

Mémoire de synthèse en vue de l'obtention de
l'Habilitation à diriger des Recherches

de l'Université Montpellier 2

Spécialité : **Biostatistique**

**Contribution à l'estimation non-paramétrique
adaptative dans les modèles de durées**

Elodie BRUNEL

soutenue le **11 Juillet 2013**

devant le jury :

Alain Berlinet	Professeur honoraire, Université Montpellier 2
Fabienne Comte	Professeur, Université Paris Descartes
Béatrice Laurent-Bonneau	Professeur, INSA Toulouse
Jean-Michel Marin	Professeur, Université Montpellier 2
André Mas	Professeur, Université Montpellier 2
Valentin Patilea	Professeur, CREST & ENSAI Rennes

après avis des rapporteurs :

Jacobo De Uña Álvarez	Professeur, Université de Vigo, Espagne
Béatrice Laurent-Bonneau	Professeur, Université Toulouse 3
Valentin Patilea	Professeur, CREST & ENSAI Rennes

Table des matières

Introduction	3
1 Estimation adaptative dans diverses situations d'échantillonnage	7
1.1 Variantes d'estimateurs du risque instantané	7
1.1.1 Le modèle de censure à droite	8
1.1.2 Estimateur du risque instantané basé sur une représentation forte de l'estimateur de Kaplan-Meier	9
1.1.3 Estimateur <i>quotient</i> du risque instantané	12
1.1.4 Estimateur <i>direct</i> du risque instantané	14
1.2 Estimation dans un modèle avec biais et censure	16
1.2.1 Echantillonnage avec biais connu	16
1.2.2 Estimation adaptative de la densité et de la fonction de risque lorsque le biais est connu	17
1.2.3 Echantillonnage dans un diagramme de Lexis	18
1.2.4 Estimation adaptative de la densité et de la fonction de risque en présence d'un biais inconnu	21
1.3 Estimation de la fonction de répartition avec censure par intervalle de type I	22
2 Espérance et lois conditionnelles pour données censurées	26
2.1 Régression en présence d'une ou plusieurs covariables	26
2.2 Lois conditionnelles	29
2.2.1 Des histoires de contrastes	30
2.2.2 Contraste pour la densité conditionnelle en présence de censure	32
2.2.3 Estimation optimale pour la fonction de répartition conditionnelle	33
3 Estimation de fonctionnelles par contraste de type moindres-carrés	37
3.1 Moyenne de vie résiduelle conditionnelle	37
3.2 Risque conditionnel	44
3.2.1 Le modèle de censure avec indicatrices manquantes	44

	2
3.2.2 Construction du contraste dans le cadre MAR	45
3.2.3 Prolongements	47
4 Travaux en cours, futurs et perspectives	49
A Choix des bases et propriétés d'approximation	55

Introduction

Ce document reprend les travaux de recherche entrepris au MAP5 à l'Université Paris Descartes, et poursuivis depuis mon arrivée en Septembre 2007, à l'I3M à l'Université Montpellier 2. Ils portent sur l'estimation non-paramétrique de diverses fonctions d'intérêt et la question du choix de modèles dans le contexte spécifique des durées de vie. L'objectif de cette synthèse est d'apporter un éclairage sur la façon de mettre à profit les outils usuels de sélection de modèles pour faire de l'estimation adaptative dans les modèles de durées. L'une des particularités de ce type de données est la présence de censure, qui peut être de différente nature, et qui dans tous les cas nécessite un traitement statistique adapté, et aussi celle de covariables.

La synthèse s'articule autour de trois thèmes qui correspondent aux trois premiers chapitres du mémoire. Le premier, dans la chronologie des premiers travaux, aborde l'estimation des fonctions de risque instantané de décès, de densité ou de répartition dans diverses conditions d'échantillonnage : censure à droite, censure par intervalle ou biais de sélection. Le second thème explore l'ajout de covariables *via* le modèle d'espérance conditionnelle et l'estimation des lois conditionnelles, la censure portant sur la variable de durée. Le troisième thème se présente comme une variation autour de fonctionnelles de régression. Enfin, la dernière partie est consacrée à l'exposition de travaux récents, certains plus aboutis que d'autres, et de pistes de recherche. Le mémoire ne reprend pas les résultats obtenus pendant ma thèse de doctorat, qui portait sur l'optimalité du choix des partitions, au sens du risque de Kulback-Leibler, pour les estimateurs de Barron de type histogramme modifiés, en densité (publications [2], [3], [4]). Toutefois, j'avais commencé à m'intéresser à l'analyse des durées de vie, dès ma thèse, pour des applications de l'estimateur de Barron, et certaines des idées développées dans le chapitre 1 de ce mémoire y ont sans doute trouvé leur point de départ.

Le contexte des durées de vie

Dans de nombreuses applications biomédicales, la variable d'intérêt est souvent une durée de vie dont l'observation directe n'est pas toujours possible. Le phénomène de *censure aléatoire à droite* constitue l'illustration la plus courante de telles observations incomplètes. En effet, toutes les durées de vies ne sont pas observées : pour certaines d'entre elles, dites censurées, on

ne connaît qu'une borne inférieure de leur valeur. Si X désigne la variable d'intérêt positive, et si au lieu d'observer X on observe C telle que $X > C$, on parle de *censure à droite*. Une telle situation peut être représentée par le modèle constitué du couple de variables, (Z, δ) où $Z = \min(X, C)$ et $\delta = \mathbf{1}_{(X \leq C)}$ est l'indicatrice de (non) censure. La censure peut dépendre ou non des observations, être fixe, occasionnée par la fin de l'étude par exemple, ou aléatoire. En général, les variables X et C sont supposées indépendantes, c'est dans ce cadre que nous travaillerons. De plus, dans toute la suite, nous considérons que les lois de probabilité qui régissent les variables X et C sont continues. Un autre type de censure des données peut être décrit par l'exemple suivant. Supposons que l'on s'intéresse à l'âge à partir duquel des enfants acquièrent une compétence, comme la lecture. La période d'observation est l'année du CP. A l'entrée en CP, certains enfants savent déjà lire, d'autres vont accéder à la lecture au cours de l'année, tandis que certains ne sauront toujours pas lire en fin de CP. Dans ce cas, on parle de *censure par intervalle*. L'événement d'intérêt n'est jamais directement observé. Ce type de censure est également privilégié dans l'étude des épidémies, où la date de contamination d'un individu ne peut pas être connue exactement. Il peut être généralisé au cas où l'on s'intéresse à la survenue d'un événement au cours de k périodes d'observations.

Dans ces contextes de censure, les problèmes d'inférence nécessitent l'introduction d'outils spécifiques. Depuis le fameux estimateur produit-limite de Kaplan & Meier (1958) pour la fonction de survie $\mathbb{P}(X \geq t) =: 1 - F(t)$, de nombreuses méthodes d'estimation non-paramétrique, en présence de données censurées ont été développées. C'est notamment avec les travaux de Aalen (1978) que l'inférence statistique dans les modèles pour données censurées (au sens large) a connu un essor considérable, grâce à un transfert de la théorie des martingales et des processus ponctuels vers la modélisation. La reformulation d'Aalen, pour le modèle de censure à droite, fait intervenir deux processus, une indicatrice $D(t)$ de présence à risque à l'instant t :

$$D(t) = \mathbf{1}_{(Z \geq t)} \text{ où } Z = X \wedge C$$

et un processus ponctuel d'événement $N(t) = \mathbf{1}_{(Z \leq t, \delta=1)}$, pour tout $t \geq 0$. La considération de ce processus fait naturellement intervenir la fonction de risque instantané h , puisque si X admet une densité f par rapport à la mesure de Lebesgue, le compensateur Λ du processus N s'écrit :

$$\Lambda(t) = \int_0^t h(u) D(u) du$$

où la fonction d'intensité $h(t)$ coïncide avec la fonction de risque instantané définie par :

$$h(t) = \frac{f(t)}{1 - F(t)}$$

L'estimateur non-paramétrique de Nelson-Aalen de la fonction de risque cumulé $H(t) = \int_0^t h(u) du$ est introduit grâce à cette formulation. Pour une présentation détaillée de cet estimateur et ses

propriétés, nous renvoyons à Andersen *et al.* (1993). Bien que la loi de la durée soit indifféremment caractérisée par la fonction de répartition (ou de survie), la densité, ou la fonction de risque instantané, cette dernière est souvent plus intéressante car sa forme reflète davantage d'éventuelles différences entre les modèles. La première partie du chapitre 1 est consacrée à la comparaison de différents estimateurs non-paramétriques de cette fonction dans le modèle de censure à droite. Notre apport est de fournir des procédures complètement automatiques (*data-driven*) et adaptatives.

Diverses situations d'échantillonnage, en particulier dans les études épidémiologiques, font parfois apparaître, en addition de la censure, des problèmes de biais de sélection des individus. Nous traitons, dans le chapitre 1, partie 1.2, l'estimation de la densité et de la fonction de risque pour des données biaisées et censurées. Là encore, nous nous attachons à construire des procédures adaptatives, et donnons des recommandations pour la calibration des constantes numériques.

Un autre aspect très important, outre les conditions de l'échantillonnage, est la prise en compte de covariables comme l'âge, ou des paramètres liés à l'état de l'individu (morphologiques ou biochimiques), connus pour avoir une influence sur la durée de vie. Ce thème est l'objet des chapitres 2 et 3 dans lesquels nous abordons l'estimation de diverses fonctionnelles de régression : espérance conditionnelle, fonction de répartition, moyenne de vie résiduelle et fonction de risque conditionnelles. Notre point de vue s'attache à fournir des alternatives plus flexibles au très populaire modèle de Cox (1972), dans lequel, sous l'hypothèse des risques proportionnels, la fonction h prend la forme :

$$h(t, x) = h_0(t) \exp(\beta_0' x), \quad t \geq 0, x \in \mathbb{R}^d$$

où h_0 est une fonction déterministe, appelée risque instantané de base, β_0 est un vecteur de \mathbb{R}^d de paramètres inconnus et le coefficient $e^{\beta_0^j}$ quantifie le risque associé à la coordonnée x^j de la covariable x .

L'estimation non-paramétrique adaptative

Le fil conducteur des travaux présentés dans ce document est donc la recherche de procédures non-paramétriques et adaptatives pour l'estimation de fonctionnelles d'intérêt dans les modèles de durées. Nous adoptons le point de vue, usuel en estimation fonctionnelle, du risque minimax sur une classe de fonctions \mathcal{F} qui donne le risque le plus pessimiste parmi toutes les fonctions de la classe \mathcal{F} . La notion d'estimateurs optimaux au sens minimax a connu un essor considérable dans les années 90 avec les travaux de Donoho *et al.* (1995) et Birgé & Massart (1997). Le plus souvent, on se contentera de donner l'ordre de grandeur du risque en terme de vitesse de convergence, les vitesses minimax étant en général déjà établies. Si d'un point de vue théorique, il est intéressant de disposer d'estimateurs optimaux pour une classe de fonctions

\mathcal{F} , de tels estimateurs dépendent fortement de la classe \mathcal{F} considérée, qui fixe la régularité des fonctions (Hölder, Sobolev, Besov,...). Pour construire un estimateur qui atteint la meilleure vitesse possible sur la classe \mathcal{F} , il faudrait connaître la régularité de la fonction à estimer. Les méthodes d'estimation dites adaptatives apportent une réponse à ce problème et fournissent des procédures automatiques de construction d'estimateurs optimaux adaptatifs à la régularité inconnue de la fonction. Les inégalités oracles, développées conjointement, constituent un outil essentiel, pour l'obtention des résultats d'adaptation. C'est le caractère non-asymptotique de ces inégalités, qui en pratique les rend très attractives, puisqu'elles garantissent les bonnes propriétés des procédures d'estimation à distance finie.

Nous nous plaçons dans le paradigme de la *sélection de modèles* de Barron *et al.* (1999), réalisée via la pénalisation d'un contraste. L'estimation d'une fonction $s \in \mathcal{S}$ inconnue à partir d'observations X_1, \dots, X_n , requiert la donnée d'un critère empirique $\gamma(X_1, \dots, X_n; t) := \gamma_n(t)$, appelé *contraste* par rapport à s , si $t \mapsto \mathbb{E}[\gamma_n(t)]$ atteint son minimum en $t = s$ sur l'ensemble \mathcal{S} . La minimisation d'un contraste sur une collection de sous-espaces d'approximation, appelés *modèles* (cf. Annexe A), fournit alors une collections d'estimateurs. L'Annexe A est consacrée à la présentation des collections de modèles et à leurs propriétés d'approximation et nous y renverrons tout au long du mémoire, pour éviter les répétitions. La sélection de modèles consiste alors à choisir un estimateur parmi tous les estimateurs de la collection. L'ajout au contraste d'un terme de pénalité, d'autant plus grand que le modèle est complexe, permet de faire ce choix, en réalisant l'usuel compromis biais au carré/variance. C'est donc sous cet angle que nous cherchons à définir des estimateurs optimaux des différentes fonctions d'intérêt. Nos travaux montrent que les techniques propres au traitement de la censure, ne dégradent pas les vitesses de convergence minimax du cas i.i.d. D'un point de vue non-asymptotique, il y a souvent un prix à payer que nous avons illustré par des études numériques dans les différentes publications présentées dans ce mémoire.

Chapitre 1

Estimation adaptative dans diverses situations d'échantillonnage

Publications [5] , [8], [10], [11], [12]

Dans ce chapitre, je présente les résultats obtenus pour l'estimation de la densité f , de la fonction de risque instantané de décès h ou de la fonction de répartition F d'une variable de durée dans diverses situations d'échantillonnage englobant la censure à droite, la censure par intervalle ou le biais de sélection. Ce sont aussi les premiers travaux parus. L'ajout de covariables, néanmoins omniprésentes en analyse des durées de vie, n'est pas ici envisagé et le sera dans les chapitres suivants. Nous noterons X la variable de durée de vie, sans confusion possible avec une éventuelle covariable. Nous avons aussi fait le choix d'unifier, autant que possible, les notations dans ce chapitre, afin d'en faciliter la lecture. Ces notations ne coïncident donc pas forcément avec celles utilisées dans les articles référencés.

Le chapitre est organisé de la façon suivante : la partie 1.1 présente des variantes d'estimateurs de la fonction de risque instantané et constitue aussi le point de départ de ma collaboration avec Fabienne Comte. Ensuite, la rencontre avec Agathe Guilloux, qui nous a soumis la problématique des données affectées par un biais de sélection, a donné lieu à différentes stratégies d'estimation dans ce contexte. La partie 1.2 expose ces travaux. Enfin, la partie 1.3 est consacrée à l'estimation de la fonction de répartition de données censurées par intervalle, et propose une alternative adaptative à l'estimateur non-paramétrique du maximum de vraisemblance étudié par Groeneboom & Wellner (1992) pour ce type de censure.

1.1 Variantes d'estimateurs du risque instantané

De nombreux estimateurs de la densité ou de la fonction de risque en présence de censure ont été proposés et étudiés dans la littérature des données de survie. Pour n'en citer que quelques

uns : Marron & Padgett (1987), Dabrowska (1987), Stute (1995), Dabrowska *et al.* (1999), Efromovich (2001) pour la densité, et Antoniadis & Grégoire (1990), Patil (1997), Antoniadis *et al.* (1999) pour la fonction de risque. Dans ces travaux, les procédures sont non adaptatives, dans le sens où une hypothèse de régularité est faite sur la loi des observations, mais les estimateurs atteignent les vitesses de convergence optimales au sens du MISE.

Enfin, l'étude générale de Reynaud-Bouret (2006) englobe l'estimation adaptative du taux de hasard comme cas particulier de l'estimation de l'intensité d'un processus de Poisson par contraste pénalisé sur des bases d'histogrammes. De notre côté, nous avons proposé plusieurs estimateurs originaux de la fonction de risque instantané ou taux de hasard. Nos différentes stratégies d'estimation adaptative se sont appuyées sur les méthodes de sélection de modèles. Leur objectif est de réaliser automatiquement le compromis non-asymptotique biais au carré/variance via la minimisation d'un contraste pénalisé de façon adéquate. Ce dernier point est un atout, qui permet de donner un cadre théorique justifiant les procédures d'estimation automatique pour des tailles d'échantillons fixées. Notons que l'équivalent, pour l'approche par noyaux, n'existait pas jusqu'aux travaux récents de Goldenshluger & Lepski (2011). Par ailleurs, je souligne ici l'élément clé du papier [8], propre aux modèles de durées, qui repose sur la représentation forte de l'estimateur de Kaplan-Meier à l'aide des courbes d'influence proposées par Reid (1981), puis utilisées dans divers contextes par Mielniczuk (1985), Delecroix & Yazourh (1991) ou Lo *et al.* (1989).

1.1.1 Le modèle de censure à droite

Je présente ici les principales notations du modèle de censure droite. On considère des variables aléatoires positives X_1, \dots, X_n , qui représentent n durées, indépendantes et de fonction de répartition (f.d.r) commune F , et C_1, \dots, C_n des variables aléatoires (dites variables de censure) de f.d.r commune G , les variables de durée et de censure étant indépendantes. On définit le taux de hasard ou risque instantané de décès h par :

$$h(x) = \lim_{u \rightarrow 0} \frac{\mathbb{P}(x \leq X \leq X + u | X \geq x)}{u}$$

qui s'interprète comme le risque instantané, pour un individu ayant survécu jusqu'à x , de décéder entre x et $x + u$. Si F admet une densité f , alors on a la relation suivante :

$$h(x) = \frac{d}{dx} H(x) = \frac{f(x)}{\bar{F}(x)}, \quad \text{si } F(x) < 1.$$

où $H = -\log(\bar{F})$ désigne le taux de hasard cumulé et $\bar{F} = 1 - F$ la fonction de survie. Lorsque les observations sont censurées à droite, on n'observe pas directement X_1, \dots, X_n mais les couples $(Z_1, \delta_1), \dots, (Z_n, \delta_n)$ où

$$Z_i = X_i \wedge C_i, \quad \delta_i = \mathbf{1}_{(X_i \leq C_i)},$$

avec la notation $a \wedge b = \min(a, b)$. Autrement dit, l'indicatrice δ_i vaut 1 si la durée de vie du i -ème individu a pu être observée. Lorsque $\delta_i = 0$, l'observation est dite censurée. On note L la f.d.r commune des minimums Z_1, \dots, Z_n et $\bar{L} := 1 - L = (1 - F)(1 - G)$ leur fonction de survie.

On estime h sur un intervalle $[0, \tau]$ avec $\tau < \tau_L = \min\{\tau_F, \tau_G\}$ et où $\tau_F = \sup\{x : F(x) < 1\}$ pour toute f.d.r F . Sans perte de généralité, on fixera $\tau = 1$ dans la suite, quitte à faire un changement d'échelle des données pour s'y ramener.

Nous ne sommes pas capable d'estimer la fonction de risque sur \mathbb{R} tout entier. De plus, pour le risque quadratique intégré que nous considérons, cela n'aurait pas de sens puisque la fonction de risque n'est pas de carré intégrable en général. On se place donc sur un intervalle borné de la forme $[0, \tau]$ avec $\tau < \tau_L$. D'après Antoniadis *et al.* (1999), puisque $Z_{(n)} \rightarrow \tau_L$ p.s lorsque $n \rightarrow \infty$ où $Z_{(n)}$ est la n -ème statistique d'ordre, la recommandation pratique est de choisir la borne τ plus grande que la plus grande valeur observée et non censurée.

L'estimateur non-paramétrique de la fonction de survie \bar{F} , introduit par Kaplan et Meier (1958) est défini à partir des observations $Z_{(j)}$ réordonnées par ordre croissant où $Z_{(j)}$ est la j -ème statistique d'ordre de l'échantillon (Z_1, \dots, Z_n) , par :

$$KM_n(x) = \begin{cases} \prod_{i=1, Z_{(i)} \leq x}^n \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} & \text{si } x \leq Z_{(n)} \\ 0 & \text{si } x > Z_{(n)}. \end{cases}$$

Par abus de notation, $\delta_{(i)}$ désigne l'indicatrice correspondant à l'observation $Z_{(i)}$. Cette définition implique cependant que la plus grande valeur observée n'est pas censurée. Ceci n'est pas toujours vrai en pratique. Par ailleurs, comme nous aurons besoin de construire un estimateur de la fonction de survie qui ne s'annule pas, nous utiliserons une version modifiée de l'estimateur de Kaplan-Meier proposée par Lo et al. (1989) :

$$\bar{F}_n(x) = \begin{cases} \prod_{i=1, Z_{(i)} \leq x}^n \left(\frac{n-i+1}{n-i+2} \right)^{\delta_{(i)}} & \text{si } x \leq Z_{(n)} \\ \bar{F}_n(Z_{(n)}) & \text{if } x > Z_{(n)}. \end{cases} \quad (1.1)$$

Cet estimateur présente des propriétés lorsqu'on le substitue à la fonction de survie inconnue \bar{F} , en particulier $\bar{F}_n(x) \geq (n+1)^{-1}$ pour tout x et $\sup_{0 \leq x \leq T} |KM_n(x) - \bar{F}_n(x)| = O(n^{-1})$ p.s pour tout $0 < T < \min\{\tau_F, \tau_G\}$.

1.1.2 Estimateur du risque instantané basé sur une représentation forte de l'estimateur de Kaplan-Meier

Nous proposons dans [8] d'estimer la fonction de risque instantané $h(x)$ sur un intervalle A (que nous prenons égal à $[0, 1]$). Etant donnée une base orthonormale $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ de $L^2([0, 1])$,

l'estimateur de la fonction h sur A est défini par :

$$\hat{h}_m = \sum_{\lambda \in \Lambda_m} \hat{a}_\lambda \varphi_\lambda \text{ et } \hat{a}_\lambda = \int_0^1 \varphi_\lambda(x) dH_n(x), \quad (1.2)$$

où $H_n = -\ln(\bar{F}_n)$, avec \bar{F}_n défini par (1.1). L'estimateur peut se réécrire, pour $x \in A$:

$$\hat{h}_m(x) = - \sum_{i/Z_{(i)} < 1} \delta_{(i)} \mathcal{K}_m(Z_{(i)}, x) \ln \left(1 - \frac{1}{n - i + 2} \right)$$

avec $\mathcal{K}_m(Z_{(i)}, x) = \sum_{\lambda \in \Lambda_m} \varphi_\lambda(Z_{(i)}) \varphi_\lambda(x)$. Cette expression permet de souligner l'analogie avec l'estimateur à noyau si l'on remplace $\mathcal{K}_m(\cdot, x)$ par un noyau usuel.

La définition des différentes bases orthonormales utilisées est donnée en détail dans l'Annexe A. Il s'agit principalement de bases de Fourier [T], de polynômes par morceaux [DP] ou de bases d'ondelettes [W]. Le vocabulaire de la sélection de modèles emploie le terme *modèle* pour désigner le sous-espace linéaire S_m de $\mathbb{L}^2([0, 1])$ engendré par la base $(\varphi_\lambda)_{\lambda \in \Lambda_m}$,

$$S_m := \text{Vect}(\varphi_\lambda, \lambda \in \Lambda_m)$$

Il est de dimension finie $D_m := |\Lambda_m|$. Pour un estimateur \hat{h}_m construit par projection sur une base orthonormale $(\varphi_\lambda)_{\lambda \in \Lambda_m}$, et avec h_m la projection orthogonale de h (restreinte à $[0, 1]$) sur S_m , on obtient le résultat préliminaire suivant qui fournit une borne supérieure de son risque \mathbb{L}^2 :

Proposition 1 *On considère l'estimateur $\hat{h}_m \in S_m$ défini par (1.2), dans l'une des collections [T], [DP] ou [W], cf. Annexe A, alors,*

$$\mathbb{E}(\|h - \hat{h}_m\|^2) \leq \|h - h_m\|^2 + \frac{2\Phi_0^2 D_m}{n} \int_0^1 \frac{h(x)}{1 - L(x)} dx + \kappa \frac{D_m^2 \ln^2(n)}{n^2}, \quad (1.3)$$

où κ est une constante qui ne dépend que du choix de la base.

