==========

République algérienne démocratique et populaire
Ministère de l'enseignement supérieur et de la recherche scientifique

Université Des Frères Mentouri
Faculté des sciences exactes

==========

Département de mathématiques

N° d'ordre : 126| DS |2014
N° de série : 09|MAT|2014

THÈSE

présentée pour l'obtention
du
diplôme de doctorat en sciences
de
mathématiques

« Inférence statistique dans le cas d'observations censurées : Estimateurs des moindres carrés et spline de lissage de la fonction de régression. »

Par
LAROUSSI ILHEM

OPTION
Probabilités et Statistique

Devant le jury :

Président	M.	Z. Mohdeb	Professeur	Université Des Frères Mentouri
Directrice de thèse	M^{me}	F. Messaci	Professeur	Université Des Frères Mentouri
Examinatrice	M^{me}.	K. Boudraa	M. C. A.	ENS Constantine
Examinatrice	M^{me}.	Z. Guessoum	M. C. A.	Université USTHB Alger
Examinatrice	M^{me}.	O. Sadki	Professeur	Université Oum El Bouaghi

Soutenue le : 16/12/ 2014.

Remerciements

Je remercie infiniment ma directrice de thèse, madame F. Messaci pour toute l'aide est les moyens qu'elle a mis a ma disposition pour avancer, je la salue particulièrement pour ses qualités humaines et sa rigueur scientifique.

J'adresse mes remerciements sincères et chaleureux à monsieur Z. Mohdeb qui me fait l'honneur de présider le jury de soutenance.
Mes sincères remerciements à O. Sadki, Z. Guessoum et K. Boudraa, pour l'honneur qu'elles me font d'examiner mon travail.
Je remercie grandement ma collègue et amie K. Kebabi pour toute l'aide qu'elle m'a apportée avec autant de gentillesse et de générosité.
Merci aussi à mes amis C. Matmat et F. Saadi, tout particulièrement pour leurs présences.

Table des matières

Introduction

L'analyse de régression, dont le nom est dû à Sir Francis Galton (1885), est un domaine très vaste et très riche par la multitude de travaux qui lui ont été consacrés. L'intérêt pour l'analyse de causalité entre plusieurs variables remonte jusqu' au $16^{\text{ème}}$ siècle. En 1757, R. Boscovich propose de minimiser la somme des valeurs absolues entre un modèle de causalité et les observations. Ensuite, en 1805 Legendre introduit la méthode des moindres carrés et lui donne son nom. Gauss, quant à lui, publie en 1809 un développement de la méthode des moindres carrés.

Le principe de cette analyse est d'étudier la relation entre une variable dépendante Y et une variable explicative X (pas obligatoirement réelle), dans le but de prédire une réalisation de Y une fois X observée. A cette fin, nous considérons le modèle $Y = r(X)+\varepsilon$ où ε est centrée et est indépendante de X. Ici ε représente l'erreur commise pendant la sélection du modèle et $r(x) = \mathbf{E}(Y|X = x)$.

Une première approche pour l'estimation de la fonction de régression est l'estimation paramétrique, dans le cadre de laquelle on suppose que la structure de la fonction de régression est connue et dépend d'un nombre fini de paramètres. L'utilisation des données sert alors à estimer les valeurs de ces paramètres, parmi les articles sur se sujet, citons Rao (1973), Seber (1977), Drapper et Smith (1981) et Farebrother (1988) qui ont tous travaillé dans le cadre de la théorie de L_2 qui consiste à minimiser la somme des carrés des résidus.

La première étape d'une étude de régression devrait être la représentation des données à l'aide d'un graphique. Ceci permet de savoir si le modèle linéaire (ou autres) est pertinent et a pour avantage la facilité du calcul et de l'interprétation. Mais il arrive que la tendance décrite par l'échantillon ne soit pas évidente. La solution à ce problème est alors apportée par l'estimation non paramétrique de la fonction de régression. Cette méthode laisse les données choisir la forme de la relation adéquate entre les variables. Les

estimateurs de r obtenus sont souvent appelés fonctions de lissage.
Il existe plusieurs méthodes d'estimation non paramétrique. La littérature s'y rapportant est si riche que nous ne pouvons prétendre tout citer. Comme exemple, les estimateurs à noyaux sont introduits dans Nadaraya (1964) et Watson (1964) et repris, entre autres, dans Devroye et Wagner (1980) et Walk (1997). Les estimateurs des plus proches voisins sont étudiés dans Stone (1977) et Devroye (1994), ceux à partition dans Walk (1997). Plus récemment la méthode des ondelettes a vu le jour et permet de mieux estimer les fonctions moins lisses, nous pouvons nous référer à Antoniadis et al.(1994), Antoniadis (1996), Donoho et Johnstone (1994, 1995, 1998) et Donoho et al. (1995).
Dans ce travail, nous nous intéressons aux méthodes des moindres carrés et des moindres carrés pénalisés dont l'introduction est motivée dans le chapitre 1. Parmi les travaux sur ce sujet, citons Vapnik et Chervonenkis (1971), Vapnik (1982, 1998), Haussler (1992), Lugosi et Zeger (1995), Devroye et al. (1996) et Kohler (1997a,1999) pour la première méthode. Quand à la seconde elle est étudiée dans Schumaker (1981), Wahba (1990), Kohler (1997, 1999), Eubank (1999) et Kohler et Krzyżak (2001). Nous commençons par expliquer comment sont construits ces estimateurs dans le cas de données complètes, autrement dit la variable expliquée est totalement observée. Puis nous passons au cas où la variable réponse est censurée, ce qui veut dire que la valeur de cette variable peut être perdue au cours de l'étude. Par exemple si la variable expliquée est le temps de survie à une maladie et la variable explicative est un traitement expérimental, dans certains cas le malade quitte l'étude (par exemple il meurt dans un accident ou change de traitement). Ici le temps de survie au traitement est censuré par le temps de départ de ce malade, on dit que la variable est censurée à droite : on sait seulement que la valeur d'intérêt est supérieure à l'observation. Parmi les travaux sur l'estimation de la fonction de survie (complément à 1 de la fonction de répartition) dans un modèle de censure à droite, en plus de l'article fondateur de Kaplan et Meier (1958), citons Peterson (1977), Stute et Wang (1993), Gill (1994). Concernant l'estimation de la fonction de régression dans ce même modèle, Beran (1981) introduit une estimation de la fonction de survie conditionnelle de laquelle se déduit, hélas laborieusement, l'estimation qui nous intéresse et prouve quelques résultats de consistance, qui ont été améliorés par Dabrowska (1987, 1989). Dans leurs travaux, ils supposent que le temps de survie et le temps de censure sont conditionnellement indépendants, sachant la covariable.
Carbonez et al. (1995) introduisent un estimateur à partitions de la fonction de régression. Kohler et al. (2002) exploitent l'idée de ce dernier travail, qui consiste à utiliser un estimateur sans biais de la moyenne de Y, pour

l'étendre à diverses méthodes (à noyau, plus proches voisins, moindres carrés et spline de lissage). Ils démontrent la consistance des estimateurs introduits en imposant l'indépendance du couple (X, Y) et de la variable de censure à droite R. Cette condition est souvent non incommodante, puisqu'elle est raisonnable lorsque la censure ne dépend pas des caractéristiques de l'individu sous étude. Ils imposent aussi la continuité de la loi de R et le fait que son support contient celui de Y.

Un phénomène de censure à gauche peut aussi empêcher l'observation du phénomène d'intérêt pour lequel on sait seulement qu'il est inférieur à la valeur observée.

Les deux types de censure peuvent se combiner et s'inviter dans un même échantillon. Ceci et le cas dans la partie constituant l'apport principal de cette thèse et qui consiste à étendre l'estimation de la fonction de régression au cas où la variable réponse est soumise à une censure mixte. Ce qui veut dire que Y est censurée à droite par une variable R (qui elle même représente un temps de survie) et que le minimum entre Y et R est censuré par une variable de censure à gauche. Toutes les variables latentes sont supposées indépendantes. C'est le modèle étudié dans l'article de Patiléa et Rolin (2006) dans lequel est proposé un estimateur produit limite de la fonction de survie fortement consistant. Dans Messaci (2010) sont construits des estimateurs à poids (à noyau, à partitions et des plus proches voisins) fortement consistants de la fonction de régression pour Y soumise à une censure mixte. L'étude de l'estimateur à noyau s'est poursuivie dans l'article de Kebabi et Messaci (2012) dans lequel des taux de convergence presque complète [1] ponctuelle et uniforme sur un compact ont été établis.

