations

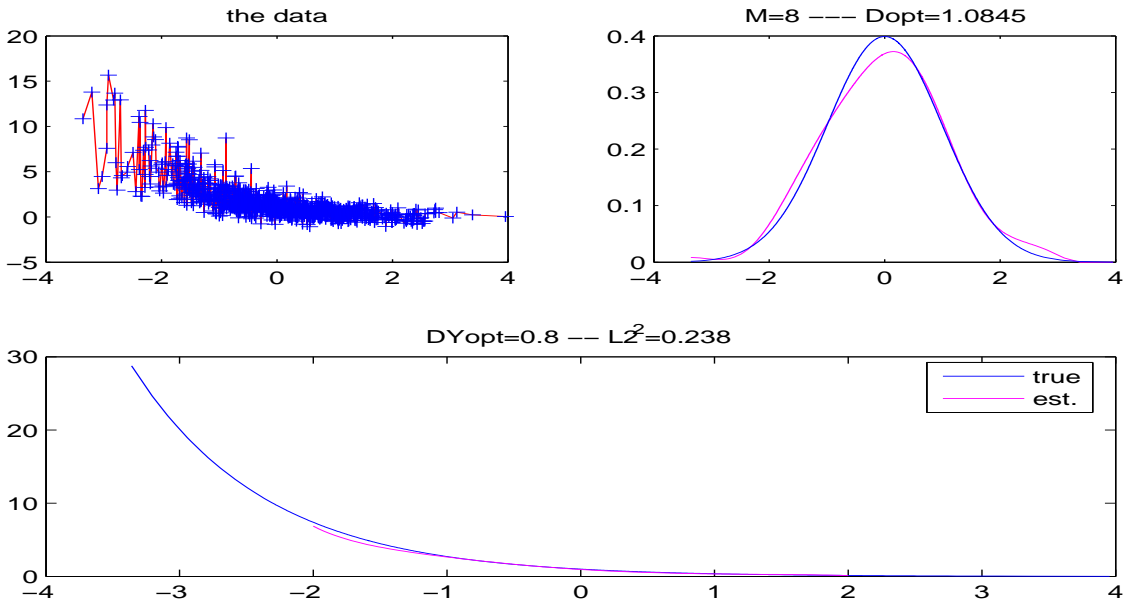
Les programmes *Matlab* sont disponibles sur la page de Yves Rozenholc à l'adresse : <http://www.math-info.univ-paris5.fr/rozen/>.

Les graphiques qui suivent représentent les estimateurs et les fonctions à estimer dans différents contextes. Pour chaque groupe de graphiques, le graphique en haut à gauche représente la vraie fonction de régression et les observations Y associées. Le graphique en haut à droite représente la vraie densité f_Z et l'estimateur optimal, choisi par la procédure d'estimation. Le graphique du bas représente la fonction de régression et l'estimateur choisi par la procédure d'estimation.

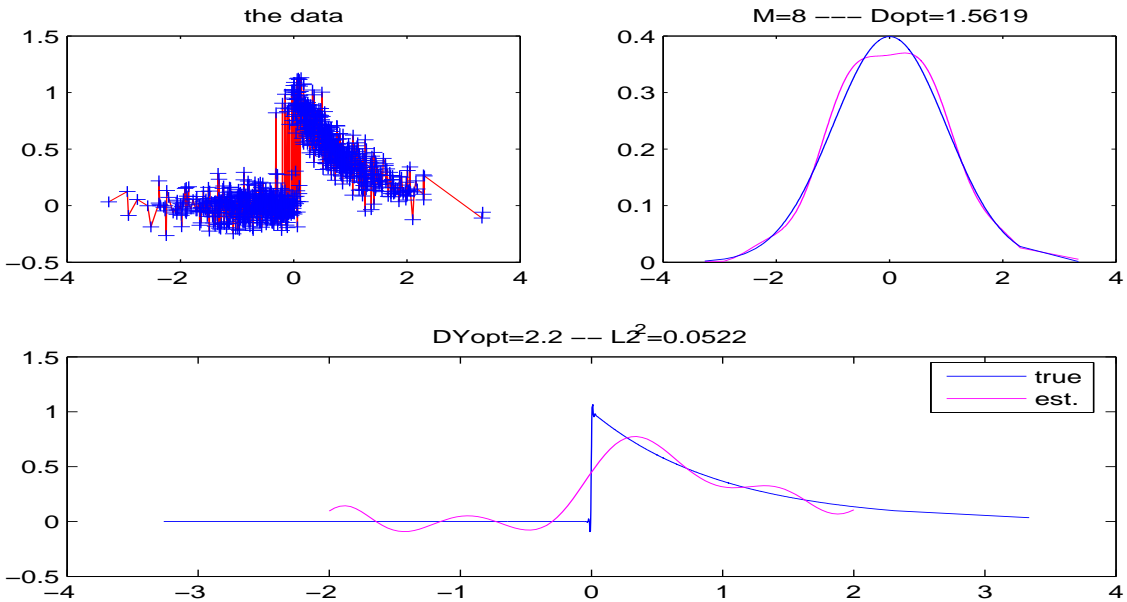
Régression uniforme On considère $f(x) = \mathbb{I}_{[-1,1]}(x)$, avec $Z \sim \text{symexp}$, $n = 1000$, $\varepsilon \sim (1/\sqrt{500})\text{symexp}$, et $\xi \sim (1/10)\mathcal{N}(0, 1)$.



Régression exponentielle On considère $f(x) = \exp(-x)$, avec $Z \sim \mathcal{N}(0, 1)$, $n = 1000$, $\varepsilon \sim (1/\sqrt{2})\text{symexp}$, et $\xi \sim (1/2)\mathcal{N}(0, 1)$.



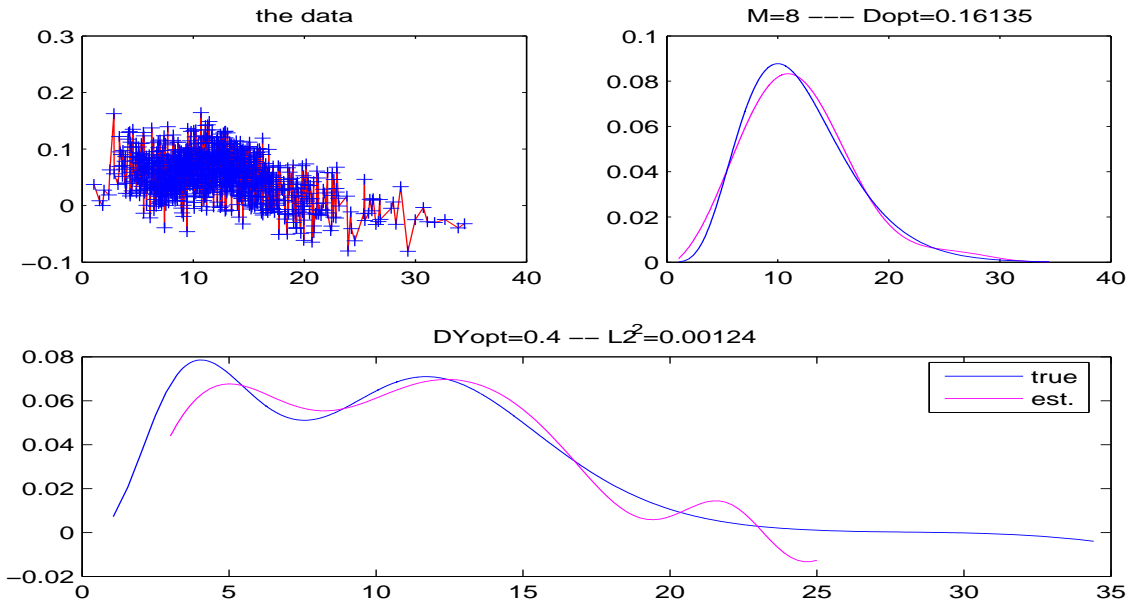
Exponentielle tronquée On considère $f(x) = \exp(-x)\mathbb{I}_{x \geq 0}$, avec $Z \sim \mathcal{N}(0, 1)$, $n = 1000$, $\varepsilon \sim (1/\sqrt{2})\text{symexp}$, et $\xi \sim (1/2)\mathcal{N}(0, 1)$.



Régression Gamma $f(x) = 0.4x^4e^{-x}/\Gamma(5) + 0.6x^{12}e^{-x}/\Gamma(13)$, avec $Z \sim \mathcal{X}(12)$, $n = 1000$, $\varepsilon \sim (1/10)\mathcal{N}(0, 1)$, $\xi \sim (1/\sqrt{1000})\mathcal{N}(0, 1)$.

3.1.6 Perspectives

La méthode d'estimation adaptative que nous avons proposé est basée sur un estimateur par quotient dont la vitesse de convergence fait intervenir à la fois la régularité de f_Z et la régularité de la fonction de régression. Dans le contexte classique d'estimation de fonction de régression (sans erreurs), des méthodes alternatives à l'estimateur de Nadaraya-Watson existent. Elles permettent notamment de s'affranchir de cette dépendance de la régularité de f_Z . Il serait très intéressant d'obtenir des méthodes d'estimation alternatives de f qui ne dépendent pas de la régularité de f_Z .



3.2 Estimation semi-paramétrique dans un modèle de régression avec erreurs sur les variables

Cette partie présente les résultats des articles Taupin (2001, 1998) et Butucea et Taupin (2008). Les deux premiers sont présentés rapidement tandis que l'article Butucea et Taupin est plus détaillé. En effet ce dernier est plus récent et améliore les résultats contenus dans Taupin (2001, 1998).

3.2.1 Le modèle

Considérons que l'on observe n copies i.i.d. de (Y, U) satisfaisant

$$\begin{cases} Y &= f_{\theta^0}(Z) + \xi \\ U &= Z + \varepsilon. \end{cases}$$

Les variables Z, ξ, ε sont indépendantes et non observées. La fonction de régression f_{θ^0} appartient à une famille paramétrique et θ^0 , le paramètre inconnu appartient à l'intérieur d'un sous-ensemble compact de $\Theta \subset \mathbb{R}^d$. Les densités marginales des Z_i et des ξ_i sont inconnues avec $\mathbb{E}(\xi) = 0$. Par contre la densité des erreurs ε_i est supposée complètement connue.

Dans ce contexte l'objectif est d'estimer le paramètre inconnu θ^0 en utilisant les observations (Y_i, U_i) pour $i = 1, \dots, n$. Le modèle est un modèle semi-paramétrique, avec $\theta^0 \in \mathbb{R}^d$ le paramètre d'intérêt et f_Z le paramètre de nuisance qui appartient à un espace fonctionnel.

3.2.2 Résultats antérieurs

Ce modèle a été largement étudié depuis les premiers résultats qui datent des années 1950 (voir par exemple Reiersøl (1950) ou Kiefer et Wolfowitz (1956)). Parmi les articles antérieurs on peut citer ceux (les plus nombreux) qui considèrent le cas linéaire, ceux qui considèrent le cas des mesures répétées ou avec des hypothèses supplémentaires (observations supplémentaires, f_Z dans une famille paramétrique,...).

Commençons par présenter les résultats antérieurs obtenus dans le cas linéaire. Le modèle s'écrit alors

$$\begin{cases} Y_i &= \theta_0^0 + \theta_1^0 Z_i + \xi_i \\ U_i &= Z_i + \varepsilon_i. \end{cases}$$

La littérature concernant l'estimation de $\theta^0 = (\theta_0^0, \theta_1^0)$ dans le cas linéaire est très abondante. Citons notamment Reiersøl (1950), Kiefer et Wolfowitz (1956), Gleser (1981, 1985), Anderson (1984), Bickel et Ritov (1987) et Cheng et Van Ness (1994). Une méthode d'estimation repose sur le maximum de vraisemblance obtenu quand Z , ξ et ε sont des variables aléatoires gaussiennes avec différentes hypothèses d'identifiabilité. Par exemple, si le rapport des variances $\lambda = \sigma_\varepsilon^2/\sigma_\xi^2$ est connu, alors cet estimateur du maximum de vraisemblance est défini par

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \sum_{i=1}^n \frac{[Y_i - \theta_0 - \theta_1 U_i]^2}{\sigma_\varepsilon^2(1 + \lambda \theta_1^2)}.$$

Quand Z , ξ et ε sont des variables aléatoires gaussiennes, cet estimateur $\hat{\theta}$ est \sqrt{n} -consistant, asymptotiquement gaussien de θ^0 et efficace. Quand Z , ξ et ε ne sont plus des variables aléatoires gaussiennes il reste un estimateur \sqrt{n} -consistant, asymptotiquement gaussien de θ^0 . Je renvoie notamment à Gleser (1981), Bickel et Ritov (1987), van der Vaart (1988), van der Vaart (1996) et Murphy et van der Vaart (1996) pour plus de détails.

Le calcul de la borne d'information ainsi que les méthodes de construction des estimateurs reposent fortement sur la linéarité en Z . Parmi les résultats plus récents sur le modèle linéaire on peut citer notamment Cui (2004), Davidov (2005), Liu et Chen (2005), Kim et Saleh (2005), Kukush *et al.* (2005), Ma (2005), Cheng et Riu (2006), You et Zhou (2007), Zhou et You (2007), Davidov et Griskin (2008). Les méthodes utilisés dans le cas linéaire ne s'appliquent pas dans le cas non linéaire.

Le cas non linéaire a été lui aussi largement étudié, mais généralement sous des hypothèses plus restrictives que celles que nous avons considérées. Pour le modèle avec mesures répétées on peut notamment citer Fuller (1987), Wolter et Fuller (1982b), (1982a). Pour des résultats avec des hypothèses supplémentaires on peut citer Hsiao (1989), Gleser (1990), Hauman *et al.* (1991, 1995), Li (2002), Schenach (2004), Kukush et Schneeweiss (2005a), (2005b), Kukush *et al.* (2005), Schneeweiss et Augustin (2006), Shklyar *et al.* (2007). Il existe également des résultats obtenus par simulation notamment sur l'algorithme SIMEX comme dans Carroll *et al.* (1996), Hsiao *et al.* (2000), (1997), Li (2000), Carroll et Stefanski (1990), Stefanski et Carroll (1993), Cook et Stefanski (1994).

De nombreuses variantes de "la méthode naïve" ont été proposées. Par exemple Gleser (1990) considère le critère des moindres carrés modifié où Z_i est remplacée par une estimation de $\mathbb{E}(Z_i|U_i)$. L'estimateur ainsi obtenu est consistant dans le cas linéaire mais ne l'est pas pour une fonction de régression générale.

À ma connaissance le premier estimateur consistant construit dans un modèle non linéaire est celui que j'ai proposé pendant ma thèse. Ce travail fait l'objet de l'article Taupin (2001) et de la note Taupin (1998) que je vais détailler dans le paragraphe suivant.