Le point essentiel de la démonstration de la proposition 1 repose sur la décomposition du terme de variance qui utilise une représentation de l'estimateur de Kaplan-Meier comme somme de variables indépendantes plus un terme de reste que nous sommes capables de contrôler. Le Lemme 3.1 dans [8] donne cette décomposition qui s'appuie sur les courbes d'influence de l'estimateur de Kaplan-Meier obtenues par Reid (1981), et aussi Lo *et al.* (1989). Un résultat analogue, aux constantes près, peut être obtenu pour des bases de splines (pour lesquelles la propriété d'orthogonalité n'est plus satisfaite), nous renvoyons également à [8] pour plus de détails. On peut ensuite déduire facilement de la Proposition 1, une vitesse de convergence pour l'estimateur en faisant une hypothèse de régularité sur la fonction h inconnue. Le Lemme 12 dans Barron *et al.* (1999) implique que $\|h - h_m\|$ est d'ordre $D_m^{-\alpha}$ dans les trois collections [T], [DP] et [W], pour tout réel $\alpha > 0$. Pour la base de splines [B], c'est le Théorème 3.3, Chapitre 12, de DeVore

& Lorentz (1993) qui assure que $\|h - h_m\|^2 = O(D_m^{-2\alpha})$, si h appartient à un espace de Besov $B_{\alpha,2,\infty}([0,1])$ avec $|h|_{\alpha,2} \leq L$ pour une constante L donnée. Le minimum dans (1.3) est alors atteint pour un modèle S_{m_n} tel que $D_{m_n} = O([n^{1/(1+2\alpha)}])$, qui est inférieur à n pour $\alpha > 0$. Ainsi, si $h \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, avec $\alpha > 0$, on retrouve la vitesse de convergence non-paramétrique usuelle $n^{-2\alpha/(1+2\alpha)}$.

L'étape suivante consiste à proposer une procédure automatique afin de garantir un choix optimal de la dimension, la régularité α de la fonction h étant inconnue. Pour cela, il est utile d'exprimer l'estimateur défini par (1.2) comme minimisant le contraste :

$$\gamma_n(t) = \|t\|^2 - 2 \int_0^1 t(x) dH_n(x) \quad (1.4)$$

avec pour toute fonction $t \in \mathbb{L}^2([0,1])$, $\|t\|^2 = \int_0^1 t^2(x) dx$. Ainsi,

$$\int_0^1 t(x) dH_n(x) = - \sum_{i/Z_{(i)} < 1} \delta_{(i)} t(Z_{(i)}) \ln \left(1 - \frac{1}{n-i+2} \right),$$

ce qui permet d'exprimer l'estimateur \hat{h}_m comme solution de la minimisation du contraste sur S_m :

$$\hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t).$$

Ensuite, pour choisir la meilleure dimension, dans une famille de modèles donnée, on pénalise le contraste de la façon suivante

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{h}_m) + \text{pen}(m)] \quad (1.5)$$

où $\text{pen}(m)$ est un terme de pénalité à choisir. Le résultat suivant donne la forme de la pénalité (théorique) adéquate :

Théorème 1 *On considère l'estimateur $\hat{h}_m = \arg \min_{t \in S_m} \gamma_n(t)$. Si $D_m \leq \sqrt{n}$, alors l'estimateur pénalisé $\hat{h}_{\hat{m}}$ avec \hat{m} défini par (1.5) et*

$$\text{pen}(m) = \kappa \Phi_0^2 \left(\int_0^1 \frac{h(x)}{1-L(x)} dx \right) \frac{D_m}{n}$$

et κ est une constante numérique qui satisfait l'inégalité

$$\mathbb{E}(\|\hat{h}_{\hat{m}} - h\|^2) \leq \inf_{m \in \mathcal{M}_n} (3\|h - h_m\|^2 + 4\text{pen}(m)) + \frac{K \ln^2(n)}{n}. \quad (1.6)$$

La constante K ne dépend que de h , F et de Φ_0 .

Le terme de pénalité dépend d'une quantité inconnue $\mathbb{E} \left(\frac{\mathbf{1}_{(\delta_1=1, Z_1 \leq 1)}}{(1-L(Z_1))^2} \right) = \int_0^1 \frac{h(x)}{1-L(x)} dx$ qui doit être estimée. On propose de lui substituer le moment empirique d'ordre 2 :

$$\hat{s}_2 = \frac{1}{n} \sum_{i=1}^n \frac{\mathbf{I}_{(Z_i \leq 1)} \mathbf{I}_{(\delta_i=1)}}{(1 - \hat{L}_n(Z_i))^2}, \quad \hat{L}_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq x)}.$$

On a alors le résultat suivant analogue au Théorème 1, pour la pénalité aléatoire, entièrement calculable cette fois-ci :

Théorème 2 *Sous les mêmes hypothèses que celles du Théorème 1, avec $\text{pen}(m)$ remplacée par*

$$\widehat{\text{pen}}(m) = \kappa \Phi_0^2 \hat{s}_2 \frac{D_m}{n}$$

pour κ une constante numérique, l'estimateur pénalisé $\hat{h}_{\widehat{m}}$ satisfait l'inégalité (1.6).

1.1.3 Estimateur *quotient* du risque instantané

Une fonction d'intérêt dans le modèle de censure à droite est la densité ψ du sous-échantillon constitué des variables non censurées $\delta_i Z_i$. On définit cette *sous-densité* par :

$$\psi(x) = f(x)(1 - G(x))$$

Par conséquent, la fonction de risque instantané h peut aussi être exprimée en fonction de la sous-densité ψ :

$$h(x) = \frac{f(x)}{1 - F(x)} = \frac{\psi(x)}{1 - L(x)} \quad \text{si } L(x) < 1.$$

L'estimation de la sous-densité ψ présente un intérêt propre mais aussi pour l'obtention d'un estimateur par quotient de la fonction de risque h qui se déduira naturellement de cette dernière expression. La première étape consiste maintenant à construire un estimateur adaptatif de la sous-densité ψ . En pratique, il n'y a donc pas de restriction à choisir la borne τ plus grande que la plus grande valeur observée non censurée. Comme précédemment, on se place sur un intervalle borné de la forme $[0, \tau]$ avec $\tau < \tau_L = \sup\{x : L(x) < 1\}$. Nous travaillons avec $\tau = 1$ sans perte de généralité. Ces considérations nous amènent à poser l'hypothèse suivante :

$$\exists c_L > 0, \quad \forall x \in [0, 1], \quad c_L \leq 1 - L(x) < 1, \quad (1.7)$$

avec $c_L = \inf_{x \in [0, 1]} (1 - L(x)) = 1 - L(1)$.

On considère le contraste suivant qui est une simple adaptation à nos données incomplètes du contraste usuel pour estimer une densité :

$$\gamma_n^\psi(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq 1)} \delta_i t(Z_i) \quad (1.8)$$

pour t une fonction dans $\mathbb{L}^2([0, 1])$. Notons $\hat{\psi}_m = \arg \min_{t \in S_m} \gamma_n^\psi(t)$.

Nous définissons également ψ_m la projection orthogonale de ψ sur S_m .

L'estimateur pénalisé permet de réaliser un choix automatique de la dimension. En effet, la dimension optimale au sens du risque \mathbb{L}^2 dépend de la régularité inconnue de la fonction ψ au travers de constantes qui apparaissent dans la décomposition biais au carré/variance. L'estimateur pénalisé est défini par :

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n^\psi(\hat{\psi}_m) + \widehat{\text{pen}}^\psi(m)] \quad (1.9)$$

où le terme de pénalité pen^ψ est choisi de façon à sélectionner la meilleure dimension. Nous avons le résultat :

Théorème 3 *Considérons l'estimateur $\hat{\psi}_{\hat{m}}$ où \hat{m} est défini par (1.9) et*

$$\widehat{\text{pen}}^\psi(m) = \kappa \Phi_0^2 \left(\frac{1}{n} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq 1)} \mathbf{I}_{(\delta_i=1)} \right) \frac{D_m}{n}$$

et κ une constante universelle. Alors l'estimateur pénalisé $\hat{\psi}_{\hat{m}}$ satisfait

$$\mathbb{E}(\|\hat{\psi}_{\hat{m}} - \psi\|^2) \leq \inf_{m \in \mathcal{M}_n} K_0 \left[\|\psi - \psi_m\|^2 + \Phi_0^2 \left(\int_0^1 \psi(x) dx \right) \frac{D_m}{n} \right] + \frac{K_1}{n}, \quad (1.10)$$

où K_0 dépend de κ et K_1 dépend de ψ et de Φ_0 .

Nous pouvons alors déduire du Théorème 3, un résultat d'adaptation à la régularité inconnue de la fonction ψ , pour ψ appartenant à un espace de Besov $\mathbb{B}_{\alpha_\psi, 2, \infty}(R)$:

On est alors en mesure d'utiliser l'estimation adaptative de la sous-densité ψ pour l'estimation de la fonction de risque h . On propose un estimateur construit par quotient de l'estimateur $\hat{\psi}_{\hat{m}}$ de la sous-densité $\psi(x)$ et de la fonction de survie empirique de $1 - L(x)$:

$$\tilde{h}_\psi = \frac{\hat{\psi}_{\hat{m}}}{1 - \hat{L}_n} \quad \text{où} \quad \hat{L}_n(x) = \frac{1}{n+1} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq x)}. \quad (1.11)$$

Grâce à la décomposition suivante :

$$\tilde{h}_\psi - h = \left(\frac{\hat{\psi}_{\hat{m}} - \psi}{1 - \hat{L}_n} + \psi \left(\frac{1}{1 - \hat{L}_n} - \frac{1}{1 - L} \right) \right)$$

on trouve la borne :

$$\mathbb{E}\|\tilde{h}_\psi - h\|^2 \leq \frac{2^4}{c_L^2} \mathbb{E}\|\hat{\psi}_{\hat{m}} - \psi\|^2 + \frac{C(c_L, \|\psi\|)}{n}, \quad (1.12)$$

où $C(c_L, \|\psi\|)$ est une constante qui dépend de c_L et de $\|\psi\|$. De l'inégalité (1.12), on déduit que \tilde{h}_ψ est un estimateur adaptatif de la fonction h si les fonctions h et ψ ont la même régularité $\alpha = \alpha_h = \alpha_\psi$. Ce résultat d'optimalité est à rapprocher de celui d'Antoniadis *et al.* (1999) (sans adaptation). Il est bien connu que cette approche par quotient implique que si la régularité α_h de la fonction h , est plus grande que celle de ψ , la vitesse de convergence de \tilde{h}_ψ est la moins bonne des deux vitesses, c'est-à-dire $n^{-\alpha_\psi/(1+2\alpha_\psi)}$, au lieu de la vitesse optimale $n^{-\alpha_h/(1+2\alpha_h)}$.

Néanmoins, on retrouve, dans le cas favorable, la vitesse minimax démontrée par Huber & MacGibbon (2004). En pratique, l'étude de simulations, réalisée dans [5], souligne que cette procédure d'estimation de la fonction h , bien que très simple, n'est pas toujours très satisfaisante même lorsque l'estimation de ψ est très bonne. Nous explorons alors une stratégie d'estimation directe de la fonction h par moindres carrés.

1.1.4 Estimateur *direct* du risque instantané

Un estimateur direct peut être obtenu par minimisation du contraste :

$$\hat{h}_m = \arg \min_{t \in S_m} \gamma_n^h(t) \quad \text{où } \gamma_n^h(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq 1)} \frac{\delta_i t(Z_i)}{1 - \hat{L}_n(Z_i)} \quad (1.13)$$

où la définition de $\hat{L}_n(x)$ est donnée dans (1.11). Pour une base orthonormale de S_m , $(\varphi_\lambda)_{\lambda \in \Lambda_m}$, on obtient alors une expression explicite de l'estimateur \hat{h}_m :

$$\hat{h}_m = \sum_{\lambda \in \Lambda_m} \hat{a}_\lambda \varphi_\lambda \quad \text{où } \hat{a}_\lambda = \frac{1}{n} \sum_{i=1}^n \mathbf{I}_{(Z_i \leq 1)} \frac{\delta_i \varphi_\lambda(Z_i)}{1 - \hat{L}_n(X_i)}. \quad (1.14)$$

Soit $h_m = \sum_{\lambda \in \Lambda_m} \left(\int_0^1 h \varphi_\lambda \right) \varphi_\lambda$ la projection orthogonale de h sur S_m . Comme dans le paragraphe précédent pour l'estimation de la sous-densité ψ , on peut montrer à l'aide de la décomposition du risque que l'estimateur \hat{h}_m atteint la vitesse minimax si la dimension D_m est choisie de façon adéquate en fonction de n et de la régularité inconnue de la fonction h . A nouveau, nous devons proposer un estimateur adaptatif capable de sélectionner cette dimension sans connaissance *a priori*. La procédure de sélection est la suivante :

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left(\gamma_n^h(\hat{h}_m) + \text{pen}^h(m) \right). \quad (1.15)$$

où

$$\widehat{\text{pen}}^h(m) = \kappa \Phi_0^2 \left(\frac{1}{n} \sum_{i=1}^n \frac{\mathbf{I}_{(Z_i \leq 1)} \delta_i}{(1 - \hat{L}_n(Z_i))^2} \right) \frac{D_m}{n}, \quad (1.16)$$

κ est une constante numérique. Nous donnons directement l'inégalité oracle pour l'estimateur pénalisé :

Théorème 4 *Soit f_Z la densité commune des Z_i . Supposons que $\sup_{x \in [0,1]} f_Z(x) \leq f_1$ et considérons les bases orthonormales [T], [DP] ou [W], définies dans l'Annexe A, et avec la condition sur la dimension maximale N_n des espaces S_m , $m \in \mathcal{M}_n$: $N_n \leq n/(16f_1K_\varphi)$ pour [DP] et [W] (bases localisées), et $N_n \leq \sqrt{n}/(4\sqrt{f_1})$ pour [T] (bases globales). On note K_φ une constante ne dépendant que de la base choisie. Alors l'estimateur $\hat{h}_{\hat{m}}$ satisfait*

$$\mathbb{E}(\|\hat{h}_{\hat{m}} - h\|^2) \leq \inf_{m \in \mathcal{M}_n} \left(7\|h - h_m\|^2 + 8\kappa\Phi_0^2 \frac{D_m}{n} \int_0^1 \frac{h(x)}{1 - L(x)} dx \right) + \frac{C\sqrt{\ln(n)}}{n}, \quad (1.17)$$

où C est une constante qui dépend de f_1 , Φ_0 , $c_L = \inf_{x \in [0,1]} (1 - L(x))$.

La preuve du Théorème 4 s'appuie sur la décomposition du contraste qui fait apparaître un processus empirique linéaire et centré, que l'on traite grâce à l'inégalité de Talagrand (1996) et un terme de reste non linéaire dont on doit contrôler l'espérance du supremum. Pour ce dernier, on utilise des arguments de Baraud (2002). Grâce à l'inégalité oracle (1.17), on peut déduire pour l'estimateur $\tilde{h} = \hat{h}_{\tilde{m}}$ un résultat d'adaptation comme pour l'estimateur quotient \tilde{h}_ψ , sans la restriction $\alpha_h = \alpha_\psi$. Ainsi, l'estimateur \tilde{h} atteint automatiquement la vitesse minimax dans tous les cas. On peut remarquer que la condition $\alpha_h > 0$ pour les bases localisées [DP] et [W] devient $\alpha_h > 1/2$ pour la base [T], car on a la restriction $N_n \leq O(\sqrt{n})$ (contraintes sur la dimension maximale N_n du Théorème 4). De façon surprenante, les simulations donnent une légère supériorité à l'estimateur quotient \tilde{h}_ψ , ce qui peut s'expliquer en partie par un choix de fonctions de risque très régulières. La dégradation en fin d'intervalle due à la censure, est observée quelle que soit la stratégie quotient ou moindres carrés. Pour l'étude complète des performances numériques de nos estimateurs, nous renvoyons aux publications [5] et [8].

Nous donnons aussi un résultat pour des collections de modèles plus riches (polynômes construits sur des partitions non-régulières) qui permettent de capter des variations de régularité et sont en particulier utilisés pour détecter des ruptures. Ces collections sont tout à fait indiquées pour l'estimation de la fonction de risque, du fait de la rareté des données en fin d'intervalle. Sous certaines restrictions, nous pouvons obtenir un résultat analogue au Théorème 4. Nous considérons une collection générale de polynômes par morceaux. Soit $K_n = N_n/(r+1)$ où N_n désigne la dimension du plus grand sous-espace de la collection. On définit l'ensemble des noeuds $\Gamma = \{\ell/K_n, \ell = 1, \dots, K_n - 1\}$. On se donne ensuite un degré maximal r pour les polynômes et l'ensemble des noeuds

$$\{a_0 = 0, a_1, \dots, a_\ell, a_{\ell+1} = 1\}$$

où $\{a_1, \dots, a_\ell\}$ est un sous-ensemble de Γ : la dimension est donc $D_m = (\ell+1)(r+1)$. Cela signifie que pour toute dimension D_m fixée, il y a $\binom{K_n - 1}{\ell}$ modèles associés qui correspondent aux différents choix possibles du sous-ensemble $\{a_1, \dots, a_\ell\}$, pour $\ell = 1, \dots, K_n - 1$. Le cardinal de \mathcal{M}_n , l'ensemble de tous les m est :

$$\sum_{\ell=1}^{K_n-1} \binom{K_n - 1}{\ell} = 2^{K_n-1} - 1 = \frac{1}{2} \exp(K_n \ln(2)) - 1.$$

Comme K_n est d'ordre n , on a donc un nombre exponentiel de modèles à explorer. La base des φ_λ avec $\lambda = (a_j, a_{j+1}; k) \in \Lambda_m$ est définie par

$$\sqrt{\frac{2k+1}{a_j - a_{j-1}}} Q_k \left(\frac{2}{a_j - a_{j-1}} x - \frac{a_j + a_{j-1}}{a_j - a_{j-1}} \right) \mathbf{1}_{[a_{j-1}, a_j]}(x)$$

pour $k = 0, 1, \dots, r$ et $j = 0, \dots, \ell + 1$, où Q_k est le k -ème polynôme de la hiérarchie de Legendre. Nous pouvons montrer une inégalité oracle analogue à (1.17), sous la contrainte

$N_n \leq Kn/\ln(n)$, avec un terme en $\log(n)$ ajouté à la pénalité qui est alors d'ordre $D_m(1 + \log(n))/n$. Ce terme implique une perte d'un logarithme dans la vitesse de convergence qui devient $O((n/\ln(n))^{-2\alpha_h/(2\alpha_h+1)})$. C'est le prix à payer pour pouvoir considérer une si grande collection de modèles. La preuve de ce résultat nécessite des arguments supplémentaires à ceux du Théorème 4, car on ne peut plus utiliser l'usuelle connection des normes $\mathbb{L}^2 - \mathbb{L}^\infty$.

1.2 Estimation dans un modèle avec biais et censure

1.2.1 Echantillonnage avec biais connu

Dans [11], nous considérons un modèle où les durées de vie sont observées avec un biais, c'est-à-dire qu'au lieu d'observer la variable d'intérêt X , on observe la variable X_w dont la densité de probabilité est donnée par :

$$f_w(x) = \frac{w(x)f(x)}{\mu}, \quad \text{avec } \mu = \int w(u)f(u)du, \quad (1.18)$$

où la fonction de poids $w(\cdot)$ déterministe est supposée connue. On dit que la variable X_w est affectée d'un biais de sélection. De plus, les données biaisées peuvent éventuellement être censurées à droite, et alors on observe seulement

$$X_w \wedge C \text{ et } \delta_w = \mathbf{I}_{(X_w \leq C)}.$$

Les variables C_i sont i.i.d. indépendantes des $X_{w,i}$. La fonction de répartition de l'échantillon biaisé est notée F_w . Pour motiver les applications de ce modèle, nous décrivons un exemple. Considérons une population de patients infectés par le VIH. Pour chaque individu i , notons D_i la durée qui s'écoule entre la contamination par le VIH et le diagnostic du SIDA et notons X_i la durée entre le diagnostic et l'éventuel décès du patient i . De plus, notons I_i la date calendaire de la contamination du patient i . Une telle population est soumise à un mécanisme de censure à droite, lorsque l'un des patients est perdu de vue avant son décès, par exemple. On introduit aussi la date calendaire de censure CT_i de l'individu i . Avec ces notations, la durée de vie que nous souhaitons étudier est celle qui s'écoule entre la contamination par le VIH et le décès du patient, soit $(D_i + X_i) \wedge (CT_i - I_i)$. A présent, examinons la façon dont les patients sont inclus dans l'étude. Supposons que seuls des patients infectés (séropositifs) mais n'ayant pas encore été diagnostiqués, soient inclus dans l'étude à la date t_0 . C'est un exemple où le modèle que nous proposons fait complètement sens ; la justification détaillée est donnée dans [11].

1.2.2 Estimation adaptative de la densité et de la fonction de risque lorsque le biais est connu

Plusieurs articles ont été consacrés à l'estimation de la loi (densité ou fonction de répartition) de X à partir d'un échantillon $X_{w,1}, \dots, X_{w,n}$ i.i.d (i.e. sans censure). Nous mentionnons Gill *et al.* (1988) et Efromovich (2004a) pour l'estimation dans le cas de données biaisées de la forme (1.18) et Vardi (1982) dans le cas particulier du biais de longueur $w(x) = x$. Le problème de l'estimation de la densité en présence d'un biais de sélection (mais sans censure) a été considéré par Efromovich (2004a) et Efromovich (2004b). Dans Efromovich (2004a), la fonction de répartition est estimée par projection sur des espaces de polynômes trigonométriques et l'estimation de la densité se fait, par corollaire, en dérivant. Les résultats d'optimalité sont obtenus pour une classe de fonctions de répartitions analytiques et dans ce contexte la vitesse minimax est d'ordre $O(\log(n)/n)$.

Le cadre de Efromovich (2004b) est moins restrictif et davantage comparable à notre approche. Un estimateur seuillé par blocs, basé à nouveau sur des bases trigonométriques, est proposé mais il ne prend pas en compte la censure. Dans [11], nous considérons l'estimation de la densité en présence d'un biais de sélection de la forme (1.18) et de censure. Comme précédemment pour l'estimation de la fonction de risque h , nous avons toujours le problème du support lié à la présence de la censure. A nouveau, notre procédure d'estimation n'est valide que sur un intervalle borné $A = [0, \tau]$ avec $\tau = \sup\{x \in \mathbb{R}^+ : (1 - F_w)(1 - G)(x) > 0\}$, où F_w est la fonction de répartition de la variable X_w . On pose $A = [0, 1]$. Nous considérons le contraste défini pour $t \in \mathbb{L}^2([0, 1])$, par

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \hat{\mu} \sum_{i=1}^n \frac{\delta_{w,i} t(X_{w,i})}{w(X_{w,i}) \hat{G}(X_{w,i})} \quad \text{avec } \hat{\mu} = \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_{w,i}}{w(X_{w,i}) \hat{G}(X_{w,i})} \right)^{-1} \quad (1.19)$$

où nous utilisons la version (1.1) de l'estimateur de Kaplan-Meier modifié \hat{G} de \bar{G} .

Nous considérons une collection $\{S_m : m \in \mathcal{M}_n\}$ d'espaces de projections. Pour chaque m , l'espace S_m , de dimension D_m est engendré par une base orthonormale à support dans $[0, 1]$, décrite dans l'Annexe A.

L'estimateur de la densité f est un estimateur par projection, défini par :

$$\hat{f}_m = \arg \min_{t \in S_m} \gamma_n(t) \quad (1.20)$$

et on procède ensuite à la sélection de modèles à l'aide du critère pénalisé suivant :

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{f}_m) + \widehat{\text{pen}}(m)] \quad \text{avec } \widehat{\text{pen}}(m) = \kappa \Phi_0^2 \hat{\mu}^2 \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_{w,i}}{w^2(X_{w,i}) \hat{G}^2(X_{w,i})} \right) \frac{D_m}{n}$$

où κ est une constante numérique et Φ_0 une constante connue qui dépend de la base choisie. Nous avons besoin de supposer que la fonction de biais w satisfait :

$$\exists w_2 > 0, \quad 0 < w(x) \leq w_2 < +\infty, \quad \forall x \in A.$$

Cette condition technique, exclut le cas du biais de sélection. Nous montrons dans [11] que l'estimateur pénalisé $\hat{f}_{\hat{m}}$ atteint la vitesse optimale minimax pour des densités appartenant à des classes de Besov.