Notre thèse est constituée de cinq chapitres. Le premier est voué à l'introduction des estimateurs des moindres carrés et des moindres carrés pénalisés de la fonction de régression pour des données complètes. Quelques propriétés de convergence, dont les preuves peuvent être trouvées dans le

1. La suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ converge presque complètement vers une variable aléatoire X lorsque $n \to \infty$ si

$$\forall \varepsilon > 0, \quad \sum_{n\in\mathbb{N}} P\left[|X_n - X| > \varepsilon\right] < \infty.$$

La vitesse de convergence presque complète de la suite de variables aléatoires réelles $(X_n)_{n\in\mathbb{N}}$ vers X est d'ordre (u_n) si

$$\exists \varepsilon_0 > 0, \quad \sum_{n\in\mathbb{N}} P[|X_n - X| > \varepsilon_0 u_n] < \infty.$$

livre de Göyorfi et al. (2002), y sont rappelées. Au second, après avoir évoqué les différents types de censure, nous rappelons les estimateurs de la fonction de survie et de la fonction de régression dans un modèle de censure à droite. Puis nous revenons à l'idée ayant permis la construction de l'estimateur de Patiléa et Rolin (2006) et nous en déduisons sa forme explicite. Ce dernier nous sert aux chapitres trois et quatre à introduire des estimateurs fortement consistants de la fonction de régression lorsque la variable réponse est soumise à une censure mixte. Finalement, au dernier chapitre, nous illustrons nos résultats théoriques par une étude de simulation incluant aussi bien des modèles linéaires que non linéaires.

1 Estimation de la fonction de régression par les méthodes M. C et M. C. P

Dans ce chapitre, nous présentons des critères de choix d'estimateurs efficaces dans le cadre classique de données complètes. Puis, nous passons à l'idée d'estimation de la fonction de régression par la méthode des moindres carrés (M. C) qui nous conduit à la construction de l'estimateur. Nous présentons aussi la deuxième méthode qui est l'estimation par moindres carrés pénalisés (M. C. P), plus précisément les spline de lissage.

1.1 Critère de sélection

On considère le couple (X, Y), où X est un vecteur aléatoire de $\mathbb{R}^d$ et Y une variable aléatoire réelle et on s'intéresse à l'influence de l'observation X sur la variable réponse Y, c'est-à-dire qu'on cherche une fonction mesurable f définie de $\mathbb{R}^d$ dans $\mathbb{R}$, telle que $f(X)$ soit une bonne approximation de Y. En d'autres termes, on cherche à avoir $f(X)$ très proche de Y et comme ces deux dernières sont dans $\mathbb{R}$, on pense à $|f(X) - Y|$ très petit, le problème est que cette quantité est aléatoire et on ne peut juger du fait qu'elle soit petite ou non. Pour résoudre ce problème nous utilisons l'erreur moyenne quadratique dite risque L_2 définie par

$$\mathbf{E}|f(X) - Y|^2, \tag{1.1}$$

qui doit être la plus petite possible. Donc, le but est de fournir une fonction r^* de $\mathbb{R}^d$ dans $\mathbb{R}$ telle que

$$\mathbf{E}|r^*(X) - Y|^2 = \min_{f:\mathbb{R}^d \to \mathbb{R}} \mathbf{E}|f(X) - Y|^2.$$

Soit la fonction de régression définie par

$$r(x) = \mathbf{E}(Y|X = x).$$

Il est connu que $r(x)$ minimise le risque dans L_2. En effet, soit $f : \mathbb{R}^d \to \mathbb{R}$, alors il vient que

$$\begin{aligned}\mathbf{E}|f(X) - Y|^2 &= \mathbf{E}|f(X) - r(X) + r(X) - Y|^2 \\ &= \mathbf{E}|f(X) - r(X)|^2 + \mathbf{E}|r(X) - Y|^2.\end{aligned}$$

Ce dernier résultat vient du fait que

$$\begin{aligned}\mathbf{E}[(f(X) - r(X))(r(X) - Y)] &= \mathbf{E}\left(\mathbf{E}\left[(f(X) - r(X))(r(X) - Y)|X\right]\right) \\ &= \mathbf{E}\left[(f(X) - r(X)\mathbf{E}\left((r(X) - Y)|X\right)\right].\end{aligned}$$

Et comme

$$\mathbf{E}\left((r(X) - Y)|X\right) = r(X) - \mathbf{E}(Y|X) = 0,$$

nous obtenons

$$\mathbf{E}[(f(X) - r(X))(r(X) - Y)] = 0.$$

Ainsi

$$\mathbf{E}|f(X) - Y|^2 = \int_{\mathbb{R}^d} |f(x) - r(x)|^2 \mu(dx) + \mathbf{E}|r(X) - Y|^2, \tag{1.2}$$

où μ est la loi de X.
Le terme $\int_{\mathbb{R}^d} |f(x) - r(x)|^2 \mu(dx)$ de cette équation est appelé l'erreur L_2 alors que le terme $\mathbf{E}|r(X) - Y|^2$ est dit l'approximation optimale par rapport au risque L_2.
En posant $f(X) = r(X)$ l'erreur est égale à zéro, ce qui démontre que $r(X)$ minimise le risque L_2.

Comme r est en général inconnue, nous allons nous intéresser à son estimation.

1.1.1 Introduction à l'estimation de la fonction de régression

En général, la fonction de régression est inconnue et la possibilité d'utiliser les observations du couple (X, Y) permet de l'estimer. Soient $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ des variables aléatoires indépendantes et identiquement distribuées (i. i. d.), avec $E(Y^2) < \infty$. Notons

$$D_n = \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$$

l'ensemble des données. Soit $r_n(X)$ définie de $\mathbb{R}^d$ dans $\mathbb{R}$ un estimateur de r tel que $r_n(x) = r_n(x, D_n)$ soit une fonction mesurable. En général un estimateur ne coincide pas avec $r(x)$. il nous faut donc mettre des critères de choix d'estimateurs qui mesurent la différence entre la vraie fonction de régression et un estimateur r_n.
Commençons par « l'erreur ponctuelle » définie par

$$|r_n(x) - r(x)|,$$

pour $x \in \mathbb{R}^d$.
Le deuxième critère est « l'erreur pour la norme supérieure » donné par

$$\|r_n - r\|_\infty = \sup_{x \in C} |r_n(x) - r(x)|,$$

où C est un compact de $\mathbb{R}^d$.
Le dernier critère donné est « l'erreur L_p» définie par

$$\int_C |r_n(x) - r(x)|^p \mu(dx).$$

En général, on choisit $p = 2$.
Remarquons que l'introduction de la fonction de régression mène naturellement vers le critère d'erreur L_2 pour mesurer la performance de l'estimation. Rappelons que nous cherchons à trouver une fonction f qui minimise le risque L_2, ce minimum est atteint par la fonction de régression r appelée approximation optimale. D'une manière similaire à la relation (1.2), nous pouvons démontrer que l'estimation r_n vérifie la relation suivante

$$\mathbf{E}\{|r_n(X) - Y|^2 | D_n\} = \int_{\mathbb{R}^d} |r_n(x) - r(x)|^2 \mu(dx) + \mathbf{E}|r(X) - Y|^2. \quad (1.3)$$

Ainsi, le risque L_2 d'un estimateur r_n est proche de la valeur optimale, si est seulement si, l'erreur L_2 tend vers zéro.
En fixant notre intérêt moyennant le troisième critère de choix de l'estimation, nous allons voir, dans la partie qui suit, l'efficacité de notre estimateur.

Mode de convergence d'un estimateur

Pour étudier l'efficacité de l'estimation de la fonction de régression, nous allons étudier les modes de convergence. Ici, nous avons besoin d'un estimateur qui converge vers la quantité estimée lorsque la taille de l'échantillon

augmente, c'est-à-dire, que l'erreur L_2 doit tendre vers zéro. L'estimation qui vérifie cette condition est dite « consistante ». Pour mesurer la différence de mesure dans l'estimation de la fonction de régression, on utilise l'erreur L_2, sans oublier que r_n dépend de l'ensemble des données, c'est pourquoi, cette erreur est une variable aléatoire, donc nous nous intéresserons à la convergence en moyenne de cette variable vers zéro, aussi bien que la convergence presque sûre. Dans ce qui suit, nous allons voir différents modes de consistance et de convergence.