Plus récemment, Hong et Tamer (2003) ont proposé un estimateur consistant de θ^0 pour des fonctions de régression non linéaires générales, dans le cas spécifique où la transformée de Fourier de la densité des erreurs ε est de la forme $f_\varepsilon^*(t) = 1/(1 + \sigma^2 t^2)$, c'est-à-dire pour des erreurs ε de loi Laplace (ou double exponentielle). Leur méthode d'estimation dépend fortement de la forme particulière de f_ε^* et l'extension de leur méthode d'estimation pour d'autres lois d'erreurs ne semble pas acquise.

3.2.3 Première méthode d'estimation

Cette partie présente les résultats de l'article Taupin (2001) et de la note Taupin (1998). Le modèle considéré est le modèle (3.2.5) dans lequel ε est une gaussienne $\mathcal{N}(0, \sigma_\varepsilon^2)$ avec σ_ε^2 connue. La méthode d'estimation est basée sur le critère des moindres carrés modifié

$$\mathbb{E} [(Y - \mathbb{E}(f_\theta(Z)|U))^2 W(U)]$$

où W est une fonction de poids à support compact et en remarquant que sous des hypothèses standards,

$$\theta^0 = \arg \min \mathbb{E} [(Y - \mathbb{E}(f_\theta(Z)|U))^2 W(U)].$$

J'ai proposé d'estimer l'espérance conditionnelle $\mathbb{E}(f_\theta(Z)|U)$ en utilisant les observations U_1, \dots, U_n et en utilisant des outils liés à la déconvolution. Plus précisément, en écrivant l'espérance conditionnelle sous la forme

$$\mathbb{E}(f_\theta(Z)|U) = \frac{\Gamma_{f_\theta}(U)}{f_U(U)}, \text{ avec } \Gamma_f(u) = \int f(z)f_Z(z)f_\varepsilon(u-z)dz, \text{ pour } u \in \mathbb{R},$$

j'ai proposé l'estimateur suivant

$$\hat{\theta} = \arg \min_{\theta \in \Theta} S_n(\theta) \text{ avec } S_n(\theta) = n^{-1} \sum_{i=1}^n W(U_i)[Y_i - \hat{\Gamma}_{f_\theta}(U_i)/\hat{f}_U(U_i)]^2,$$

où

$$\hat{\Gamma}_{f,n}(y_0) = \int f(x)f_\varepsilon(y_0-x)\hat{f}_Z(x)dx, \quad \text{avec} \quad \hat{f}_Z(x) = \frac{1}{n} \sum_{i=1}^n K_{n,C_n}(x-U_i),$$

K_{n,C_n} étant un noyau de déconvolution défini par (2.4.9) et C_n une suite qui tend vers l'infini.

Dans l'article Taupin (2001), je montre que cet estimateur est consistant et que sa vitesse de convergence qui n'a pas de forme explicite dépend de la régularité de f_ε et de celle de la fonction de régression (comme fonction de z) au travers du ratio

$$(f_\theta f_\varepsilon(z - \cdot))^*(t)/f_\varepsilon^*(t), \text{ quand } t \text{ tend vers l'infini.}$$

La vitesse paramétrique est atteinte pour certaines fonctions de régression comme la régression polynomiale ou exponentielle.

Cette méthode a l'avantage considérable de fournir un estimateur consistant dans un cadre général, ce qui n'existait pas avant. Elle fournit également une majoration de la vitesse de convergence dans un certain nombre d'exemples classiques. Elle présente néanmoins le défaut d'être un peu complexe et surtout de ne pas fournir de vitesse de convergence explicite pour des classes de fonctions de régression.

3.2.4 Deuxième méthode d'estimation

La méthode d'estimation que je présente ici est celle de l'article Butucea et Taupin (2008) écrit en collaboration avec Cristina Butucea. Les résultats qui suivent sont issus de cet article également.

Le point de départ de notre méthode d'estimation est le critère des moindres carrés en remarquant que sous des hypothèses raisonnables

$$\theta^0 = \arg \min S_{\theta^0, f_Z}(\theta) \text{ avec } S_{\theta^0, f_Z}(\theta) = \mathbb{E}[(Y - f_\theta(Z))^2 w(Z)],$$

où w est une fonction de poids positive choisie de telle façon que pour tout θ dans Θ , wf_θ , wf_θ^2 et leur dérivées jusqu'à l'ordre 2 par rapport à θ soient intégrables.

Quand les variables $(Z_i)_{1 \leq i \leq n}$ sont observées, on peut estimer θ^0 par

$$\arg \min_{\theta \in \Theta} \frac{1}{n} \sum_{k=1}^n (Y_k - f_\theta(Z_k))^2 w(Z_k).$$

Autrement dit on estime θ^0 par l'argument minimum d'un estimateur consistant de $\mathbb{E}[(Y - f(Z))^2 w(Z)]$. On va donc estimer ce critère en utilisant les observations $(Y_1, U_1), \dots, (Y_n, U_n)$. Ceci passe par l'estimation de fonctionnelles linéaires intégrales de f_Z , de la forme

$$\mathbb{E}(\Phi(Z)) = \int \Phi(z) f_Z(z) dz \quad \text{ou} \quad \mathbb{E}(\Psi(Y)\Phi(Z)) = \iint \Psi(y)\Phi(z) f_{Y,Z}(y, z) dy dz.$$

Nous proposons d'injecter un estimateur de f_Z construit avec le noyau de déconvolution et défini par

$$\widehat{f}_Z(x) = \frac{1}{n} \sum_{k=1}^n K_{n,C_n}(x - U_k),$$

où K_{n,C_n} est un noyau de déconvolution défini par (2.4.9) avec K^* à support compact. Ainsi

$$\mathbb{E}(\widehat{\Phi(Z)}) = \langle \Phi, \widehat{f}_Z \rangle \quad \text{et} \quad \mathbb{E}(\widehat{\Psi(Y)\Phi(Z)}) = \frac{1}{n} \sum_{k=1}^n \Psi(Y_k) \int \Phi(x) K_{n,C_n}(x - U_k) dx.$$

Dès que pour tout $x \in \mathbb{R}$, $f_\varepsilon^*(x) \neq 0$, on a que pour toute fonction intégrable Φ ,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \Phi \star K_{n,C_n}(U_i) = \mathbb{E}(\Phi(Z)).$$

En effet l'étude de la partie biais nous permet décrire que

$$\lim_{n \rightarrow \infty} \mathbb{E} \left[n^{-1} \sum_{i=1}^n \Phi \star K_{n,C_n}(U_i) \right] = \lim_{n \rightarrow \infty} \frac{1}{2\pi} \int_{|t| \leq C_n} \Phi^*(t) f_Z^*(-t) dt = \mathbb{E}(\Phi(Z)).$$

De la même façon, pour toute ψ telle que $\mathbb{E}(|\psi(Y)|) < \infty$ et $\Phi \in \mathbb{L}_1(\mathbb{R})$,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \psi(Y_i) \Phi \star K_{n,C_n}(U_i) = \mathbb{E}(\psi(Y)\Phi(Z)).$$

Le critère $S_{\theta^0, f_Z}(\theta)$ est estimé par

$$S_n(\theta) = \frac{1}{n} \sum_{k=1}^n \int (Y_k - f_\theta(x))^2 w(x) K_{n,C_n}(x - U_k) dx = \frac{1}{n} \sum_{k=1}^n ((Y_k - f_\theta)^2 w) \star K_{n,C_n}(U_k).$$

Sous des hypothèses raisonnables $S_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} S_{\theta^0, f_Z}(\theta)$ pour tout $\theta \in \Theta$ et on estime θ^0 par

$$\widehat{\theta} = \arg \min_{\theta \in \Theta} S_n(\theta).$$

Pour cette construction il est nécessaire de supposer que

$$\text{la densité } f_\varepsilon \text{ appartient à } \mathbb{L}_\infty(\mathbb{R}) \text{ et que pour tout } x \in \mathbb{R}, f_\varepsilon^*(x) \neq 0. \quad (\mathbf{N}_1)$$

3.2.5 Propriétés asymptotiques

Sous des hypothèses classiques de régularité, d'identifiabilité et de moments et pour une fonction de poids judicieusement choisie, $\hat{\theta}$ est un estimateur consistant de θ^0 . En effet on montre que

$$\hat{\theta} = \arg \min_{\theta \in \Theta} S_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \arg \min_{\theta \in \Theta} S_{\theta^0, f_Z}(\theta) = \theta^0.$$

Sa vitesse de convergence dépend de la régularité de f_ε et de celle de $(f_\theta w)(z)$ comme fonction de z au travers du comportement des ratios des transformées de Fourier $(f_\theta w)^*(t)/f_\varepsilon^*(-t)$ et $(f_\theta^2 w)^*(t)/f_\varepsilon^*(-t)$ quand t tend vers l'infini. Plus précisément, on établit la majoration du risque quadratique suivante $\mathbb{E}(\|\hat{\theta} - \theta^0\|_{\ell^2}^2) = O(\varphi_n^2)$ avec $\varphi_n = \|(\varphi_{n,j})\|_{\ell^2}$, $\varphi_{n,j}^2 = B_{n,j}^2(\theta^0) + V_{n,j}(\theta^0)/n$, $j = 1 \dots, d$, où

$$B_{n,j}(\theta) = \min \left\{ B_{n,j}^{[1]}(\theta), B_{n,j}^{[2]}(\theta) \right\} \text{ et } V_{n,j}(\theta) = \min \left\{ V_{n,j}^{[1]}(\theta), V_{n,j}^{[2]}(\theta) \right\}$$

et pour $q = 1, 2$

$$B_{n,j}^{[q]}(\theta) = \left\| \left(\frac{\partial(w f_\theta)}{\partial \theta_j} \right)^* (K_{C_n}^* - 1) \right\|_q^2 + \left\| \left(\frac{\partial(w f_\theta^2)}{\partial \theta_j} \right)^* (K_{C_n}^* - 1) \right\|_q^2,$$

et

$$V_{n,j}^{[q]}(\theta) = \left\| \left(\frac{\partial(w f_\theta)}{\partial \theta_j} \right)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2 + \left\| \left(\frac{\partial(w f_\theta^2)}{\partial \theta_j} \right)^* \frac{K_{C_n}^*}{f_\varepsilon^*} \right\|_q^2.$$

Les termes $B_{n,j}^2$ et $V_{n,j}/n$ sont respectivement le carré du biais et la variance. Comme en déconvolution, le terme de variance $V_{n,j}$ est d'autant plus gros que la densité de l'erreur ε est régulière. Et le biais est lui comme d'habitude, d'autant plus petit que $(w f_\theta)$ est régulière.

On constate donc que la vitesse de convergence pour estimer θ^0 est régie par la régularité de f_ε et par celle de $f_\theta w$, avec les vitesses de convergence les plus lentes obtenues pour les lois d'erreur ε les plus régulières, par exemple pour ε gaussienne. Ces vitesses vont être améliorées par un choix judicieux de la fonction de poids w . Le choix de la fonction de poids est donc crucial. En effet la fonction de poids assure l'existence de l'estimateur au travers de l'hypothèse d'intégrabilité de $w f_\theta$. Mais elle permet aussi d'améliorer la vitesse de convergence en rendant $w f_\theta$ plus régulière que f_θ seule. En particulier l'estimateur $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 , dès qu'on peut trouver une fonction de poids w telle que $w f_\theta$ est plus régulière que f_ε au sens où les fonctions $w f_\theta$ et $w f_\theta^2$, ainsi que leurs dérivées jusqu'à l'ordre 2 par rapport à θ soient telles que

$$\sup_{\theta} w^*/f_\varepsilon^*, \quad \sup_{\theta} (f_\theta w)^*/f_\varepsilon^*, \quad \text{et} \quad \sup_{\theta} (f_\theta^2 w)^*/f_\varepsilon^* \text{ appartiennent à } \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R}). \quad (3.2.5)$$

Pour les fonctions de régression qui satisfont (3.2.5), il n'est pas forcément judicieux d'utiliser un noyau de déconvolution puisque les fonctionnelles linéaires s'estiment directement. Pour de nombreuses fonctions de régression, il est facile d'exhiber une fonction de poids w telle que (3.2.5) soit vérifiée. Ce sera notamment le cas des fonctions de régression classiques polynomiales, exponentielles,....

Néanmoins, pour certaines fonctions de régression comme $f_\theta(z) = \theta \exp(-|z|)$ il semble difficile de trouver une fonction de poids w qui assure que la vitesse de convergence paramétrique soit atteinte pour toutes les lois d'erreur f_ε . Pour ces dernières, nous établissons la consistance de l'estimateur et fournissons un ordre de grandeur du risque quadratique pour des classes de régularité classiques.

3.2.6 Vitesses de convergence pour des classes de régularité

Précisons les vitesses de convergence pour des classes de régularité usuelles, comme celles décrites par $(\mathbf{A}_1^\varepsilon)$ et (2.1.8).

Soit f_ε telle que

il existe des constantes strictement positives $\underline{C}(f_\varepsilon), \overline{C}(f_\varepsilon)$, et positives δ, α, u_0 et $\rho \leq 2$ (\mathbf{N}_2) telles que $\underline{C}(f_\varepsilon) \leq |f_\varepsilon^*(u)| |u|^\alpha \exp(\delta |u|^\rho) \leq \overline{C}(f_\varepsilon)$ pour tout $|u| \geq u_0$.

Considérons Wr_β (et ses dérivées) satisfaisant la condition (\mathbf{R}_1) suivante.

Une fonction f satisfait (\mathbf{R}_1) si f appartient à $\mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R})$ et si il existe $a, d, u'_0, r \geq 0$ tels que $0 < \underline{L}(f) \leq |f^*(u)| |u|^a \exp(d|u|^r) \leq \overline{L}(f) < \infty$ pour tout $|u| \geq u'_0$, (\mathbf{R}_1)

avec la convention que $d = 0$ si et seulement si $r = 0$.

On obtient alors le tableau de vitesses de $\hat{\theta} - \theta^0$:

		f_ε	
		$\rho = 0$ dans (\mathbf{N}_2) ordinary smooth	$\rho > 0$ dans (\mathbf{N}_2) super smooth
$w f_{\theta^0}$	$d = r = 0$ (\mathbf{R}_1) Sobolev	$\frac{a < \alpha + 1/2}{a \geq \alpha + 1/2} \quad n^{-\frac{2a-1}{2\alpha}}$ n^{-1}	$(\log n)^{-\frac{2a-1}{\rho}}$
	$r > 0$ dans (\mathbf{R}_1) \mathcal{C}^∞	n^{-1}	$r < \rho \quad (\log n)^{A(a,r,\rho)} \exp \left\{ -2d \left(\frac{\log n}{2\delta} \right)^{r/\rho} \right\}$ <hr/> $r = \rho \quad \begin{array}{ll} d < \delta & (\log n)^{A(a,r,\rho)+2\alpha d/(\delta r)} n^{-d/\delta} \\ d = \delta, a < \alpha + 1/2 & (\log n)^{(2\alpha-2a+1)/r} n^{-1} \\ d = \delta, a \geq \alpha + 1/2 & n^{-1} \\ d > \delta & n^{-1} \end{array}$ <hr/> $r > \rho \quad n^{-1}$

où $A(a, r, \rho) = (-2a + 1 - r + (1 - r)_-)/\rho$.