1.2.3 Echantillonnage dans un diagramme de Lexis

Le processus d'inclusion des individus dans un échantillon peut être représenté par un diagramme de Lexis (1875), cf. Lexis (1977). Considérons une population \mathcal{I} . Pour chaque individu $i \in \mathcal{I}$, notons σ_i sa date de naissance et X_i sa durée de vie. Comme décrit par Keiding (1990) et Lund (2000), une telle population peut être représentée par le système de coordonnées avec en abscisse la date calendaire et en ordonnée l'âge. Dans un tel diagramme, une *ligne de vie* $L(\sigma, X)$ est définie par :

$$L(\sigma, X) = \{(\sigma + y, y) : 0 \leq y \leq X\}.$$

Ainsi, la ligne de vie $L(\sigma, X)$ est un segment de droite de pente 1 qui joint les points de coordonnées $(\sigma, 0)$ de la date de naissance et $(\sigma + X, X)$ de la date de mort. La population \mathcal{I} est représentée par le diagramme de Lexis comme l'ensemble de toutes les lignes de vie $L(\sigma_i, X_i)$ pour $i \in \mathcal{I}$, cf. Figure (1.1).

Considérons un ensemble mesurable $\mathcal{S} \in \mathbb{R} \times \mathbb{R}^+$ qui décrit le mécanisme de l'échantillonnage, comme sur la Figure (1.1). Nous supposons qu'un individu i est inclus dans l'échantillon si sa ligne de vie $L(\sigma_i, X_i)$ coupe l'ensemble \mathcal{S} . Soit $a_{\mathcal{S}}(s)$ l'âge de l'individu à l'inclusion dans l'étude, pour une date de naissance $s \in \mathbb{R}$, défini par :

$$\begin{cases} a_{\mathcal{S}}(s) = \inf\{y \geq 0, (s + y, y) \in \mathcal{S}\} \\ a_{\mathcal{S}}(s) = \infty \text{ si l'infimum n'existe pas.} \end{cases}$$

L'individu i , dont la date de naissance est σ_i , et la durée de vie X_i , est inclus dans l'échantillon si :

$$\mathcal{L}(\sigma_i, X_i) \cap \mathcal{S} \neq \emptyset \Leftrightarrow a_{\mathcal{S}}(\sigma_i) < \infty \text{ and } X_i \geq a_{\mathcal{S}}(\sigma_i). \quad (1.21)$$

Sur la Figure 1.1, on peut voir que l'individu 1 est inclus dans l'échantillon, tandis que l'individu 2 aurait pu l'être s'il n'était pas mort avant son inclusion puisque $x_2 < a_{\mathcal{S}}(s_2)$. Quant à l'individu 3, il ne peut être inclus car $a_{\mathcal{S}}(s_3) = \infty$.

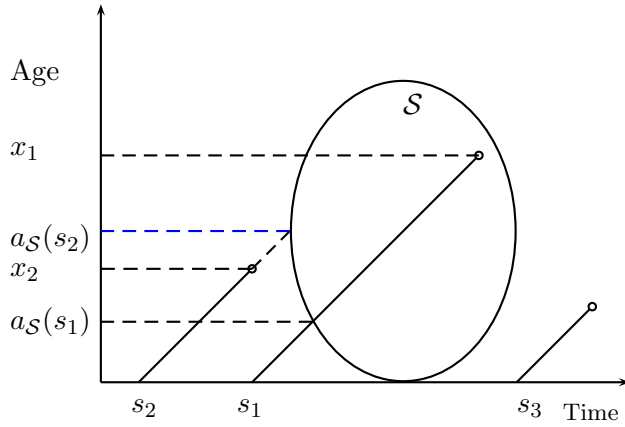


FIGURE 1.1 – Diagramme de Lexis

Nous utilisons la modélisation suivante du diagramme de Lexis et nous renvoyons à Guilloux (2007) pour une présentation plus détaillée. Soit $\eta = \sum_{i \in \mathcal{I}} \delta_{\sigma_i}$ le processus ponctuel de Poisson non-homogène sur \mathbb{R} d'intensité φ . Supposons que les durées de vie X_i , pour $i \in \mathcal{I}$ sont i.i.d. de densité commune f et de fonction de répartition associée F et qu'elles sont indépendantes de $(\sigma_i)_{i \in \mathcal{I}}$. Les propriétés des processus de Poisson assurent que le processus ponctuel $\mu = \sum_{i \in \mathcal{I}} \delta_{(\sigma_i, X_i)}$ est aussi un processus de Poisson non-homogène d'intensité φf .

Comme seuls sont inclus les individus dont la ligne de vie intersecte l'ensemble \mathcal{S} , seule la restriction $\mu|_{\mathcal{S}}$ du processus μ à l'ensemble \mathcal{S} nous intéresse. Etant donné le nombre $\mu(\mathcal{S})$ de points dans le borélien \mathcal{S} , les points du processus de Poisson $\mu|_{\mathcal{S}}$ se comportent comme des variables aléatoires indépendantes, de mesure de probabilité commune : $\mathbb{P}(\cdot) = \int_{\cdot \cap \mathcal{S}} \varphi f / \int_{\mathcal{S}} \varphi f$. On peut alors exprimer la distribution jointe du couple $(X_{\mathcal{S}}, \sigma_{\mathcal{S}})$, pour $x \in \mathbb{R}_+$ et $s \in \mathbb{R}$, par :

$$\mathbb{P}(X_{\mathcal{S}} \leq x, \sigma_{\mathcal{S}} \leq s) = \frac{\iint_{]-\infty, s] \times [0, x]} \mathbf{1}_{\{(u,v) \in \mathcal{S}\}} \varphi(u) f(v) dudv}{\mu_{\mathcal{S}}}$$

où

$$\mu_{\mathcal{S}} = \iint_{\mathbb{R} \times \mathbb{R}_+} \mathbf{1}_{\{a_{\mathcal{S}}(u) < \infty\}} \mathbf{1}_{\{a_{\mathcal{S}}(u) \leq v\}} \varphi(u) f(v) dudv.$$

On en déduit alors la loi marginale de la v.a. $X_{\mathcal{S}}$, pour tout $x \in \mathbb{R}_+$:

$$F_{\mathcal{S}}(x) = \mathbb{P}(X_{\mathcal{S}} \leq x) = \frac{1}{\mu_{\mathcal{S}}} \int_0^x w(v) f(s) ds, \text{ avec } w(x) = \int_{-\infty}^{\infty} \mathbf{1}_{\{a_{\mathcal{S}}(u) \leq x\}} \varphi(u) du.$$

Dans les études de type "time-window", la fonction de poids w est donnée par $w(x) = \int_{t_1-x}^{t_2} \varphi(u) du$, pour $x \geq 0$. Lorsque $t_1 = t_2$ et $\varphi = cste$, on parle de biais de longueur, cf. Asgharian *et al.* (2002), Asgharian (2003) et de Uña-Álvarez (2002).

Censure

Etant données la date de naissance σ_S et la durée de vie X_S , nous supposons que l'âge à l'inclusion $a_S(\sigma_S)$ est observé. La durée de vie X_S s'écrit alors :

$$X_S = \underbrace{a_S(\sigma_S)}_{\text{age à l'inclusion}} + \underbrace{(X_S - a_S(\sigma_S))}_{\text{durée passée dans l'étude}}.$$

La durée passée dans l'étude est $X_S - a_S(\sigma_S)$, et elle peut potentiellement être censurée. Pour prendre en compte cette éventuelle censure, nous introduisons la v.a. C positive, de fonction de répartition G , indépendante de X_S et $a_S(\sigma_S)$, telle que la durée observée est :

$$Z = a_S(\sigma_S) + (X_S - a_S(\sigma_S)) \wedge C.$$

Les observations sont alors constituées de :

$$\begin{cases} \sigma_{S,i} \\ Z_i = a_S(\sigma_{S,i}) + (X_{S,i} - a_S(\sigma_{S,i})) \wedge C_i, \quad \text{pour } i=1, \dots, n. \\ \mathbf{I}_{\{X_{S,i} - a_S(\sigma_{S,i}) \leq C_i\}} \end{cases} \quad (1.22)$$

Notre objectif est alors l'estimation de la densité f et de la fonction de risque instantané h de la v.a. X à partir de ces observations.

Processus de comptage pour l'estimation

Guilloux (2007) a introduit le processus de comptage $D(x)$ défini par :

$$D(x) = \sum_{i=1}^n \mathbf{I}_{\{Z_i \leq x, X_{S,i} - a_S(\sigma_{S,i}) \leq C_i\}}, \quad \text{pour } x \geq 0. \quad (1.23)$$

Pour $x \geq 0$, la v.a. $D(x)$ est le nombre de décès observés avant l'âge x dans l'échantillon. Nous introduisons aussi O et D les processus définis par :

$$\begin{aligned} O(x) &= \sum_{i=1}^n \mathbf{I}_{\{a_S(\sigma_{S,i}) \leq x \leq Z_i\}} \\ &= \sum_{i=1}^n \mathbf{I}_{\{a_S(\sigma_{S,i}) \leq x \leq X_{S,i}, x \leq a_S(\sigma_{S,i}) + C_i\}} \quad \text{pour } x \geq 0 \end{aligned} \quad (1.24)$$

$$D(x) = \sum_{i=1}^n \mathbf{I}_{\{Z_i \leq x, X_{S,i} - a_S(\sigma_{S,i}) \leq C_i\}} \quad (1.25)$$

La variable $O(x)$ représente le nombre d'individus à risque de décès à l'âge x . Ici, être à risque à l'âge x signifie que l'individu a été inclus dans l'étude à un âge antérieur et qu'il n'est ni mort, ni censuré avant l'âge x .

Enfin, nous avons également besoin d'introduire des estimateurs de la fonction de répartition F et de la fonction de risque cumulé Λ de la v.a. X . Leur construction imite celles des estimateurs de Nelson-Aalen et de Kaplan-Meier dans le contexte usuel de données censurées à droite :

$$\widehat{\Lambda}_n(x) = \int_0^x \frac{dD(s)}{O(s) + n\epsilon_n}, \text{ et } \widehat{F}_n(x) = 1 - \prod_{i: Z_{(i)} \leq x} \left(1 - \frac{\mathbf{1}_{\{X_{S,i} - a_S(\sigma_{S,i}) \leq C_i\}}}{O(Z_{(i)}) + n\epsilon_n} \right),$$

pour tout $x \geq 0$ où $(\epsilon_n)_{n \geq 1}$ est une suite de réels positifs tels que $\epsilon_n \rightarrow 0$ et $\sqrt{n}\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$, et qui garantit que le dénominateur ne s'annule pas. Guilloux (2007) a établi la convergence faible de ces estimateurs et obtenu l'inégalité de déviation suivante, analogue à celle de Bitouzé *et al.* (1999) pour l'estimateur de Kaplan-Meier.

Théorème 5 (Guilloux, 2007) *On suppose qu'il existe w_1 tel que, pour tout $x \geq 0$, $w_1 \leq w(x)$. Alors, pour tout $u > 0$:*

$$\mathbb{P} \left(\sqrt{n} \sup_{x \geq 0} \left| \left(\widehat{F}_n(x) - F(x) \right) (1 - G)(x) w_1 \right| > u \right) \leq 2.5 \exp(-2u^2 + Cu),$$

où C est une constante universelle.

Le lemme suivant est une conséquence du Théorème 5, et nous permet de contrôler les termes de reste qui apparaissent dans la décomposition du contraste pour l'estimation de la fonction de risque dans [10].

Lemme 1 *On suppose qu'il existe w_1 tel que, pour tout $x \geq 0$, $w_1 \leq w(x)$. Pour tout $k \in \mathbb{N}^*$, il existe une constante $C_F(k)$ telle que*

$$\mathbb{E} \left(\sup_{x \in A} |\widehat{F}_n(x) - F(x)|^{2k} \right) \leq C_F(k) n^{-k}.$$

1.2.4 Estimation adaptative de la densité et de la fonction de risque en présence d'un biais inconnu

Dans [10], nous proposons des estimateurs de la fonction de risque instantané h et de la densité f de X . Nous considérons des espaces S_m de dimension finie $D_m = m$ engendrés par la base trigonométrique $[T]$ (cf. Annexe A). Les estimateurs sont définis sur un intervalle $A = [0, \tau]$, où $\tau = \sup\{t > 0 : (1 - G(t))(1 - F(t)) > 0\}$. Nous posons $A = [0, 1]$.

Pour estimer la fonction de risque h , nous considérons le contraste suivant, pour $t \in S_m$:

$$\gamma_n(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \frac{\delta_i t(Z_i)}{O(Z_i)/n}. \quad (1.26)$$

Le contraste $\gamma_n(t)$ est la version empirique de $\|t\|^2 - 2\langle t, \lambda \rangle = \|t - \lambda\|^2 - \|\lambda\|^2$, où $\|t\|^2 = \int_0^1 t^2(x) dx$.

En le réécrivant

$$\gamma_n(t) = \|t\|^2 - 2 \int_0^1 t(x) d\widehat{\Lambda}_n(x)$$

avec $\hat{\Lambda}_n$, défini par (1.26) en prenant $\epsilon_n = 0$. Ce contraste est à rapprocher de celui que nous avons proposé, dans le contexte de la censure droite dans [5], et avec celui de Reynaud-Bouret (2006) pour le modèle d'intensité multiplicative de Aalen.

L'estimateur par projection sur un espace S_m est alors défini par $\hat{\lambda}_m = \arg \min_{t \in S_m} \gamma_n(t)$, et on construit l'estimateur pénalisé $\hat{\lambda}_{\hat{m}}$ de λ sur l'espace $S_{\hat{m}}$ avec

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left\{ \gamma_n(\hat{\lambda}_m) + \text{pen}(m) \right\} \text{ et } \text{pen}(m) = \kappa \left(\frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{O^2(Z_i)/n^2} \right) \frac{m}{n}.$$

Nous donnons dans le Théorème 3.1 de [10] une inégalité oracle pour le risque \mathbb{L}^2 de cet estimateur, de laquelle on peut déduire la vitesse de convergence minimax pour des fonctions de risque dans des classes de Besov. De même, nous renvoyons à l'article [10] pour une étude d'un estimateur adaptatif de la densité dans ce contexte, qui n'est pas présenté ici.

1.3 Estimation de la fonction de répartition avec censure par intervalle de type I

Un schéma d'échantillonnage, que je présente dans cette dernière partie du chapitre, concerne la censure par intervalle de type I. On considère la variable de durée X , de fonction de répartition F inconnue. Dans [12], nous nous intéressons à l'estimation de F . La variable X représente la durée écoulée jusqu'à ce que se produise un événement d'intérêt \mathcal{I} . Dans le modèle de censure par intervalle de type I, la variable X n'est pas observée. On observe le couple (U, δ) où U est la durée qui s'écoule jusqu'à une date E , où l'on évalue l'état de l'individu et δ est l'indicatrice de $(X \leq U)$. La seule connaissance que nous avons de la variable de durée X est la survenue ou non-survenue de l'événement \mathcal{I} avant la date E . On retrouve ce type d'observations, sous le terme de *current status data* dans la littérature anglo-saxonne. Les applications sont nombreuses et diverses : en épidémiologie, ce schéma de censure est intéressant dans les études sur le SIDA ou les maladies infectieuses pour lesquelles, il est difficile de connaître avec exactitude la date de contamination. Dans ce contexte, l'événement d'intérêt \mathcal{I} peut, par exemple, correspondre à la séroconversion d'un individu. Des exemples ont été donnés en démographie (cf. Diamond & McDonald (1991)) ou en sciences de l'éducation, si l'on s'intéresse, par exemple, à l'acquisition d'une compétence comme la lecture pour un jeune enfant. On suppose que U est indépendante de X . On note respectivement G et g la fonction de répartition et la densité de probabilité de la variable U . Avec ces notations, les observations sont constituées de $(U_1, \delta_1), \dots, (U_n, \delta_n)$ un échantillon du couple (U, δ) où, $\delta_i = \mathbf{1}_{(X_i \leq U_i)}$ pour $i = 1, \dots, n$.

La stratégie d'estimation pour de telles données est très différente de celle qui est mise en œuvre dans le contexte de la censure à droite. Dans le modèle avec censure droite, l'estimateur de Kaplan & Meier (1958) de la fonction de survie a été étudié depuis longtemps et il est établi

qu'il converge en loi vers une gaussienne à la vitesse $1/\sqrt{n}$. Les données censurées par intervalle ont connu un intérêt plus récent, mais ont néanmoins été étudiées depuis la fin des années 80 : on peut consulter par exemple Jewell & van der Laan (2004) pour un état de l'art sur le sujet. Dans le modèle de censure par intervalle, Groeneboom & Wellner (1992) ont montré que l'estimateur non-paramétrique du maximum de vraisemblance (NPMLE) de la fonction de survie était uniformément convergent, et qu'il converge en loi vers une distribution non gaussienne à la vitesse $n^{-1/3}$. De son côté, van de Geer (1993) a établi que le NPMLE converge au sens de la norme \mathbb{L}^2 à la vitesse $n^{-1/3}$. Citons également, Hudgens *et al.* (2007) qui construisent divers estimateurs à partir du NPMLE. Birgé (1999) propose également un estimateur par histogramme, très facile à implémenter, et qui atteint la vitesse de convergence optimale pour le risque \mathbb{L}^1 .

Les travaux récents prennent désormais deux directions. Tout d'abord, l'un des axes de recherche se concentre sur l'amélioration de la vitesse de convergence, dans le cas où la fonction de survie est plus régulière, pour proposer des procédures adaptatives. L'estimateur localement linéaire proposé par Yang (2000), peut ne pas être monotone, ce qui n'est pas le cas du NPMLE. En revanche, il atteint de meilleures vitesses de convergence sous certaines hypothèses de régularité de la fonction de densité f . Dans la même veine, Ma & Kosorok (2006) ont introduit un estimateur de type moindre carrés pour des bases de splines, dans un contexte semi-paramétrique.

Dans [12], nous proposons des estimateurs adaptatifs de la fonction de répartition F . Nous adaptons les techniques de sélection de modèles en régression au contexte de la censure par intervalle, qui rendent notre solution attrayante, tant du point de vue théorique qu'algorithmique.

Enfin, il faut aussi mentionner un second axe de recherche qui élargit le présent cadre d'étude, en introduisant dans le modèle des covariables possiblement multivariées ou dépendantes du temps. C'est le cas des travaux de van der Vaart & van der Laan (2006) ou très récemment, de Plancade (2011) qui propose une généralisation de notre contraste en présence d'une covariable.

Stratégie quotient

On définit la densité ψ de U tels que $\delta = 1$. Comme X et U sont indépendantes, on a

$$\mathbb{P}(U \leq x, \delta = 1) = \mathbb{P}(U \leq x, X \leq U) = \int_0^x F(u)g(u)du$$

et donc,

$$\psi(x) = F(x)g(x). \tag{1.27}$$

Cette expression suggère de construire un estimateur de la fonction ψ , un estimateur de la fonction g , puis d'en faire le quotient. Nous considérons des familles de modèles $\{S_m, m \in \mathcal{M}_n\}$, cf Annexe A.

Etape 1 : On construit un estimateur adaptatif de la densité g :

$$\hat{g}_m = \arg \min_{t \in S_m} \gamma_n^g(t) \quad \text{with } \gamma_n^g(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n t(U_i).$$

puis $\tilde{g} = \hat{g}_{\hat{m}_g}$ avec $\hat{m}_g = \arg \min_{m \in \mathcal{M}_n} [\gamma_n^g(\hat{g}_m) + \kappa \Phi_0^2 D_m/n]$.

Etape 2 : On construit un estimateur adaptatif $\tilde{\psi}$ de la densité ψ :

$$\hat{\psi}_m = \arg \min_{t \in S_m} \gamma_n^\psi(t) \quad \text{avec } \gamma_n^\psi(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \delta_i t(U_i),$$

puis on définit $\tilde{\psi} = \hat{\psi}_{\hat{m}}$ où $\hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n^\psi(\hat{\psi}_m) + \widehat{\text{pen}}^\psi(m)]$, et

$$\widehat{\text{pen}}^\psi(m) = \kappa \Phi_0^2 \left(\frac{1}{n} \sum_{i=1}^n \delta_i \right) \frac{D_m}{n}.$$

Etape 3 : On définit l'estimateur de la fonction de répartition F par :

$$\tilde{F}(x) = \begin{cases} 0 & \text{si } \tilde{\psi}(x)/\tilde{g}(x) < 0 \\ \frac{\tilde{\psi}(x)}{\tilde{g}(x)} & \text{si } 0 \leq \tilde{\psi}(x)/\tilde{g}(x) \leq 1 \\ 1 & \text{si } \tilde{\psi}(x)/\tilde{g}(x) > 1 \end{cases}$$

Proposition 2 *Si la densité g de U est bornée inférieurement par une constante $g_0 > 0$ et sous les hypothèses du Lemme 3.1 dans [12],*

$$\mathbb{E} \|\tilde{F} - F\|^2 \leq \frac{2^4}{g_0^2} \left(\mathbb{E} \|\tilde{\psi} - \psi\|^2 + \mathbb{E} \|\tilde{g} - g\|^2 \right) + \frac{C(g_0, \|\psi\|)}{n}, \quad (1.28)$$

avec $C(g_0, \|\psi\|)$ qui dépend de g_0 et de $\|\psi\|$.

Si l'on suppose que $g \in \mathcal{B}_{\alpha_g, 2, \infty}([0, 1])$ et $\psi \in \mathcal{B}_{\alpha_\psi, 2, \infty}([0, 1])$, on déduit de l'inégalité (1.12), la vitesse de convergence de l'estimateur adaptatif \tilde{F} . Si la régularité de F , α_F , est plus grande que celle de $\psi = Fg$, α_ψ , alors la vitesse de convergence de l'estimateur \tilde{F} est donnée par $n^{-\alpha_\psi/(1+2\alpha_\psi)}$ au lieu de la vitesse optimale $n^{-\alpha_F/(1+2\alpha_F)}$. C'est pourquoi nous proposons une stratégie directe.

Stratégie directe par moindres carrés

Le mécanisme de censure est tel que la loi conditionnelle de $\delta = \mathbb{I}_{(X \leq U)}$ étant donné $U = u$ est une loi de Bernoulli de paramètre $F(u)$. Ainsi, on a :

$$\mathbb{E}(\delta|U = u) = F(u) \quad (1.29)$$

Un estimateur direct de la fonction de répartition F peut donc être défini en considérant le contraste de moindres carrés en régression :

$$\gamma_n^{\text{MS}}(t) = \frac{1}{n} \sum_{i=1}^n [\delta_i - t(U_i)]^2$$

Puis, on pénalise le contraste et on obtient l'estimateur $\hat{F}_{\hat{m}_0}$ avec :

$$\hat{m}_0 = \arg \min_{m \in \mathcal{M}_n} \{ \gamma_n^{MS}(\hat{F}_m) + \kappa_0 \frac{D_m}{n} \}.$$

où κ_0 est une constante numérique. On obtient une inégalité oracle pour l'estimateur $\hat{F}_{\hat{m}_0}$ et un résultat d'adaptation valable quelle que soit la régularité de ψ . Globalement, la stratégie par moindres carrés est préférable à celle du quotient, tant du point de vue théorique que numérique, comme le montre l'étude de simulations dans [12]. Notre procédure fournit un estimateur qui n'est pas nécessairement une fonction de répartition. Nous appliquons une transformation quantile proposée par Chernozhukov et al. (2009) pour monotoniser l'estimateur $\hat{F}_{\hat{m}_0}$, qui sera présentée plus en détail dans la partie 2.2 du chapitre 2. L'estimateur résultant possède les mêmes propriétés en terme de risque que l'estimateur pénalisé $\hat{F}_{\hat{m}_0}$ (cf. [12]).

Chapitre 2

Espérance et lois conditionnelles pour données censurées

Publications [6], [7], [9] et [13]

Dans le chapitre précédent, j'ai présenté différents estimateurs de courbes associées à la loi marginale de la durée de vie sans prendre en compte la présence de covariables. Je m'attache désormais, dans cette partie, à décrire les modèles d'espérance et de lois conditionnelles de la variable de durée Y , censurée à droite, étant donnée une covariable continue \vec{X} . Dans notre étude, les prédicteurs \vec{X} sont complètement observés. L'apport des résultats obtenus par Claire Lacour (2007) pour la partie 2.2 est essentiel.