Définitions

Soit $\{r_n\}$ une suite d'estimations de la fonction de régression, r_n est dit faiblement consistant, pour une certaine répartition de (X, Y) si

$$\lim_{n\to\infty} \mathbf{E} \int |r_n(x) - r(x)|^2 \mu(dx) = 0.$$

Ce qui veut dire que la moyenne de l'erreur L_2 tend vers zéro.
Si cette suite vérifie la relation suivante

$$\mathbf{P}\left\{ \lim_{n\to\infty} \int |r_n(x) - r(x)|^2 \mu(dx) = 0 \right\} = 1,$$

on dit que cet estimateur est fortement consistant.
Le problème est que ces deux définitions donnent la consistance de notre estimateur pour une classe de lois et pas pour la totalité des lois. Pour généraliser ces notions à toutes les lois, nous allons parler de consistance universelle. C'est-à-dire que, pour un estimateur r_n qui est faiblement consistant (fortement consistant), pour toutes les répartitions de (X, Y) avec $\mathbf{E}(Y^2) < \infty$, est dit faiblement consistant universellement (fortement consistant universellement).
Cette notion est importante pour la régression non-paramétrique, puisque son utilisation est une conséquence directe d'un manque partiel ou total d'informations sur la loi de (X, Y).

1.2 Estimation de la fonction de régression par la méthode des moindres carrés

Dans ce qui va suivre, nous allons construire un estimateur de la fonction de régression non paramétrique, en utilisant la méthode des moindres carrés,

surnommée aussi, modélisation globale. Nous verrons aussi la consistance de cet estimateur.

1.2.1 Construction de l'estimateur

Dans cette partie, nous allons voir que l'idée principale de la construction de l'estimation de la fonction de régression vient du fait qu'on doit minimiser l'erreur L_2 pour toutes les fonctions mesurables de $\mathbb{R}^d$ dans $\mathbb{R}$. On a vu précédemment que la fonction de régression r vérifiait cette condition, le problème est qu'elle-même dépend de la loi du couple (X, Y) qui est inconnue à son tour. Pour contourner ce problème on revient à l'idée de minimiser l'erreur L_2 pour se retrouver avec une fonction mesurable f qui représente un estimateur de r. Le risque L_2 est estimé par le risque empirique qui s'écrit comme suit

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

et c'est sur ce dernier que va porter la minimisation. L'idée de minimiser le risque empirique, pour toutes les fonctions mesurables, n'est pas raisonnable, puisque, on se retrouve avec la fonction d'interpolation des données $(X_1, Y_1), \ldots, (X_n, Y_n)$, qui n'est pas un estimateur adéquat, puisque sa consistance est douteuse dans un cadre aléatoire.
Pour surmonter ce problème, on choisit une classe de fonctions $\mathcal{F}_n$, qui dépend des données ou tout au moins de la taille de l'échantillon n. Aussi ces fonctions minimisent le risque empirique que sur $\mathcal{F}_n$, ce qui revient à dire, qu'on définit un estimateur des moindres carrés r_n par

$$r_n(x) = \arg\min_{f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^{n}|f(X_i)-Y_i|^2. \tag{1.4}$$

Un exemple de classe $\mathcal{F}_n$ qui assure l'existence de ce minimum est donné dans le chapitre 10 du livre de Györfi et al. [21].
Remarquons que si $\mathcal{F}_n = \{\sum_j a_j 1_{A_{n,j}} : a_j \in \mathbb{R}\}$, où $\{A_{n,1}, A_{n,2}, ..., A_{n,j}, ...\}$ est une partition de $\mathbb{R}^d$, alors il est aisé de voir que l'estimateur des moindres carrés n'est autre que l'estimateur à partitions donné par

$$r_n(x) = \sum_{i=1}^{n}\frac{1_{\{X_i\in A_{n,j}\}}Y_i}{\sum_{i=1}^{n}1_{\{X_i\in A_{n,j}\}}},\ \text{pour } x\in A_{n,j}. \tag{1.5}$$

La classe de fonction $\mathcal{F}_n$ affecte de deux manières l'erreur d'estimation. En premier temps, si elle n'est pas trop riche le risque empirique s'approche

du risque L_2 d'une manière uniforme sur $\mathcal{F}_n$. Ainsi, l'erreur produite par minimisation du risque empirique devient petite. D'autre part, comme $r_n \in \mathcal{F}_n$, il ne peut pas dépasser en performance le meilleur choix dans $\mathcal{F}_n$. Ceci est formulé dans le lemme suivant.

Lemme 1 *Soit r_n un estimateur vérifiant la relation* (1.4)*, alors*

$$\int |r_n(x) - r(x)|^2 \mu(dx)$$

$$\leq 2 \sup_{f \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 - \mathbf{E}\{(f(X) - Y)^2\} \right|$$

$$+ \inf_{f \in \mathcal{F}_n} \int |f(x) - r(x)|^2 \mu(dx).$$

Preuve 1 *Nous avons d'après la formule* (1.3) *que*

$$\begin{aligned} \int_{\mathbb{R}^d} |r_n(x) - r(x)|^2 \mu(dx) &= \mathbf{E}\{|r_n(X) - Y|^2 | D_n\} - \mathbf{E}|r(X) - Y|^2 \\ &= \left(\mathbf{E}\{|r_n(X) - Y|^2 | D_n\} - \inf_{f \in \mathcal{F}_n} \mathbf{E}|f(X) - Y|^2 \right) \\ &\quad + \left(\inf_{f \in \mathcal{F}_n} \mathbf{E}|f(X) - Y|^2 - \mathbf{E}|r(X) - Y|^2 \right). \end{aligned}$$

En vertu de l'équation (1.2), le second terme du membre de droite de l'égalité précédente vaut

$$\inf_{f \in \mathcal{F}_n} \int |f(x) - r(x)|^2 \mu(dx).$$

Il reste à s'occuper du premier terme. Or

$$\begin{aligned} &\mathbf{E}\{|r_n(X) - Y|^2 | D_n\} - \inf_{f \in \mathcal{F}_n} \mathbf{E}|f(X) - Y|^2 \\ &= \sup_{f \in \mathcal{F}_n} \Big\{ \mathbf{E}\{|r_n(X) - Y|^2 | D_n\} - \frac{1}{n} \sum_{i=1}^n |r_n(X_i) - Y_i|^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n |r_n(X_i) - Y_i|^2 - \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 \\ &\quad + \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 - \mathbf{E}|f(X) - Y|^2 \Big\} \end{aligned}.$$

De la définition de r_n, il vient

$$\frac{1}{n} \sum_{i=1}^n |r_n(X_i) - Y_i|^2 - \frac{1}{n} \sum_{i=1}^n |f(X_i) - Y_i|^2 \leq 0.$$

Donc

$$\begin{aligned}
&\mathbf{E}\{|r_n(X)-Y|^2|D_n\} - \inf_{f\in\mathcal{F}_n}\mathbf{E}|f(X)-Y|^2 \\
&\leq \sup_{f\in\mathcal{F}_n}\left\{\mathbf{E}\{|r_n(X)-Y|^2|D_n\} - \tfrac{1}{n}\textstyle\sum_{i=1}^n |r_n(X_i)-Y_i|^2\right. \\
&\qquad \left. + \tfrac{1}{n}\textstyle\sum_{i=1}^n |f(X_i)-Y_i|^2 - \mathbf{E}|f(X)-Y|^2\right\}.
\end{aligned}$$

Il en découle que

$$\begin{aligned}
&\mathbf{E}\{|r_n(X)-Y|^2|D_n\} - \inf_{f\in\mathcal{F}_n}\mathbf{E}|f(X)-Y|^2 \\
&\leq 2\sup_{f\in\mathcal{F}_n}\left|\tfrac{1}{n}\textstyle\sum_{i=1}^n |f(X_i)-Y_i|^2 - \mathbf{E}|f(X)-Y|^2\right|.
\end{aligned}$$

La quantité

$$\mathbf{E}(|r_n(X)-Y|^2|D_n) - \inf_{f\in\mathcal{F}_n}\mathbf{E}(|f(X)-Y|^2), \tag{1.6}$$

est appelée erreur d'estimation et représente la différence entre le risque L_2 de l'estimateur et le meilleur risque L_2 qu'on peut obtenir dans la famille $\mathcal{F}_n$.
Quant à la quantité

$$\inf_{f\in\mathcal{F}_n}\int |f(x)-r(x)|^2\mu(dx), \tag{1.7}$$

elle s'interprète comme étant l'erreur d'approximation qu'on obtient en remplaçant la fonction de régression par un élément de $\mathcal{F}_n$.
Pour obtenir un estimateur consistant, il suffit donc de montrer que les deux erreurs tendent vers zéro.
En général, le choix d'une famille $\mathcal{F}_n$ permettant d'avoir l'erreur d'approximation tendant vers zéro est assez simple, il suffit que $\bigcup_n \mathcal{F}_n$ soit dense dans L_2. Par contre pour l'erreur d'estimation, c'est plus compliqué et on doit avoir recours aux résultats données dans l'appendice.