TAB. 3.1 – Vitesses φ_n^2

3.2.7 Exemples

L'objectif de ce paragraphe est d'illustrer sur quelques exemples les propriétés asymptotiques de l'estimateur $\hat{\theta}$. Dans tous les exemples la loi de ε est connue. Ces exemples illustrent comment trouver une fonction de poids w qui assure que $\hat{\theta}$ converge vers θ^0 à la vitesse paramétrique.

Exemple 1 Régression polynomiale. Soient f_θ de la forme $f_\theta(x) = \sum_{k=1}^p \theta_k x^k$ et $w(x) = \exp\{-x^2/(4\beta)\}$. Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

Dans le cas linéaire, il existe d'autres estimateurs qui convergent à la vitesse paramétrique vers θ^0 . (voir l'introduction de cette partie pour les références). Dans le cas polynomial la \sqrt{n} -consistance était déjà connue (voir Taupin (2001), Comte et Taupin (2001), Kukush *et. al.* (2005), Shklyar *et. al.* (2007) ou aussi dans le modèle dit "fonctionnel" ou les $(Z_i)_{1 \leq i \leq n}$ sont fixes et non plus aléatoires (voit Hausman *et al.* (1995), (1991) ou Chan et Mak (1985)).

Exemple 2 Régression exponentielle. Soient f_θ de la forme $f_\theta(x) = \exp(\theta x)$ et $w(x) = \exp\{-x^2/(4\beta)\}$. Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

Exemple 3 Combinaison linéaire de cosinus. Soient f_θ de la forme $f_\theta(x) = \sum_{j=1}^d \theta_j \cos(jx)$ et $w(x) = \exp\{-x^2/(4\beta)\}$. Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 . Ici la vitesse de convergence de $\hat{\theta}$ est plus rapide que celle obtenue dans Taupin (2001).

Exemple 4 Régression Cauchy Soient $f_\theta(x) = \theta/(1+x^2)$ et $w(x) = (1+x^2)^4 \exp\{-x^2/(4\beta)\}$. Avec ce choix de w , $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

Cet exemple souligne l'importance du choix de la fonction de poids. En effet sans la fonction de poids (c'est-à-dire en prenant $w \equiv 1$), pour ε gaussienne, la vitesse de convergence prédite est $\exp(-\sqrt{\log n})$ au lieu de la vitesse \sqrt{n} que nous obtenons. Cet exemple est à relier aux conditions données dans Fazekas et Kukush (1998), Fazekas *et al.* (1999) et Baran (2000) dans un contexte légèrement différent. Il considèrent le modèle "fonctionnel" où les variables $(Z_i)_{1 \leq i \leq n}$ ne sont pas aléatoires et construisent un estimateur \sqrt{n} -consistant sous l'hypothèse qu'il existe deux fonctions ϕ_1 et ϕ_2 telles que

$$\mathbb{E}(\phi_1(x + \varepsilon, \theta)) = f_\theta(x) \text{ et } \mathbb{E}(\phi_2(x + \varepsilon, \theta)) = f_\theta^2(x). \quad (3.2.1)$$

Nos conditions pour assurer que la vitesse paramétrique est atteinte sont nettement moins restrictives du fait de la présence de la fonction de poids w . En effet dans l'exemple ci-dessus, les conditions (3.2.1) ne sont pas satisfaites et donc l'estimateur qu'ils proposent n'est pas valide ici, alors que notre méthode fournit un estimateur \sqrt{n} -consistant.

Exemple 5 Fonction de régression exponentielle symétrique Soit $f_\theta(x) = \theta \exp(-|x|/2)$.

- Si ε satisfait (\mathbf{N}_2) avec $0 \leq \rho < 1$ alors on peut trouver une fonction de poids w telle que $\hat{\theta}$ soit un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

- Si ε satisfait (\mathbf{N}_2) avec $\rho \geq 1$, alors $\mathbb{E} \|\hat{\theta} - \theta^0\|_{\ell^2}^2 = O(1) (\log n)^{\frac{1-2a-r}{\rho}} \exp\{-2d(\log n/(2\beta))^{r/\rho}\}$, avec r aussi proche que l'on veut de 1. Il est important de noter que dans cet exemple, la vitesse obtenue est nettement plus rapide que la vitesse logarithmique obtenue par Taupin (2001).

3.2.8 Perspectives

La procédure d'estimation repose sur un "lissage" de la fonction de régression en la multipliant par une fonction de poids. Or pour certaines fonctions de régression, on ne peut pas exhiber de fonction de poids qui rende $w f_\theta$ plus régulière que f_ε . Pour ces fonctions de régression, la vitesse de convergence pour l'estimation de θ^0 est plus lente que la vitesse paramétrique. Néanmoins cette vitesse est nettement améliorée par rapport aux méthodes existantes et ce grâce à la fonction de poids. La question qui reste ouverte est donc de savoir si la vitesse paramétrique est atteignable pour toutes les fonctions de régression, sous des conditions raisonnables. Une étude des propriétés pratiques de l'estimateur fait l'objet d'une travail en cours avec Cristina Butucea et Jean-Michel Marin. Elle montre déjà les bonnes propriétés des estimateurs dans des cadres variés.

3.3 Estimation dans un modèle autorégressif avec erreurs sur les variables

3.3.1 Le modèle, les résultats antérieurs

Cette partie présente les résultats de l'article Comte et Taupin (2001). Considérons le modèle autorégressif suivant

$$\begin{cases} Z_i = f_{\theta^0}(Z_{i-1}) + \xi_i, \\ U_i = Z_i + \varepsilon_i, \end{cases} \quad (3.3.2)$$

avec ε_i de densité f_ε connue et indépendante de Z_i et avec ξ admettant une densité inconnue. Dans ce modèle, la fonction de régression appartient à une famille paramétrique

$$\mathcal{F} = \{f_\theta, \theta \in \Theta \text{ compact de } \mathbb{R}^d\}.$$

L'objectif est d'estimer le paramètre θ^0 , en utilisant les observations U_1, \dots, U_n . On suppose que les variables ξ_i et ε_i sont i.i.d. et que les suites $(\xi_i)_{i \in \mathbb{Z}}$ et $(\varepsilon_i)_{i \in \mathbb{Z}}$ sont mutuellement indépendantes. On suppose également que les suites $(Z_i)_{i \in \mathbb{Z}}$ et $(\varepsilon_i)_{i \in \mathbb{Z}}$ sont mutuellement indépendantes. Ce modèle autorégressif avec erreurs sur les variables est un cas particulier de chaîne de Markov cachée, à espace d'état continu et non compact, dont la loi n'est pas complètement paramétrique du fait de la non connaissance de la loi de ξ_i .

Quand la chaîne de Markov est à espace d'états compact avec une transition paramétrique, Leroux (1992), Bickel *et al.* (1998), Jensen et Petersen (1999), Douc et Matias (2001), Douc *et al.* (2004) obtiennent des résultats de \sqrt{n} -consistance et de normalité asymptotique. Ces résultats ne s'appliquent pas au modèle (3.3.2) car la non connaissance de la loi de ξ implique que la transition de la chaîne $(Z_i)_{i \in \mathbb{N}}$ n'est pas complètement paramétrique et que la chaîne n'est pas forcément à espace d'états compact.

Des résultats plus récents comme ceux de Genon-Catalot et Laredo (2006) considèrent un espace d'état non compact.

Quand la fonction de régression est linéaire, on peut citer les travaux de Hannan (1963); Nowak (1985), Anderson et Deistler (1984) ou Chanda (1995). Il n'y a pas à ma connaissance d'estimateur consistant dans le modèle (3.3.2). Notre objectif est donc de proposer un estimateur consistant de θ^0 , pour une fonction de régression générale, avec ξ de loi inconnue et pour une loi de ε connue générale.

3.3.2 Procédure d'estimation

La procédure d'estimation est basée sur le critère

$$\mathbb{E} [(U_i - \mathbb{E}(f_\theta(Z_{i-1})|U_{i-1}))^2 W(U_{i-1})] \quad (3.3.3)$$

où W est une fonction de poids à support compact et en remarquant que sous des hypothèses standards,

$$\theta^0 = \arg \min \mathbb{E} [(U_i - \mathbb{E}(f_\theta(Z_{i-1})|U_{i-1}))^2 W(U_{i-1})].$$

Si f_Z était connue, on pourrait estimer θ^0 par

$$\arg \min \left[\frac{1}{n} \sum_{i=1}^n W(U_{i-1}) (U_i - \mathbb{E}(f_\theta(Z_{i-1})|U_{i-1}))^2 \right]. \quad (3.3.4)$$

Comme la densité f_Z est inconnue et que les variables $(Z_i)_{1 \leq i \leq n}$ ne sont pas observées nous proposons d'estimer l'espérance conditionnelle $\mathbb{E}(f_\theta(Z_{i-1})|U_{i-1})^2 W(U_{i-1})$ en utilisant des outils liés à la déconvolution et des idées analogues à celles décrites dans la partie 3.2.3. Plus précisément, θ^0 est estimée

par

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left[n^{-1} \sum_{i=1}^n W(U_{i-1}) [U_i - \hat{\Gamma}_{f_\theta}(U_{i-1}) / \widehat{f_U}(U_{i-1})]^2 \right],$$

où

$$\widehat{\Gamma}_{f,n}(U_{i-1}) = \int f(x) f_\varepsilon(U_{i-1} - x) \widehat{f_Z}(x) dx, \quad \text{avec} \quad \widehat{f_Z}(x) = \frac{1}{n} \sum_{k=1}^n K_{n,C_n}(x - U_k),$$

K_{n,C_n} étant un noyau de déconvolution défini par (2.4.9) et C_n une suite qui tend vers l'infini et où les sommes portent sur les $j \neq i - 1$.

3.3.3 Résultats

Nous montrons que l'estimateur ainsi construit est consistant et que sa vitesse de convergence est la même que celle obtenue dans Taupin (2001), dans le cas i.i.d. Les hypothèses sous lesquelles nous démontrons ces résultats sont des hypothèses classiques de moments, de régularité et d'identifiabilité, auxquelles nous ajoutons des hypothèses assurant la stationnarité de la suite $(Z_i)_{i \in \mathbb{Z}}$, l'existence d'une solution stationnaire unique et assurant que la suite $(Z_i)_{i \in \mathbb{Z}}$ est β -mélangeante au sens de Rozanov et Volkonskii (1960).

3.3.4 Perspectives

Cette méthode a l'avantage considérable de fournir un estimateur consistant dans un cadre autorégressif général, ce qui n'existait pas avant dans ce modèle qui a des applications évidents notamment en pharmacocinétique. Elle présente néanmoins l'inconvénient d'être un peu compliquée et assez peu explicite, surtout au niveau de l'expression des vitesses de convergence. Il serait donc intéressant de proposer une méthode plus directe, qui fournisse des vitesses plus explicites et ceci sous des conditions de dépendance plus faibles.

Chapitre 4

Modèles de survie avec erreurs sur les variables

Ce chapitre présente l'article en collaboration Martin-Magniette et Taupin (2008).

4.1 Introduction, résultats antérieurs et modèle considéré

Introduction

On s'intéresse à la relation entre un temps de survie (ou durée de vie) T et une covariable $Z \in \mathbb{R}$. On appelle usuellement durée de vie ou temps de survie la durée qui s'écoule depuis un instant donné jusqu'à la survenue d'un évènement particulier. Cet évènement particulier peut être la mort, la guérison, la rémission, ... L'instant de départ peut être le début de l'étude, le début du traitement, le début de la maladie, la date de diagnostic de la maladie, ... Cette durée de vie, notée T , est une variable aléatoire continue et positive. En analyse de survie sa loi est souvent décrite soit par la fonction de survie, par exemple quand on fait de l'estimation non paramétrique, soit par le risque instantané. Ce risque instantané, aussi appelé taux d'incidence ou fonction de hasard est la probabilité que l'évènement survienne dans un petit intervalle de temps après t , conditionnellement au fait qu'il n'ait pas eu lieu jusqu'à l'instant t . Il est formellement défini par

$$R(t) = \lim_{\Delta_t \rightarrow 0} \frac{1}{\Delta_t} \mathbb{P}(t < T \leq t + \Delta_t | T > t) = \frac{f_T(t)}{S(t)} = \frac{f_T(t)}{1 - F_T(t)},$$

où S désigne la fonction de survie et F_T et f_T désignent respectivement la fonction de répartition et la densité de T . La relation entre le temps de survie T et la covariable Z est souvent décrite par le risque instantané conditionnel à la valeur de la covariable Z . On peut l'écrire formellement

$$R(t|z) = \lim_{\Delta_t \downarrow 0} \frac{1}{\Delta_t} \mathbb{P}(t < T \leq t + \Delta_t | T > t, Z = z).$$

Modèles à risque proportionnel Dans ce chapitre je m'intéresse à l'estimation des paramètres dans un modèle à risque proportionnel c'est-à-dire où le risque instantané s'écrit comme le produit d'un risque de base et d'un risque relatif :

$$R(t|Z) = \alpha_0(t)r_\beta(Z).$$

Le risque de base α_0 ne dépend que du temps et il correspond au risque pour un individu dans les conditions "standards". Le risque relatif $r_\beta(Z)$ ne dépend lui pas du temps, mais uniquement de la covariable Z et d'un paramètre inconnu β . Pour assurer que α_0 est le risque de base, on impose que

$r_\beta(0) = 1$. Dans ce contexte $\alpha_0(t)$ représente donc le risque instantané de T conditionnellement à $Z = 0$. Pour assurer que le risque instantané est positif on impose que $r_\beta(z) \geq 0$ pour tous les z appartenant au support de la loi de Z et pour tout β dans un certain compact \mathbb{B} .

Citons en particulier deux modèles à risque proportionnel classiquement utilisés en analyse de survie. Le premier est le modèle de Cox, introduit par Cox (1972) (voir aussi Kalbfleisch et Prentice (1973) ou Prentice et Kalbfleisch (1979)). Ce modèle s'écrit $R(t|Z) = \alpha_0(t) \exp(\beta_0 Z)$. La structure exponentielle du modèle de Cox assure la positivité du risque instantané sans hypothèse supplémentaire. De plus, le rapport des risques instantanés ne fait intervenir que l'écart $Z - Z'$.