2.1 Régression en présence d'une ou plusieurs covariables

Le modèle

On considère le vecteur d'observations $(\vec{X}_i, Y_i), i = 1, \dots, n$ où les durées de Vie Y_i , pour $i = 1, \dots, n$ sont des variables aléatoires positives de même loi de fonction de répartition F et le vecteur $\vec{X}_i \in A = [0, 1]^d$ pour $i = 1, \dots, n$ désigne les covariables. Pour $T > 0$ une durée d'étude fixée, on définit la variable

$$Y_{i,T} = Y_i \wedge T, \text{ pour } i = 1, \dots, n.$$

On considère que la variable $Y_{i,T}$ peut potentiellement être censurée à droite et on introduit les variables de censure C_i , pour $i = 1, \dots, n$ indépendantes et de fonction de répartition commune notée G . Avec ces notations, les observations sont constituées du triplet $(\vec{X}_i, Z_i, \delta_i)$ où

$$Z_i = Y_{i,T} \wedge C_i \quad \text{et} \quad \delta_i = \mathbf{1}_{(Y_{i,T} \leq C_i)} \quad \text{pour } i = 1, \dots, n.$$

Dans ce modèle, on s'intéresse à l'espérance conditionnelle $r_T(\vec{X}) := \mathbb{E}(Y_T | \vec{X})$. On suppose que les variables C_i sont indépendantes du couple $(\vec{X}_i, Y_{i,T})$ pour $i = 1, \dots, n$ et que les fonctions

de répartitions F et G sont à support dans \mathbb{R}^+ tout entier. Cette dernière hypothèse n'est pas très restrictive car la plupart des modèles paramétriques utilisés en analyse des durées de vie la vérifie et elle a pour conséquence :

$$\mathbb{P}(Y_i \geq T) = \mathbb{P}(Y_{i,T} = T) > 0 \text{ et } \mathbb{P}(C_i > T) > 0.$$

Ainsi, on a :

$$\begin{cases} \forall i = 1, \dots, n, 1 - G(Y_{i,T}) \geq 1 - G(T) := c_G, \\ \forall t \in [0, T], 1 - F(t) \geq 1 - F(T) := c_F > 0. \end{cases}$$

Pour corriger la présence de la censure, on cherche une transformation des observations $\varphi_a(Z)$, qui laisserait inchangée l'espérance conditionnelle dans le sens où

$$\mathbb{E}(\varphi_a(Z)|\vec{X}) = \mathbb{E}(Y_T|\vec{X}).$$

Soit $\bar{G}(t) = 1 - G(t)$ la survie de la variable de censure C , on peut poser

$$\varphi_a(Z) = (1 + a) \int_0^Z \frac{dt}{1 - G(t)} - a \frac{\delta Z}{1 - G(t)}, \quad \alpha \in \mathbb{R}.$$

Puis, on remplace \bar{G} inconnue par la version modifiée de son estimateur de Kaplan-Meier $\hat{\bar{G}}$ donnée par (1.1),

$$\hat{\varphi}_a(Z) = (1 + a) \int_0^Z \frac{dt}{\hat{\bar{G}}(t)} - a \frac{\delta Z}{\hat{\bar{G}}(t)}.$$

Pour optimiser le choix de la constante a , on propose de minimiser la variance des données transformées $\text{Var}(\hat{\varphi}_a(Z))$:

$$\hat{a} = \arg \min_a \text{Var}(\hat{\varphi}_a(Z)) = -\text{cov} \left(\int_0^Z \frac{dt}{\hat{\bar{G}}(t)}; \frac{\delta Z}{\hat{\bar{G}}(t)} \right) / \text{var} \left(\frac{\delta Z}{\hat{\bar{G}}(t)} \right),$$

Ce type de transformation très générale a été introduit par Fan & Gijbels (1996) en combinant les transformations antérieures proposées par Leurgans (1987), pour $a = 0$ et Koul *et al.* (1981) pour $a = -1$. Des estimateurs non-paramétriques ont été proposés par, Zheng (1987), Dabrowska (1987) et Zheng (1988). Györfi *et al.* (2002) ont étudié la consistance de l'estimateur de Stone de la régression en présence de censure. Heuchenne & Van Keilegom (2007) considèrent un modèle semi-paramétrique de régression pour données censurées.

Récemment, un modèle à directions révélatrices ou *single index* a été proposé par Lopez *et al.* (2013) pour estimer la loi du couple (Y, \vec{X}) . Cette approche permet une réduction de la dimension sous l'hypothèse qu'il existe une fonction $g : \mathbb{R}^d \rightarrow \mathbb{R}$ telle que Y et C sont indépendantes sachant $g(\vec{X})$ et que $\mathbb{P}(Y \leq C|\vec{X}, Y) = \mathbb{P}(Y \leq C|g(\vec{X}), Y)$. Ce sont actuellement des approches, avec des développements importants pour traiter la grande dimension et qui sont des alternatives au modèle additif que nous traitons dans [6] et [9].

Nous proposons un estimateur obtenu par la minimisation du contraste des moindres carrés où l'on substitue aux Z_i les $\hat{\varphi}_{\hat{a}}(Z_i)$, sur une famille d'espaces vectoriels S_m , de dimension finie D_m , engendrés par des bases orthonormales (cf. Annexe A) :

$$\gamma_n(t) = \frac{1}{n} \sum_{i=1}^n \left[\hat{\varphi}_{\hat{a}}(Z_i) - t(\vec{X}_i) \right]^2$$

avec $\dim(S_m) = D_m < +\infty$. La norme empirique associée au design $\|t\|_n^2 = \frac{1}{n} \sum_{i=1}^n t^2(\vec{X}_i)$ apparaît naturellement avec ce contraste de type moindres-carrés. Nous étudions le risque associé à cette norme empirique. Il est toujours possible de calculer l'estimateur aux points $\vec{X}_1, \dots, \vec{X}_n$, même si la résolution du problème de minimisation général requiert l'inversion d'une matrice. En effet, le vecteur $[\hat{r}_m(\vec{X}_1), \dots, \hat{r}_m(\vec{X}_n)]'$ est la projection orthogonale dans \mathbb{R}^n du vecteur $[\hat{\varphi}_{\hat{a}}(Z_1), \dots, \hat{\varphi}_{\hat{a}}(Z_n)]$ sur le s.e.v engendré par $\{t(\vec{X}_1), \dots, t(\vec{X}_n), t \in S_m\}$. L'estimateur calculé aux points du design est toujours bien défini.

Pour le contrôle du risque, nous fixons la valeur de $\hat{a} = -1$. Le critère $\gamma_n(t)$ est alors défini avec $\hat{\varphi}_{-1}(Z_i) = \frac{\delta_i Z_i}{1 - \hat{G}(Z_i)}$ et l'étude théorique de l'estimateur n'est proposée que dans ce cas.

En dimension 1, si la fonction r_T est de régularité α (par exemple r_T appartient à un espace de Besov $\mathcal{B}_{2,\infty}^\alpha$ et $r > \alpha > 0$, alors pour un modèle S_m satisfaisant $m = m(n)$ tel que $D_{m(n)} = O(n^{1/(2\alpha+1)})$, on a

$$\mathbb{E}(\|r_T - \hat{r}_{m(n)}\|_n^2) = O\left(n^{-\frac{2\alpha}{2\alpha+1}}\right).$$

En dimension supérieure, le modèle additif permet de donner un cadre utilisable en pratique pour un nombre raisonnable de covariables (si \vec{X} est un vecteur de taille 2 ou 3, cf. [9]). On suppose que $\vec{X}_i = (X_i^{(1)}, \dots, X_i^{(d)})$, $i = 1, \dots, n$ et que l'espérance conditionnelle s'exprime :

$$\mathbb{E}(Y_{i,T} | \vec{X}_i) = r_T(\vec{X}_i) = r_{T,1}(\vec{X}_i^{(1)}) + \dots + r_{T,d}(\vec{X}_i^{(d)}).$$

Pour l'identifiabilité, on suppose que $\mathbb{E}(r_{T,j}(X^{(j)})) = 0$, $j = 2, \dots, d$ et $\forall i$.

On démontre que :

$$\mathbb{E}(\|r_T - \hat{r}_{m(n)}\|_n^2) = O\left(n^{-\frac{2\alpha}{2\alpha+1}}\right).$$

avec $\underline{\alpha} = \min(\alpha_1, \dots, \alpha_d)$. Dans le cas isotrope, la vitesse de convergence est d'ordre $n^{-2\underline{\alpha}/(2\underline{\alpha}+d)}$. Dans le cas anisotrope, c'est-à-dire si on a des régularités différentes selon les directions, soit $\alpha = (\alpha_1, \dots, \alpha_d)'$, la vitesse est d'ordre $n^{-2\bar{\alpha}/(2\bar{\alpha}+d)}$ où $\bar{\alpha} = d / \sum_{i=1}^d \alpha_i^{-1}$ est la moyenne harmonique des régularités α_i dans chacune des directions. Nous renvoyons à [6] pour plus de détails. On obtient ensuite une inégalité oracle pour l'estimateur des moindres carrés pénalisé $\hat{r}_{\hat{m}}$ avec :

$$\hat{r}_m = \arg \min_{t \in S_m} \gamma_n(t) \quad \text{et} \quad \hat{m} = \arg \min_{m \in \mathcal{M}_n} [\gamma_n(\hat{r}_m) + \widehat{\text{pen}}(m)] \quad (2.1)$$

où

$$\widehat{\text{pen}}(m) = \kappa \Phi_0^2 \left(\frac{1}{n} \sum_{i=1}^n \hat{\varphi}_{-1}^2(Z_i) \right) \frac{D_m}{n}$$

κ est une constante universelle dont la valeur numérique est calibrée par simulations.

Théorème 6 *On considère les collections d'espaces emboîtés [T], [DP] ou [W] (cf. annexe A), et on suppose que $|\mathcal{M}_n| \leq n$, $D_m \leq N_n \leq \sqrt{n}$, $\forall m \in \mathcal{M}_n$ où N_n désigne la dimension maximale des espaces S_m . Si de plus, $0 < f_0 \leq f \leq f_1 < +\infty$, où f est la densité des \vec{X}_i , et si $\mathbb{E}(Y_1^8) < \infty$, alors, l'estimateur $\hat{r}_{\hat{m}}$ défini par (2.1) satisfait,*

$$\mathbb{E}(\|\hat{r}_{\hat{m}} - r_T\|_n^2) \leq K \inf_{m \in \mathcal{M}_n} (\|r_T - r_m\|^2 + \text{pen}(m)) + \frac{K' \sqrt{\ln(n)}}{n}.$$

L'un des points clé de la démonstration de ce résultat s'appuie sur la substitution des $\hat{\varphi}_{-1}(Z_i)$ aux $\varphi_{-1}(Z_i)$ dans le contraste. Pour cela, nous avons besoin de contrôler la probabilité de déviation en norme uniforme de l'estimateur de la survie \hat{G} . C'est ce qui est établi dans le lemme suivant :

Lemme 2 *Pour tout $k \in \mathbb{N}^*$, il existe une constante $C_k > 0$, qui dépend de k , c_G et c_F , et qui vérifie :*

$$\mathbb{E} \left(\sup_{y \in [0, T]} |\hat{G}(y) - \bar{G}(y)|^{2k} \right) \leq \frac{C_k}{n^k}.$$

La preuve du lemme s'obtient grâce à une borne exponentielle non-asymptotique pour l'estimateur de Kaplan-Meier démontrée par Bitouzé *et al.* (1999).

2.2 Lois conditionnelles

Publications [7] et [13] en collaboration avec F. Comte et C. Lacour

Dans cette partie, j'explique comment on aboutit à des stratégies d'estimation par minimum de contraste pénalisé pour la densité et la fonction de répartition (f.d.r) conditionnelles. Ces travaux ne sont pas spécifiques au cadre où la variable d'intérêt Y est censurée. Cependant, nous avons toujours en tête ce type applications et les deux papiers abordent ce problème en proposant les corrections adéquates pour prendre en compte une éventuelle censure.

Nous cherchons à estimer la densité conditionnelle $\pi(x, y)$ sur un compact $A = A_1 \times A_2$ à partir d'un échantillon i.i.d $(X_1, Y_1), \dots, (X_n, Y_n)$. L'estimation de la densité conditionnelle de Y sachant $X = x$ apporte une information qualitative plus riche que le (très étudié) modèle de régression, qui ne porte que sur l'espérance conditionnelle. Cette densité conditionnelle s'exprime comme un quotient, donné par :

$$\pi(x, y) = \frac{f_{X,Y}(x, y)}{f_X(x)} \quad \text{si } f_X(x) > 0$$

avec les notations naturelles $f_{X,Y}$ et f_X pour respectivement la densité du couple (X, Y) et la densité marginale de X . On se rend compte que finalement assez peu de références concernent l'estimation de la densité conditionnelle, même dans un cadre non adaptatif. Pour des observations i.i.d., on peut citer Györfi & Kohler (2007) pour des estimateurs par histogrammes ou Faugeras (2009) pour une approche par noyaux utilisant les copules. Pour des observations dépendantes, citons De Gooijer & Zerom (2003) et nous renvoyons à Lacour (2007) pour une revue sur l'estimation de la probabilité de transition d'une chaîne de Markov qui correspond au cas particulier où $Y_i = X_{i+1}$. Cependant, des estimateurs adaptatifs de la densité conditionnelle n'ont été proposés que très récemment. Pour des données indépendantes, Efromovich (2007), Efromovich (2008) a donné des inégalités oracles et des résultats d'adaptativité au sens mini-max. Tandis que les travaux d'Efromovich considèrent une décomposition en base de Fourier pour construire un estimateur à seuillage par blocs, nous proposons, dans [7], une procédure de sélection de modèles basé sur un critère de moindres carrés pénalisés. Toutes les méthodes adaptatives que nous venons d'évoquer prennent en compte une possible anisotropie de la densité conditionnelle, c'est-à-dire des régularités différentes dans chacune des deux directions. Dans notre proposition, nous considérons des fonctions appartenant à des espaces de Besov $\mathcal{B}_{2,\infty}^\alpha$.

2.2.1 Des histoires de contrastes

Contraste pour une densité

Si l'on veut estimer la densité univariée f_Y de Y (sur un compact A_1) à partir d'observations Y_1, \dots, Y_n , le contraste classique associé à la fonction de perte $\ell(f_Y, t) = \|f_Y - t\|^2$ est

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

En effet, la fonction \hat{f}_{m_1} qui minimise $\gamma_n^{(1)}$ sur le sous-espace $S_{m_1}^{(1)}$, décrit ci-dessus, s'obtient directement

$$\hat{f}_{m_1} = \sum_{j \in J_{m_1}} \hat{a}_j \varphi_j^{m_1}, \quad \text{où } \hat{a}_j = \frac{1}{n} \sum_{i=1}^n \varphi_j^{m_1}(Y_i).$$

Notons $f_{m_1} = \sum_{j \in J_{m_1}} a_j \varphi_j^{m_1}$ la projection orthogonale au sens du produit scalaire dans \mathbb{L}^2 de f_Y sur le sous-espace S_{m_1} . Le Théorème de Pythagore nous permet de voir facilement que

$$\begin{aligned} \|f_Y - \hat{f}_{m_1}\|^2 &= \|f_Y - f_{m_1}\|^2 + \|f_{m_1} - \hat{f}_{m_1}\|^2 \\ &= \|f_Y - f_{m_1}\|^2 + \sum_{j \in J_{m_1}} (\hat{a}_j - a_j)^2, \end{aligned}$$

et il s'ensuit

$$\begin{aligned}
\mathbb{E}(\|f_Y - \hat{f}_{m_1}\|^2) &= \|f_Y - f_{m_1}\|^2 + \frac{1}{n} \sum_{j \in J_{m_1}} \text{Var}(\varphi_j^{m_1}(Y_1)) \\
&\leq \|f_Y - f_{m_1}\|^2 + \frac{1}{n} \sum_{j \in J_{m_1}} \mathbb{E}[(\varphi_j^{m_1}(Y_1))^2] \\
&\leq \|f_Y - f_{m_1}\|^2 + \phi_1 \frac{D_{m_1}}{n}.
\end{aligned}$$

où la constante ϕ_1 ne dépend que de la base $(\varphi_j)_{j \in J_{m_1}}$.

Contraste de régression

Maintenant, examinons le cas où nous sommes intéressés par l'estimation d'une fonction de régression r donnée par $\mathbb{E}(Y|X = x) =: r(x)$ à partir d'observations i.i.d. $(X_i, Y_i)_{1 \leq i \leq n}$, le contraste de moindres carrés associé à la perte $\|b - t\|_{f_X}^2$ où la norme de référence du problème de régression $\|t\|_{f_X}^2 = \int t^2(x) f_X(x) dx$ est pondérée par la densité marginale f_X du *design* est donné par :

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

ou bien

$$\gamma_n^{(3)}(t) = \frac{1}{n} \sum_{i=1}^n (t^2(X_i) - 2Y_i t(X_i)).$$

La décomposition du risque fournit (comme dans le cas du contraste de densité) un terme de biais au carré et un terme de variance d'ordre $\phi_1 D_{m_1}^{(1)}/n$.

Contraste mixte densité/régression

Pour l'estimation de la densité conditionnelle $\pi(x, y)$ de Y sachant $X = x$, une stratégie mixte est mise en place. Puisqu'il s'agit d'estimer une fonction définie dans \mathbb{R}^2 , nous avons besoin d'introduire une famille de modèles adéquate $\{S_m, m \in \mathcal{M}_n\}$ d'espaces de projection, avec \mathcal{M}_n un ensemble d'indices multiples (cf. Annexe A). Pour chaque couple d'indices $m = (m_1, m_2)$, S_m est un espace de fonctions à support dans $A = A_1 \times A_2$ défini comme un espace produit de $S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$ avec $S_{m_i}^{(i)} \subset (L^2 \cap L^\infty)(\mathbb{R})$, pour $i = 1, 2$, engendrés par deux bases orthonormales *a priori* différentes $(\varphi_j^{m_1})_{j \in J_{m_1}}$ avec $|J_{m_1}| = D_{m_1}^{(1)}$ et $(\psi_k^{m_2})_{k \in K_{m_2}}$ avec $|K_{m_2}| = D_{m_2}^{(2)}$. Ainsi, on a :

$$S_{m_1}^{(1)} = \{t / t(x) = \sum_{j \in J_{m_1}} a_j^{m_1} \varphi_j^{m_1}(x)\},$$

$$S_{m_2}^{(2)} = \{t / t(y) = \sum_{k \in K_{m_2}} a_k^{m_2} \psi_k^{m_2}(y)\}$$

et

$$S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)} = \{T / T(x, y) = \sum_{j \in J_{m_1}} \sum_{k \in K_{m_2}} A_{j,k}^m \varphi_j^{m_1}(x) \psi_k^{m_2}(y), A_{j,k}^m \in \mathbb{R}\}.$$

A partir d'observations i.i.d. $(X_i, Y_i)_{1 \leq i \leq n}$, on définit le contraste

$$\gamma_n^{(4)}(T) = \frac{1}{n} \sum_{i=1}^n \left(\int T^2(X_i, y) dy - 2T(X_i, Y_i) \right).$$

Le contraste $\gamma_n^{(4)}$ s'interprète comme un contraste de densité par rapport à la coordonnée y et comme un contraste de régression vis à vis de la coordonnée x .

2.2.2 Contraste pour la densité conditionnelle en présence de censure

Dans le cas où la variable Y est censurée, on modifie le contraste en utilisant une transformation analogue à celle de la partie 2.1 qui devient :

$$\gamma_n^{(4)}(T) = \frac{1}{n} \sum_{i=1}^n \left(\int T^2(X_i, y) dy - 2\hat{w}_i T(X_i, Z_i) \right) \text{ avec } \hat{w}_i = \begin{cases} 1 & \text{si aucune censure} \\ \frac{\delta_i}{\hat{G}(Z_i)} & \text{si censure} \end{cases}$$

L'existence du minimiseur du contraste n'est pas toujours garantie. Cependant la proposition 2.1 dans [7] assure qu'il est toujours possible de définir une fonction $\hat{\pi}_m(\cdot, \cdot)$ dans S_m aux points (X_i, y) :

$$(\hat{\pi}_m(X_i, y))_{1 \leq i \leq n} = P_{\mathcal{W}} \left(\left(\sum_k \psi_k(Z_i) \psi_k(y) \right)_{1 \leq i \leq n} \right)$$

où $P_{\mathcal{W}}$ désigne le projecteur orthogonal sur $\mathcal{W} = \{(t(X_i, y))_{1 \leq i \leq n}, t \in S_m\}$ pour le produit scalaire euclidien dans \mathbb{R}^n . Ainsi, la norme qui apparaît naturellement dans ce problème est la norme empirique suivante :

$$\|t\|_n = \left(\frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} t^2(X_i, y) dy \right)^{1/2}$$

On montre dans [7] que la décomposition du risque empirique donne un terme de biais au carré et un terme de variance majoré par $\phi_1 \phi_2 D_{m_1}^{(1)} D_{m_2}^{(2)} / n$ où la constante ϕ_1 (resp. ϕ_2) ne dépend que de la base $(\varphi_j)_{j \in J_{m_1}}$ (resp. $(\psi_k)_{k \in K_{m_2}}$).

Les dimensions $D_{m_1}^{(1)}$ et $D_{m_2}^{(2)}$ doivent être choisies de façon optimale et la vitesse résultante dépend de la régularité de la fonction inconnue $\pi(x, y)$ selon les deux directions en x et en y . L'estimateur pénalisé $\tilde{\pi} = \hat{\pi}_{\hat{m}}$ est défini par $\hat{\pi}_m = \operatorname{argmin}_{t \in S_m} \gamma_n^{(4)}(t)$, et $\hat{m} = \operatorname{argmin}_{m \in \mathcal{M}_n} \gamma_n^{(4)}(\hat{\pi}_m) + \operatorname{pen}(m)$ avec

$$\operatorname{pen}(m) = \kappa \frac{\|\pi\|_{\infty}}{c_G} D_{m_1} D_{m_2} / n \text{ ou bien } \operatorname{pen}(m) = \kappa \frac{\Phi_0}{f_0} \mathbb{E} \left(\frac{\delta_1}{\hat{G}^2(Z_1)} \right) D_{m_1} D_{m_2} / n$$

où l'on pose

$$f_0 = \inf_{x \in A_1} f_X(x) \quad \text{et} \quad c_G = \inf_{y \in A_2} \bar{G}(y). \quad (2.2)$$

Dans les deux cas, on montre une inégalité oracle avec des conditions légèrement plus restrictives sur les dimensions maximales pour la seconde pénalité, qui en pratique est préférable car elle fait intervenir le moment d'ordre deux de $\delta_1/\bar{G}(Z_1)$ au lieu de la constante c_G . Nous renvoyons à la publication [7] pour l'énoncé détaillé du résultat. Notre résultat est donné pour le risque associé à la norme empirique, mais si nous étions intéressés par le risque \mathbb{L}^2 , nous pourrions seuller notre estimateur cf. Remarque 1 dans [7]. On en déduit que notre estimateur pénalisé atteint la vitesse optimale et minimax, dont la borne inférieure a été obtenue par Lacour (2007).

Les travaux récents de Akakpo & Lacour (2011) ont permis de traiter, pour le même contraste, le cas de fonctions avec régularité inhomogène (c'est-à-dire $\pi \in \mathcal{B}_{p,\infty}^\alpha$ et $0 < p < 2$), grâce à un choix de collections de modèles plus riches (bases de polynômes par morceaux sur des partitions irrégulières). Cohen & Lepennec (2011) explorent une approche par sélection de modèles pour un critère de maximum de vraisemblance et étudient un risque de type divergence de Kullback-Leibler. Les outils reposent sur des contrôles d'entropie. Enfin, citons aussi le travail de Chagny (2011) qui propose un estimateur de la densité conditionnelle construit à partir de bases déformées. Ces contributions montrent l'intérêt renouvelé de la communauté autour de la densité conditionnelle.