Pour démontrer la consistance de cet estimateur, il est plus facile d'avoir une fonction bornée d'où le choix de tronquer notre estimateur. C'est-à-dire, pour $\tilde{r_n}$ définie par

$$\tilde{r_n}\in\mathcal{F}_n \text{ et } \frac{1}{n}\sum_{i=1}^n |\tilde{r_n}(X_i)-Y_i|^2 = \min_{f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^n |f(X_i)-Y_i|^2. \tag{1.8}$$

On tronque $\tilde{r_n}$ par r_n, pour $|Y| \leq B_n$ p.s., comme suit

$$r_n(x) = \mathbb{T}_{B_n} \tilde{r_n}(x), \tag{1.9}$$

où $\mathbb{T}_L$ est l'opérateur de troncature défini par

$$\mathbb{T}_L(x) = \begin{cases} x & \text{si } |x| \leq L \\ L\ sign(x) & \text{si non} \end{cases}. \tag{1.10}$$

Nous présentons ci dessous un résultat de consistence forte de l'estimateur des moindres carrés

Théorème 1 *(cf théorème 10.1 dans Györfi et al. [21]) Soient* $\Psi_1, \Psi_2, \ldots : \mathbb{R}^d \to \mathbb{R}$ *des fonctions vérifiant* $|\Psi_j(x)| \leq 1$. *Supposons que l'ensemble des fonctions*

$$\bigcup_{K=1}^{\infty} \left\{ \sum_{j=1}^{K} a_j \Psi_j(x) : a_1, \ldots, a_K \in \mathbb{R} \right\},$$

est dense dans $L_2(\mu)$ *pour toute mesure de probabilité* μ *sur* $\mathbb{R}^d$.

Soit r_n *l'estimateur de la fonction de régression qui minimise le risque empirique* L_2 *donné par*

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

par rapport aux fonctions $f(x) = \sum_{j=1}^{K_n} a_j \Psi_j(x)$ *avec* $\sum_{j=1}^{K_n} |a_j| \leq B_n$.
Si $\mathbf{E}\{Y^2\} < \infty$, $K_n \to \infty$, $B_n \to \infty$, $\frac{K_n B_n^4 \log B_n}{n} \to 0$ *et* $\frac{B_n^4}{n^{1-\delta}} \to 0$, *pour un certain* $\delta > 0$, *alors*

$$\int (r_n(x) - r(x))^2 \mu(dx) \to 0 \ \ p.s.$$

On peut également trouver des taux de convergence de cet estimateur en se reportant au chapitre 12 de [21].

Dans la section qui suit, nous allons présenter une alternative à l'estimateur des moindres carrés qui est l'estimateur des moindres carrés pénalisés.

1.3 Estimation de la fonction de régression par la méthode des moindres carrés pénalisés

1.3.1 Construction de l'estimateur

L'estimateur des moindres carrés se base sur le choix de la classe de fonctions sur laquelle on prend le minimum dans L_2. Ce choix nous évite d'avoir un estimateur qui dépend d'une manière excessive des données, ce qui le rend non efficace dans une optique de prévision. Dans le cas de l'estimateur des moindres carrés pénalisés, au lieu de restreindre la classe de fonctions sur laquelle porte la minimisation, nous rajoutons un terme de "pénalité". Le plus souvent le choix se porte sur un terme de pénalité proportionnel a une intégrale du carré d'une dérivée de la fonction et correspond alors à ce qu'on appelle spline de lissage. Donc l'estimation de la fonction de régression combine la mesure classique de la qualité de l'ajustement, la somme des résidus au carré, et une mesure de la qualité de lissage. Aussi, nous allons présenter deux termes de pénalités, le premier s'applique dans le cas unidimensionnel et le second dans le cas multidimensionnel. Remarquons que cette méthode a l'avantage de conduire à un estimateur lisse (au moins continue), par opposition à la méthode des moindres carrés dont le choix de la classe $\mathcal{F}_n$ ne le permet pas généralement.

Cas unidimensionnel

Nous allons supposer, dans le suite, que la variable X est bornée et même que $X \in]0,1[$ p.s. sans perte de généralité. Dans le cas unidimensionnel le terme de pénalité est donnée par $J_{n,p}(f) \geq 0$ et

$$J_{n,p}(f) = \lambda_n \int_{-\infty}^{+\infty} |f^{(p)}(x)|^2 dx, \tag{1.11}$$

où $f^{(p)}$ dénote la p-ième dérivée de f et $\lambda_n > 0$ est une constante qui dépend de la taille de l échantillon, appelée paramètre de lissage. Elle permet de réaliser un compromis entre l'adéquation aux données exprimé par $\sum_{i=1}^{n}(f(x_i)-y_i)^2$ et les fluctuations locales exprimées par $\int_{-\infty}^{+\infty} |f^{(p)}(x)|^2 dx$. L'estimateur des moindres carrés pénalisés r_n de la fonction de régression est alors donnée par la relation suivante

$$r_n(.) = \arg\, min \left\{ \frac{1}{n} \sum |f(X_i) - Y_i|^2 + J_{n,p}(f) \right\},$$

où le minimum est pris sur l'ensemble des fonctions mesurables.
En traitant ce minimum sur toute les fonctions mesurables de p-ième dérivée carré intégrable, on se retrouve pour $p > 1$, avec des fonctions p fois continuellement différentiables et nous allons noter cette classe de fonctions par $C^p(\mathbb{R})$, donc l'écriture finale de notre estimateur est

$$r_n(.) = \underset{f \in C^p(\mathbb{R})}{\arg\, min} \left\{ \frac{1}{n} \sum |f(X_i) - Y_i|^2 + J_{n,p}(f) \right\}. \tag{1.12}$$

En fait, il est prouvé au chapitre 20 de [21]que la solution de (5.2) est une spline de degré $2p-1$ (pour $p \geq 1$) (c'est-à-dire une fonction polynomiale par morceaux, de degré $2p-1$ au plus et qui est $2p-2$ fois continûment différentiable). Ceci justifie l'appellation de spline de lissage donnée à cet estimateur. En particulier pour $p = 2$, on obtient une spline cubique.

Cas multidimensionnel

La généralisation de la partie précédente à l'espace $\mathbb{R}^d$ se fait de la même manière en cherchant une fonction définie de $\mathbb{R}^d$ dans $\mathbb{R}$ et qui minimise le risque empirique L_2. Dans ce cas le terme de pénalité s'écrit sous la forme suivante

$$\lambda_n \mathbf{J}_p^2(f) = \lambda_n \sum_{\substack{\alpha_1, ..., \alpha_d \in \mathbb{N} \\ \alpha_1 + ... + \alpha_d = p}} \frac{p!}{\alpha_1! ... \alpha_d!} \int_{\mathbb{R}^d} \left| \frac{\partial^p f}{\partial {x_1}^{\alpha_1} ... \partial {x_d}^{\alpha_d}}(x) \right|^2 dx, \tag{1.13}$$

où $\lambda_n > 0$ est un paramètre de l'estimation.
On remarque que ce terme de pénalité dépend de l'existence des dérivées partielles de f. La démonstration de l'existence de la fonction qui minimise le risque empirique L_2 est basée sur des fonctions qui appartiennent à un espace de Hilbert, ce qui ne requière pas l'existence des dérivées au sens classique, mais plutôt dans le sens des dérivées faibles, donc le minimum sera recherché dans l'espace de Sobolev $W^p(\mathbb{R}^d)$ qui contient toutes les fonctions qui admettent des dérivées faibles d'ordre p, ces dernières sont dans $L^2(\mathbb{R})$. Notre estimateur est donné par

$$\underset{f \in \mathcal{W}^p(\mathbb{R}^d)}{\arg\, min} \left(\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_n J_p^2(f) \right). \tag{1.14}$$

Remarque 1 *Dans la suite, nous allons supposer que $2p > d$, cela garantit l'existence du minimum. Une solution de ce problème peut être trouvée par résolution d'un système d'équations linéaires (cf [15]).*

1.3.2 Consistance de L'estimateur

Les outils utilisés pour démontrer la consistance de cet estimation sont les mêmes que ceux relatifs à l'estimation moindres carrés sans terme de pénalité, le changement qui intervient concerne le terme de pénalité.
Pour $2p > d$, $\tilde{r}_n$ est donné par

$$\tilde{r}_n = \underset{f \in \mathcal{W}^p(\mathbb{R}^d)}{\arg\, min} \left(\frac{1}{n} \sum_{i=1}^{n} |f(X_i) - Y_i|^2 + \lambda_n J_p^2(f) \right). \tag{1.15}$$

L'estimateur tronqué est alors défini par

$$r_n(x) = \mathbb{T}_{B_n} \tilde{r_n}(x), \tag{1.16}$$

où $|Y| \leq B_n$ p.s. Le théorème qui suit donne la consistance de l'estimateur spline de lissage dans le cas multidimensionnel.