Le deuxième modèle est le modèle d'excès de risque relatif. Il s'écrit $R(t|Z) = \alpha_0(t)(1 + \beta_0 Z)$. Pour le modèle d'excès de risque relatif, il est nécessaire de faire une hypothèse supplémentaire. On va imposer que $(1 + \beta_0 z) \geq 0$ pour tout β dans un compact de \mathbb{R} et pour tous les z dans le support de la loi de Z .

L'intérêt des modèles à risque proportionnel est de séparer l'influence du temps de l'influence des covariables. Ainsi pour deux individus de vecteurs de covariables respectifs Z et Z' , le rapport des risques instantanés ne dépend pas du temps et est égal au ratio des risques relatifs.

La censure Notons $i = 1, \dots, n$ et T_i le temps de survie de l'individu i . Il est très fréquent que les temps de survie des n individus ne soient pas tous disponibles. On parle alors de données censurées. Il existe plusieurs types de censure, mais nous nous centrerons ici sur le type de censure le plus classique : la censure aléatoire à droite. Pour décrire mathématiquement cette censure aléatoire à droite, on peut introduire une variable C_i représentant le temps de censure de l'individu i . Pour $i = 1, \dots, n$ on observe

$$X_i = \min(T_i, C_i) \text{ et } D_i = \mathbb{1}_{T_i \leq C_i}.$$

Ainsi, pour un individu i , soit $C_i \geq T_i$, et on observe son temps de survie T_i et $D_i = 1$; la donnée n'est pas censurée. Soit $T_i > C_i$, la censure est arrivée avant l'évènement d'intérêt et on n'observe pas son temps de survie, mais uniquement son temps de censure C_i . L'indicateur de censure D_i vaut alors $D_i = 0$; la donnée i est alors dite censurée. Dans ce cas là, on sait juste que son temps de survie est d'au moins C_i . Les causes de censure aléatoires à droite sont multiples. Si T représente la durée jusqu'au décès d'une maladie précise, on peut citer par exemple : les patients perdus de vue ou sortis de l'étude, les patients décédés d'une autre cause que la maladie considérée, et les patients qui ne sont pas décédés à la fin de l'étude. On suppose souvent, que la censure est non informative. Autrement dit que le processus de censure n'est pas lié à l'évènement d'intérêt. Cette hypothèse faite pour simplifier le modèle a évidemment ses limites. Malgré cette réserve, nous allons considérer ici une censure aléatoire à droite, et non informative.

La première idée qui vient à l'esprit est d'enlever les données censurées et de ne considérer que les données non censurées c'est-à-dire les données telles que $D_i = 1$. Or il ne faut pas enlever les données censurées, et ce pour plusieurs raisons. La première raison est une éventuelle perte considérable d'information. La deuxième raison est que l'on risque de sous ou surévaluer l'effet d'une covariable. La troisième raison est une raison plus technique qui renforce les deux premières raisons : les estimateurs construits uniquement à partir des données non censurées sont biaisés. Par exemple on peut montrer que si S_C est la fonction de survie de la variable "censure" C ,

$$\widehat{S}_b(t) = \frac{1}{\sum_{i=1}^n D_i} \sum_{i=1}^n \mathbb{1}_{T_i > t, D_i = 1} \xrightarrow{n \rightarrow \infty} \int_t^\infty f_T(x) S_C(x) dx \neq S(t).$$

Cas ou les covariables sont observées Supposons que l'on observe Z_1, \dots, Z_n , i.i.d. sur un intervalle d'étude $[0, \tau]$ avec $\tau < \infty$. L'objectif est alors d'estimer α_0 et β_0 en utilisant les observations

$$(X_i, D_i, Z_i) = (\min(T_i, C_i), \mathbb{1}_{T_i \leq C_i}, Z_i), \quad i = 1, \dots, n.$$

Quand le risque de base α_0 appartient à une famille paramétrique, on peut estimer tous les paramètres du risque instantané par maximum de vraisemblance. Quand on ne s'intéresse qu'à l'estimation du paramètre du risque relatif β^0 ou que le risque de base n'a pas de forme paramétrique connue, on utilise la vraisemblance partielle de Cox, appelée aussi pseudo-vraisemblance. En notant $Z^{(n)} = (Z_1, \dots, Z_n)$, cette vraisemblance partielle s'écrit

$$L_n(\beta, Z^{(n)}) = \prod_{i=1}^n \left(\frac{r_\beta(Z_i)}{\sum_{j: X_j \geq X_i} r_\beta(Z_j)} \right)^{D_i}.$$

Et la log-vraisemblance partielle s'écrit

$$\begin{aligned} \ell_n(\beta, Z^{(n)}) &= \sum_{i=1}^n D_i \ln(r_\beta(Z_i)) - \sum_{i=1}^n D_i \ln\left[\sum_{j: X_j \geq X_i} r_\beta(Z_j) \right] \\ &= \sum_{i=1}^n \ln\left(\frac{r_\beta(Z_i)}{S_n(\beta, X_i)} \right) N_i(\tau) = \sum_{i=1}^n \int_0^\tau \ln\left(\frac{r_\beta(Z_i)}{S_n(\beta, s)} \right) dN_i(s), \end{aligned} \quad (4.1.1)$$

avec $N_i(t) = \mathbb{1}_{X_i \leq t, D_i=1}$, $Y_i(t) = \mathbb{1}_{X_i \geq t}$, et pour $0 \leq s \leq \tau$, $S_n(\beta, s) = n^{-1} \sum_{j=1}^n r_\beta(Z_j) Y_j(s)$. Le paramètre β_0 peut être estimé par

$$\hat{\beta} = \arg \max_{\beta \in \mathbb{R}^p} \ell_n(\beta, Z^{(n)}),$$

et sous des conditions standards de régularité, on peut montrer que $\hat{\beta}$ est solution de $\ell_n^{(1)}(\hat{\beta}, Z^{(n)}) = 0$,

$$\text{avec} \quad \ell_n^{(1)}(\beta, Z^{(n)}) = \frac{\partial}{\partial \beta} \ell_n(\beta, Z^{(n)}) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau \left(\frac{r_\beta^{(1)}(Z_i)}{r_\beta(Z_i)} - \frac{\sum_{j=1}^n Y_j(t) r_\beta^{(1)}(Z_j)}{\sum_{j=1}^n Y_j(t) r_\beta(Z_j)} \right) dN_i(t), \quad (4.1.2)$$

où $r_\beta^{(1)}$ est la dérivée première de r_β par rapport à β . L'estimateur ainsi construit est \sqrt{n} -consistant, asymptotiquement gaussien et efficace. Je renvoie à Gill et Andersen (1982) et Andersen *et al.* (1993) pour plus de détails.

La vraisemblance partielle permet donc de séparer l'estimation de β^0 de celle du risque de base (ou des paramètres dont il dépend) en fournissant une critère d'estimation de paramètre β^0 dans lequel le risque de base n'intervient pas.

Le modèle

Considérons le problème de l'estimation de $\theta^0 = (\beta^0, \gamma^0)^\top$ dans un modèle de risque instantané paramétrique de la forme

$$R_{\theta^0}(t|Z) = \eta_{\gamma^0}(t) r_{\beta^0}(Z),$$

quand les covariables Z_1, \dots, Z_n sont mesurées avec erreurs. Plus précisément, on observe U_1, \dots, U_n où Z_i et U_i sont liées par la relation

$$U_i = Z_i + \varepsilon_i,$$

avec ε de densité f_ε connue. L'objectif est alors d'estimer les paramètres en utilisant les observations

$$(X_i, D_i, U_i) = (\min(T_i, C_i), \mathbb{1}_{T_i \leq C_i}, U_i), \text{ avec } U_i = Z_i + \varepsilon_i.$$

La covariable Z , non observée et indépendante du temps est de densité f_Z inconnue. Nous allons supposer ici que les fonctions η_γ et r_β appartiennent toutes les 2 à des familles paramétriques et que $\theta^0 = (\beta^0, \gamma^0)^\top$ appartient à l'intérieur d'un compact $\Theta = \mathbb{B} \times \Gamma \subset \mathbb{R}^{m+p}$. Pour assurer la positivité du risque instantané on suppose que $\eta_\gamma(t) \geq 0$ pour tout $\gamma \in \Gamma$ et tous $t \in [0, \tau]$, $\tau < \infty$, et aussi que $r_\beta(Z) \geq 0$ pour

tous $\beta \in \mathbb{B}$ et Z de densité f_Z . La modélisation paramétrique du risque instantané présente certains avantages. En particulier elle permet d’avoir une interprétation de la valeur des coefficients.

Parmi les modèles paramétriques les plus connus, on peut citer les modèles exponentiels où $R_\theta(t|Z) = \gamma r(\beta^\top Z)$, les modèles de Weibull avec $R_\theta(t|Z) = \gamma_1 \gamma_2 t^{\gamma_2 - 1} r(\beta^\top Z)$, les modèles avec risque de base constants par morceaux et les modèles de Gompertz-Makeham avec $R_\theta(t|Z) = (\gamma_1 + \gamma_2 (\gamma_3)^t) r(\beta^\top Z)$. Ce dernier est souvent utilisé dans les analyses de cause de mortalité (voir Tyurin *et. al.* (1995)). On renvoie à Andersen *et. al.* (1993), Cox et Oakes (1984) et Hosmer et Lemeshow (1999) pour des discussions sur les modèles paramétriques et leurs avantages en analyse de survie.

Résultats antérieurs

L’intérêt pour les modèles de survie quand les covariables sont mesurées avec erreurs est plus récent que pour le modèle de régression. La littérature sur le sujet est néanmoins relativement abondante. Mais il n’y a pas à ma connaissance de résultat de consistance pour l’estimation des paramètres dans un modèle à risque relatif paramétrique général quand les covariables sont mesurées avec erreur, que le risque de base soit paramétrique ou non paramétrique. Tous les résultats antérieurs dans ce domaine concernent le modèle de Cox semi-paramétrique.

La première idée qui vient à l’esprit pour prendre en compte que Z est observée avec erreur consiste à remplacer $Z^{(n)}$ par $U^{(n)}$ dans la vraisemblance partielle (4.1.1) ou dans sa dérivée (4.1.2). Cette idée, analogue de la méthode naïve en régression fournit un estimateur biaisé non consistant (voir par exemple Nakamura (1990, 1992)). On peut également citer Li et Ryan (2004) qui ont étudié le biais dans le cas de covariables hétérogènes mesurées avec erreurs. Une seconde idée proche de la calibration est de proposer des corrections du critère d’estimation (4.1.1) (ou de sa dérivée) obtenues en remplaçant Z_i par une approximation de $\mathbb{E}(Z_i|U_i)$. Une troisième idée proposée par Prentice (1982), Tsiatis (1995) et Xie *et. al.* (2001) est d’approximer la vraisemblance partielle liée à la filtration générée par les observations $\mathcal{E}_t = \sigma\{U, N(s), \mathbb{I}_{X>s}, 0 \leq s \leq t\}$. Toutes ces méthodes fournissent des estimateurs biaisés et non consistants, à la fois dans le modèle de Cox, mais aussi dans des modèles plus généraux. On peut également citer des résultats obtenus dans le modèle de Cox, sous des hypothèses différentes (répétitions, loi de Z paramétrique,...). Parmi ceux là citons Hu *et. al.* (1998) et Dupuy (2004).

Une dernière idée a été beaucoup utilisée dans le modèle de Cox semi-paramétrique. Elle consiste à corriger la fonction de score partielle (4.1.2) dans laquelle on a remplacé Z_i par U_i . L’estimateur $\tilde{\beta}$ tel que $\ell_n^{(1)}(\tilde{\beta}, U^{(n)}) = 0$ n’est pas consistant mais, dans le modèle de Cox, on peut construire des corrections de $\ell^{(1)}(\beta, U^{(n)})$ permettant d’avoir la consistance. Parmi les travaux basés sur cette idée, on peut citer Kong (1999), Buzas (1998), Stefanski (1989), Nakamura (1990, 1992), Kong et Gu (1999), Hu et Lin (2002, 2004), Huang et Wang (2000), Augustin (2004), Song et Huang (2005), et plus récemment Li et Ryan (2006). Toutes ces corrections sont très liées à la forme exponentielle du risque relatif dans le modèle de Cox. En effet quand $U = Z + \varepsilon$, avec Z indépendante de ε , alors $\lim_{n \rightarrow \infty} \mathbb{E}[\ell_n^{(1)}(\beta, Z^{(n)})]$ ne dépend que de $\mathbb{E}(Z)$ et de $\mathbb{E}[\exp(\beta Z)]$, avec $\mathbb{E}(Z) = \mathbb{E}(U)$ et $\mathbb{E}[\exp(\beta U)] = \mathbb{E}[\exp(\beta Z)]\mathbb{E}[\exp(\beta \varepsilon)]$. Ces méthodes adaptées au modèle de Cox, nécessitent que ε admette une transformée de Laplace et ne peuvent à priori pas s’étendre à des risques relatifs généraux. Par exemple, si on considère le modèle d’excès de risque relatif, $r_\beta(Z) = (1 + \beta Z)$, on peut montrer que $\lim_{n \rightarrow \infty} \mathbb{E}[\ell_n^{(1)}(\beta, Z^{(n)})]$ dépend de $\mathbb{E}[Z/(1 + \beta Z)]$ tandis que $\lim_{n \rightarrow \infty} \mathbb{E}[\ell_n^{(1)}(\beta, U^{(n)})]$ dépend de $\mathbb{E}[U/(1 + \beta U)]$. Or le modèle $U = Z + \varepsilon$ ne fournit pas d’expression de $\mathbb{E}[Z/(1 + \beta Z)]$ en fonction de $\mathbb{E}[U/(1 + \beta U)]$. Par conséquent une correction analogue à celle proposée dans le modèle de Cox ne semble pas pouvoir être mise en évidence. En d’autres termes, il ne semble pas possible d’exhiber une fonction $\Psi_n(\beta, U)$ indépendante de la densité de Z et qui satisfasse $\lim_{n \rightarrow \infty} \mathbb{E}(\Psi_n(\beta, U)) = \mathbb{E}[Z/(1 + \beta Z)]$.

Les méthodes existantes qui fournissent des estimateurs consistants sont basées sur des corrections de la vraisemblance partielle qui n’imposent donc pas de forme particulière au risque de base, mais qui ne fonctionnent que dans le modèle de Cox et avec une hypothèse assez contraignante sur la loi des erreurs.