2.2.3 Estimation optimale pour la fonction de répartition conditionnelle

Les méthodes non-paramétriques existantes pour estimer la fonction de répartition conditionnelle $F(y|x)$ ne sont pas adaptatives la plupart du temps. Stute (1986), dans le contexte de données complètement observées ou Dabrowska (1989), pour des observations censurées étudient les propriétés de l'estimateur de Beran, qui généralise l'estimateur de Kaplan-Meier en présence de covariables. Ces travaux mettent en œuvre des stratégies par noyaux, sous l'hypothèse que la fonction de répartition est deux fois différentiable par rapport à x et le choix de la fenêtre optimale n'est pas adaptatif. Dans [13], nous proposons une procédure d'estimation adaptative en considérant des fonctions de répartition de régularité quelconque.

Commençons par examiner le cas de l'estimation de la fonction de répartition d'une variable Y , sans covariable X . En modifiant le contraste de densité $\gamma_n^{(1)}$, on définit :

$$\gamma_n^{(5)}(t) = \|t\|^2 - \frac{2}{n} \sum_{i=1}^n \int t(y) \mathbf{I}_{(Y_i \leq y)} dy,$$

qui donne en minimisant par rapport à $t \in \mathcal{S}_{m_2}^{(2)}$ l'estimateur

$$\hat{F}_{Y,m_2}(y) = \sum_{j \in J_{m_2}} \hat{a}_j \psi_j^{m_2}(y) \quad \text{où} \quad \hat{a}_j = \frac{1}{n} \sum_{i=1}^n \int \psi_j^{m_2}(y) \mathbf{I}_{(Y_i \leq y)} dy = \int \psi_j^{m_2}(y) F_n(y) dy.$$

Cette expression fait apparaître l'estimateur $\hat{F}_{Y,m_2}(y)$ comme la projection orthogonale de la fonction de répartition empirique $F_n(y) = (1/n) \sum_{i=1}^n \mathbf{I}_{(Y_i \leq y)}$ sur $S_{m_2}^{(2)}$. Par conséquent, on obtient un estimateur plus lisse, si la base de projection est régulière, en revanche aucun gain n'est à attendre sur la vitesse de convergence, qui est déjà la vitesse paramétrique de la fonction de répartition empirique. L'estimateur \hat{F}_{Y,m_2} atteint donc la vitesse paramétrique si le biais est assez petit, puisque la variance ne dépend pas de la dimension de $S_{m_2}^{(2)}$. Ce comportement caractéristique reste inchangé avec l'ajout d'une covariable. Comme pour l'estimation de la densité conditionnelle, nous proposons un contraste *mixte* de densité/régression, pour toute fonction $T \in S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$ défini comme suit :

$$\Gamma_n^0(T) = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} (T^2(X_i, y) - 2T(X_i, y) \mathbf{I}_{(Y_i \leq y)}) dy.$$

Ce contraste est de type moindres carrés selon la direction en x . Le terme de variance dépend uniquement de la dimension $D_{m_1}^{(1)}$ associé à la covariable. Donc, seule la dimension $D_{m_1}^{(1)}$ intervient dans le compromis biais au carré/variance, la dimension $D_{m_2}^{(2)}$, doit être choisie la plus grande possible exactement comme dans le cas où il n'y a pas de covariable. Un calcul de l'espérance du contraste donne :

$$\mathbb{E}(\Gamma_n(T)) = \|T - F\|_{f_X}^2 - \|F\|_{f_X}^2$$

avec $\|T\|_{f_X}^2 = \iint_A T^2(x, y) f_X(x) dx dy$ et la perte associée à ce contraste est $\|T - F\|_{f_X}^2$.

Comme nous avons à l'esprit d'appliquer notre procédure à des durées de vie, nous considérons que la variable Y est positive. Mais la procédure proposée dans [13] reste valable pour des variables réelles. Nous introduisons la suite des variables de censure positives (C_i) , $i = 1 \dots, n$, i.i.d. de fonction de répartition commune G (inconnue), et indépendantes des couples (X_i, Y_i) , $i = 1 \dots, n$. L'échantillon observé est constitué de (X_i, Z_i, δ_i) où $Z_i = Y_i \wedge C_i$ et $\delta_i = \mathbf{I}_{(Y_i \leq C_i)}$ pour $i = 1, \dots, n$. Pour corriger l'effet de la censure, nous utilisons, comme pour l'estimation de l'espérance conditionnelle de la partie 2.1, la transformation des données proposée par Fan & Gijbels (1996) qui se traduit par l'ajout d'une pondération dans le contraste Γ_n^0 :

$$\hat{w}_i = \frac{\delta_i}{\hat{G}(Z_i)}$$

où l'estimateur \hat{G} est la version modifiée de l'estimateur de Kaplan & Meier (1958) de la fonction de survie $\bar{G} = 1 - G$ de la variable de censure C , donnée par (1.1). Le contraste devient :

$$\Gamma_n(T) = \frac{1}{n} \sum_{i=1}^n \int_{\mathbb{R}} (T^2(X_i, y) - 2\hat{w}_i T(X_i, y) \mathbf{I}_{(Z_i \leq y)}) dy.$$

Lorsqu'il n'y a pas de censure, les poids \hat{w}_i sont tous égaux à 1 et le contraste coïncide alors avec le contraste Γ_n^0 .

Nous définissons l'estimateur pénalisé $\tilde{F} = \hat{F}_{\hat{m}}$ avec

$$\hat{m} = \arg \min_{m_1 \in \mathcal{M}_n} \{ \Gamma_n(\hat{F}_{m_1}) + \text{pen}(m_1) \} \quad (2.3)$$

où

$$\text{pen}(m) \geq K_0 \ell(A_2) \xi \frac{D_{m_1}^{(1)}}{n}, \text{ avec } \xi = \frac{\phi_1}{f_0} \mathbb{E} \left(\frac{\delta_1}{\bar{G}^2(Z_1)} \right) \text{ ou } \xi = \frac{1}{c_G},$$

K_0 est une constante, et les constantes f_0 et c_G sont définies ci-dessous dans [A1] et [A2]. Nous avons les conditions suivantes pour établir le Théorème 7 :

[A1] La densité marginale f_X vérifie $\|f_X\|_\infty := \sup_{x \in A_1} |f_X(x)| < \infty$ et il existe une constante $f_0 > 0$ telle que, pour tout x in A_1 , $f_X(x) \geq f_0$,

[A2] Pour tout $y \in A_2$, $1 - G(y) \geq c_G > 0$.

Pour une fonction h et un sous-espace S , on pose

$$d(h, S) = \inf_{g \in S} \|h - g\| = \inf_{g \in S} \left(\iint |h(x, y) - g(x, y)|^2 dx dy \right)^{1/2}.$$

Théorème 7 (cf. Théorème 4.1 dans [13]) *Sous des hypothèses [A1] et [A2], on considère \tilde{F} l'estimateur pénalisé de la f.d.r. conditionnelle $F\mathbf{1}_A$, construit sur des collections de modèles (cf. Annexe A) tels que la dimension maximale $\mathcal{D}_n^{(1)} \leq \sqrt{n}$.*

$$\mathbb{E} \|F\mathbf{1}_A - \tilde{F}\|_n^2 \leq C \inf_{m_1 \in \mathcal{M}_n} \{ d^2(F\mathbf{1}_A, S_{m_1}^{(1)} \otimes S_n^{(2)}) + \text{pen}(m_1) \} + \frac{C'}{n}$$

avec C et C' des constantes.

Quelle que soit la forme retenue pour ξ , elle fait intervenir des quantités inconnues – f_0 et $\mathbb{E}(\delta_1/\bar{G}^2(Z_1))$ pour l'une ou c_G pour l'autre – qui doivent être remplacées par des estimateurs. Notre méthode ne nous garantit pas l'obtention d'un estimateur strict de la fonction de répartition conditionnelle. Pour y remédier, nous utilisons une procédure *a posteriori* de réarrangement proposée par Chernozhukov et al. (2009). Pour tout $X_i, i = 1, \dots, n$,

$$\tilde{F}^*(X_i, y) = \inf \left\{ z \in \mathbb{R}, \int \mathbf{1}_{\{\tilde{F}(X_i, u) \leq z\}} du \geq y \right\}.$$

Cette transformation quantile fournit un estimateur croissant par rapport à la variable y . Par conséquent, l'estimateur \tilde{F}^* est une fonction de répartition :

$$\tilde{F}^*(x, y) = \begin{cases} 0 & \text{si } \tilde{F}^*(x, y) < 0 \\ \tilde{F}^*(x, y) & \text{si } 0 \leq \tilde{F}^*(x, y) \leq 1 \\ 1 & \text{si } \tilde{F}^*(x, y) > 1 \end{cases}$$

Nous avons montré dans [13] que l'estimateur pénalisé de la fonction de répartition, obtenu par transformation quantile, vérifie les mêmes propriétés que l'estimateur brut \tilde{F} (avant réarrangement). L'inégalité oracle du Théorème 7 pour \tilde{F} s'étend à F^* .

Nous sommes en mesure de donner la vitesse de convergence, la preuve est établie dans le cas non-censuré :

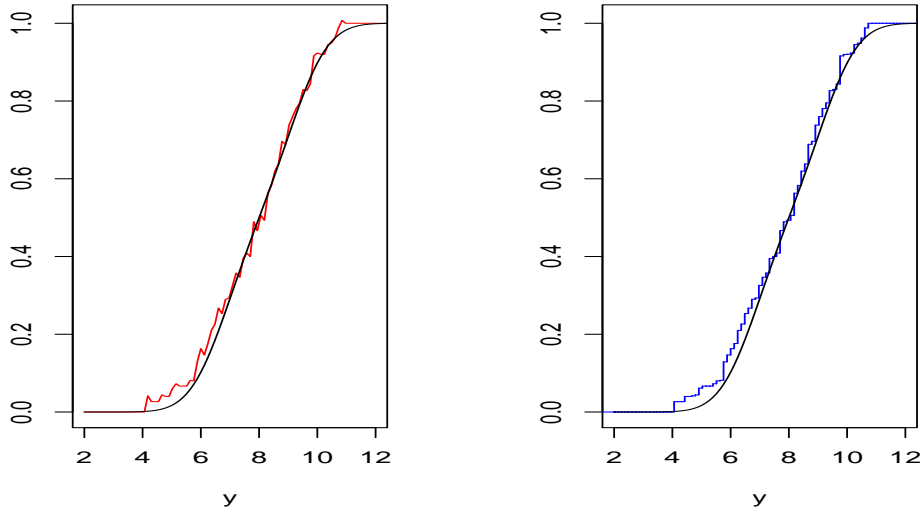


FIGURE 2.1 – Estimateur pénalisé $y \mapsto \tilde{F}(y|x)$ (à gauche en rouge) en bases d’histogramme et fonction de répartition conditionnelle $y \mapsto F(y|x)$ (en noir), et estimateur monotone par réarrangement quantile $y \mapsto \tilde{F}^*(x, y)$ (à droite en bleu) pour $x = 0.29$ et $Y_i|X_i = x \sim 0.5\mathcal{N}(8 - 4x, 1) + 0.5\mathcal{N}(8 + 4x, 1)$ avec $n = 150$ observations et aucune censure.

Corollaire 1 *Supposons que F restreinte à A appartienne à un espace de Besov $B_{2,\infty}^\alpha(A)$ anisotropique, de régularité $\alpha = (\alpha_1, \alpha_2)$ avec $\alpha_1 > 1/2$ et $\alpha_2 > 1$. Considérons les bases $[T]$, $[DP]$ ou $[W]$ (de degré r pour les polynômes et les ondelettes plus grand que $\alpha_i - 1$). Alors, pour $\mathcal{D}_n^{(2)} \geq \sqrt{n}$, $\text{pen}(m_1) = K_0 \ell(A_2) D_{m_1}^{(1)}/n$, et avec les hypothèses du Théorème 7,*

$$\mathbb{E} \|F \mathbf{1}_A - \tilde{F}^*\|_n^2 = O(n^{-\frac{2\alpha_1}{2\alpha_1+1}}).$$

On obtient donc une vitesse qui correspond à la vitesse non-paramétrique usuelle pour l’estimation d’une fonction univariée de régularité α_1 . Nous montrons que cette vitesse est minimax grâce au Théorème 3.2 dans [13] qui donne une borne inférieure pour tous les estimateurs de F et $F \in \mathcal{B} = \{F : \text{f.d.r conditionnelles dans } \mathbb{R}^2 \text{ t.q. } \|F\|_{B_{2,\infty}^\alpha(A)} \leq L\}$ et $\|F\|_{B_{2,\infty}^\alpha(A)}$ est la norme de F dans l’espace de Besov $B_{2,\infty}^\alpha(A)$:

$$\|F\|_{B_{2,\infty}^\alpha(A)} = \|F\|_A + |F|_{B_{2,\infty}^\alpha(A)}$$

où $\|F\|_A = (\int_A F^2)^{1/2}$ et $|F|_{B_{2,\infty}^\alpha(A)}$ est la semi-norme de Besov, cf. paragraphe 4.5 dans Massart (2007). La preuve est une adaptation des résultats de Lacour (2007).

Chapitre 3

Estimation de fonctionnelles par contraste de type moindres-carrés

Publications [14], [15]

Après le travail sur les lois conditionnelles, de nouvelles perspectives se sont ouvertes pour étendre les techniques à d'autres fonctionnelles de régression. Nous avons alors considéré deux fonctions de régression : la moyenne de vie résiduelle et la fonction de risque conditionnelles.

3.1 Moyenne de vie résiduelle conditionnelle

Dans les essais cliniques, les durées de vie sont souvent observées à partir du début du traitement et jusqu'à la survenue du décès ou de la censure. Mais l'étude des fonctions de survie ou de risque instantané ne permet pas de savoir si le traitement améliore l'espérance de vie d'un patient au cours de l'essai. Pour répondre à cette question l'espérance de vie doit être considérée comme une fonction dépendant du temps. C'est pourquoi les praticiens utilisent comme indicateur la fonction moyenne de vie résiduelle, c'est-à-dire la durée de vie moyenne restante sachant que l'individu a déjà survécu au delà d'une durée t . Nous proposons un nouvel estimateur purement non-paramétrique de la moyenne de vie résiduelle en présence d'une covariable. Nous utilisons les mêmes notations que dans le chapitre 2 partie 2.2 pour la fonction de répartition conditionnelle.

On introduit la fonction moyenne de vie résiduelle (MRL) définie par :

$$e(y) = \mathbb{E}(Y - y | Y > y), \quad y > 0,$$

avec $\mathbb{E}(Y) < +\infty$.

Si on note la fonction de survie $\bar{F}(y) = 1 - F(y)$ alors, on montre que la MRL $e(y)$ s'exprime aussi :

$$e(y) = \begin{cases} \int_y^{+\infty} \bar{F}(u) du / \bar{F}(y) & \text{si } \bar{F}(y) > 0 \\ 0 & \text{sinon.} \end{cases}$$

Cette égalité a conduit à diverses stratégies d'estimation non-paramétrique par *plug-in* de la fonction de répartition empirique ou de l'estimateur de Kaplan-Meier en présence de censure cf. Hall & Wellner (1981), Csörgo & Zitikis (1996). On trouve aussi des versions régularisées de ces estimateurs par lissage avec des noyaux : Chaubey & Sen (1999), Abdous & Berred (2005).

Souvent, dans les applications, l'effet combiné du temps et d'une covariable X , par exemple l'âge du patient, sur la moyenne de vie restante d'un individu, présente un intérêt. On est alors plutôt intéressé par la MRL conditionnelle définie par :

$$\begin{aligned} e(y|x) &= \mathbb{E}(Y - y | Y > y, X = x) \\ &= \begin{cases} \int_y^{+\infty} \bar{F}(u|x) du / \bar{F}(y|x) & \text{si } \bar{F}(y|x) > 0 \\ 0 & \text{sinon,} \end{cases} \end{aligned}$$

où $\bar{F}(y|x) = \mathbb{P}(Y > y | X = x)$ est la fonction de survie de Y conditionnelle à $X = x$.

Le modèle de MRL proportionnelle proposé par Oakes & Dasu (1990) postule :

$$e(y|X) = e_0(y) \exp(\beta^t X)$$

où $e_0(y)$ représente la MRL en l'absence de covariable et β est le paramètre de régression.

L'inférence dans ce modèle semi-paramétrique a été étudiée par Maguluri & Zhang (1994) et revisitée par Chen & Cheng (2005) lorsque la variable de durée est censurée à droite.

Notre objectif est plutôt de développer une alternative purement non-paramétrique pour estimer la MRL conditionnelle $e(y|x)$ sur un compact $A = A_1 \times A_2$. La stratégie d'estimation repose sur la définition d'un contraste de type moindre-carrés. Soit $T : (x, y) \mapsto T(x, y)$ une fonction bivariable, mesurable à support compact $A = A_1 \times A_2$. Le contraste candidat pour estimer la MRL conditionnelle $e(y|x)$ est :

$$\Gamma_n(T) = \frac{1}{n} \sum_{i=1}^n \left(\int T^2(X_i, y) \mathbf{1}_{(Y_i \geq y)} dy - 2\Psi_T(X_i, Y_i) \right)$$

où

$$\Psi_T(x, y) = \int_0^y (y - u) T(x, u) du.$$

On introduit la densité jointe $f_{(X,Y)}$ du couple (X, Y) et on suppose que :

$$\bar{F}_1(x, y) = \int_y^{+\infty} f_{(X,Y)}(x, u) du \quad \text{et} \quad \bar{F}_2(x, y) = \int_y^{+\infty} \bar{F}_1(x, u) du$$

sont mesurables. Alors, on peut réécrire la MRL conditionnelle :

$$e(y|x) = \frac{\bar{F}_2(x, y)}{\bar{F}_1(x, y)} \quad \text{si } \bar{F}_1(x, y) > 0.$$

Soient les fonctions S et T telles que

$$\iint S^2(x, y) \bar{F}_1(x, y) dx dy < +\infty \quad \text{et} \quad \iint T^2(x, y) \bar{F}_1(x, y) dx dy < +\infty.$$

On définit le μ -produit scalaire entre les fonctions S et T par :

$$\langle S, T \rangle_\mu = \iint S(x, y)T(x, y)d\mu(x, y) \text{ où } d\mu(x, y) = \bar{F}_1(x, y)dx dy$$

et par $\|\cdot\|_\mu$ la norme associée $\|T\|_\mu^2 = \langle T, T \rangle_\mu$.

$$\mathbb{E}(\Gamma_n(T)) = \mathbb{E} \left(\int T^2(X_1, y)\mathbf{1}_{(Y_1 \geq y)} dy - 2\Psi_T(X_1, Y_1) \right). \quad (3.1)$$

On calcule alors séparément les espérances des termes qui interviennent dans (3.1) :

$$\begin{aligned} \mathbb{E} \left[\int T^2(X_1, y)\mathbf{1}_{(Y_1 \geq y)} dy \right] &= \iint \left(\int T^2(x, y)\mathbf{1}_{(u \geq y)} dy \right) f_{(X, Y)}(x, u) dx du \\ &= \iint T^2(x, y)\bar{F}_1(x, y) dx dy = \|T\|_\mu^2, \end{aligned}$$

puis, après avoir vérifié que $\mathbb{E}[\Psi_T(X_1, Y_1)] < +\infty$

$$\begin{aligned} \mathbb{E}[\Psi_T(X_1, Y_1)] &= \iint \int_0^y (y-u)T(x, u)du f_{(X, Y)}(x, y) dx dy \\ &= \iint \left(\int \mathbf{1}_{(u \leq y)}(y-u)f_{(X, Y)}(x, y) dy \right) T(x, u) dx du \end{aligned}$$

Ici, avec l'hypothèse que $\lim_{y \rightarrow +\infty} y\bar{F}_1(x, y) = 0$, pour tout $x \in A_1$, on peut faire une intégration par parties et on aboutit à :

$$\mathbb{E}[\Psi_T(X_1, Y_1)] = \iint T(x, u)\bar{F}_2(x, u) dx du = \iint T(x, u)e(u|x)\bar{F}_1(x, u) dx du = \langle T, e \rangle_\mu.$$

Finalement, on a montré que $\mathbb{E}(\Gamma_n(T)) = \|T\|_\mu^2 - 2\langle T, e \rangle_\mu$ et par suite

$$\mathbb{E}(\Gamma_n(T)) = \|T - e\|_\mu^2 - \|e\|_\mu^2.$$

L'hypothèse $\lim_{y \rightarrow +\infty} y\bar{F}_1(x, y) = 0$, pour tout $x \in A_1$, n'apparaît pas trop restrictive ; elle est vérifiée pour la plupart des modèles paramétriques classiques. Nous reprenons les mêmes notations que dans la partie 2.2 du chapitre 2 à quelques différences près. On introduit une famille de modèles $\{S_m : m \in \mathcal{M}_n\}$ et pour chaque m , l'espace de fonctions S_m à support dans $A = A_1 \times A_2$ est défini par :

$$S_m = S_{m_1} \otimes \mathcal{H}_n = \left\{ T, \quad T(x, z) = \sum_{j \in J_m} \sum_{k \in \mathcal{K}_n} a_{j,k} \varphi_j^m(x) \psi_k(z), \quad a_{j,k} \in \mathbb{R} \right\},$$

où S_{m_1} et $\mathcal{H}_n \subset (\mathbb{L}^2 \cap \mathbb{L}^\infty)(\mathbb{R})$ sont engendrés respectivement par deux bases orthonormales $(\varphi_j^m)_{j \in J_m}$ et $(\psi_k)_{k \in \mathcal{K}_n}$ avec $|J_m| = D_m$ varie et $|\mathcal{K}_n| = \mathcal{D}_n^{(2)}$ est fixée (la plus grande possible). Ici, comme pour la fonction de répartition conditionnelle, il n'y a pas de sélection de modèles à réaliser selon la dimension du sous-espace \mathcal{H}_n (de projection en y). La dimension $\mathcal{D}_n^{(2)}$ est fixée et correspond à la dimension du plus grand sous-espace \mathcal{H}_n possible. En revanche, la dimension D_m

du sous-espace S_{m_1} varie et doit être sélectionnée de façon optimale. Dans [14], nous présentons les résultats pour des collections de polynômes par morceaux mais il n'y aurait pas de difficultés à considérer d'autres bases de l'annexe A, à quelques modifications près des contraintes sur les dimensions maximales.

Nous posons les hypothèses d'existence de bornes inférieures pour certaines quantités qui semblent inévitables avec les techniques de sélection de modèle en régression que nous utilisons.

$$(A0) \quad \forall x \in A_1, \quad \lim_{y \rightarrow +\infty} y \bar{F}_1(x, y) = 0,$$

$$(A1) \quad \exists \bar{F}_0, f_1 > 0 \text{ t.q. } \forall (x, y) \in A_1 \times A_2, \quad \bar{F}_1(x, y) \geq \bar{F}_0 \text{ and } f_X(x) \leq f_1.$$

$$(A2) \quad \forall (x, y) \in A_1 \times A_2, \quad e(y|x) \leq \|e\|_{\infty, A} < +\infty.$$

La construction de l'estimateur se fait alors en trois étapes.

Première étape : on définit un estimateur sur chaque modèle S_m par $\hat{e}_m = \arg \min_{T \in S_m} \Gamma_n(T)$.

Précisément, soit $T(x, y) = \sum_{j \in J_m} \sum_{k \in \mathcal{K}_n} a_{j,k} \varphi_j^m(x) \psi_k(y)$ une fonction de S_m .

$$\forall j_0 \in J_m, \forall k_0 \in \mathcal{K}_n, \quad \frac{\partial \Gamma_n(T)}{\partial a_{j_0, k_0}} = 0$$

que l'on peut résumer matriciellement $G_m \hat{A}_m = \Upsilon_m$, où

- \hat{A}_m représente le vecteur $vec((\hat{a}_{j,k})_{j \in J_m, k \in \mathcal{K}_n})$ des coefficients de l'estimateur dans la base,
- $G_m := \left(\frac{1}{n} \sum_{i=1}^n \varphi_j^m(X_i) \varphi_\ell^m(X_i) \int \psi_k(z) \psi_p(z) \mathbf{1}_{\{Y_i \geq z\}} dz \right)_{(j,k), (\ell,p) \in (J_m \times \mathcal{K}_n)^2}$
- $\Upsilon_m := vec \left(\left(\frac{1}{n} \sum_{i=1}^n \varphi_j^m(X_i) \int_0^{Y_i} (Y_i - u) \psi_k(u) du \right)_{j \in J_m, k \in \mathcal{K}_n} \right)$.

$vec(\cdot)$ est l'opérateur qui transforme les colonnes d'une matrice en un vecteur.