Théorème 2 *Soit $p \in \mathbb{N}$ avec $2p > d$. Pour $n \in \mathbb{N}$, choisissons $\lambda_n > 0$ tel que*

$$\lambda_n \underset{n\to\infty}{\longrightarrow} 0. \tag{1.17}$$

$$\frac{n\lambda_n^{\frac{d}{2p}}}{(\log n)^7} \underset{n\to\infty}{\longrightarrow} \infty. \tag{1.18}$$

Soit r_n l'estimateur donné par (1.16), alors

$$\int |r_n(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\longrightarrow} 0 \ p.s.$$

Pour toute loi de (X, Y) avec $||X||_2$ bornée p.s. et $\mathbf{E}(Y^2) \leq \infty$.

La démonstration est omise (se reporter au chapitre 20 de Györfi [21]).

Les splines de lissage permettent toutes de contrôler la flexibilité de l'estimation via un paramètre de lissage. Toutefois, cette flexibilité a un

prix et toutes les méthodes non paramétriques doivent composer avec la dualité biais variance. En effet, le fait de suivre plus fidèlement les données augmente la variance et diminue le biais. La relation entre le lissage et la flexibilité de l'estimation est identifiée comme la dualité biais variance. Ainsi en augmentant la flexibilité de l'estimation, il est possible de suivre plus fidèlement les données, ce qui fait diminuer le biais. La courbe obtenue a tendance à plus osciller, ce qui implique que la variance augmente. Quand on diminue la flexibilité de l'estimation, on suit moins fidèlement les données, et donc on augmente le biais. Par conséquent, tout utilisateur d'une méthode de régression non paramétrique doit composer avec cette dualité quand il choisit la valeur du paramètre de lissage. Le biais correspond à la différence entre la vraie fonction de régression et l'estimation. Ce terme décroît quand le paramètre de lissage (λ) diminue. Le terme de variance correspond à la variance de l'estimation, qui augmente quand λ diminue. La sélection du paramètre de lissage est ainsi un problème difficile qui demande à trouver un bon compromis entre le biais et la variance, nous en donnons un aperçu ci dessous.

1.3.3 Choix du paramètre de lissage

Il existe plusieurs méthodes de sélection du paramètre de lissage. Chacune essaye d'estimer la valeur optimale, mais le critère d'optimalité retenu peut différer d'une méthode à l". Il faut choisir un critère mesurant la qualité du modèle grâce à une distance entre les observations prévues et les vraies observations. Le plus souvent, la valeur optimale est définie comme celle minimisant le $MISE$.

$$MISE(\lambda) = \int \mathbf{E}(r_n(x,\lambda) - r(x))^2 dx,$$

ou le $MASE$ donné par,

$$MASE(\lambda) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{E}(r_n(x,\lambda) - r(x))^2.$$

Aucun de ces critères ne peut être évalué directement car la fonction de régression r est inconnue. On cherche le paramètre de lissage qui réalise le meilleur compromis entre biais et variance, c'est-à-dire, entre la fidélité aux données (et à l'extrême l'interpolation) et le lissage. Si on cherche λ qui minimise la somme des carrés des résidus, on choisirait $\lambda = 0$ et on interpolerait les données ce qui n'est forcément pas optimum dans une optique de

prévision, pour de nouvelles données. Une première solution au problème de la sélection du paramètre de lissage consiste à diviser l'échantillon en deux sous-ensembles : en utiliser une partie (l'ensemble d'apprentissage) pour l'estimation des fonctions et la partie restante (l'ensemble de test) pour la sélection du paramètre de lissage. La procédure d'apprentissage-validation consiste donc à séparer de manière aléatoire les données en deux parties distinctes.
Néanmoins cette solution n'est pas réalisable quand le nombre d'observations n'est pas suffisamment élevé. En effet, il faut suffisamment de données pour estimer le modèle, mais il faut aussi beaucoup d'observation dans la validation.
Un critère très souvent utilisé est le critère de validation croisée (cross validation). Dans ce cas, les n ensembles de validation sont constitués d'un seul élément.

$$CV(\lambda) = \frac{1}{n}\sum_{i=1}^{n}(r_n^{-i}(x,\lambda) - y_i)^2,$$

où, $r_n^{-i}(x,\lambda)$ est l'estimateur engendré sans la i-ème donnée, pour chaque i, l'estimation est construite à partir de l'ensemble d'apprentissage $x_{j\neq i}$ et évaluée ensuite en x_i.
Donc, on cherche λ qui minimise le critère de validation croisée. On a

$$\mathbf{E}(CV(\lambda)) = \sigma^2 + \mathbf{E}(\frac{1}{n}\sum_{i=1}^{n}(r_n^{-i}(x,\lambda) - r(x_i))^2,$$

$$\mathbf{E}(CV(\lambda)) = \sigma^2 + MASE^(\lambda).$$

On peut alors dire que $CV(\lambda)$ est minimum en λ là où $MASE(\lambda)$ est minimum.

2 Estimation dans un modèle de censure

Ce chapitre servira de passage entre l'estimation de la fonction de régression par les méthodes des moindres carrés et spline de lissage pour des données complètes aux données censurées. Il présentera un avant propos sur les données de survie et la censure avec ces différents types ainsi que l'estimation de la fonction de répartition aussi bien dans un modèle de censure à droite que de censure mixte (qui est nécessaire à la construction de nos estimateurs aux deux chapitres suivants).

2.1 Données de survie et censure

Une donnée de survie représente le temps écoulé entre le début d'une observation et l'arrivée d'un événement. Historiquement, cette théorie a démarré dans Le cadre biomédical, d'où l'utilisation du terme décès. Cependant, le terme "données de survie" couvre d'autres événements, comme l'apparition d'une maladie ou une épidémie. Dans l'industrie, il peut s'agir de la panne d'une machine. En économie, du temps écoulé pour qu'une personne trouve un travail. Dans plusieurs cas, l'événement est la transition d'un état à un autre. Par exemple, le décès est la transition de l'état "vivant" vers l'état "mort". L'apparition d'une maladie est la transition de l'état "en santé" vers l'état "malade".
Selon le contexte, les termes décès, événement, échec ou transition peuvent être utilisés pour désigner l'événement d'intérêt.
Une caractéristique importante de données de survie est la présence de données censurées. Cette caractéristique, source de difficulté, a nécessité le développement de techniques alternatives à l'inférence usuelle.
Les données censurées sont des observations ne correspondant pas à de vraies valeurs de la variable d'intérêt. Cependant, nous disposons tout de

même d'une information partielle permettant par exemple de fixer une borne inférieure (censure à droite) ou une borne supérieure (censure à gauche).
Les raisons de cette censure peuvent être le fait que le patient soit toujours vivant ou non malade à la fin de l'étude, ou qu'il se soit retiré de l'étude pour des raisons personnelles (immigration, mutation, ...).
Il existe différents types de censures : censure de type I, si le temps de censure est fixé par le chercheur comme étant la fin de l'étude. La censure de type II, se caractérise par le fait que l'étude cesse aussitôt que se produit un nombre d'événements prédéterminés par l'expérimentateur. Une autre possibilité de censure aléatoire est que la censure n'est plus du tout sous le contrôle du chercheur et (ou) que le temps d'entrée varie aléatoirement. Quelques détails et exemples de différents cas de censure suivent.

Censure à droite La durée de vie est dite censurée à droite si l'individu n'a pas subi l'événement à sa dernière observation. En présence de censure à droite, les durées de vie (Y) ne sont pas toutes observées ; pour certaines d'entre elles, on sait seulement qu'elles sont supérieures à une certaine valeur connue. Soit R une variable aléatoire de censure, au lieu d'observer la variable Y qui nous intéresse, on observe le couple de variables (Z, δ) avec $Z = \min(Y, R)$ et $\delta = 1_{\{Y \leq R\}}$. δ est appelé indicateur de censure puisque ses valeurs nous informent sur le fait que l'observation est complète (si $\delta = 1$) ou censurée à droite (si $\delta = 0$).
Un exemple illustratif est lorsqu'on s'intéresse à la durée de vie d'un genre de machines précis mais que ces dernières tombent en panne s'il se produit une surtension d'électricité. Ici, la durée de vie de la machine est censurée à droite par l'instant auquel se produit la surtension.
C'est le type de censure la plus fréquente dans la pratique.