Nous proposons une méthode d'estimation de $\theta^0 = (\beta^0, \gamma^0)^\top$ quand le risque relatif et le risque de base appartiennent à des familles paramétriques, quand f_Z est inconnue et que la densité f_ε est connue mais générale. Comme je l'ai mentionné, le critère d'estimation usuel est la pseudo-vraisemblance, en particulier parce qu'il présente l'avantage de séparer l'estimation du risque de base de celle du paramètre de régression β^0 . Néanmoins d'autres critères d'estimation existent et nous proposons d'utiliser le critère des moindres carrés. Notre méthode d'estimation, relativement inhabituelle en survie puisqu'elle utilise à la fois le critère des moindres carrés et des outils liés à la déconvolution, fournit un estimateur consistant pour un modèle à risque proportionnel paramétrique général avec f_ε générale connue, et sans restriction sur f_Z . Cet estimateur est \sqrt{n} -consistant et asymptotiquement gaussien pour des risques relatifs classiques comme dans le modèle de Cox ou le modèle d'excès de risque relatif. Dans les autres cas, nous fournissons une majoration du risque quadratique de l'estimateur pour un risque relatif et une densité f_ε appartenant à des classes de régularité classiques. Un tel résultat dans ce modèle est nouveau et ne semble pas pouvoir être obtenu par les méthodes qui ont été développées antérieurement dans le modèle de Cox avec covariable mesurée avec erreur.

4.2 Procédure d'estimation

Le point de départ de notre procédure d'estimation est le critère des moindres carrés

$$S_{\theta^0, f_Z}(\theta) = \mathbb{E}\left((r_\beta^2 W)(Z) \int_0^\tau Y(t) \eta_\gamma^2(t) dt\right) - 2\mathbb{E}\left((r_\beta W)(Z) \int_0^\tau \eta_\gamma(t) dN(t)\right)$$

avec W une fonction de poids positive à choisir, $N(t) = \mathbb{I}_{X \leq t, D=1}$ et $Y(t) = \mathbb{I}_{X \geq t}$. Comme l'intensité du processus censuré $N(t)$ par rapport à $\mathcal{F}_t = \sigma\{Z, U, N(s), \mathbb{I}_{X \geq s}, 0 \leq s \leq t \leq \tau\}$ est égale à $\lambda(t, \theta^0, Z) = \eta_{\gamma^0}(t)Y(t)r_{\beta^0}(Z)$, on peut réécrire $S_{\theta^0, f_Z}(\theta)$ comme

$$S_{\theta^0, f_Z}(\theta) = \int_0^\tau \left\{ \mathbb{E}\left[(\eta_\gamma(t)r_\beta(Z) - \eta_{\gamma^0}(t)r_{\beta^0}(Z))^2 Y(t)W(Z)\right] - \mathbb{E}\left[(\eta_{\gamma^0}(t)r_{\beta^0}(Z))^2 Y(t)W(Z)\right] \right\} dt.$$

Ceci établit que sous des hypothèses raisonnables, $S_{\theta^0, f_Z}(\theta)$ est minimum quand $\theta = \theta^0$ dès que W est positive. Ce critère a été, bien que rarement, utilisé en analyse de survie ou dans des domaines proches. Voir notamment Reynaud-Bouret (2003) pour l'estimation de l'intensité d'un processus de Poisson.

Quand les $(Z_i)_{1 \leq i \leq n}$ sont observées, on peut estimer θ^0 par

$$\arg \min \left\{ -\frac{2}{n} \sum_{i=1}^n (r_\beta W)(Z_i) \int_0^\tau \eta_\gamma(t) dN_i(t) + \frac{1}{n} \sum_{i=1}^n (r_\beta^2 W)(Z_i) \int_0^\tau \eta_\gamma^2(t) Y_i(t) dt \right\}. \quad (4.2.3)$$

On peut donc estimer θ^0 par l'argument minimum d'un estimateur consistant de $S_{\theta^0, f_Z}(\theta)$.

L'étape suivante consiste donc à construire un estimateur de $S_{\theta^0, f_Z}(\theta)$ en utilisant les observations (X_i, D_i, U_i) , pour $i = 1, \dots, n$. En utilisant des outils liés à la déconvolution nous proposons d'estimer S_{θ^0, f_Z} par

$$S_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left[(r_\beta^2 W) \star K_{n, C_n}(U_i) \int_0^\tau \eta_\gamma^2(t) Y_i(t) dt - 2(r_\beta W) \star K_{n, C_n}(U_i) \int_0^\tau \eta_\gamma(t) dN_i(t) \right]$$

avec $K_{n, C_n}(\cdot) = C_n K_n(C_n \cdot)$ défini par (2.4.9), K^* à support compact, et C_n une suite qui tend vers l'infini. La fonction de poids W sera choisie de telle sorte que pour tout β dans B , Wr_β , Wr_β^2 et leur dérivées jusqu'à l'ordre 2 par rapport à β soient intégrables. Pour cette construction on suppose que f_ε satisfait (\mathbf{N}_1) .

Sous des hypothèses raisonnables $S_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} S_{\theta^0, f_Z}(\theta)$ pour tout $\theta \in \Theta$ et on estime θ^0 est par

$$\hat{\theta} = (\hat{\beta}, \hat{\gamma})^\top = \arg \min_{\theta = (\beta, \gamma)^\top \in \Theta} S_n(\theta). \quad (4.2.4)$$

4.3 Propriétés asymptotiques

Sous des hypothèses classiques de régularité, d'identifiabilité et de moments et pour une fonction de poids judicieusement choisie, $\hat{\theta}$ est un estimateur consistant de θ^0 . En effet on montre que

$$\hat{\theta} = (\hat{\beta}, \hat{\gamma})^\top = \arg \min_{\theta \in \Theta} S_n(\theta) \xrightarrow[n \rightarrow \infty]{\mathbb{P}} \arg \min_{\theta \in \Theta} S_{\theta^0, f_Z}(\theta) = \theta^0.$$

Sa vitesse de convergence dépend de la régularité de f_ε et de celle de $(r_\beta W)(z)$ comme fonction de z . Plus précisément, la vitesse de convergence de $\hat{\theta}$ dépend du comportement du ratio des transformées de Fourier $(r_\beta W)^*(t)/f_\varepsilon^*(-t)$ et $(r_\beta^2 W)^*(t)/f_\varepsilon^*(-t)$ quand t tend vers l'infini. Nous établissons une borne supérieure du risque quadratique $\mathbb{E} \|\hat{\theta} - \theta^0\|_{\ell^2}^2$ de la forme

$$\mathbb{E}(\|\hat{\theta}_1 - \theta^0\|_{\ell^2}^2) = O(\varphi_n^2) \quad \text{avec } \varphi_n^2 = \|(\varphi_{n,j})\|_{\ell^2}^2, \quad \varphi_{n,j}^2 = B_{n,j}^2(\theta^0) + V_{n,j}(\theta^0)/n,$$

$$B_{n,j}^2(\theta^0) = \min\{B_{n,j}^{[1]}(\theta^0), B_{n,j}^{[2]}(\theta^0)\}, \quad V_{n,j}(\theta^0) = \min\{V_{n,j}^{[1]}(\theta^0), V_{n,j}^{[2]}(\theta^0)\}, \quad j = 1, \dots, m$$

$$\text{où } B_{n,j}^{[q]}(\theta^0) = \|(r_{\beta^0}^2 W)^*(K_{C_n}^* - 1)\|_q^2 + \|(r_{\beta^0} W)^*(K_{C_n}^* - 1)\|_q^2 + \|(r_{\beta^0, j}^{(1)} W)^*(K_{C_n}^* - 1)\|_q^2 \\ + \|(r_{\beta^0, j}^{(1)} r_{\beta^0} W)^*(K_{C_n}^* - 1)\|_q^2,$$

$$\text{et } V_{n,j}^{[q]}(\theta^0) = \|(r_{\beta^0}^2 W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\|_q^2 + \|(r_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\|_q^2 + \|(r_{\beta^0, j}^{(1)} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\|_q^2 + \|(r_{\beta^0, j}^{(1)} r_{\beta^0} W)^* \frac{K_{C_n}^*}{f_\varepsilon^*}\|_q^2.$$

Les termes $B_{n,j}^2$ et $V_{n,j}$ sont le carré du biais et la variance. Comme dans un contexte d'estimation non paramétrique, le biais est d'autant plus petit que $(Wr_\beta)(z)$ et $\partial(r_\beta W)(z)/\partial\beta$ sont régulières comme fonction de z . Comme en déconvolution la variance est d'autant plus importante que la densité des erreurs ε est régulière. Par conséquent, les vitesses de convergence les plus lentes vont être obtenues pour les densité de bruit les plus régulières, par exemple pour ε gaussienne. Par contre un choix judicieux de la fonction de poids W va permettre d'améliorer la vitesse de convergence de $\hat{\theta}$.

4.4 Choix de la fonction de poids et \sqrt{n} -consistance

Le choix de W est crucial, mais elle doit au minimum satisfaire les conditions suivantes

$$\sup_{\beta \in \mathbb{B}} (Wr_\beta), \quad W \text{ et } \sup_{\beta \in \mathbb{B}} (Wr_\beta^2) \text{ appartient à } \mathbb{L}_1(\mathbb{R}). \quad (\mathbf{A}_3)$$

$$\sup_{\beta \in \mathbb{B}} (Wr_\beta^{(1)}) \text{ et } \sup_{\beta \in \mathbb{B}} (Wr_\beta r_\beta^{(1)}) \text{ appartient à } \mathbb{L}_1(\mathbb{R}). \quad (\mathbf{A}_4)$$

Ces conditions assurent l'existence de l'estimateur. En réalité la fonction de poids W va être choisie de façon à ce que Wr_β soit plus régulière que la fonction de régression seule r_β . En particulier, l'estimateur $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 dès qu'on peut trouver une fonction de poids qui assure que Wr_β est plus régulière que f_ε au sens que Wr_β , Wr_β^2 et leurs dérivées jusqu'à l'ordre 2 par rapport à θ soient telles que

$$\sup_{\beta \in \mathbb{B}} (r_\beta W)^*/f_\varepsilon^* \text{ et } \sup_{\beta \in \mathbb{B}} (r_\beta^2 W)^*/f_\varepsilon^* \text{ appartiennent à } \mathbb{L}_1(\mathbb{R}) \cap \mathbb{L}_2(\mathbb{R}).$$

Pour certains risques relatifs, on peut aisément trouver une fonction de poids assurant que la vitesse paramétrique est atteinte. Ceci est notamment vérifié pour les risques relatifs usuels, comme le risque relatif exponentiel (modèle de Cox), le risque relatif linéaire (modèle d'excès de risque relatif) ou les

risques relatifs polynomiaux, et ce quelque soit la densité f_ε . Ceci est particulièrement intéressant et nouveau. Rappelons notamment que dans le modèle d'excès de risque relatif, aucun estimateur consistant n'avait été construit. Alors que notre procédure d'estimation fournit un estimateur consistant, \sqrt{n} -consistant et asymptotiquement gaussien. Ces exemple de risques relatifs polynomiaux sont d'un intérêt pratique particulier puisqu'ils correspondent aux risques relatifs considérés dans les modèles qui étudient l'influence de la dose de rayons ionisants mesurée avec erreur (voir par exemple Prentice (1982), Preston *et al.* (2004) et Martin-Magniette (2005)).

Néanmoins il reste des risques relatifs pour lesquels on ne sait pas exhiber de fonction de poids assurant que la vitesse paramétrique est atteinte. Pour ceux là nous établissons la consistance de l'estimateur et fournissons un ordre de grandeur du risque quadratique pour des classes de régularité classiques.

4.5 Vitesses de convergence pour des classes de régularité classiques

Comme pour le modèle de régression semi-paramétrique, on peut préciser la vitesse de convergence en quantifiant la décroissance des transformées de Fourier. Pour f_ε satisfaisant (\mathbf{N}_2) et pour Wr_β (et ses dérivées) satisfaisant (\mathbf{R}_1) , on obtient les vitesses décrites dans le tableau (3.1).

4.6 Exemples

L'objectif de ce paragraphe est d'illustrer sur quelques exemples les propriétés asymptotiques des estimateurs que nous avons construits. Dans tous les exemples, le risque de base est paramétrique, de forme non spécifiée et la densité f_ε arbitraire connue.

Exemple 6 Risque relatif exponentiel (modèle de Cox). Soient $r_\beta(z) = \exp(\beta z)$ et $W(z) = \exp\{-z^2/(4\delta)\}$ avec δ comme dans (\mathbf{N}_2) . Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

L'objectif de cet exemple est de montrer que notre méthode d'estimation fournit un estimateur qui atteint la vitesse paramétrique. Ce résultat est à relier aux résultats antérieurs, notamment ceux de Nakamura (1992), Kong et Gu (1999) et Augustin (2004). Ces derniers avaient déjà proposé un estimateur convergeant à la vitesse paramétrique, dans un modèle légèrement différent. En effet dans les articles mentionnés, le risque de base n'avait pas de forme paramétrique, et ε est supposée gaussienne.

Exemple 7 Modèle d'excès de risque relatif. Soit r_β de la forme $r_\beta(z) = 1 + \beta z$ avec β et f_Z dans \mathbb{B} et \mathcal{G} tels que pour $Z \sim g$, $1 + \beta Z$ est positif ou nul. Soit $W(z) = \exp\{-z^2/(4\delta)\}$ avec δ comme dans (\mathbf{N}_2) . Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 . Comme nous l'avons déjà mentionné, il n'existait pas d'estimateur consistant dans le modèle d'excès de risque relatif avec covariables mesurées avec erreurs.

Exemple 8 Risque relatif polynomial. Soit r_β de la forme $r_\beta(z) = 1 + \sum_{k=1}^m \beta_k z^k$ avec $m \geq 1$, et avec β et f_Z dans \mathbb{B} et \mathcal{G} tels que $r_\beta(Z)$ est positive pour $Z \sim g$. Soit $W(z) = \exp\{-z^2/(4\delta)\}$ avec δ comme dans (\mathbf{N}_2) . Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

Exemple 9 Risque relatif Cauchy. Soit r_β de la forme $r_\beta(z) = 1 - \beta + \beta/(1 + z^2)$ avec β et f_Z dans \mathbb{B} et \mathcal{G} tels que $r_\beta(Z)$ est positive pour $Z \sim g$. Soit $W(z) = (1 + z^2)^2 \exp\{-z^2/(4\delta)\}$ avec δ comme dans (\mathbf{N}_2) . Alors $\hat{\theta}$ est un estimateur \sqrt{n} -consistant et asymptotiquement gaussien de θ^0 .

Ce dernier exemple illustre l'importance du choix de la fonction de poids. En effet si $W \equiv 1$, la vitesse prédite par le tableau des vitesses est nettement plus lente. Par exemple si ε est gaussienne, le choix $W \equiv 1$ nous donne une vitesse d'ordre $\exp(-2\sqrt{\log n})$.