Deuxième étape : la résolution du problème de minimisation passe par l'inversion de la matrice G_m . On vérifie facilement que ces valeurs propres sont positives ou nulles. Nous renvoyons à [14] pour l'algèbre détaillée du problème. On modifie donc la définition de l'estimateur comme suit :

$$\hat{e}_m := \begin{cases} \arg \min_{T \in S_m} \Gamma_n(T) & \text{on } \hat{H}_m \\ 0 & \text{on } \hat{H}_m^c \end{cases},$$

où $\hat{H}_m := \left\{ \min \text{Sp}(G_m) \geq \max(\hat{F}_0/3, n^{-1/2}) \right\}$ et $\text{Sp}(G_m)$ désigne le spectre de G_m . Ainsi, pour tout m , sur l'ensemble \hat{H}_m , la matrice G_m est inversible. L'estimateur \hat{F}_0 de la borne \bar{F}_0 (le minimum de \bar{F}_1 dans l'hypothèse (A1)) doit satisfaire la condition

$$(A3) \quad \text{Pour tout entier } k \geq 1, \quad \mathbb{P}(|\hat{F}_0 - \bar{F}_0| > \bar{F}_0/2) \leq C_k/n^k.$$

Ici, nous devons faire un commentaire sur l'hypothèse (A3). Elle permet de substituer \hat{F}_0 à \bar{F}_0 et de se placer sur l'ensemble aléatoire où toutes les valeurs propres de G_m sont strictement positives. Nous donnons la définition d'un tel estimateur. Soit

$$\hat{F}_m(x, y) = \sum_{j \in J_m, k \in \mathcal{K}_n} \hat{b}_{j,k} \varphi_j^m(x) \psi_k(y), \quad \text{avec } \hat{b}_{j,k} = \frac{1}{n} \sum_{i=1}^n \varphi_j^m(X_i) \int \psi_k(y) \mathbf{1}_{\{Y_i \geq y\}} dy.$$

Un calcul direct donne $\mathbb{E}(\hat{b}_{j,k}) = \iint \varphi_j^m(x) \psi_k(y) \bar{F}_1(x, y) dx dy = \langle \varphi_j^m \otimes \psi_k, \bar{F}_1 \rangle$, si bien que $\hat{F}_m(x, y)$ est l'estimateur par projection de \bar{F}_1 . Ensuite, en prenant

$$\hat{F}_0 = \inf_{(x,y) \in A} \hat{F}_{m^*}(x, y)$$

où m^* est choisi de façon à vérifier $\log(n) \leq D_{m^*} \leq n^{1/4}/\sqrt{\log(n)}$, et $\mathcal{D}_n^{(2)} = [n^{1/4}/\sqrt{\log(n)}]$, si $\bar{F}_1 \in \mathcal{B}_{2,\infty}^\beta(A)$ avec $\beta = (\beta_1, \beta_2)$ et $\bar{\beta} > 1$ (la moyenne harmonique de β_1 et β_2) alors pour n assez grand, la condition (A3) est vérifiée. La preuve de ce résultat suit les arguments de la proposition 1 dans Comte *et al.* (2011) adaptée à notre contexte.

Notons, que l'introduction de cette condition fait perdre le caractère non-asymptotique de la suite de la procédure d'estimation. Mais c'est le prix à payer pour construire un estimateur de la MRL e sur tout l'intervalle A , plutôt qu'un estimateur défini seulement aux points (X_i, y) comme nous l'avons fait pour l'estimateur de la densité ou la fonction de répartition conditionnelle dans la partie 2.2.

Troisième étape : On pénalise le contraste afin de réaliser le compromis usuel biais au carré/variance en choisissant un modèle de la collection \hat{m} défini par

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left(\Gamma_n(\hat{e}_m) + \text{pen}(m) \right),$$

où $\text{pen}(m)$ est une pénalité à déterminer. Notre estimateur de e sur A est alors $\tilde{e} := \hat{e}_{\hat{m}}$.

Nous avons aussi les hypothèses sur les modèles, en particulier sur la dimension maximale des modèles :

(M1) $\mathcal{D}_n^{(1)} \leq n^{1/4}/\sqrt{\log n}$, $\mathcal{D}_n^{(2)} \leq n^{1/4}/\sqrt{\log n}$ et $\forall m, F_m \subset \mathcal{F}_n$.

(M2) Il existe une constante $\phi_1 > 0$ t.q. $\forall m \in \mathcal{M}_n$, on a :

$$\forall x \in A_1, \sum_{j \in J_m} (\varphi_j^m(x))^2 \leq \phi_1 D_m$$

(M3) Condition d'emboîtement :

$$D_m \leq D_{m'} \Rightarrow F_m \subset F_{m'}.$$

Nous pouvons démontrer une inégalité-oracle pour le risque \mathbb{L}^2 .

Soit e_m la \mathbb{L}^2 -projection orthogonale sur S_m de e restreinte à A .

Théorème 8 *Sous les hypothèses (A0)–(A3) et si la collection de modèles satisfait (M1)–(M2)–(M3). Si $\mathbb{E}(Y_1^6) < +\infty$, alors, avec la pénalité*

$$\text{pen}(m) = \kappa \phi_1 \frac{\mathbb{E}(Y_1^3) + \ell(A_2) \mathbb{E}(Y_1^2)}{\bar{F}_0} \frac{D_m}{n},$$

où κ est une constante numérique, on montre que l'estimateur pénalisé $\tilde{e} = \hat{e}_{\hat{m}}$ vérifie :

$$\mathbb{E}(\|\tilde{e} - e\|_A^2) \leq C \inf_m (\|e - e_m\|_A^2 + \text{pen}(m)) + \frac{C'}{n},$$

où C est une constante qui dépend de \bar{F}_0 et C' dépend de $\mathbb{E}(Y_1^6)$, \bar{F}_0 , $\|e\|_{\infty, A}$.

Les constantes dans la pénalité n'ont pas toutes le même statut :

$$\text{pen}(m) = \kappa \phi_1 \frac{\mathbb{E}(Y_1^3) + \ell(A_2) \mathbb{E}(Y_1^2)}{\bar{F}_0} \frac{D_m}{n},$$

La constante κ est *universelle* dans le sens où elle ne dépend d'aucune quantité inconnue, mais nous ne connaissons pas sa valeur numérique. Elle est donc calibrée par simulations pour différentes fonctions $e(y|x)$. La constante ϕ_1 est connue et dépend de la base utilisée. Les quantités $\mathbb{E}(Y_1^2)$ et $\mathbb{E}(Y_1^3)$ doivent être remplacées par des moments empiriques, La borne \bar{F}_0 doit être remplacée par $\hat{\bar{F}}_0$ – qui vérifie l'hypothèse (A_3) – et cette substitution appelle un commentaire. On définit

$$\hat{\bar{F}}_m(x, y) = \sum_{j \in J_m, k \in \mathcal{K}_n} \hat{b}_{j,k} \varphi_j^m(x) \psi_k(y), \quad \text{avec} \quad \hat{b}_{j,k} = \frac{1}{n} \sum_{i=1}^n \varphi_j^m(X_i) \int \psi_k(y) \mathbf{I}_{(Y_i \geq y)} dy.$$

Alors, un simple calcul permet de vérifier que

$$\mathbb{E}(\hat{b}_{j,k}) = \iint \varphi_j^m(x) \psi_k(y) \bar{F}_1(x, y) dx dy = \langle \varphi_j^m \otimes \psi_k, \bar{F}_1 \rangle.$$

Ainsi, $\hat{\bar{F}}_m(x, y)$ est l'estimateur par projection naturel de \bar{F}_1 .

Ensuite, on montre que $\hat{\bar{F}}_0 = \inf_{(x,y) \in A} \hat{\bar{F}}_{m^*}(x, y)$ vérifie (A_3) où m^* est choisi de façon à satisfaire $\log(n) \leq D_{m^*} \leq n^{1/4} / \sqrt{\log(n)}$, et $\mathcal{D}_n^{(2)} = n^{1/4} / \sqrt{\log(n)}$.

FIGURE 3.1 – Estimateurs de la MRL en fonction de la durée de vie et de l'âge (haut-gauche) et en fonction de la durée de vie et du score d'autonomie (haut-droite). En bas à gauche, estimatur de la MRL pour trois valeurs de l'âge fixées 41, 54 et 68 ans et en bas à droite pour deux valeurs du score d'autonomie 35 et 79.

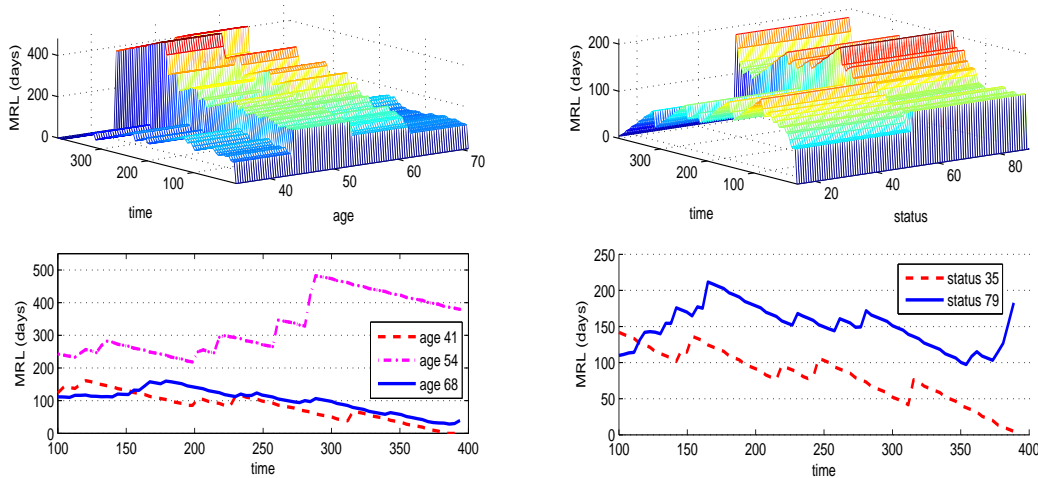
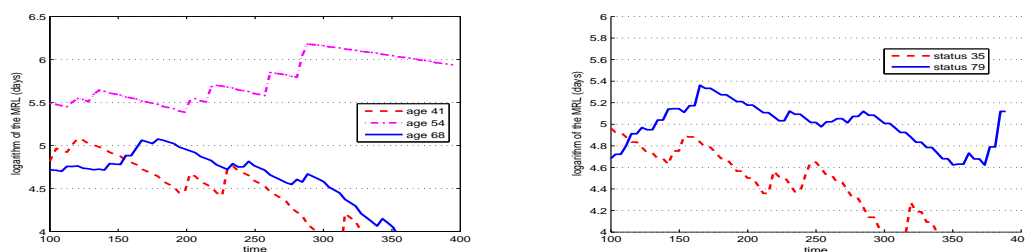


FIGURE 3.2 – Estimateur de la MRL en échelle logarithmique, en fonction de l'âge (gauche) et du score d'autonomie (droite).



Application

En guise d'illustration, nous appliquons notre procédure à un jeu de données réelles de la littérature étudié par Prentice (1973) où l'on s'intéresse à la durée de vie en jours de 128 hommes atteints d'un cancer du poumon inopérable, et qui reçoivent un traitement de chimiothérapie. Les durées observées vont de 1 à 587 jours et ne sont pas censurées, les deux covariables d'intérêt sont ici l'âge et un score de performance (performance status) allant de 1 à 100 qui évalue l'autonomie du patient. L'influence de cette covariable a été étudiée par Chen & Cheng (2005) avec un modèle de Oakes-Dasu. La Figure 3.1 en haut à gauche montre l'estimateur pénalisé de la MRL comme fonction de deux variables : la durée de vie et l'âge du patient. Nous renvoyons à [14] pour le choix des constantes dans la pénalité pour ce jeu de données. On montre trois coupes de la fonction $e(y|x)$ pour trois âges $x = 41, 54$ et 68 sur la figure 3.1 en bas à gauche. Alors que les courbes sont très similaires pour les âges 41 et 68, la MRL semble environ 2 fois plus longue pour l'âge $x = 54$. Cela laisse supposer qu'il y a un âge optimal pour recevoir le traitement. Nous donnons aussi une estimation de la MRL conditionnellement à la variable du score de performance (considérée comme une variable continue) sur la figure 3.1-droite. On peut détecter, comme pour la variable âge deux comportements qui correspondent respectivement à un score d'autonomie faible ou important. Ainsi, la perte d'autonomie peut être identifiée comme de mauvais pronostic pour la durée de vie restante. L'aspect qualitatif de ces courbes est donc intéressant en soi et l'approche purement non-paramétrique donne une méthode graphique qui permet de valider ou d'infirmer l'hypothèse de proportionnalité du modèle de Oakes-Dasu, comme l'illustre la figure 3.2. En échelle logarithmique, les courbes de la figure 3.2 devraient, sous l'hypothèse du modèle de Oakes-Dasu être approximativement parallèles, pour une valeur fixée de la covariable. Mais cela n'est pas vérifié pour la covariable âge. En revanche, pour le score d'autonomie, après une période initiale de 0 à 150 jours, l'hypothèse de proportionnalité semble être satisfaite.

3.2 Risque conditionnel

Un autre type de fonctionnelle de régression est le risque instantané de décès en présence de covariables. Comte *et al.* (2011) traitent de façon générale l'estimation de l'intensité conditionnelle d'un processus ponctuel, le risque instantané constituant un cas particulier.

3.2.1 Le modèle de censure avec indicatrices manquantes

Nous nous concentrons ici sur le risque conditionnel en présence de censure et avec le problème supplémentaire des indicatrices de censure manquantes. Soit Y une variable aléatoire qui représente la durée écoulée jusqu'au décès. Soit C la variable aléatoire qui représente la censure à droite. Dans le modèle classique de censure à droite, nous avons déjà vu que l'observation est constituée du couple $Z = Y \wedge C$ et $\delta = \mathbf{I}(Y \leq C)$. On note X une covariable réelle d'intérêt. Dans la suite, on suppose que Y , C et X admettent des densités par rapport à la mesure de Lebesgue, respectivement notées f_Y , g et f_X . De plus, la variable C est supposée indépendante de Y conditionnellement à X .

Lorsque la cause du décès n'est pas connue ou ne peut être attribuée avec certitude à la pathologie d'intérêt, l'indicatrice de censure est manquante : c'est le modèle de censure manquante décrit par Subramanian (2006) dont nous rappelons à présent la formulation. Soit ξ l'indicatrice de censure manquante définie par :

$$\xi = \begin{cases} 1 & \text{si } \delta = \mathbf{I}(Y \leq C) \text{ est observée} \\ 0 & \text{sinon.} \end{cases}$$

L'observation pour un individu i est alors constituée de :

$$(Z_i, X_i, \delta_i, \xi_i = 1) \quad \text{ou} \quad (Z_i, X_i, \xi_i = 0).$$

Nous dirons alors que le modèle est :

- MCAR (Missing Completely At Random) sous l'hypothèse que ξ est indépendante de Y , C et X .
- MAR (Missing At Random) sous l'hypothèse que ξ et δ sont indépendantes conditionnellement à Z , X .

Ce modèle de censure a été étudié par différents auteurs, la plupart du temps c'est la fonction de survie ou de risque cumulé qui est considérée. van der Laan and McKeague (1998) proposent une amélioration du papier de Lo (1991) en construisant un estimateur non paramétrique par maximum de vraisemblance de la fonction de survie dans le contexte MAR. Leur estimateur généralise l'estimateur de Kaplan & Meier (1958). Subramanian (2004) propose aussi un estimateur efficace. Gijbels *et al.* (2007) s'intéresse à des modèles de régression semi-paramétriques de type modèles de Cox dans différents contextes de censure manquante. Les approches par noyaux

ont aussi été largement explorées pour des données MAR. Subramanian (2006) propose un estimateur à noyaux pour la fonction de risque cumulé. Il montre un Théorème Central Limite et en déduit des résultats pour la fonction de survie. Dans la même veine, on pourra aussi consulter les travaux de Wang and Ng (2008). Wang *et al.* (2009) s'intéressant à la densité et fournissent une méthodologie pour le choix de la fenêtre. Des extensions de ce papier au cadre de la fonction de risque ou de la survie conditionnelles sont développées par Wang & Shen (2008).

Nous pouvons également citer (Dikta (1998), ou plus récemment Subramanian (2009), Subramanian (2011) qui considèrent des modèles semi-paramétriques pour le mécanisme de censure manquante et des méthodes d'imputation. L'avantage de notre approche est de fournir un estimateur qui minimise un contraste et donc évite les écueils rencontrés avec les stratégies par noyaux, qui font intervenir un quotient d'estimateurs, comme Wang & Shen (2008) et Subramanian (2006). Notre approche, toutefois, nécessite aussi l'estimation d'une fonction auxiliaire, qui doit être injectée dans le contraste. Mais cette étape est également présente dans les approches par noyaux. Nous sommes capables de faire une étude très précise de l'estimateur *plug in* et de contrôler son risque \mathbb{L}^2 . Nous obtenons des vitesses anisotropiques qui correspondent à la régularité de la fonction de risque conditionnelle, plus un terme pour l'estimation de la fonction auxiliaire.

3.2.2 Construction du contraste dans le cadre MAR

Nous décrivons maintenant un peu plus en détail, la construction du contraste. Nous nous plaçons tout d'abord dans le cadre MAR qui est le contexte le plus général. Nous souhaitons estimer la fonction de risque conditionnelle sur un compact $A = A_1 \times [0, \tau]$:

$$\lambda(x, y) = \lambda_{Y|X}(x, y) = \frac{f_{Y|X}(x, y)}{1 - F_{Y|X}(x, y)},$$

où $f_{Y|X}$ et $F_{Y|X}$ désignent la densité et la fonction de répartition de Y conditionnelle à X . Des quantités d'intérêt sont les espérances conditionnelles de ξ et δ :

$$\begin{aligned}\pi(x, y) &= \mathbb{E}(\xi|X = x, Z = y) \\ \zeta(x, y) &= \mathbb{E}(\delta|X = x, Z = y).\end{aligned}$$

Notre stratégie d'estimation repose sur la minimisation d'un contraste de type moindres-carrés sur des espaces fonctionnels d'approximation linéaires et de dimension finie.

Construction du contraste sans covariable

Pour simplifier la présentation du contraste, nous commençons par étudier le problème sans covariable. La fonction de risque est réduite à

$$\lambda(y) = \frac{f_Y(y)}{1 - F_Y(y)}.$$

Si $\zeta(y) = \mathbb{E}[\delta|Z = y]$ était connue, un contraste pour estimer la fonction de risque λ serait :

$$\Gamma_n^{th}(h) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau h^2(y) \mathbf{I}_{(Z_i \geq y)} dy - \frac{2}{n} \sum_{i=1}^n (\xi_i \delta_i + (1 - \xi_i) \zeta(Z_i)) h(Z_i).$$

En effet, avec l'hypothèse MAR et la définition de la fonction ζ , on a

$$\begin{aligned} \mathbb{E}(\delta_i \xi_i + (1 - \xi_i) \zeta(Z_i) | Z_i) &= \mathbb{E}(\delta_i | Z_i) \mathbb{E}(\xi_i | Z_i) + \mathbb{E}[(1 - \xi_i) \mathbb{E}(\delta_i | Z_i) | Z_i] \\ &= \mathbb{E}(\mathbb{E}(\delta_i | Z_i) (\xi_i + (1 - \xi_i)) | Z_i) \\ &= \zeta(Z_i) \end{aligned}$$

et l'espérance du contraste est égale à :

$$\begin{aligned} \mathbb{E}(\Gamma_n^{th}(h)) &= \int h^2(y) d\mu(y) - 2 \int h(y) \lambda(y) d\mu(y) \\ &= \|h\|_\mu^2 - 2\langle h, \lambda \rangle_\mu = \|h - \lambda\|_\mu^2 - \|\lambda\|_\mu^2. \end{aligned}$$

avec $d\mu(y) = (1 - L)(y) dy$ et $(1 - L)(y) = (1 - F_Y)(1 - G)(y) = \mathbb{P}(Z \geq y)$.

Comme ζ est inconnue, nous devons lui substituer un estimateur $\tilde{\zeta}$ et le contraste Γ_n^{th} devient :

$$\Gamma_n^0(h) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau h^2(y) \mathbf{I}_{(Z_i \geq y)} dy - \frac{2}{n} \sum_{i=1}^n (\xi_i \delta_i + (1 - \xi_i) \tilde{\zeta}(Z_i)) h(Z_i). \quad (3.2)$$

Construction du contraste avec covariable

Le contraste (3.2) se généralise en présence d'une covariable, au contraste de type moindres carrés suivant :

$$\Gamma_n(h) = \frac{1}{n} \sum_{i=1}^n \int_0^\tau h^2(X_i, y) \mathbf{I}_{(Z_i \geq y)} dy - \frac{2}{n} \sum_{i=1}^n (\xi_i \delta_i + (1 - \xi_i) \tilde{\zeta}(X_i, Z_i)) h(X_i, Z_i), \quad (3.3)$$

pour la norme de référence $\iint h^2(x, y) d\mu(x, y) = \iint h^2(x, y) (1 - L_{Z|X}(y, x)) f_X(x) dx dy$ et

$$1 - L_{Z|X}(y, x) := \mathbb{P}(Z \geq y | X = x) = (1 - F_{Y|X}(x, y))(1 - G_{C|X}(x, y)).$$

Estimation de la fonction auxiliaire ζ

Revenons à l'estimation de la fonction auxiliaire ζ . C'est la clé importante de notre stratégie d'estimation. Nous verrons dans la conclusion, que d'autres propositions pourraient être choisies. Dans [15], nous proposons un critère de type moindres carrés pour estimer $\zeta(x, y) = \mathbb{E}(\delta | X = x, Z = y)$ sur A . L'estimateur $\hat{\zeta}_m$, pour un modèle S_m , est défini comme minimisant le critère :

$$\tilde{\gamma}_n(T) = \frac{1}{n} \sum_{i=1}^n [\xi_i T^2(X_i, Z_i) - 2\xi_i \delta_i T(X_i, Z_i)],$$

pour $T \in S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$. Puis, on obtient l'estimateur pénalisé $\tilde{\zeta} := \hat{\zeta}_{\tilde{m}}$ avec :

$$\tilde{m} = \arg \min_{m \in \mathcal{M}_n} \tilde{\gamma}_n(\hat{\zeta}_m) + \tilde{\kappa} \frac{\dim(S_m)}{n}.$$

L'estimateur $\tilde{\zeta}$ peut alors être substitué à ζ dans le critère Γ_n pour l'estimation de la fonction de risque λ . On est capable de contrôler en norme \mathbb{L}^2 (ou équivalente) l'estimateur $\tilde{\zeta}$.

Le défaut de cette procédure d'estimation en deux temps, avec *plug-in* de l'estimateur de la fonction ζ dans le contraste Γ_n , est de produire une inégalité oracle où les erreurs d'approximation de la fonction de risque λ et de la fonction auxiliaire ζ se cumulent. Ainsi, la vitesse résultante est

$$\mathbb{E} \left(\|\lambda - \hat{\lambda}_{\tilde{m}}\|_A^2 \right) = O(n^{-\frac{2\tilde{\alpha}}{2\tilde{\alpha}+2}}) + O(n^{-\frac{2\tilde{\beta}}{2\tilde{\beta}+2}}),$$

c'est-à-dire que l'estimateur $\hat{\lambda}_{\tilde{m}}$ n'atteint la vitesse optimale que si la fonction ζ est plus régulière que la fonction λ .

Une variante dans le cas MCAR

Sous l'hypothèse (MCAR) que ξ est indépendante de T , C (et de la covariable éventuelle X), on montre qu'un contraste adéquat pour estimer la fonction de risque (conditionnelle) est :

$$\gamma_n^{(1)}(h) = \frac{1}{n} \sum_{i=1}^n \int_0^1 h^2(y) \xi_i \mathbf{I}(Y_i \geq y) dy - \frac{2}{n} \sum_{i=1}^n \delta_i \xi_i h(Y_i),$$

pour la mesure de référence $d\mu(y) = \mathbb{E}(\xi)(1 - L(y))dy$. Ce contraste ne nécessite pas l'étape d'estimation de la fonction auxiliaire ζ mais il n'utilise que les observations *complètes* c'est-à-dire celles pour lesquelles l'indicatrice $\xi_i = 1$.