Remarque 2 *Dans beaucoup de cas, l'hypothèse d'indépendance entre Y et R peut être admise dans la pratique. C' est le cas pour l'exemple précédent ou dans le cas de malades qui sortent de l'étude suite à un déménagement.*

Censure à gauche La censure à gauche correspond au cas où l'individu a déjà subi l'événement avant qu'il ne soit observé. On sait uniquement que la valeur d'intérêt est inférieure à une certaine valeur connue représentée par une variable aléatoire L. Pour chaque individu, on peut associer un couple de variables aléatoires (Z, δ) telles que $Z = \max(Y, L)$ et $\delta = \mathbf{1}_{\{Y \geq L\}}$. Un

des premiers exemples de censure à gauche rencontré dans la littérature concerne le cas d'observateurs qui s'intéressent à l'horaire où les babouins descendent de leurs arbres pour aller manger (les babouins passent la nuit dans les arbres). Le temps d'événement (descente de l'arbre) est observé si le babouin descend de l'arbre après l'arrivée des observateurs. Par contre, la donnée est censurée si le babouin est descendu avant l'arrivée des observateurs : dans ce cas on sait uniquement que l'horaire de descente est inférieur à l'heure d'arrivée des observateurs. On observe donc le maximum entre l'heure de descente des babouins et l'heure d'arrivée des observateurs.

Remarque 3 *Très peu de travaux s'intéressent à la seule censure à gauche car beaucoup moins fréquente et constitue un phénomène symétrique à celui de la censure à droite. Certains auteurs ont proposé de renverser l'échelle de temps.*

Dans un même échantillon peuvent être présentes des données censurées à droite et d'autres censurées à gauche, comme c'est le cas dans ce qui suit.

Censure double Par exemple, une étude s'est intéressée à l'âge auquel les enfants d'une communauté africaine apprennent à accomplir certaines tâches. Au début de l'étude, certains enfants savaient déjà effectuer les tâches étudiées, on sait seulement alors que l'âge où ils ont appris est inférieur à leur âge à la date du début de l'étude. A la fin de l'étude, certains enfants ne savaient pas encore accomplir ces tâches et on sait alors seulement que l'âge auquel ils apprendront éventuellement ont appris est supérieur à leur âge à la fin de l'étude. L'âge au début de l'étude (variable de censure à gauche L) est évidemment inférieure à l'âge à la fin de l'étude (variable de censure à droite R). L'âge d'intérêt est observé ssi il se trouve dans la période d'étude. Nous observons $Z = \max(\min(Y, R), L)$ avec un indicateur de censure. Ce modèle a été étudié dans Turnbul [46] qui a introduit un estimateur implicite de la fonction de survie de Y donné comme solution d'une équation de self-consistance.

Censure par intervalle Une observation est censurée par intervalle si au lieu d'observer avec certitude le temps de l'événement, la seule information disponible est qu'il se trouve dans un intervalle. Par exemple, dans le cas d'un suivi médical du diabète, les personnes sont souvent suivies par intermittence (pas en continu), on sait alors uniquement que l'événement (pic de

la glycémie) s'est produit entre deux temps d'observations (deux analyses). Ce modèle généralise ceux de la seule censure à droite (ou à gauche).
Dans certains cas, il n'y a pas de raison de supposer que la variable de censure à gauche est inférieure à celle de censure à droite. De plus, dans le cas de données censurées on ne peut pas toujours déterminer un intervalle auquel appartient la valeur d'intérêt. Ceci est le cas dans le modèle de Patiléa et Rolin [37] qui suit.

Censure mixte Nous disons qu'il y a censure mixte lorsque deux phénomènes de censure (l'un à gauche et l'autre à droite) peuvent empêcher l'observation du phénomène d'intérêt sans qu'on puisse nécessairement déterminer un intervalle auquel il appartient. Dans le modèle I décrit dans l'article de [Patilea et Rolin (2006)], au lieu d'observer un échantillon de la variable d'intérêt Y, on observe un échantillon du couple $(Z; A)$ avec $Z = \max(\min(Y, R), L)$ et

$$A = \begin{cases} 0 & \text{si} \quad L < Y \leq R \\ 1 & \text{si} \quad L < R < Y \\ 2 & \text{si} \quad \min(Y, R) \leq L \end{cases} .$$

où L et R sont des variables de censure et A est l'indicateur de censure. Un exemple de ce modèle est donné par un système formé par trois composants, dont deux sont placés en série (le composant dont le temps de fonctionnement nous intéresse et un autre). Un troisième est placé en parallèle avec ce système en série. Ici, il est clair qu'il n'est pas raisonnable de supposer que le temps de fonctionnement d'un composant soit inférieur à un autre.

Remarque 4 *Il ne faut pas confondre censure et troncature. La troncature diffère complètement de la censure. Une donnée tronquée ne figure pas du tout dans l'échantillon. Si une maison de retraite n'accepte que des personnes âgées d'au moins soixante ans, aucun individu décédé avant cet âge n'a la possibilité d'y avoir été admis et est de ce fait tronqué. Le traitement mathématique est alors complètement différent de celui de la censure.*

L'analyse de survie a connu un développement important dans la seconde moitié du vingtième siècle après que Kaplan et Meier [25] aient introduit leur célèbre estimateur de la fonction de survie pour des données censurées à droite. Estimateur qui généralise le complément à un de la fonc-

tion de répartition empirique et que nous rappelons ci dessous.

2.2 Estimation de la fonction de survie

Le développement de l'analyse de la survie a d'abord, et de manière prédominante, porté sur l'estimation de la fonction de survie $S(z)$, qui représente la probabilité qu'un individu ait une durée de vie X supérieure à un temps z, et s'exprime par

$$S(z) = \mathbf{P}(X > z),$$

où X est une variable aléatoire non négative.
Il est clair que S est une fonction décroissante, continue à droite avec $\lim_{z\to 0} S(z) = 1$ et $\lim_{z\to+\infty} S(z) = 0$.
Si nous disposons d'un échantillon de données complètes de X, S est classiquement estimée par le complément à 1 de la fonction de répartition empirique, qui n'est plus calculable si les données sont censurées. Nous donnons ci dessous, l'expression d'estimateurs non paramétriques de S dans le cas de deux modèles de censure qui nous intéressent particulièrement dans ce travail, puisque nous en ferons usage.

2.2.1 Estimateur de Kaplan-Meier

Kaplan et Meier [25] ont proposé un estimateur très efficace de la fonction de survie quand les observations sont censurées à droite, nommé estimateur produit limite vu son expression.
Soit $X_1, \ldots, X_n$ un échantillon de v.a. indépendantes représentant les durées d'intérêt), de fonction de répartition F, et $C_1, \ldots, C_n$ un échantillon représentant les temps de censure, que l'on suppose indépendants des durées d'intérêt, de fonction de répartition G. Dans le modèle de censure aléatoire à droite, on observe non pas la durée d'intérêt X_i mais plutôt la plus petite des deux valeurs $Z_i = \min(X_i, C_i)$, ainsi que l'indicateur de censure δ_i qui vaut 1 si la durée d'intérêt est observée, et 0 si elle est censurée, i.e. $\delta_i = 1_{\{X_i \le C_i\}}$.

Dans ce cas, la fonction de répartition F est estimée par l'estimateur introduit par [25] , donné pour $z < Z_{(n)}$ où $Z_{(n)} = \max\{Z_1, \ldots, Z_n\}$ par

$$F_n(z) = 1 - \prod_{i:Z_i \le z} \left(\frac{N_n(Z_i) - 1}{N_n(Z_i)} \right)^{\delta_i}$$

avec $N_n(x) = \sum_{i=1}^{n} 1_{\{Z_i \geq x\}}$. Pour $z \geq Z_{(n)}$, il y a plusieurs conventions pour définir $F_n(z)$: Soit on le définit par $F_n(Z_{(n)})$, ce qui a pour conséquence non souhaitable que F_n peut ne pas être une fonction de répartition (c'est le cas si $Z_{(n)}$ est une donnée censurée), soit on le définit par 0 pour y remédier, soit on le laisse non défini.