4.7 Une étude par simulation

J'ai choisi d'illustrer les propriétés de l'estimateur au travers d'une étude par simulation. Cette partie est extraite de l'article "Estimation of the hazard function in semi-parametric models with mismeasured covariates", qui je vais soumettre prochainement.

Loi des erreurs Dans ce jeu de simulation je considère deux lois d'erreurs ε :

- Cas 1 : ε Double exponentielle (ou Laplace). La densité de ε est alors

$$f_\varepsilon(x) = \frac{\sigma_\varepsilon}{\sqrt{2}} \exp\left(-\frac{\sqrt{2}}{\sigma_\varepsilon}|x|\right), \text{ et } f_\varepsilon^*(x) = \frac{1}{1 + \sigma_\varepsilon^2 x^2/2}.$$

- Cas 2 : ε gaussienne.

$$f_\varepsilon(x) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma_\varepsilon^2}\right), \text{ et } f_\varepsilon^*(x) = \exp(-\sigma_\varepsilon^2 x^2/2).$$

Risques instantanés considérés :

- Modèle de Cox avec risque de base constant : $R(t|Z) = \lambda^0 \exp(\theta^0 Z)$, $\theta = (\beta, \lambda)^T$.
- Modèle de Cox avec risque de base Weibull : $R(t|Z) = \lambda^0 \gamma^0 t^{\gamma^0 - 1} \exp(\theta^0 Z)$, $\theta = (\beta, \lambda, \gamma)^T$.

Estimateurs comparés Les estimateurs comparés sont $\hat{\theta}$ construit par notre méthode et défini par (4.2.4), l'estimateur naïf, l'estimateur construit quand la covariable Z est observée et l'estimateur construit en utilisant une correction de la pseudo-vraisemblance. Plus précisément, $\hat{\beta}_{naïf}$ est obtenu en minimisant la pseudo-vraisemblance $\ell_n(U^{(n)})$ défini par (4.1.1). Les estimateurs naïfs des paramètres du risque de base sont obtenus en mimisant le critère des moindres carrés $\tilde{S}_n(\theta)$ défini par (4.2.3) où Z_i est remplacée par U_i . L'estimateur $\hat{\theta}_{ssbruit}$ est obtenu par minimisation de la pseudo-vraisemblance (4.1.1) tandis que les estimateurs (sans bruit) du risque de base sont obtenus par minimisation du critère des moindres carrés (4.2.3). Nous comparons notre estimateur de β^0 à celui obtenu par correction de la pseudo-vraisemblance de Nakamura (voir (1992), Kong et Gu (1999) et Augustin (2004)) pour des erreurs ε gaussiennes. Cet estimateur $\hat{\beta}_{Nakamura}$ est obtenu en mimisant le critère

$$\frac{1}{n} \sum_{k=1}^n \int_0^\tau \left(\beta U_k + \beta^2 \sigma_\varepsilon^2 / 2 - \frac{\sum_{j=1}^n Y_j(t) U_j \exp(\beta U_j)}{\sum_{j=1}^n Y_j(t) \exp(\beta U_j)} \right) dN_i(t).$$

Les tableaux suivants présentent la moyenne des estimations sur $N_e = 100$ échantillons de tailles $n = 1000$ et $n = 5000$, le risque quadratique estimé sur ces 100 échantillons ainsi que le carré du biais :

$$m(T_n) = \frac{1}{N_e} \sum_{k=1}^{N_e} T_n^{(k)}, \quad MSE(T_n) = \frac{1}{N_e} \sum_{k=1}^{N_e} (T_n^{(k)} - \theta^0)^2, \text{ et } \text{Biais}^2(T_n) = (m(T_n) - \theta^0)^2,$$

avec $T_n^{(k)}$ l'estimation de θ^0 obtenue sur l'échantillon k . Pour chacun des cas considérés, la covariable Z est choisie comme $Z \sim \mathcal{N}(2, 1)$. Ainsi σ_ε^2 indique bien la quantité de bruit sur Z .

Résultats pour le modèle de Cox avec risque de base constant

Dans les deux tableaux suivants, nous présentons les résultats quand le risque de base est constant et pour deux tailles d'échantillon $n = 1000$ et $n = 5000$. Les simulations sont effectuées sous les conditions suivantes : $Z \sim \mathcal{N}(2, 1)$; $\sigma_\varepsilon^2 = 0.5$; $N_e = 100$ et $\theta^0 = (\beta^0, \lambda^0) = (0.55, 2)$.

- Résultats pour le modèle de Cox avec risque de base constant- $n = 1000$

ε Laplace $\sigma_\varepsilon^2 = 0.5$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5483	0.33	0.5514	-
$MSE(\hat{\beta})$	0.009	0.0457	0.0014	-
Bias ²	3.10^{-6}	0.0447	2.1564e-06	-
$m(\hat{\lambda})$	2.0261	2.9809	1.9738	-
$MSE(\hat{\lambda})$	0.1028	1.0354	0.0516	-
Bias ²	$6.8153e - 04$	0.9621	6.8521e-04	-
$\varepsilon \sim \mathcal{N}(0; 0.5)$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5519	0.3514	0.5441	0.5539
$MSE(\hat{\beta})$	0.0374	0.0413	0.0023	0.0065
Bias ²	3.5228e-06	0.0395	3.4246e-05	1.5343e-05
$m(\hat{\lambda})$	2.0941	3.0154	2.0248	-
$MSE(\hat{\lambda})$	0.4679	1.2158	0.1048	-
Bias ²	0.0089	1.0309	6.1377e-04	-

- Résultats pour le modèle de Cox avec risque de base constant- $n = 5000$

ε Laplace $\sigma_\varepsilon^2 = 0.5$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\theta})$	0.5480	0.3350	0.5572	-
$MSE(\hat{\theta})$	0.0045	0.0469	0.0011	-
Bias ²	3.82e-06	0.0462	5.24e-05	-
$m(\hat{\lambda})$	2.0173	3.1539	1.9707	-
$MSE(\hat{\lambda})$	0.0560	1.3718	0.0274	-
Bias ²	3e-04	1.3314	8.55e-04	-
$\varepsilon \sim \mathcal{N}(0; 0.5)$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5506	0.3451	0.5535	0.5559
$MSE(\hat{\beta})$	0.0024	0.042	0.00032	7.7500e-04
Bias ²	3.2517e-07	0.041	1.3e-05	3.4390e-05
$m(\hat{\lambda})$	2.0067	2.20	2.0002	-
$MSE(\hat{\lambda})$	0.0262	0.65	0.0091	-
Bias ²	4.4572e-05	0.646	<1e-7	-

L'estimateur $\hat{\theta}_{Nakamura}$ est basé sur des corrections de la pseudo-vraisemblance dans le cas de ε gaussienne. Il n'existe pas dans le cas d'erreur Laplace.

Résultats pour le modèle de Cox avec risque de base Weibull - $n = 2000$

Il est clair que la variance du bruit ε a une influence. Plus cette variance est importante plus il y a du bruit sur la covariable. Cette influence est mise en évidence par les résultats suivants, dans le modèle de Cox, avec les deux types de loi d'erreur et avec deux valeurs pour la variance de ε , $\sigma_\varepsilon^2 = 0.1$ et $\sigma_\varepsilon^2 = 0.5$. Dans cet exemple, le risque de base est un risque de Weibull. Les simulations sont faites sous les conditions suivantes : $Z \sim \mathcal{N}(2, 1)$; $n = 2000$, $N_e = 100$, $\theta^0 = (\beta^0, \lambda^0, \gamma^0) = (0.55, 1, 2)$.

ε Laplace $\sigma_\varepsilon^2 = 0.5$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5613	0.3395	0.5531	-
$MSE(\hat{\beta})$	0.0071	0.0447	6.2390e-04	-
$Bias^2$	1.2768e-04	0.0443	9.5033e-06	-
$m(\hat{\lambda})$	0.9929	1.3015	0.9964	-
$MSE(\hat{\lambda})$	0.0119	0.0893	0.0037	-
$Bias^2$	5.0759e-05	0.0875	1.2984e-05	-
$m(\hat{\gamma})$	2.0288	1.8803	2.0082	-
$MSE(\hat{\gamma})$	0.0108	0.0207	0.0025	-
$Bias^2$	8.2975e-04	0.0143	6.7005e-05	-
$\varepsilon \sim \mathcal{N}(0; 0.5)$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5581	0.3421	0.5533	0.5577
$MSE(\hat{\beta})$	0.0178	0.0436	5.0209e-04	0.0016
$Bias^2$	6.6277e-05	0.0432	1.1051e-05	5.9700e-05
$m(\hat{\lambda})$	0.9988	1.2908	1.0034	-
$MSE(\hat{\lambda})$	0.0259	0.0889	0.0040	-
$Bias^2$	1.4234e-06	0.0846	1.1652e-05	-
$m(\hat{\gamma})$	2.0614	1.8950	2.0055	-
$MSE(\hat{\gamma})$	0.0221	0.0163	0.0026	-
$Bias^2$	0.0038	0.0110	2.9819e-05	-

ε Laplace $\sigma_\varepsilon^2 = 0.1$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5449	0.4896	0.5520	-
$MSE(\hat{\beta})$	0.0477	0.0042	5e-04	-
$Bias^2$	2e-05	0.0036	4e-06	-
$m(\hat{\lambda})$	1.0047	1.0715	0.9936	-
$MSE(\hat{\lambda})$	0.0203	0.0165	0.0040	-
$Bias^2$	2e-05	0.0051	4e-05	-
$m(\hat{\gamma})$	2.0625	2.0072	2.0108	-
$MSE(\hat{\gamma})$	0.0325	0.0237	0.0022	-
$Bias^2$	0.0039	5e-05	1e-04	-
$\varepsilon \sim \mathcal{N}(0; 0.1)$	$\hat{\theta}$	$\hat{\theta}_{naif}$	$\hat{\theta}_{ssbruit}$	$\hat{\theta}_{Nakamura}$
$m(\hat{\beta})$	0.5693	0.4872	0.5463	0.5485
$MSE(\hat{\beta})$	0.0847	0.0045	5e-04	7e-04
$Bias^2$	3e-04	0.0039	1e-05	2e-06
$m(\hat{\lambda})$	0.9853	1.0653	1.0021	-
$MSE(\hat{\lambda})$	0.0305	0.0159	0.0040	-
$Bias^2$	2e-04	0.0043	4e-06	-
$m(\hat{\gamma})$	2.1162	2.0268	2.0172	-
$MSE(\hat{\gamma})$	0.0658	0.0249	0.0033	-
$Bias^2$	0.0135	7e-04	3e-04	-

Commentaires

Ces tableaux nous permettent de constater les propriétés de consistance de l'estimateur que nous avons construit. On constate sans ambiguïté la non consistance de l'estimateur naïf. En effet dès que la variance σ_ε^2 augmente, on constate que l'estimateur naïf ne converge pas vers la vraie valeur θ^0 . Ce tableau met aussi en évidence l'influence du niveau de bruit et que la loi des erreurs a peu d'influence sur la qualité de notre estimateur. On constate ensuite que les propriétés de l'estimateur que nous avons proposé sont très comparables à celles de l'estimateur proposé par Nakamura (1990, 1992), lorsque ce dernier existe c'est-à-dire pour des erreurs ε gaussiennes. Plus précisément le biais (carré) de $\hat{\theta}$ est plutôt plus faible que celui de $\hat{\theta}_{nakamura}$ mais la variance est elle plus élevée. Ceci vient de l'utilisation de la déconvolution, notamment dans le cas où ε est gaussienne.

Chapitre 5

Projets de recherche

Dans le document de synthèse qui précède, j'ai mentionné quelques questions ouvertes ou extensions possibles. Dans ce chapitre, je présente mes projets de recherche de façon plus précise. Certains travaux font l'objet d'articles qui seront soumis prochainement, d'autres sont des travaux en cours et les autres sont plus des projets.

Mes projets de recherche s'articulent autour de plusieurs axes.

Le modèle de régression avec erreurs sur les variables

Même si le modèle de régression avec erreurs sur les variables a déjà été largement étudié, il reste porteur de nombreuses questions.

Dans ce modèle, un premier travail fait l'objet d'un article en collaboration avec Cristina Butucea et Jean-Michel Marin. Il porte sur l'illustration pratique par simulations des résultats obtenus dans le modèle de régression avec erreurs sur les variables. Cette illustration met déjà en évidence les bonnes propriétés de convergence de l'estimateur. Nous la poursuivons par une comparaison des propriétés de notre estimateur avec celles d'estimateurs obtenus par d'autres méthodes pour les modèles dans lesquels des estimateurs alternatifs existent comme dans le cas linéaire, polynomial ou basés sur l'algorithme SIMEX,....

Dans le modèle que nous avons considéré avec Cristina Butucea, la variable explicative était réelle. Une extension de ces travaux au cas d'une variable explicative dans \mathbb{R}^d avec éventuellement certaines coordonnées correctement observées m'intéresse particulièrement.

Un autre projet porte sur l'extension des résultats obtenus en régression dans Butucea et Taupin (2008) à des contextes markoviens ou dépendants. La méthode d'estimation proposée dans le modèle de régression dans un cadre i.i.d. avec Cristina Butucea devrait s'étendre au cadre dépendant, sous des conditions de dépendance faibles, comme celles considérées dans Comte *et. al.* (2008a, 2008b). Cette méthode devrait améliorer les résultats de Comte et Taupin (2001), aussi bien pour les vitesses de convergence que pour les conditions de dépendance. Ce modèle considéré dans un cadre autorégressif est une chaîne de Markov cachée particulière à espace d'états non compact, qui a de nombreuses applications notamment en pharmacocinétique.

L'estimation non paramétrique de la fonction de régression dans un modèle avec erreurs sur les variables laisse une question importante ouverte. Notre méthode d'estimation par un ratio induit que le risque pour l'estimation de f dépend du risque pour l'estimation de la densité f_Z des variables explicatives. La question est donc de savoir si effectivement la vitesse d'estimation de f dépend de celle pour l'estimation de f_Z ou si, comme dans le cas sans erreurs, cette dépendance n'est induite que par la méthode d'estimation. En effet l'estimateur de Nadaraya-Watson est connu pour avoir ce défaut alors que dans le cas sans erreurs, d'autres méthodes existent (projection, polynômes locaux,...) qui ne l'ont pas. La question est donc de développer des méthodes analogues dans le cas avec erreurs.

Estimation du risque instantané dans un modèle avec covariables mesurées avec erreur

Les travaux sur modèles d'analyse de données de survie avec covariables mesurées avec erreurs portent essentiellement sur le modèle de Cox. Dans ce domaine, de nombreuses questions associées à des modèles variés restent ouvertes.