3.2.3 Prolongements

D'autres propositions de contraste ont été évoquées dans [15] mais non complètement exploitées. Pourtant différents prolongements de ce travail sont possibles. Nous pourrions par exemple porter notre attention sur les contrastes basés sur l'espérance $\pi(x, y) = \mathbb{E}(\xi|X = x, Z = y)$:

$$\Gamma_n^{\text{EST}}(h) = \frac{1}{n} \sum_{i=1}^n \int_0^1 h^2(X_i, y) \tilde{\pi}(X_i, y) \mathbf{I}(Z_i \geq y) dy - \frac{2}{n} \sum_{i=1}^n \delta_i \xi_i h(X_i, Y_i). \quad (3.4)$$

Là encore la substitution à la fonction π inconnue d'un estimateur $\tilde{\pi}$ est nécessaire. L'estimation de π peut être obtenue directement par un contraste de régression pour $\mathbb{E}(\xi|X, Z)$, les triplets (X_i, Z_i, ξ_i) , $i = 1, \dots, n$ étant complètement observés.

La littérature des données manquantes est foisonnante, Little & Rubin (2002) y ont consacré une monographie. En particulier, les techniques d'imputation stochastique, qu'ils présentent,

sont souvent utilisées et implémentées dans les logiciels, pour traiter les données manquantes. Les idées d'imputation pourraient être adaptées à nos estimateurs. En l'absence de covariable, pour simplifier la présentation, cela consisterait à substituer aux termes $\zeta(Z_i)$, une quantité aléatoire de même espérance. Différentes imputations pourraient être envisagées :

Imputation naïve : soit γ_i une variable aléatoire de Bernoulli de paramètre

$$\#\{i : \xi_i = 1\}^{-1} \sum_{i:\xi_i=1} \delta_i.$$

Imputation/régression : soit γ_i une variable aléatoire de Bernoulli de paramètre $\zeta(Y_i)$ étant donné $\xi_i = 0$. La quantité $\zeta(Y_i)$ doit être estimée par une première étape de régression.

Alors le contraste par imputation peut être défini par :

$$\Gamma_n^{\text{IMP}}(h) = \frac{1}{n} \sum_{i=1}^n \int_0^1 h^2(y) \mathbf{I}(Z_i \geq y) dy - \frac{2}{n} \sum_{i=1}^n (\xi_i \delta_i + (1 - \xi_i) \gamma_i) h(Z_i). \quad (3.5)$$

Une étude approfondie tant du point de vue théorique que numérique serait utile pour donner des recommandations sur la meilleure stratégie à adopter.

Chapitre 4

Travaux en cours, futurs et perspectives

Dans ce dernier chapitre, j'expose les travaux en cours, certains plus aboutis que d'autres, et les perspectives de recherche. Le thème des données fonctionnelles et plus largement de la grande dimension est à l'origine du groupe de travail "ADONF" (Analyse des Données Fonctionnelles), initié par André Mas. Ce GT favorise les interactions entre les chercheurs d'I3M et nos collègues de l'INRA/SupAgro. C'est dans cet environnement que je cherche à orienter mes travaux. Depuis Septembre 2011, je co-encadre avec André Mas la thèse d'Angelina Roche sur le thème *Sélection de modèles pour données fonctionnelles*. Les connexions avec les modèles de régression, que j'ai pu étudier jusque là, sont nombreuses. Au-delà des premiers résultats obtenus dans le cadre du modèle linéaire fonctionnel, d'autres pistes de recherche sont ouvertes pour prendre en compte des covariables fonctionnelles dans les modèles de durées ; je les évoque dans ce qui suit. Je continue néanmoins à m'intéresser aux modèles spécifiques aux durées de vie et je présente aussi les projets en cours, notamment sur l'inférence des événements récurrents ou la construction de bandes de confiance.

Données fonctionnelles

Prépublications [16], [17]

La statistique des données fonctionnelles modélise des courbes. Il s'agit donc de données qui appartiennent à des espaces de fonctions : la modélisation de la température en un point donné, la consommation d'électricité ou le cours d'une action sont autant d'exemples d'applications de ce type de données. Nous renvoyons aux monographies de Ferraty & Vieu (2006) et Ramsay & Silverman (2005) pour des exemples détaillés en médecine, linguistique ou chimiométrie, et à Preda & Saporta (2005) pour des applications en économétrie. Le modèle de régression auquel nous nous intéressons, considère une variable réponse réelle et des prédicteurs qui sont

des courbes aléatoires : un exemple qui illustre cette situation de référence est la prédiction journalière de la consommation d'électricité en fonction de la courbe de température.

Dans le modèle linéaire fonctionnel, la variable réponse réelle Y est liée à un prédicteur fonctionnel X qui appartient à un espace de Hilbert $(\mathbb{H}, \langle \cdot, \cdot \rangle, \|\cdot\|)$, par la relation suivante :

$$Y = \langle \beta, X \rangle + \varepsilon, \quad (4.1)$$

où la variable centrée ε représente un terme d'aléa indépendant du prédicteur X . On suppose sans perte de généralité que X est centrée, c'est-à-dire que la fonction $t \mapsto \mathbb{E}[X(t)]$ est identiquement nulle. En multipliant l'équation (4.1) par $X(s)$ et en prenant l'espérance, le problème de l'estimation de la fonction de pente β apparaît clairement relié à l'inversion de l'opérateur de covariance Γ de X défini par :

$$\Gamma f := \mathbb{E}[\langle f, X \rangle X]. \quad (4.2)$$

Il s'agit d'un problème inverse mal-posé, on pourra pour plus de détails se référer au Chapitre 2.1, Engl *et al.* (1996).

Il existe de nombreuses méthodes d'estimation de la fonction de pente dans le modèle linéaire fonctionnel. Les méthodes consistent généralement à développer la fonction β dans une base (déterministe) connue de fonctions de \mathbb{L}^2 . Li & Hsing (2007) ont proposé une procédure d'estimation par minimisation d'un critère de moindres-carrés sur un espace de fonctions périodiques. Crambes *et al.* (2009) ont généralisé le populaire estimateur spline de régression. Certains auteurs ont considéré des estimateurs par projection sur des bases fixes, comme des bases de B-splines (Ramsay & Dalzell (1991)), bases de Fourier (Li & Hsing (2007)) ou des bases orthogonales quelconques (Cardot & Johannes (2010)). Mais la méthode la plus populaire demeure sans doute la régression sur composantes principales fonctionnelles qui consiste à développer la fonction de pente sur l'espace de fonctions aléatoire engendré par les éléments propres de l'opérateur de covariance empirique,

$$\Gamma_n : f \in \mathbb{H} \mapsto \frac{1}{n} \sum_{i=1}^n \langle f, X_i \rangle X_i, \quad (4.3)$$

associés aux plus grandes valeurs propres. Des versions lissées par spline ont été proposées par Cardot *et al.* (2003). Cai & Hall (2006) ont montré que l'estimateur par ACP fonctionnelle atteint les vitesses de convergence optimales associées au risque de l'erreur de prédiction et Hall & Horowitz (2007) pour le risque \mathbb{L}^2 . Toutes les méthodes envisagées reposent sur le choix d'un paramètre de lissage qui est crucial pour la qualité de l'estimation. Le choix optimal de ce paramètre dépend simultanément de la régularité inconnue de la fonction de pente et de celle du prédicteur X , cf. Cai & Hall (2006); Crambes *et al.* (2009); Cardot & Johannes (2010), et en pratique seule la validation croisée permet de le choisir automatiquement, sans garantie d'optimalité.

Avec Angelina et André, nous cherchons à aborder le problème d'un point de vue non-asymptotique. Dans [16], une première étape a été de proposer une méthode de sélection de modèles dans le modèle linéaire fonctionnel pour l'estimateur des moindres carrés étudié par Cardot & Johannes (2010) et pour le risque associé à l'erreur de prédiction. Cette approche utilise les résultats de Baraud en régression et s'appuie sur l'hypothèse (forte) de la connaissance des vecteurs propres de l'opérateur de covariance et illustrée par le contexte des données circulaires comme dans Comte & Johannes (2010). Un estimateur original de la variance du bruit inconnu est substitué à la variance inconnue dans le terme de pénalité et on montre par simulation que ses performances sont excellentes. L'approche est comparée avec la méthode de validation croisée de Cardot *et al.* (2003). Les résultats numériques sont comparables en terme d'erreur de prédiction. Un gain de temps est à noter pour notre approche.

Cai & Yuan (2012) ont proposé un choix automatique du paramètre de lissage, dans un cadre d'espace à noyaux reproduisant. Leur procédure d'estimation est optimale et ne requiert aucune connaissance a priori sur la structure de covariance. Dans le même esprit, mais avec des techniques très différentes, Comte & Johannes (2012) définissent un critère de sélection de la dimension par pénalisation de contraste, utilisant la méthode de Lepski, et sans hypothèse restrictive sur X . Leurs estimateurs sont adaptatifs et atteignent les vitesses de convergence optimales pour des risques de type \mathbb{L}^2 pondérés. Leur objectif est d'estimer non seulement la fonction de pente mais aussi ses dérivées. Cependant, leur procédure de sélection ne permet pas de traiter l'erreur de prédiction, puisque dans ce cas les poids qui définissent le risque sont précisément les valeurs propres inconnues de l'opérateur de covariance de X .

Partant de cette constatation, un deuxième volet du travail, plus ambitieux, a consisté à utiliser la base des vecteurs propres de l'opérateur de covariance empirique. Dans [17], on ne travaille plus alors avec des bases fixées mais avec des bases aléatoires et le contrôle du biais de l'estimateur n'est plus aussi aisé. Grâce à des résultats de théorie de la perturbation, on obtient une inégalité oracle pour l'estimateur pénalisé de la régression en composantes principales fonctionnelles. L'estimateur pénalisé est adaptatif et atteint des vitesses minimax optimales pour des classes de fonctions de Sobolev.

Les perspectives de travail autour du thème des données fonctionnelles, sont très nombreuses. En particulier, les fonctionnelles de régression, qui apparaissent dans les modèles de durées de vie peuvent très souvent se présenter avec une covariable fonctionnelle (spectrométrie de masse de la fonction rénale, imagerie IRM, courbes de tension artérielle, de glycémie, etc). Je pense, par exemple, que le modèle *single index* en régression, que je n'ai fait qu'évoquer dans le chapitre 2, pour des covariables réelles, peut trouver de nouveaux développements dans le cas où la variable X est fonctionnelle : $Y = g(\langle \beta, X \rangle) + \varepsilon$. Les travaux récents de Chen *et al.* (2011) envisagent cette généralisation du modèle linéaire fonctionnel. Mais, à ma connaissance, il n'existe pas de

procédure adaptative pour estimer simultanément les fonctions g ou β .

Événements récurrents (en cours)

Avec Segolen Geffray, nous avons commencé à étudier une procédure d'estimation totalement non-paramétrique des fonctions de répartitions conditionnelles d'événements récurrents en présence de censure. Dans les études épidémiologiques, les patients peuvent être sujet à une succession de plusieurs événements, comme les récives de tumeurs cancéreuses ou les infections opportunistes pour les malades du SIDA. Les événements récurrents en sont un exemple. Nous renvoyons à Schaubel & Cai (2004) ou à Cook & Lawless (2008) pour une présentation de ce type de données et des exemples.

Nous considérons des situations où les événements récurrents sont associés à un risque élevé de décès. Le décès des patients peut donc survenir pendant l'étude. Par ailleurs, l'observation d'un événement récurrent ou du décès peut être censuré. Différentes causes de censure indépendantes sont possibles : perdu de vue, fin de l'étude avant le décès ou décès indépendant de la maladie ou du traitement.

Un exemple d'événements récurrents en présence de censure est illustré par la figure 4.1 pour 6 patients S_1 à S_6 . La période de suivi d'un patient est représentée par une ligne. On s'intéresse à la survenue du deuxième événement récurrent.

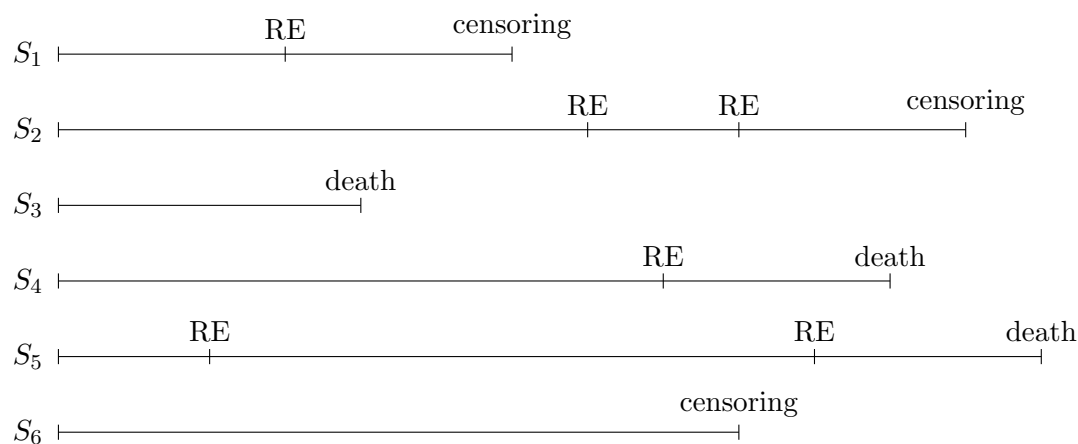


FIGURE 4.1 – Événements récurrents (RE = événement d'intérêt).

Nous supposons que les réalisations des événements récurrents, terminaux (décès) ou de censure ne peuvent se produire aux mêmes instants.

Dans le contexte des risques compétitifs, Allignol *et al.* (2011) considèrent un modèle de

régression pour la probabilité conditionnelle d'un événement concurrent. Van Keilegom *et al.* (2011) proposent des estimateurs non-paramétriques de la fonction de répartition conditionnelle :

$$\mathbb{P} \left[Y^{[2]} \leq y_2 | Y^{[1]} = y_1 \right]$$

lorsque le modèle position/échelle est satisfait par les deux premiers événements récurrents :

$$Y^{[2]} = m(Y^{[1]}) + \sigma(Y^{[1]})\varepsilon$$

où les fonctions m et σ sont des fonctions suffisamment régulières et le bruit ε est indépendant de $Y^{[1]}$. Cependant, leurs travaux ne prennent pas en compte la survenue d'un événement terminal.

L'échantillon est constitué de la façon suivante :

- On note K le nombre total d'événements observés pour un individu donné.
- Pour $k = 1, \dots, K$, on note $Y^{[k]}$ la durée entre le $(k - 1)$ -ème et le k -ème événement.
- Pour $k = 1, \dots, N$, on note $\mathcal{C}^{[k]}$ la variable qui indique la nature du k -ème événement observé :

$$\mathcal{C}^{[k]} = \begin{cases} 1 & \text{s'il s'agit d'un événement récurrent,} \\ 2 & \text{s'il s'agit d'un décès.} \end{cases}$$

- Pour $k = 1, \dots, K$, on pose $Z^{[k]} = \min(Y^{[k]}, C - \sum_{l=0}^{k-1} Y^{[l]})$ qui représente la durée écoulée entre le $(k - 1)$ -ème et le k -ème événement observé (avec la convention $Y^{[0]} = 0$).
- Pour $k = 1, \dots, K$, la variable $J^{[k]} = \mathcal{C}^{[k]} I(\sum_{l=1}^k Y^{[l]} \leq C)$ indique la nature du k -ème événement observé.

Notre objectif est d'adapter la procédure d'estimation adaptative du chapitre 2, paragraphe 2.2 pour proposer un estimateur purement non-paramétrique de la fonction de répartition conditionnelle $\mathbb{P} [Y^{[2]} \leq y_2 | Y^{[1]} = y_1]$ dans le contexte des événements récurrents.

Intervalles de confiance (en cours)

Avec O. Bouaziz et F. Comte, nous nous penchons sur la construction de bandes de confiance pour les fonctionnelles de survie. C'est un écueil des méthodes que j'ai présentées tout au long de ce mémoire, qui sont surtout adaptées à des contrôles de risque de type \mathbb{L}^2 (ou \mathbb{L}^p). En effet, l'approche globalisée dans le sens où l'on estime les coefficients de la fonction cible dans une base fixée, ne rend pas immédiate l'obtention d'intervalles de confiance. Les méthodes par noyaux qui donnent des résultats en risque ponctuel et pour lesquelles on dispose souvent de théorèmes limites de convergence en loi, donnent plus facilement accès à des intervalles de confiance (au moins ponctuels). Les travaux récents très généraux de Giné and Nickl (2010) ouvrent une voie pour la construction de bandes de confiance pour la densité puisque leurs résultats théoriques

englobent à la fois l'approche par noyaux et par ondelettes. L'outil clé, de ces travaux est de contrôler l'erreur en norme uniforme. D'autres approches ont également été proposées, dans le cadre des ondelettes par Picard & Tribouley (2000) ou Tribouley (2004) et s'appuie sur les propriétés d'approximation locales des ondelettes. Nous proposons d'adapter ces techniques à l'estimateur étudié dans [5], pour des bases de polynômes par morceaux, pour obtenir des bandes de confiance adaptatives de la fonction de risque instantané.

Annexe A

Choix des bases et propriétés d'approximation

Les bases $(\varphi_\lambda)_{\lambda \in \Lambda_m}$ orthonormales utilisées pour définir les collections de modèles $(S_m)_{m \in \mathcal{M}_n}$ sont construites sur $A = [0, 1]$. On peut facilement se ramener à n'importe quel intervalle $A = [a, b]$ par un changement d'échelle, puisque si $\{\varphi_\lambda\}_{\lambda \in \Lambda_m}$ est une base orthonormale de $\mathbb{L}^2([0, 1])$ alors $\left\{ \frac{1}{\sqrt{b-a}} \varphi_\lambda \left(\frac{\cdot - a}{b-a} \right) \right\}$ est une base orthonormale de $\mathbb{L}^2([a, b])$.

La notation Λ_m de l'ensemble d'indices, possiblement multiples, est une notation générique qui permet d'uniformiser l'écriture des bases. On explicite dans les exemples ci-dessous la signification de cette notation et on donne la dimension $D_m = |\Lambda_m|$ de chacun des espaces S_m engendré par la base $(\varphi_\lambda)_{\lambda \in \Lambda_m}$.

[T] Base Trigonométrique : $\{\varphi_0, \dots, \varphi_{m-1}\}$ avec $\varphi_0(x) = \mathbf{I}_{[0,1]}(x)$,

$$\varphi_{2j}(x) = \sqrt{2} \cos(2\pi jx) \mathbf{I}([0, 1])(x) ; \varphi_{2j-1}(x) = \sqrt{2} \sin(2\pi jx) \mathbf{I}([0, 1])(x)$$

pour $j \geq 1$. Cette base engendre un modèle S_m de dimension $D_m = m$.

[DP] Base de Polynômes par morceaux dyadiques : l'ensemble d'indices $\Lambda_m = \{\lambda = (k, d), 1 \leq k \leq 2^m, 0 \leq d \leq r\}$ est un ensemble de couples indexant des polynômes de degré $0, \dots, r$ (le degré maximal r est fixé), définis sur les intervalles $[(k-1)/2^m, k/2^m]$ avec $k = 1, \dots, 2^m$ d'une partition (dyadique) de $[0, 1]$. L'espace S_m engendré est alors de dimension $D_m = (r+1)2^m$.

[W] Base d'ondelettes sur $[0, 1]$: $\{\Psi_{j,k}, j = -1, \dots, m \text{ et } k \in \mathcal{K}(j)\}$, avec $\Psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k)$, ψ est l'ondelette mère. L'ensemble $\mathcal{K}(j) = \{0, \dots, K2^j - 1\}$, pour $j \geq 0$, où K est une constante qui ne dépend que de ψ . Par exemple, pour la base de Haar, construite avec $\psi(x) = \mathbf{I}_{[0,1/2]}(x) - \mathbf{I}_{[1/2,1]}(x)$, on a $K = 1$. L'espace S_m engendré par la base d'ondelettes est de dimension $D_m = \sum_{j=-1}^m |\mathcal{K}(j)| = |\mathcal{K}(-1)| + K(2^{m+1} - 1)$.

Par construction, les éléments de cette base ont leur support inclus dans $[0, 1]$.

[H] Base d'histogramme : $\varphi_1, \dots, \varphi_{2^m}$ avec $\varphi_j = 2^{m/2} \mathbf{1}([(j-1)/2^m, j/2^m[)$ pour $j = 1, \dots, 2^m$. Dans ce cas, la dimension est $D_m = 2^m$. La base d'histogramme est un exemple particulier de bases [DP] et [W] qui sont des bases localisées.

Toutes ces bases engendrent, pour chaque $m \in \mathcal{M}_n$, un sous-espace vectoriel de $\mathbb{L}^2(A) \cap L^\infty(A)$ (ou $\mathbb{L}^2([0, 1]) \cap L^\infty([0, 1])$), qui possède les propriétés suivantes :

$$(\mathcal{M}_1) \forall m \in \mathcal{M}_n, \dim(S_m) = D_m \leq n.$$

$$(\mathcal{M}_2) \text{ Il existe une constante } \Phi_0 > 0 \text{ telle que } \forall t \in S_m, \|t\|_\infty \leq \Phi_0 \sqrt{D_m} \|t\|.$$

C'est une propriété importante qui permet de relier la norme infinie à la norme dans $\mathbb{L}^2(A)$, introduite sous cette forme par Barron *et al.* (1999).

Enfin, pour les résultats d'adaptation, nous avons besoin que les modèles soient emboîtés :

$$(\mathcal{M}_3) \forall m, m' \in \mathcal{M}_n, D_m \leq D_{m'} \implies S_m \subset S_{m'}.$$

Extension des modèles en dimension 2

Dans les chapitres 2 et 3, nous estimons des fonctions de deux variables sur $A = A_1 \times A_2$ un compact de \mathbb{R}^2 . Une base orthonormale de $\mathbb{L}^2(A_1 \times A_2)$ peut être construite en prenant simplement le produit tensoriel de deux bases de $\mathbb{L}^2(A_1)$ et de $\mathbb{L}^2(A_2)$.

Pour chaque couple d'indices $m = (m_1, m_2)$, S_m est un espace de fonctions à support dans $A = A_1 \times A_2$ défini comme un espace produit de $S_{m_1}^{(1)} \otimes S_{m_2}^{(2)}$ avec $S_{m_i}^{(i)} \subset (L^2 \cap L^\infty)(\mathbb{R})$, pour $i = 1, 2$, engendrés par deux bases orthonormales $(\varphi_j^{m_1})_{j \in J_{m_1}}$ avec $|J_{m_1}| = D_{m_1}^{(1)}$ et $(\psi_k^{m_2})_{k \in K_{m_2}}$ avec $|K_{m_2}| = D_{m_2}^{(2)}$. Les indices j et k peuvent désigner des couples d'entiers comme nous l'avons décrit pour la base de polynômes par morceaux. Ainsi, on a :

$$S_{m_1}^{(1)} = \{t / t(x) = \sum_{j \in J_{m_1}} a_j^{m_1} \varphi_j^{m_1}(x)\},$$

$$S_{m_2}^{(2)} = \{t / t(y) = \sum_{k \in K_{m_2}} a_k^{m_2} \psi_k^{m_2}(y)\}$$

et

$$S_m = S_{m_1}^{(1)} \otimes S_{m_2}^{(2)} = \{T / T(x, y) = \sum_{j \in J_{m_1}} \sum_{k \in K_{m_2}} A_{j,k}^m \varphi_j^{m_1}(x) \psi_k^{m_2}(y), A_{j,k}^m \in \mathbb{R}\}.$$

On peut utiliser la même base dans les deux directions ou au contraire autoriser des bases différentes. La dimension de chaque modèle $S_m = S_{m_1} \otimes S_{m_2}$ est alors égale à $D_{m_1} D_{m_2}$.

Propriétés d'approximation

Les méthodes d'estimation par projection sur des bases orthonormales, consistent à estimer une fonction s (ou $s \mathbf{1}_A$ sa restriction sur un compact A) par approximation de sa projection

orthogonale S_m sur un espace de dimension finie, le modèle S_m . Le résultat suivant nous fournit l'ordre de l'erreur d'approximation et se déduit de Hochmuth (2002) pour les bases localisées (polynômes ou ondelettes) et de Nikol'skii (1975) pour les bases trigonométriques.