Cet estimateur a des propriétés assez similaires à celles la fonction de répartition empirique comme la convergence uniforme presque sûre [45, 53] et la normalité asymptotique [4, 18]. Dans la suite, nous allons nous intéresser à l'extension de l'estimateur de Kaplan-Meier au cas de la censure mixte puisque ce nouveau estimateur interviendra dans l'expression de nos estimateurs ultérieurs.

2.2.2 Estimateur de la fonction de survie dans un modèle de censure mixte

Considérons trois variables aléatoires positives indépendantes X, L et R de fonctions de répartition respectives F_X, F_L et F_R, et de fonctions de survie respectives S_X, S_L et S_R, où X représente la durée d'intérêt et L et R sont les durées de censure à gauche et à droite respectivement. Dans le modèle I de Patiléa et Rolin [37], rappelons qu' au lieu d'observer un échantillon de X nous diposons seulement d' un échantillon du couple (Z, A) où $Z = \max(\min(X, R), L)$ et

$$A = \begin{cases} 0 & \text{si } L < X \leq R \\ 1 & \text{si } L < R < X \\ 2 & \text{si } \min(X, R) \leq L \end{cases}$$

Soit H la fonction de répartition de Z, elle peut s'écrire $\sum_{k=0}^{2} H^{(k)}(t)$ où

$$H^{(k)}(t) = \mathbf{P}(Z \leq t, A = k), \quad \text{pour } k = 0, 1, 2.$$

En notant pour toute application R de $\mathbb{R}$ dans $\mathbb{R}$, $R^-(t)$ la limite de R à gauche de t, lorsque cette limite existe, ces fonctions s'écrivent

$$H^{(0)}(t) = \int_0^t F_L^-(u) S_R^-(u)\, dF_X(u),$$
$$H^{(1)}(t) = \int_0^t F_L^-(u) S_X(u)\, dF_R(u),$$
$$H^{(2)}(t) = \int_0^t \{1 - S_X(u) S_R(u)\}\, dF_L(u),$$

et c'est à partir de ces équations que l'estimateur est obtenu.

L'idée est de considérer dans un premier temps $Y = \min(X, R)$ et L dans un modèle de censure à gauche (c'est-à-dire que l'on considère une donnée complète si $A = 0$ ou $A = 1$ et censurée à gauche si $A = 2$), et d'estimer la fonction de répartition de Y, puis l'utiliser pour estimer la fonction de répartition de la variable d'intérêt X en considérant un modèle de censure à droite.

L'estimateur de la fonction de survie S_X ainsi obtenu, en remplaçant à la fin les fonctions $H^{(0)}$, $H^{(1)}$ et $H^{(2)}$ par leurs estimateurs empiriques, obtenus à partir d'un échantillon $(Z_i, A_i)_{1\leq i\leq n}$, est donné par

$$\hat{S}_n(Z'_j) = 1 - F_n(Z'_j) = \prod_{1\leq l\leq j} \left\{1 - \frac{D_{0l}}{U_{l-1} - N_{l-1}}\right\},$$

où $(Z'_j)_{1\leq j\leq M}$ sont les valeurs distinctes des Z_i prises dans l'ordre croissant, et

$$D_{kj} = \sum_{1\leq i\leq n} 1_{\{Z_i = Z'_j, A_i = k\}}, \quad N_j = \sum_{1\leq i\leq n} 1_{\{Z_i \leq Z'_j\}},$$

$$U_{j-1} = n \prod_{j\leq l\leq M} \left\{1 - \frac{D_{2l}}{N_l}\right\}$$

pour $0 \leq l \leq 2$ et $1 \leq j \leq M$.

Faison remarquer que si $L \equiv 0$ (pas de censure à gauche), $\hat{S}_n$ se réduit à l'estimateur de Kaplan-Meier qui lui-même se réduit au complément à 1 de la fonction de répartition empirique si $R \equiv \infty$ (pas de censure).

Patiléa et Rolin [37] ont introduit cet estimateur et montré sa convergence uniforme presque sûre et sa convergence en tant que processus vers un gaussien sous des conditions d'identifiabilité du modèle.
Pour nos besoins ultérieurs ($\hat{S}_n$, va figurer au dénominateur dans l'expression de nos estimateurs de la fonction de régression aux chapitres suivants), donnons une condition nécessaire et suffisante pour que $\hat{S}_n$ s'annule.
Dans un souci de clarté, nous notons dans la suite de ce chapitre D_{kj} par $D_{k,j}$.

Lemme 2 *i) Une condition nécessaire et suffisante pour que* $\hat{S}_n(Z'_{k_0}) = 0$ *pour la première fois et reste nul est :* $D_{0,k_0} \neq 0$, $D_{1,k_0} = 0$ *et* $\forall j > k_0, D_{0,j} = D_{1,j} = 0$ *si* $k_0 \neq M$.

ii) $\hat{S}_n$ *s'annule pour la première fois en* Z'_M *si et seulement si* $D_{0,M} \neq 0$ *et* $D_{1,M} = 0$

Preuve 2 *1. Commençons par montrer que pour tout* $k : 0 \leq k \leq M-1$, *nous avons*

$$U_k \geq N_k. \tag{2.1}$$

En effet

$$\begin{aligned} U_k &= n\left(\frac{N_{k+1} - D_{2,(k+1)}}{N_{k+1}}\right)\left(\frac{N_{k+2} - D_{2,(k+2)}}{N_{k+2}}\right)\cdots\left(\frac{N_M - D_{2,M}}{N_M}\right) \\ &= \left(\frac{N_k + D_{0,(k+1)} + D_{1,(k+1)}}{N_{k+1}}\right)\cdots\left(\frac{N_{M-2} + D_{0,(M-1)} + D_{1,(M-1)}}{N_{M-1}}\right) \\ &\quad \times (N_{M-1} + D_{0,M} + D_{1,M}) \\ &\geq \frac{N_k}{N_{k+1}} \times \cdots \times \frac{N_{M-2}}{N_{M-1}} \times N_{M-1} = N_k. \end{aligned}$$

Remarquons que s'il existe j *tel que* $j > k$ *avec* $D_{1,j} \neq 0$ *ou* $D_{0,j} \neq 0$ *alors* $U_k > N_k$.

2. Soit k_0 *le premier indice* k *tel que* $D_{0,k} = U_{k-1} - N_{k-1}$ *(le premier* k *pour lequel* $\hat{S}_n(Z'_k) = 0$*). Nous avons alors :*

$$D_{0,k_0} \neq 0 \text{ et } D_{0,k_0} = U_{k_0-1} - N_{k_0-1}. \tag{2.2}$$

Par ailleurs

$$U_{k_0-1} = n\left(1 - \frac{D_{2,k_0}}{N_{k_0}}\right)\cdots\left(1 - \frac{D_{2,M}}{N_M}\right) = \left(1 - \frac{D_{2,k_0}}{N_{k_0}}\right)U_{k_0}. \tag{2.3}$$

D'après (2.2) *et* (2.3), *il vient*

$$\begin{aligned} D_{0,k_0} + N_{k_0-1} &= \left(\frac{N_{k_0} - D_{2,k_0}}{N_{k_0}}\right)U_{k_0} \\ &= \left(\frac{N_{k_0-1} + D_{0,k_0} + D_{1,k_0}}{N_{k_0}}\right)U_{k_0}, \end{aligned}$$

et en vertu de (2.1)*, nous devons avoir* $D_{1,k_0} = 0$*, donc* $U_{k_0=N_{k_0}}$ *alors* $D_{0,k_0} \neq 0$*,* $D_{1,k_0} = 0$ *et* $\forall j > k_0, D_{1,j} = D_{0,j} = 0$*, (au-delà de* k_0*, ce qui montre que la condition énoncée est nécessaire Montrons que la condition nécessaire est aussi suffisante. Supposons que* $D_{1,k_0} = 0$*,* $D_{0,k_0} \neq 0$ *et* $\forall j > k_0, D_{1,j} = D_{0,j} = 0$*, et montrons que* $\hat{S}_n(Z'_{k_0}) = 0$*. Nous avons*

$$\begin{aligned} U_{k_0-1} &= n\left(\frac{N_{k_0} - D_{2,k_0}}{N_{k_0}}\right)\left(\frac{N_{k_0+1} - D_{2,(k_0+1)}}{N_{k_0+1}}\right)\cdots\left(\frac{N_M - D_{2,M}}{N_M}\right) \\ &= n\left(\frac{N_{k_0-1} + D_{0,k_0}}{N_{k_0}}\right)\left(\frac{N_{k_0}}{N_{k_0+1}}\right)\cdots\left(\frac{N_{M-1}}{N_M}\right) \\ &= N_{k_0-1} + D_{0,k_0}, \end{aligned}$$

avec $D_{0,k_0} \neq 0$*, autrement dit* $\hat{S}_n(Z'_{k_0}) = 0$*.*

Passons maintenant à l'estimation de la fonction de régression lorsque la variable réponse est censurée à droite.