Je travaille actuellement sur l'illustration pratique des propriétés de l'estimateur construit dans l'article Martin-Magniette et Taupin (2008) pour l'estimation du risque instantané dans un modèle où la covariable est mesurée avec erreur. Les résultats obtenus par simulation montrent d'ores et déjà les bonnes propriétés pratiques de convergence de l'estimateur que nous avons construit. Ce travail fait l'objet d'un article qui sera soumis prochainement.

Dans l'article Martin-Magniette et Taupin (2008) le modèle ne considère qu'une covariable Z réelle. Il me semble particulièrement intéressant d'étendre la méthode proposée au cas d'un risque instantané de forme général faisant intervenir une covariable de \mathbb{R}^p dont certaines coordonnées seraient correctement observées et d'autres seraient mesurées avec erreur. Par exemple considérons le temps de survie T , depuis l'explosion de la bombe sur Hiroshima jusqu'au décès par cancer "solides" (cancer du sein, prostate, etc..., par opposition par exemple à la leucémie). On note T ce temps de survie et on observe $(X_i, D_i) = (\min(T_i, C_i), \mathbb{1}_{T_i \leq C_i})$. Les covariables sont typiquement $\mathbb{Z} = (Z, S, A) = (\text{dose, sexe, age})$. Un des modèle considéré est le suivant

$$R_{\theta^0}(t|\mathbb{Z}) = \eta^0(t) [1 + r_{\beta^0}(Z) \exp(\mu_1^0 \mathbb{1}_{S=2} + \mu_2^0(A - 30))]. \quad (5.0.1)$$

Il est maintenant acquis que la dose de rayonnement reçue est observée avec erreur. Par contre l'âge et le sexe sont des covariables qui vont être correctement observées. L'un des objectifs est d'estimer $(\eta^0, \beta^0, \mu_1^0, \mu_2^0)$ en utilisant $(X_i, D_i, U_i, S_i, A_i)$, $i = 1, \dots, n$ avec

$$U_i = Z_i + \varepsilon_i \quad \text{ou} \quad \log(U_i) = \log(Z_i) + \varepsilon_i.$$

La littérature sur l'analyse de survie pour des personnes soumises à des rayons ionisants, notamment les personnes rescapées de la bombe Hiroshima est très abondante. Je renvoie notamment à Prentice (1982), Preston *et al.* (2004) et Martin-Magniette (2005), pour la modélisation. Dans ce modèle (5.0.1), il n'existe pas à ma connaissance d'estimateur consistant des paramètres $(\eta^0, \beta^0, \mu_1^0, \mu_2^0)$.

Un des mes travaux en cours porte sur l'estimation du risque instantané dans un modèle avec plusieurs covariables dont certaines sont observées et d'autre sont observées avec erreur. Dans ce contexte je propose une version modifiée du critère des moindres carrés qui utilise des outils liés à la déconvolution et qui fournit un estimateur consistant des paramètres.

Toujours dans ce contexte de survie, je n'ai mentionné jusque là que des modèles où le risque de base est paramétrique. Il serait intéressant d'avoir une méthode d'estimation qui sépare l'estimation du paramètre de régression β^0 de l'estimation du risque de base α^0 .

Tous ces modèles et leurs extensions ont des applications évidentes dans le domaine biomédical. Suite à ma nouvelle intégration au laboratoire MAP5 de l'Université Paris Descartes, des discussions sont en cours pour l'application de ces méthodes sur des données réelles.

Optimalité des vitesses et efficacité

D'un point de vue plus théorique, pour chacun des deux modèles mentionnés ci-dessus, deux questions au moins restent ouvertes. La première concerne la vitesse de convergence pour l'estimation du paramètre θ^0 . Plus précisément, la vitesse paramétrique \sqrt{n} est-elle atteignable pour toutes les fonctions de régression (respectivement tous les risques relatifs) ? Le deuxième porte sur la question de l'efficacité pour l'estimation de θ^0 . Il est clair que le choix du critère des moindres carrés pour l'estimation ne

fournira probablement pas un estimateur efficace. Ce choix du critère des moindres carrés a été guidé avant tout par le souci de la vitesse de convergence. En effet la vitesse de convergence étant régie par la régularité de wf_θ (ou Wr_β) comparée à celle de f_ε , il semble plus simple d'exhiber une fonction de poids qui assure des bonnes propriétés de régularité wf_θ (ou Wr_β) que de celles de fonctions qui apparaîtraient dans une vraisemblance. L'étape suivante serait de construire une borne d'efficacité dans ces modèles semi-paramétriques puis de construire un estimateur efficace par une méthode à un pas.

Modèles de mélange discrets

Tous les modèles que j'ai considéré jusqu'à présent peuvent être vus comme des modèles de mélange continus. Ceci est dû au modèle de convolution avec l'hypothèse que le bruit et le signal admettent une densité et que les variables sont indépendantes. Quand l'une ou l'autre des variables est discrète, le modèle de convolution fait apparaître un modèle de mélange discret. Les outils sont alors complètement différents.

L'un de mes projets, en collaboration avec Catherine Matias (CR, CNRS Evry), porte sur l'estimation dans les modèles de mélange discrets et plus précisément sur des minorations pour l'estimation dans un modèle de mélange discret.

Le modèle "single-index"

Un autre de mes projets, en collaboration avec Cristina Butucea et Marian Hristache (MCF, ENSAI) concerne le modèle "single-index". Dans ce modèle "single index", très utilisé en économétrie, on modélise la fonction de régression $m(z) = \mathbb{E}(Y|Z = z)$ sous la forme $m(z) = r_0(z^T)\theta_0$ où θ_0 appartient à \mathbb{R}^d est le paramètre à estimer et où $r_0(\cdot)$, est un paramètre de nuisance fonctionnel. Les travaux antérieurs sur ce modèle considèrent le problème de l'estimation de θ_0 , basée sur n couples d'observations $(Y_1, Z_1^T), \dots, (Y_n, Z_n^T)^T$ où $(Y, Z^T)^T$ est un vecteur de \mathbb{R}^{d+1} . Nous proposons de construire un estimateur de θ_0 quand Z n'est pas complètement observé, parce que certaines variables explicatives sont mesurées avec erreurs. Ce problème est lié à la déconvolution avec la différence qu'il s'agit de vecteurs, que certaines variables explicatives sont par nature des variables discrètes et que certaines sont correctement observées.

Estimation de densité dans un cadre faiblement dépendant

Je mentionnerai un dernier projet en collaboration avec Jérôme Dedecker. Il concerne l'estimation adaptative d'une densité dans des contextes de dépendance faible. On considère le problème de l'estimation de la densité de variables X_1, \dots, X_n quand la suite $(X_i)_{i \geq 0}$ satisfait des conditions de dépendance faible comme celles introduites dans la partie 2.2. L'estimation adaptative d'une densité dans un contexte mélangeant a déjà fait l'objet de plusieurs travaux. L'objectif serait d'avoir une procédure adaptative ne dépendant pas des coefficients de dépendance, dans un contexte de dépendance faible.

Références

- Andersen, P. K., Borgan, Ø., Gill, R. D., and Keiding, N. (1993). *Statistical models based on counting processes*. Springer Series in Statistics.
- Anderson, B. D. O. and Deistler, M. (1984). Identifiability in dynamic errors-in-variables models. *J. Time Ser. Anal.* 5(1), 1–13.
- Anderson, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* 12, 1–45.
- Ango Nzé, P. (1992). Critères d’ergodicité de quelques modèles à représentation markovienne. *C. R. Acad. Sci. Paris Sér. I Math.* 315(12), 1301–1304.
- Augustin, T. (2004). An exact corrected log-likelihood function for Cox’s proportional hazards model under measurement error and some extensions. *Scand. J. Stat.* 31(1), 43–50.
- Baran, S. (2000). A consistent estimator in general functional errors-in-variables models. *Metrika* 51(2), 117–132 (electronic).
- Bickel, P. J. and Ritov, A. J. C. (1987). Efficient estimation in the errors-in-variables model. *Ann. Statist.* 15, 513–540.
- Bickel, P. J., Ritov, Y., and Rydén, T. (1998). Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.* 26(4), 1614–1635.
- Birgé, L. and Rozenholc, Y. (2006). How many bins should be put in a regular histogram. *ESAIM Probab. Stat.* 10, 24–45 (electronic).
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *J. Econometrics* 31(3), 307–327.
- Bougerol, P. and Picard, N. (1992a). Stationarity of GARCH processes and of some nonnegative time series. *J. Econometrics* 52(1-2), 115–127.
- Bougerol, P. and Picard, N. (1992b). Strict stationarity of generalized autoregressive processes. *Ann. Probab.* 20(4), 1714–1730.
- Bradley, R. C. (2007). *Introduction to strong mixing conditions. Vol. 1*. Kendrick Press, Heber City, UT.
- Butucea, C. (2004). Deconvolution of supersmooth densities with smooth noise. *Canadian J. Statist.* 32, 181–192.
- 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. Technical report.
- Butucea, C. and Taupin, M. L. (2008). New m-estimators in semiparametric regression with errors-in-variables. *Annales de l’Institut Henri Poincaré* 44(3), 393–421.
- Butucea, C. and Tsybakov, A. B. (2007a). Sharp optimality in density deconvolution with dominating bias. I. *Theory of Probability and Its Applications* 52(1), 111–128.
- Butucea, C. and Tsybakov, A. B. (2007b). Sharp optimality in density deconvolution with dominating bias. II. *Teor. Veroyatn. Primen.* 52(to appear).

- Buzas, J. S. (1998). Unbiased scores in proportional hazards regression with covariate measurement error. *J. Statist. Plann. Inference* 67, 247–257.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* 83, 1184–1186.
- Carroll, R. J., Küchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Measurement error models. *J. Amer. Statist. Assoc.* 91, 242–250.
- Carroll, R. J., Ruppert, D., and Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman and Hall, London.
- Carroll, R. J. and Stefanski, L. (1990). Approximate quaslikelihood estimation in models with surrogate predictors. *J. Amer. Statist. Assoc.* 85, 652–663.
- Cator, E. A. (2001). Deconvolution with arbitrarily smooth kernels. *Statist. Probab. Lett.* 54(2), 205–214.
- Chan, L. K. and Mak, T. K. (1985). On the polynomial functional relationship. *J. R. Statist. Soc. B* 47, 510–518.
- Chanda, K. C. (1995). Large sample analysis of autoregressive moving-average models with errors in variables. *J. Time Ser. Anal.* 16(1), 1–15.
- Cheng, C. H. and Van Ness, J. W. (1994). On estimating linear relationships when both variables are subject to errors. *J.R. Statist. Soc. B* 56, 167–183.
- Cheng, C.-L. and Riu, J. (2006). On estimating linear relationships when both variables are subject to heteroscedastic measurement errors. *Technometrics* 48(4), 511–519.
- Comte, F., Dedecker, J., and Taupin, M. L. (2008a). Adaptive density deconvolution with dependent inputs. *Math. Methods Statist.* 17(2), 87–112.
- Comte, F., Dedecker, J., and Taupin, M. L. (2008b). Adaptive density estimation for general arch models. *Econometric Theory* 24(6), 1628–1662.
- Comte, F. and Merlevède, F. (2002). Adaptive estimation of the stationary density of discrete and continuous time mixing processes. *ESAIM Probab. Statist. (electronic)* 6, 211–238.
- Comte, F., Rozenholc, Y., and Taupin, M.-L. (2006). Penalized contrast estimator for adaptive density deconvolution. *Canad. J. Statist.* 34(3), 431–452.
- Comte, F., Rozenholc, Y., and Taupin, M. L. (2007). Finite sample penalization in adaptive density deconvolution. *Journal of Statistical Computation and Simulation* 77, 977–1000.
- Comte, F. and Taupin, M.-L. (2001). Semiparametric estimation in the (auto)-regressive β -mixing model with errors-in-variables. *Math. Methods Statist.* 10(2), 121–160.
- Comte, F. and Taupin, M.-L. (2007). Adaptive estimation in a nonparametric regression model with errors-in-variables. *Statist. Sinica* 17(3), 1065–1090.
- Cook, J. R. and Stefanski, L. A. (1994). Simulation-extrapolation estimation in parametric measurement error models. *J. Amer. Statist. Assoc.* 89, 1314–1328.
- Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B* 34, 187–220. With discussion by F. Downton, Richard Peto, D. J. Bartholomew, D. V. Lindley, P. W. Glassborow, D. E. Barton, Susannah Howard, B. Benjamin, John J. Gart, L. D. Meshalkin, A. R. Kagan, M. Zelen, R. E. Barlow, Jack Kalbfleisch, R. L. Prentice and Norman Breslow, and a reply by D. R. Cox.
- Cox, D. R. and Oakes, D. (1984). *Analysis of survival data*. Monographs on Statistics and Applied Probability. London : Chapman & Hall.
- Cui, H. (2004). Estimation in partial linear EV models with replicated observations. *Sci. China Ser. A* 47(1), 144–159.
- Davidov, O. (2005). Estimating the slope in measurement error models—a different perspective. *Statist. Probab. Lett.* 71(3), 215–223.

- Davidov, O. and Griskin, V. (2008). A note on constrained estimation in the simple linear measurement error model. *Statist. Probab. Lett.* 78(5), 508–517.
- de Vries, C. G. (1991). On the relation between GARCH and stable processes. *J. Econometrics* 48(3), 313–324.
- Dedecker, J. and Priour, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Relat. Fields* 132(2), 203–236.
- Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *J. R. Stat. Soc. Ser. B Stat. Methodol.* 64(4), 869–886.
- Delaigle, A. and Gijbels, I. (2004a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Statist. Math.* 56(1), 19–47.
- Delaigle, A. and Gijbels, I. (2004b). Practical bandwidth selection in deconvolution kernel density estimation. *Comput. Statist. Data Anal.* 45(2), 249–267.
- Delaigle, A. and Gijbels, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statist. Sinica* 16(3), 773–788.
- Delaigle, A. and Hall, P. (2006). On optimal kernel choice for deconvolution. *Statist. Probab. Lett.* 76(15), 1594–1602.
- Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* 20, 2037–2056.
- Douc, R. and Matias, C. (2001). Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli* 7(3), 381–420.
- Douc, R., Moulines, É., and Rydén, T. (2004). Asymptotic properties of the maximum likelihood estimator in autoregressive models with Markov regime. *Ann. Statist.* 32(5), 2254–2304.
- Doukhan, P. (1994). *Mixing : Properties and examples*. Lecture Notes in Statistics (Springer). New York : Springer-Verlag.
- Duan, J.-C. (1997). Augmented GARCH(p, q) process and its diffusion limit. *J. Econometrics* 79(1), 97–127.
- Dupuy, J.-F. (2004). Nonparametric maximum likelihood estimation in the proportional hazards model with covariate measurement error. In *Parametric and semiparametric models with applications to reliability, survival analysis, and quality of life*, Stat. Ind. Technol., pp. 13–25. Boston, MA : Birkhäuser Boston.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50(4), 987–1007.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Stat.* 19(3), 1257–1272.
- Fan, J. and Masry, E. (1992). Multivariate regression estimation with errors-in-variables : Asymptotic normality for mixing processes. *J. Multivariate Anal.* 43(2), 237–271.
- Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* 21(4), 1900–1925.
- Fan, J., Truong, Y. K., and Wang, Y. (1991). Nonparametric function estimation involving errors-in-variables. *Nonparametric Functional Estimation and Related Topics*, 613–627.
- Fazekas, I., Baran, S., Kukush, A., and Lauridsen, J. (1999). Asymptotic properties in space and time of an estimator in nonlinear functional errors-in-variables models. *Random Oper. Stochastic Equations* 7(4), 389–412.
- Fazekas, I. and Kukush, A. G. (1998). Asymptotic properties of estimators in nonlinear functional errors-in-variables with dependent error terms. In *Proceedings of the 18th Seminar on Stability Problems for Stochastic Models, Part I (Hajdúszoboszló, 1997)*, Volume 92, pp. 3890–3895.