Proposition 3 *Soit s une fonction de $\mathcal{B}_{2,\infty}^\alpha(A)$, avec $A \subset \mathbb{R}$, et $\alpha > 0$. Considérons un modèle S_m décrit ci-dessus de dimension D_m . Si s_m est la projection orthogonale de $s\mathbf{1}_A$ sur S_m , alors il existe une constante $C > 0$ telle que*

$$\|s\mathbf{1}_A - s_m\|_A = \left(\int_A (s - s_m)^2 \right)^{1/2} \leq CD_m^{-\alpha}$$

où la constante C dépend uniquement de la base et de la norme de s dans l'espace de Besov.

En dimension 2, le résultat de la proposition précédente devient :

Proposition 4 *Soit s une fonction de $\mathcal{B}_{2,\infty}^\alpha(A)$, avec $A = A_1 \times A_2 \subset \mathbb{R}^2$ et $\alpha = (\alpha_1, \alpha_2)$. Considérons un modèle $S_m = S_{m_1} \otimes S_{m_2}$ décrit ci-dessus de dimension $D_{m_1}D_{m_2}$. Si s_m est la projection orthogonale de $s\mathbf{1}_A$ sur S_m , alors il existe une constante $C' > 0$ telle que*

$$\|s\mathbf{1}_A - s_m\|_A = \left(\int_A (s - s_m)^2 \right)^{1/2} \leq C'[D_{m_1}^{-\alpha_1} + D_{m_2}^{-\alpha_2}]$$

où la constante C dépend uniquement de la base et de la norme de s dans l'espace de Besov.

Présentation des travaux

Les articles sont cités dans le manuscrit par leur numéro entre crochets. Les publications [1]–[4] correspondent aux travaux de la thèse de doctorat. Le chapitre 1 présente les articles [5], [8], [10], [11] et [12], le chapitre 2 les articles [6], [7], [9] et [13], le chapitre 3 les articles [14] et [15].

- [1] Brunel E., (1998) Applications d’Estimateurs de la Densité à la Simulation d’Episodes Pluvieux Extrêmes en Languedoc-Roussillon, *Rev. Statistique Appliquée*, XLVI(4), 45-58.
- [2] Berline A., Brunel E., (2000) Choix optimal du nombre de classes pour l’estimateur de Barron. *Comptes Rendus de l’Académie des Sciences de Paris*, 331, no 9, 713-716.
- [3] Berline A., Brunel E., (2002) Kullback-Leibler optimization of density estimates. *Recent Advances in Statistical Methods (Montreal, QC, 2001)*, 31-41, Imp. Coll. Press, London.
- [4] Berline A., Brunel E., (2004) Cross-validated density estimates based on Kullback-Leibler information, *Journal of Nonparametric Statistics*, 16, 3-4, 493-513.
- [5] Brunel E., Comte F., (2005) Penalized contrast estimation of density and hazard rate with censored data, *Sankhya*, 67, Part 3, 441-475.
- [6] Brunel E., Comte F., (2006) Adaptive nonparametric regression estimation in presence of right censoring, *Math. Methods Statist.* , 15, 3, 233–255.
- [7] Brunel E., Comte F., Lacour C., (2007) Adaptive estimation of the conditional density in presence of censoring, *Sankhya* 69, 4, 734-763.
- [8] Brunel E., Comte F., (2008) Adaptive estimation of hazard rate with censored data, *Comm. Statist. Theory Methods* 37, no. 8, 1284-1305.
- [9] Brunel E., Comte F., (2008) Model selection for additive regression models in the presence of censoring, Chapitre 1, *Mathematical Methods in Survival Analysis, Reliability and Quality of Life*. Edité par C. Huber, N. Limnios, M. Mesbah and M. Nikulin, ISTE and Wiley, p. 17-31.
- [10] Brunel E., Comte F., Guillaou A., (2008) Estimation strategies for censored lifetimes with a Lexis-diagram type model, *Scand. J. Statist.* 35, 3, 557-576.
- [11] Brunel E., Comte F., Guillaou A., (2009) Nonparametric density estimation in presence of bias and censoring, *Test* , 18, no. 1, 166-194.
- [12] Brunel E., Comte F., (2009) Cumulative distribution function estimation under interval censoring case 1, *Electron. J. Stat.* 3, 1-24.
- [13] Brunel E., Comte F., Lacour C., (2010) Minimax estimation of the conditional cumulative distribution function, *Sankhya*. 72 , no. 2, 293-330.
- [14] Brunel E., Comte F., (2011) Conditional mean residual life estimation. *J. Nonparam. Statist.*, 23, no. 2, 471-495.

- [15] Brunel, E., Comte F., Guilloux A., (2013) Nonparametric estimation for survival data with censoring indicators missing at random. *J. Statist. Planning Inf.*, 143, 1653-1671.

Travaux soumis pour publication :

- [16] Brunel, E., Roche, A., (2012) Penalized contrast estimation in functional linear models with circular data. *soumis*. HAL <http://hal.archives-ouvertes.fr/hal-00651399>
- [17] Brunel, E., Mas, A., Roche, A. (2013) Non-asymptotic Adaptive Prediction in Functional Linear Models. *soumis*. HAL <http://hal.archives-ouvertes.fr/hal-00763924>
- [18] Brunel E., Comte, F., Guilloux, A., (2013) Estimation/Imputation strategies for missing data in survival analysis. *soumis*. HAL <http://hal.archives-ouvertes.fr/hal-00818859>

Bibliographie

- Aalen, O. (1978). Nonparametric inference for a family of counting processes. *Ann. Statist.*, **6**, 701-726.
- Allignol, A., Latouche, A., Yan, J., Fine, J.P. (2011). A regression model for the conditional probability of a competing event : application to monoclonal gammopathy of unknown significance, *Applied Statistics* **60**, 135-142.
- Andersen, P.K. , Borgan, O., Gill, R. D., Keiding, N. (1993). Statistical models based on counting processes. Springer Series in Statistics. Springer-Verlag, New York.
- Akakpo, N., Lacour, C. (2011). Inhomogeneous and anisotropic conditional density estimation from dependent data. *Electron. J. Stat.* **5**, 1618–1653.
- Antoniadis, A., Grégoire, G. (1990). Penalized likelihood estimation for rates with censored survival data. *Scand. J. Statist.* **17**, 43-63.
- Antoniadis, A., Grégoire, G., Nason, G. (1999). Density and hazard rate estimation for right-censored data by using wavelet methods. *J. R. Stat. Soc., Ser. B Stat. Methodol.* **61**, 63-84.
- Asgharian, M., M'Lan, C.E., Wolfson, D.B. (2002). Length-biased sampling with right censoring : an unconditional approach. *J. Amer. Statist. Assoc.* **97**, 201-209.
- Asgharian, M. (2003). Biased sampling with right censoring : a note on Sun, Cui & Tiwari (2002). *Canad. J. Statist.* **30**, 475-490.
- Azzalini, A., Bowman, A.W. (1990). A look at some data on the Old Faithfull geyser. *Applied Statistics* **39**, 357-365.
- Baraud, Y. (2002) Model selection for regression on a random design. *ESAIM P&S* **6**, 127-146.
- Baraud, Y., Comte, F., Viennet, G. (2001). Adaptive estimation in autoregression or β -mixing regression via model selection. *Ann. Statist* **29**, 839-875.
- Barron, A.R., Cover, T. M. (1991). Minimum complexity density estimation. *IEEE Trans. Inform. Theory* **37**, 1034-1054.

- Barron, A.R., Birgé, L., Massart, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301-413.
- Bhattacharya, P.K., Gangopadhyay, A.K. (1990). Kernel and nearest-neighbor estimation of a conditional quantile. *Ann. Stat.* **18**, 1400-1415.
- Birgé, L. (1999). Interval censoring : a nonasymptotic point of view. *Math. Methods Statist.* **8**, 285–298.
- Birgé, L., Massart, P. (1993). Rates of convergence for minimum contrast estimators. *Probab. Theory Relat. Fields* **97**, 113-150.
- Birgé, L., Massart, P. (1997). From model selection to adaptive estimation, in *Festschrift for Lucien Le Cam : Research Papers in Probability and Statistics* (D. Pollard, E. Torgersen and G. Yang, eds), 55-87, Springer-Verlag, New-York.
- Birgé, L., Massart, P. (1998). Minimum contrast estimators on sieves : exponential bounds and rates of convergence. *Bernoulli* **4**, 329-375.
- Bitouzé, D., Laurent, B., Massart, P. (1999). A Dvoretzky-Kiefer-Wolfowitz type inequality for the Kaplan-Meier estimator. *Ann. Inst. Henri Poincaré* **32**, 735-763.
- Cai, T., Hall, P. (2006). Prediction in functional linear regression. *Ann. Statist.* **34**, 2159–2179.
- Cai, T., Yuan, M. (2012). Minimax and Adaptive Prediction for Functional Linear Regression, *J. Amer. Statist. Assoc.* **107**, 1201-1216.
- Cardot, H. , Ferraty, F., Sarda, P. (2003). Spline estimator for the functional linear model, *Statistica Sinica*, 571-591.
- Cardot, H., Johannes, J. (2010). Thresholding projection estimators in functional linear models. *J. Multivariate Anal.* **101**, 395–408.
- Comte, F., Johannes, J. (2010). Adaptive estimation in circular functional linear models. *Math. Methods Statist.* **19** 42–63.
- Comte, F., Johannes, J. (2012). Adaptive functional linear regression, *Ann. Statist.* **40**, 2765-2797.
- Cook, R.J., Lawless, J.F., (2008) The statistical analysis of recurrent events, Springer, New-York.
- Chagny, G. (2011). Warped bases for conditional density estimation. Preprint. hal-00641560, version 2.

- Chen, Dong., Hall, P., Müller, H-G. (2011). Single and multiple index functional regression models with nonparametric link. *Ann. Statist.* **39**, 1720-1747.
- Chen, Y. Q., Cheng, S. (2005). Semiparametric regression analysis of mean residual life with censored survival data, *Biometrika* **92**, 19-29.
- Chernozhukov, V., Fernández-Val, I., Galichon, A. (2009). Improving point and interval estimates of monotone functions by rearrangement. *Biometrika* **96**, 559-575.
- Comte, F., Gaïffas, S., Guillaou, A. (2011). Adaptive estimation of the conditional intensity of marker-dependent counting processes, *Ann. Inst. H. Poincaré Probab. Statist.* **47**, 4, 1171-1196.
- Cohen, A., Daubechies, I., Vial, P.B. (1993). Wavelets on the interval and fast wavelet transforms. *Appl. Comput. Harmon. Anal.* **1**, 54-81.
- Cohen, S., Lepennec, E. (2011). Conditional density estimation by penalized likelihood model selection and applications. Preprint. arXiv :1103.2021v5.
- Crambes, C., Kneip, A., Sarda, P. (2009). Smoothing Splines Estimators for Functional Linear Regression, *Ann. Statist.*, **1**, 35–72.
- Cox, D. R. (1972). Regression models and life-tables. *J. Roy. Statist. Soc. Ser. B*, **34**, 187-220.
- Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data. *Scand. J. Statist.* **14**, 181-197.
- Dabrowska, D.M., Duffy, D.L., Zhang, Z.D. (1999). Hazard and density estimation from bivariate censored data. *J. Nonparametr. Statist.* **10**, 67-93.
- de Boor, C. (1978). *A practical guide to splines*. Applied Mathematical Science, Springer, New York.
- De Gooijer, J. G., Zerom, D. (2003). On conditional density estimation. *Statist. Neerlandica*, **57**, 159-176.
- Delecroix, M., Yazourh, Y. (1991). Nonparametric inference for failure rates with random right censorship : the method of orthogonal functions [French]. *Statist. Anal. Données* **16**, 39-62.
- DeVore, R.A., Lorentz, G.G. (1993). *Constructive approximation*. Springer-Verlag.
- Diamond, I.D., McDonald, J.W. (1991). The analysis of current status data. *Demographic applications of event history analysis*.(eds J. Trussell, R. Hankinson & J. Tilton). Oxford University Press, Oxford.

- Dikta, M.K., (1998). On semiparametric random censorship models. *J. Stat. Plann. Inference* **66**, 253-279.
- Donoho, D.L., Johnstone, I.M. (1998). Minimax estimation with wavelet shrinkage. *Ann. Statist.*, **26**, 879-921.
- Donoho, D. L., Johnstone, I. M., Kerkyacharian, G., Picard, D. (1995). Wavelet shrinkage : asymptopia? *J. Roy. Statist. Soc. Ser. B* **57**, 301-369.
- Efromovich, S. (2001). Density estimation under random censorship and order restrictions : from asymptotic to small sample sizes. *J. Amer. Statist. Assoc.* **96**, 667-685.
- Efromovich, S. (2004). Density estimation for biased data. *Ann. Statist.* **32**, 1137-1161.
- Efromovich, S. (2004). Distribution estimation for biased data. *J. Stat. Plann. Inference* **124**, 1-43.
- Efromovich, S. (2007). Conditional density estimation in a regression setting. *Ann. Statist.* **35**, 2504-2535.
- Efromovich, S. (2008). Oracle inequality for conditional density estimation and an actuarial example. *Ann. Inst. Statist. Math.*, 1-27.
- Engl, H. W., Hanke, M., Neubauer, A. (1996). Regularization of inverse problems, Mathematics and its Applications, 375 Kluwer Academic Publishers Group.
- Faugeras, O.P. (2009). A quantile-copula approach to conditional density estimation. *J. Multivariate Anal.* **100**, 2083-2099.
- Fan, J., Gijbels, I. (1994). Censored regression : local linear approximations and their applications. *J. Amer. Statist. Assoc.* **89**, 560-570.
- Fan, J., Hu, T.-C., Truong, Y. (1994) Robust nonparametric function estimation. *Scand. J. Statist.*, **21**, 433-446.
- Ferraty, F., Vieu, P. (2006). Nonparametric functional data analysis, Springer Series in Statistics, Theory and practice, New York, Springer.
- Gijbels, I., Lin, D., Ying, Z., (2007). Non- and semi-parametric analysis of failure time data with missing failure indicators. Complex datasets and inverse problems, 203-223, IMS Lecture Notes Monogr. Ser., 54, Inst. Math. Statist., Beachwood, OH.
- Gill, R.D., Vardi, Y., Wellner, J. A. (1988). Large sample theory of empirical distributions in biased sampling models. *Ann. Statist.* **16**, 1069–1112.

- Giné, E. and Nickl, R. (2010). Confidence bands in density estimation. *Ann. Statist.* **38**, 1122-1170.
- Goldenshluger, A. and Lepski, O. (2011). Bandwidth selection in kernel density estimation : oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39**, 1608-1632.
- Groeneboom, P., Wellner, J.A. (1992). Information bounds and nonparametric maximum likelihood estimation. Boston, Birkhäuser Verlag.
- Guilloux, A. (2007). Non-parametric estimation for censored lifetimes suffering from an unknown selection bias. *Math. Methods Statist.* **16**, 202-216.
- Györfi, L., Kohler, M., Krzyzak, A., Walk, H. (2002). A distribution-free theory of nonparametric regression. Springer Series in Statistics, Springer 2002.
- Györfi, L., Kohler, M. (2007). Nonparametric estimation of conditional distributions. *IEEE Trans. Inform. Theory* **53**, 1872-1879.
- HALL, P. and HOROWITZ, J. L. (2007). Methodology and convergence rates for functional linear regression. *Ann. Statist.* **35**, 70–91.
- He, X., Shi, P. (1994). Convergence rate of B -spline estimators of nonparametric conditional quantile functions. *J. Nonparam. Statist.*, **3**, 299-308.
- Heuchenne, C., Van Keilegom, I. (2007). Nonlinear regression with censored data. *Technometrics* **49**, 34-44.
- Hochmuth, R. (2002) Wavelet characterizations for anisotropic Besov spaces. *Appl. Comput. Harmon. Anal.* **2**, 179-208.
- Huber, C., MacGibbon, B. (2004). Lower bounds for estimating a hazard. *Advances in survival analysis*, 209–226, Handbook of Statist., 23 Elsevier, Amsterdam.
- Hudgens, M.G., Maathuis, M.H., Gilbert, P.B. (2007). Nonparametric estimation of the joint distribution of a survival time subject to interval censoring and a continuous mark variable. *Biometrics* **63**, 372–380.
- Jewell, N. P., van der Laan, M. (2004). *Current status data : review, recent developments and open problems*. Advances in survival analysis, 625–642, Handbook of Statist., 23, Elsevier, Amsterdam.
- Kaplan, E.L., Meier, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53**, 457-481.

- Keiding, N. (1990). Statistical inference in the Lexis diagram. *Philos. Trans. Roy. Soc. London Ser. A* **332**(1627) :487–509.
- Koenker, R. (2005). *Quantile regression*. Econometric Society Monographs, Cambridge.
- Koenker, R. and Bassett, G. (1978). Regression quantiles. *Econometrica*, **46**, 33-50.
- Koenker, R., Ng, P., Portnoy, S. (1994). Quantile smoothing splines. *Biometrika*, **81**, 673-680.
- Koul, H., Susarla, V., Van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276–1288.
- Lacour, C. (2007). Adaptive estimation of the transition density of a Markov chain. *Ann. Inst. Henri Poincaré Probab. Statist.* **43**, 571-597.
- Lejeune, M., Sarda, P. (1988). Quantile regression : a nonparametric approach. *Comput. Statist. Data Anal.* , **6**, 229-239.
- Leurgans, S. (1987). Linear models, random censoring and synthetic data. *Biometrika* **74**, 301–309.
- Lexis, W. (1875). Einleitung in die Theorie der Bevölkerung-Statistik. In *Mathematical demography* (édition D. Smith and N. Keyfitz), *Biomathematics* 6, 39-41, 1977. Springer-Verlag, Berlin.
- Li, Y., Hsing, T. (2007). On rates of convergence in functional linear regression, *J. Multivariate Anal.* **98**, 1782–1804.
- Little, R.J.A Rubin, D.B. (2002). *Statistical analysis with missing data*. Second edition. Wiley Series in Probability and Statistics.
- Lo, S.-H., (1991). Estimating a survival function with incomplete cause-of-death data. *J. Multivariate Anal.* **39**, 217-235.
- Lo, S.H., Mack, Y.P., Wang, J.L. (1989). Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator. *Probab. Theory Relat. Fields* **80**, 461-473.
- Lopez, O., Patilea, V. , Van Keilegom, I.(2013). Single index regression models in the presence of censoring depending on the covariates. à paraître *Bernoulli*.
- Lund, J. (2000). Sampling bias in population studies - how to use the Lexis diagram. *Scand. J. Statist.* **27**, 589-604.

- Ma, S., Kosorok, M.R. (2006). Adaptive penalized M-estimation with current status data. *Ann. Inst. Statist. Math.* **58**, 511–526.
- Marron, J., Padgett, W.J. (1987). Asymptotically optimal bandwidth selection for kernel density estimators from randomly right-censored samples. *Ann. Statist.* **15**, 1520-1535.
- Massart, P. (2007). Concentration inequalities and model selection. Ecole d'été de Probabilités de Saint-Flour 2003. Lecture Notes in Mathematics 1896, Springer Berlin Heidelberg.
- Mielniczuk, J. (1985). Properties of some kernel estimators and of the adapted Loftsgarden- Quesenberry estimator of a density function for censored data. *Period. Math. Hungar.* **16**, 69-81.
- Nikol'skii, S.M. (1975). Approximations of functions of several variables and imbedding theorems, Springer-Verlag, New York, translated from the Russian by John M. Danskin, Jr., Die Grundlehren der Mathematischen Wissenschaften, Band 205.
- Patil, P. (1997). Nonparametric hazard rate estimation by orthogonal wavelet method. *J. Statist. Plann. Inference* **60**, 153-168.
- Park, J. (2004). Optimal global rate of convergence in nonparametric regression with left-truncated and right-censored data. *J. Multivariate Anal.* **89**, 70–86.
- Picard, D., Tribouley, K.(2000) Adaptive confidence interval for pointwise curve estimation. *Ann. Statist.* **28**, 298-335.
- Placade, S. (2011). Estimation of conditional cumulative distribution function from current status data. Preprint. arXiv :1110.5927v1.
- Poiraud-Casanova, S., Thomas-Agnan, C. (1998). Quantiles conditionnels. *J. Soc. Française Stat.*, **139**, 31-44.
- Prentice, R.I. (1973). Exponential survivals with censoring and explanatory variables, *Biometrika* **60**, 279-288.
- Preda, C., Saporta, G. (2005). PLS regression on a stochastic process, *Comput. Statist. Data Anal.*, **48**, 149–158.
- Ramsay, J.O., Silverman, B.W., (2005). Functional Data Analysis, Springer Series in Statistics, 2nd edition Springer.
- Ramsay, J.O., Dalzell, C.J., (1991). Some Tools for Functional Data Analysis, *J. R. Statist. Soc.* **53**, 539–572.

- Reid, N. (1981) Influence functions for censored data. *Ann. Statist.* **9**, 78–92.
- Reynaud-Bouret, P. (2006). Penalized projection estimators of the Aalen multiplicative intensity. *Bernoulli*, **12**, 633-661.
- Ruppert, D., Carroll, J. (1988). Transformation and weighting in regression. Chapman and Hall, New York.
- Schaubel, D.E., Cai, J. (2004). Regression Methods for Gap Time Hazard Functions of Sequentially Ordered Multivariate Failure Time Data , *Biometrika* **91**, 291–303.
- Stute, W. (1995). The central limit theorem under random censorship. *Ann. Statist.* **23**, 422-439.
- Talagrand, M. (1996). New concentration inequalities in product spaces. *Invent. Math.*, **126**, 505-563.
- Subramanian, S., (2004). Asymptotically efficient estimation of a survival function in the missing censoring indicator model. *J. Nonparam. Statist.* **16**, 797-817.
- Subramanian, S., (2006). Survival analysis for the missing censoring indicator model using kernel density estimation techniques. *Stat. Methodol.* **3**, 125-136.
- Subramanian, S., (2009). The multiple imputations based Kaplan-Meier estimator. *Statist. Probab. Lett.* **79**, 1906-1914.
- Subramanian, S., (2011). Multiple imputations and the missing censoring indicator model. *J. Multivariate Anal.* **102**, 105-117.
- Triboulay, K. (2004). Adaptive simultaneous confidence interval in nonparametric estimation. *Statist. Probab. Letters* **69** , 37-51.
- de Uña-Álvarez, J. (2002). Product-limit estimation for length-biased censored data. *Test* **11**, 109-125.
- Vardi, Y. (1982). Nonparametric estimation in renewal processes. *Ann. Statist.* **10**, 772–785.
- van de Geer, S. (1993). Hellinger-consistency of certain nonparametric likelihood estimators. *Ann. Statist.* **21**, 14-44.
- van der Laan, M. J., McKeague, I. W., 1998. Efficient estimation from right-censored data when failure indicators are missing at random. *Ann. Statist.* **26**, 164-182.
- van der Vaart, A., van der Laan, M. J. (2006). Estimating a survival distribution with current status data and high-dimensional covariates. *Int. J. Biostat.* **2**, Art 9, 42pp.

- Van Keilegom, I., de Una-Álvarez, J., Meira-Machado, L. (2011). Nonparametric location-scale models for successive survival times under dependent censoring, *J. Statist. Plann. Infer.* **141**, 1118-1131.
- Wang, Q., Shen, J., (2008). Estimation and confidence bands of a conditional survival function with censoring indicators missing at random. *J. Multivariate Anal.* **99**, 928-948.
- Wang, Q., Ng, K. W., (2008). Asymptotically efficient product-limit estimators with censoring indicators missing at random. *Statist. Sinica* **18**, 749-768.
- Wang, Q. , Liu, W., Liu, C., (2009). Probability density estimation for survival data with censoring indicators missing at random. *J. Multivariate Anal.* **100**, 835-850.
- Yang, S. (2000). Functional estimation under interval censoring case 1. *J. Statist. Plann. Inference* **89**, 135-144.
- Zheng, Z. K. (1987). A class of estimators of the parameters in linear regression with censored data, *Acta Math. Appl. Sinica* **3**, 231-341.
- Zheng, Z. K. (1988). Strong consistency of nonparametric regression estimates with censored data, *J. Math. Res. Exposition* **8**, 307-313.