- Fuller, W. A. (1987). *Measurement error models*. Wiley Series in Probability and Mathematical Statistics.
- Genon-Catalot, V. and Laredo, C. (2006). Leroux’s method for general hidden Markov models. *Stochastic Process. Appl.* 116(2), 222–243.
- Gill, R. D. and Andersen, P. K. (1982). Cox’s regression model for counting processes : a large sample study. *Ann. Statist.* 10, 1100–1120.
- Giraitis, L., Kokoszka, P., and Leipus, R. (2000). Stationary ARCH models : dependence structure and central limit theorem. *Econometric Theory* 16(1), 3–22.
- Gleser, L. J. (1981). Estimation in a multivariate “errors in variables” regression model : large sample result. *Ann. Statist.* 9, 24–44.
- Gleser, L. J. (1985). A note on g.r. dolby’s unreplicated ultrastructural model. *Biometrika* 72, 117–124.
- Gleser, L. J. (1990). Improvements of the naive approach to estimation in nonlinear errors-in-variables regression models. *Contemp. Math.* 112, 99–114.
- Hannan, E. J. (1963). Regression for time series with errors of measurement. *Biometrika* 50, 293–302.
- Hausman, J. A., Newey, W. K., Ichimura, I., and Powell, J. L. . (1991). Identification and estimation of polynomial errors-in-variables models. *J. Econometrics* 50, 273–295.
- Hausman, J. A., Newey, W. K., and Powell, J. L. . (1995). Nonlinear errors in variables estimation of some engel curves. *J. Econometrics* 65, 205–233.
- Hesse, C. H. (1999). Data-driven deconvolution. *J. Nonparametr. Statist.* 10(4), 343–373.
- Hong, H. and Tamer, E. (2003). A simple estimator for nonlinear error in variable models. *J. Econometrics* 117(1), 1–19.
- Hosmer, D. W. J. and Lemeshow, S. (1999). *Applied survival analysis. Regression modeling of time to event data*. Wiley Series in Probability and Mathematical Statistics.
- Hsiao, C. (1989). Consistent estimation for some nonlinear errors-in-variables models. *Journal of Econometrics* 41, 159–185.
- Hsiao, C., Wang, L., and Wang, Q. (1997). Estimation of nonlinear errors-in-variables models : an approximate solution. *Statist. Papers* 38(1), 1–25.
- Hsiao, C. and Wang, Q. K. (2000). Estimation of structural nonlinear errors-in-variables models by simulated least-squares method. *Internat. Econom. Rev.* 41(2), 523–542.
- Hu, C. and Lin, D. Y. (2002). Cox regression with covariate measurement error. *Scand. J. Stat.* 29, 637–655.
- Hu, C. and Lin, D. Y. (2004). Semiparametric failure time regression with replicates of mismeasured covariates. *J. Am. Stat. Assoc.* 99, 105–118.
- Hu, P., Tsiatis, A. A., and Davidian, M. (1998). Estimating the parameters in the Cox model when covariate variables are measured with error. *Biometrics* 54(4), 1407–1419.
- Huang, Y. and Wang, C. y. (2000). Cox regression with accurate covariates unascertainable : A nonparametric-correction approach. *J. Am. Stat. Assoc.* 95, 1209–1219.
- Ioannides, D. and Alevizos, P. (1997). Nonparametric estimation in time series with measurement errors. *Nonlinear Anal., Theory Methods Appl.* 30(6), 3539–3546.
- Jensen, J. L. and Petersen, N. V. (1999). Asymptotic normality of the maximum likelihood estimator in state space models. *Ann. Statist.* 27(2), 514–535.
- Kalbfleisch, J. D. and Prentice, R. L. (1973). Marginal likelihoods based on Cox’s regression and life model. *Biometrika* 60, 267–278.
- Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many nuisance parameters. *Annals math. Statist.* 27, 887–906.

- Kim, H. M. and Saleh, A. K. M. E. (2005). Improved estimation of regression parameters in measurement error models. *J. Multivariate Anal.* 95(2), 273–300.
- Kong, F. H. (1999). Adjusting regression attenuation in the Cox proportional hazards model. *J. Statist. Plann. Inference* 79, 31–44.
- Kong, F. H. and Gu, M. (1999). Consistent estimation in Cox proportional hazards model with covariate measurement errors. *Statistica Sinica* 9, 953–969.
- Koo, J.-Y. (1999). Log spline deconvolution in Besov space. *Scand. J. Statist.* 26(1), 73–86.
- Koo, J.-Y. and Lee, K.-W. (1998). *B*-spline estimation of regression functions with errors in variable. *Statist. Probab. Lett.* 40(1), 57–66.
- Kukush, A., Markovsky, I., and Van Huffel, S. (2005). Consistency of the structured total least squares estimator in a multivariate errors-in-variables model. *J. Statist. Plann. Inference* 133(2), 315–358.
- Kukush, A. and Schneeweiss, H. (2005a). Comparing different estimators in a nonlinear measurement error model. I. *Math. Methods Statist.* 14(1), 53–79.
- Kukush, A. and Schneeweiss, H. (2005b). Comparing different estimators in a nonlinear measurement error model. II. *Math. Methods Statist.* 14(2), 203–223.
- Kukush, A., Schneeweiss, H., and Wolf, R. (2005). Relative efficiency of three estimators in a polynomial regression with measurement errors. *J. Statist. Plann. Inference* 127(1-2), 179–203.
- Leroux, B. G. (1992). Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.* 40(1), 127–143.
- Li, Y. Ryan, L. (2006). Inference on survival data with covariate measurement error - an imputation-based approach. *Scand. J. Stat.* 33.
- Li, T. (2000). Estimation of nonlinear errors-in-variables models : a simulated minimum distance estimator. *Statist. Probab. Lett.* 47(3), 243–248.
- Li, T. (2002). Robust and consistent estimation of nonlinear errors-in-variables models. *J. Econometrics* 110(1), 1–26.
- Li, Y. and Ryan, L. (2004). Survival analysis with heterogeneous covariate measurement error. *J. Amer. Statist. Assoc.* 99.
- Liu, J. and Chen, X. (2005). Consistency of LS estimator in simple linear EV regression models. *Acta Math. Sci. Ser. B Engl. Ed.* 25(1), 50–58.
- Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canad. J. Statist.* 17(4), 427–438.
- Ma, J.-l. (2005). Empirical likelihood for errors-in-variables partially linear models. *Math. Appl. (Wuhan)* 18(1), 136–143.
- Martin-Magniette, M.-L. (2005). Nonparametric estimation of the hazard function by using a model selection method : estimation of cancer deaths in Hiroshima atomic bomb survivors. *J. Roy. Statist. Soc. Ser. C* 54, 317–331.
- Martin-Magniette, M. L. and Taupin, M. L. (2008). Semi-parametric estimation of the hazard function in a model with covariate measurement error. *ESAIM P&S (to appear)*.
- Masry, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Trans. Inf. Theory* 37(4), 1105–1115.
- Masry, E. (1993a). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *J. Multivariate Anal.* 44(1), 47–68.
- Masry, E. (1993b). Multivariate regression estimation with errors-in-variables for stationary processes. *Nonparametric statistics* 3, 13–36.
- Masry, E. (1993c). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes Appl.* 47(1), 53–74.

- Masry, E. (1994). Probability density estimation from dependent observations using wavelets orthonormal bases. *Stat. Probab. Lett.* 21(3), 181–194.
- Masry, E. (2003). Deconvolving multivariate kernel density estimates from contaminated associated observations. *IEEE Trans. Inform. Theory* 49(11), 2941–2952.
- Matias, C. (2002). Semiparametric deconvolution with unknown noise variance. *ESAIM Probab. & Stat* 6, 271–292.
- Matias, C. and Taupin, M.-L. (2004). Minimax estimation of linear functionals in the convolution model. *Math. Methods Statist.* 13(3), 282–328.
- Meister, A. (2004). Deconvolution density estimation with a testing procedure for the error distribution. Technical report, Universitat Stuttgart, <http://www.mathematik.uni-stuttgart.de/preprints/listen/listen.php>.
- Murphy, S. A. and van der Vaart, A. W. (1996). Likelihood inference in the errors-in-variables model. *Journal of Multivariate Analysis* 59, 81–108.
- Nakamura, T. (1990). Corrected score function for errors-in-variables models : methodology and application to generalized linear models. *Biometrika* 77, 127–137.
- Nakamura, T. (1992). Proportional hazards model with covariates subject to measurement error. *Biometrics* 48, 829–838.
- Neumann, M. H. (1997). On the effect of estimating the error density in nonparametric deconvolution. *J. Nonparametr. Statist.* 7(4), 307–330.
- Nowak, E. (1985). Global identification of the dynamic shock-error model. *J. Econometrics* 27(2), 211–219.
- Pensky, M. and Vidakovic, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Stat.* 27(6), 2033–2053.
- Prentice, R. L. (1982). Covariate measurement errors and parameter estimation in a failure time regression model. *Biometrika*, 331–342.
- Prentice, R. L. and Kalbfleisch, J. D. (1979). Hazard rate models with covariates. *Biometrics* 35(1), 25–39.
- Preston, D. L., Pierce, D. A., Shimizu, Y., Cullings, H. M. Fujita, S., Funamoto, S., and Kodama, K. (2004). Effect of recent changes in atomic bomb survivor dosimetry on cancer mortality risk estimates. *Radiat. Res.* 162, 377–389.
- Reiersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica.* 18, 375–389.
- Reynaud-Bouret, P. (2003). Adaptive estimation of the intensity of inhomogeneous Poisson processes via concentration inequalities. *Probab. Theory Relat. Fields* 126 (1), 103–153.
- Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Natl. Acad. Sci. USA* 42, 43–47.
- Schennach, S. M. (2004). Estimation of nonlinear models with measurement error. *Econometrica* 72(1), 33–75.
- Schneeweiss, H. and Augustin, T. (2006). Some recent advances in measurement error models and methods. *Allg. Stat. Arch.* 90(1), 183–197.
- Shklyar, S., Schneeweiss, H., and Kukush, A. (2007). Quasi score is more efficient than corrected score in a polynomial measurement error model. *Metrika* 65(3), 275–295.
- Song, X. and Huang, Y. (2005). On corrected score approach for proportional hazards model with covariate measurement error. *Biometrics* 61(3), 702–714.
- Stefanski, J. H. and Carroll, R. J. (1993). Semi-parametrics quasilielihood and variance function estimation in measurement error models. *J. Econometrics* 58, 223–256.

- Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Commun. Stat. -Theory Meth.* 18, 4335–4358.
- Taupin, M.-L. (1998). Estimation in the nonlinear errors-in-variables model. *C. R. Acad. Sci. Paris Sér. I Math.* 326(7), 885–890.
- Taupin, M.-L. (2001). Semi-parametric estimation in the nonlinear structural errors-in-variables model. *Ann. Statist.* 29(1), 66–93.
- Tribouley, K. and Viennet, G. (1998). \mathbb{L}_p adaptive density estimation in a β mixing framework. *Ann. Inst. Henri Poincaré, Probab. Stat.* 34(2), 179–208.
- Truong, Y. K. (1995). Survival time regression involving covariate measurement error. *Bull. Inform. Cybernet.* 27(1), 31–51.
- Tsiatis, T. T., DeGruttola, V., and Wulfsohn, M. S. (1995). Modeling the relationship of survival to longitudinal data measured with error. Application to survival and cd4 counts in patients with aids. *J. Amer. Statist. Assoc.* 90, 27–37.
- Tyurin, Y. N., Yakovlev, A., Shi, J., and Bass, L. (1995). Testing a model of aging in animal experiments. *Biometrics* 51(1), 363–372.
- Van der Vaart, A. (1996). Efficient maximum likelihood estimation in semiparametric mixture models. *Ann. Statist.* 24(2), 862–878.
- van der Vaart, A. W. (1988). Estimating a real parameter in a class of semiparametric models. *Ann. Statist.* 16, 1450–1474.
- van Es, B., Spreij, P., and van Zanten, H. (2005). Nonparametric volatility density estimation for discrete time models. *J. Nonparametr. Stat.* 17(2), 237–251.
- Viennet, G. (1997). Inequalities for absolutely regular sequences : application to density estimation. *Probab. Theory Relat. Fields* 107(4), 467–492.
- Volkonskii, V. and Rozanov, Y. (1960). Some limit theorems for random functions. I. *Theor. Probab. Appl.* 4, 178–197.
- Wolter, K. M. and Fuller, W. A. (1982a). Estimation of nonlinear errors-in variables models. *Ann. Statist.* 10, 539–548.
- Wolter, K. M. and Fuller, W. A. (1982b). Estimation of the quadratic errors-in-variables model. *Ann. Statist.* 69, 175–182.
- Xie, S. X., Wang, C. Y., and Prentice, R. L. (2001). A risk set calibration method for failure time regression by using a covariate reliability sample. *J. R. Stat. Soc., Ser. B, Stat. Methodol.* 63(4), 855–870.
- You, J. and Zhou, H. (2007). On semiparametric EV models with serially correlated errors in both regression models and mismeasured covariates. *Scand. J. Statist.* 34(2), 365–383.
- Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Ann. Statist.* 18, 806–831.
- Zhou, H. and You, J. (2007). Statistical inference for a semiparametric measurement error regression model with heteroscedastic errors. *J. Statist. Plann. Inference* 137(7), 2263–2276.