2.3 Estimation de la fonction de régression en présence de la censure à droite

Le but de l'analyse de régression est d'estimer la moyenne conditionnelle de la variable réponse réelle Y sachant une variable explicative X à valeurs dans $\mathbb{R}^d$, autrement dit $r(x) = E(Y|X = x)$. Lorsque l'observation du couple (X, Y) est possible, l'estimation est basée sur un échantillon de variables de même loi que ce couple. Cependant, dans quelques applications un tel échantillon n'est pas disponible. Lorsque la variable réponse Y est censurée à droite, on ne peut qu'observer l'échantillon composé par $(X, \min(Y, R), \delta)$, où R est une variable de censure et δ est l'indicateur de censure qui désigne laquelle des variables Y ou R est réellement observée. Et comme il a été introduit dans le chapitre précédent, il est connu que la fonction de régression r atteint le minimum de l'erreur moyenne quadratique, et cela d'après la relation $\mathbf{E}|f(X)-Y|^2 = \mathbf{E}|r(X)-Y|^2+\int |f(x)-r(x)|^2\mu(dx)$, qui est vérifiée pour toute fonction mesurable f. On s'intéresse à la convergence dans L_2 de l'estimateur r_n, mesurée par $\int |r_n(x) - r(x)|^2\mu(dx)$ où μ

représente la mesure de probabilité de X, basé sur l'échantillon $(X_i, Z_i = \min(Y_i, R_i); \delta_i = 1_{\{Y_i \leq R_i\}})_{1 \leq i \leq n}$ de n couples de variables aléatoires i.i.d. et de même loi que $(X, Z = \min(Y, R); \delta)$. Cette partie reprend des résultats de l'article Kohler et al. [31], dans lequel ils introduisent , parmi d'autres, des estimateurs des moindres carrés et spline de lissage, auxquels nous nous intéressons plus particulièrement dans cette thèse.
Dans toute la suite de ce document, pour toute variable aléatoire réelle V, de fonction de répartition notée F_V, notons $T_V := T_{F_V} = \sup\{t : F_V(t) < 1\}$ et $I_V := I_{F_V} = \inf\{t : F_V(t) \neq 0\}$; les points terminaux du support de V.
Appelons H_D l'hypothèse englobant les conditions suivantes

- R et (X, Y) sont indépendants.
- T_Y est finie.
- S_R est continue.
- $S_R(T_Y) > 0$.

La première hypothèse est très utile d'un point de vue mathématique et est très souvent utilisée. Il est donc important de voir si elle se justifie en application. Par exemple, dans le cas d'une censure causée par la fin de l'étude, cette hypothèse est naturelle. La seconde condition concerne le support de notre variable d'intérêt et est généralement vérifié dans la pratique. La dernière laisse la possibilité d'observation de toutes les valeurs de notre variable d'intérêt malgré la censure.

Commençons par la méthode des moindres carrés puis passons à la méthode spline de lissage.

2.3.1 La méthode des moindres carrés

L'idée de construction de cet estimateur a été introduite dans le travail de Carbonez et al.[5], qui se sont intéressés à l'estimateur à partitions . Et en s'inspirant de cette idée, Kohler et Mathé [31] ont repris et étendu la travail à d'autres types d'estimateurs (à noyau, des plus roches voisins, des moindres carrés et spline de lissage).
Soit h une fonction de $\mathbb{R}^d \times \mathbb{R} \longrightarrow \mathbb{R}$. Ils proposèrent comme "estimateur" de $\mathbf{E}[h(X, Y)]$, la quantité

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i h(X_i, Z_i)}{S_R(Z_i)},$$

qui, sous l' hypothèse H_1, est un "estimateur" sans biais. En effet

$$\begin{aligned}\mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\delta_i h(X_i,Z_i)}{S_R(Z_i)}\right] &= \mathbf{E}\left[\frac{\delta_1 h(X_1,Z_1)}{S_R(Z_1)}\right]\\ &= \mathbf{E}\left[E\left(\frac{\delta_1 h(X_1,Z_1)}{S_R(Z_1)}|X_1,Y_1\right)\right]\\ &= \mathbf{E}\left[\frac{h(X_1,Y_1)}{S_R(Y_1)}\mathbf{E}\left(\delta_1|X_1,Y_1\right)\right]\\ &= \mathbf{E}\left[\frac{h(X_1,Y_1)}{S_R(Y_1)}\mathbf{E}\left(1_{\{Y_1\leq R_1\}}|X_1,Y_1\right)\right].\end{aligned}$$

Soit A un élément de la tribu engendrée par (X_1, Y_1), alors $A = (X_1, Y_1)^{-1}(B)$, où B est un borélien de $\mathbb{R}^{d+1}$, et nous avons

$$\int_A 1_{\{Y_1\leq R_1\}} d\mathbf{P} = \int_{B\times\mathbb{R}} 1_{\{y\leq u\}} d\mathbf{P}_{(X_1,Y_1,R_1)}(x,y,u).$$

De l'indépendance du couple (X, Y) de R, nous obtenons

$$\begin{aligned}\int_{B\times\mathbb{R}} 1_{\{y\leq u\}} d\mathbf{P}_{(X_1,Y_1,R_1)}(x,y,u) &= \int_B \left(\int_{\{y\leq u\}} d\mathbf{P}_{R_1}(u)\right) d\mathbf{P}_{(X_1,Y_1)}(x,y)\\ &= \int_B S_{R_1}(y) d\mathbf{P}_{(X_1,Y_1)}(x,y)\\ &= \int_A S_{R_1}(Y_1) d\mathbf{P}.\end{aligned}$$

Le problème qui se pose est que la fonction de survie S_R, est dans la plupart des cas inconnue. On doit donc la remplacer par un estimateur $\hat{S}_n$, en l'occurrence celui de Kaplan- Meier de la fonction de survie de R. Ce qui conduit à

$$\tilde{r}_{n,1}(x) = \arg\min_{f\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n} \frac{\delta_i |f(X_i) - Z_i|^2}{\hat{S}_n(Z_i)}. \tag{2.4}$$

Pour $0 \leq t < \infty$ et $x \in \mathbb{R}$, définissons

$$T_{[0,t]}(x) = \begin{cases} t & \text{si } x > t \\ x & \text{si } 0 \leq x \leq t \\ 0 & \text{si } x < 0, \end{cases} \tag{2.5}$$

et pour $f : \mathbb{R}^d \to \mathbb{R}$, définisons $(T_{[0,t]}f)(x) := T_{[0,t]}(f(x))$.
Puisque Y est bornée, il est naturel de borner aussi $\tilde{r}_{n,1}$, ce qui donne en définitive l'estimateur

$$r_{n,1}(x) = T_{[0,M_n]}(\tilde{r}_{n,1}(x)), \tag{2.6}$$

où $M_n := \max\{Z_1, \ldots, Z_n\}$.

Ici $\hat{S}_n$ s'écrit

$$\hat{S}_n(z) = \begin{cases} \prod\limits_{j=1,\ldots,n} \left(1 - \frac{1-\delta_j}{n-j+1}\right)^{1_{\{Z_{(i)} \leq z\}}} & \text{si} \quad z \leq M_n \\ \lim\limits_{s \to M_n, s < M_n} \hat{S}_n(s) & \text{si} \quad z \geq M_n \end{cases},$$

La deuxième relation(pour $z \geq M_n$) est donnée pour éviter que $\hat{S}_n$ ne s'annule du fait qu'il apparaît au dénominateur dans la relation (2.4).

Le théorème de type Glivenko-Cantelli pour le cas censuré à droite, obtenu par Stute et Wang [45]), implique que puisque S_R est continue

$$\sup_{z \leq T_Y} \left|\hat{S}_n(z) - S_R(z)\right| \underset{n \longrightarrow \infty}{\longrightarrow} 0 \text{ p.s.} \tag{2.7}$$

En combinant cette dernière formule à des résultas de la théorie de Vapnik-Chervonenkis, Kohler et al. [31] ont établi que l'estimateur défini par les relations (2.4) et (2.6) est universellement convergent, résultat que nous présentons ci dessous.
Soit $\mathcal{F}_n$ une famille de fonctions $f : \mathbb{R}^d \to \mathbb{R}$, notons

$$\mathcal{B}_n^* \mathcal{F}_n = \{f \in \mathcal{F}_n : 0 \leq f(x) \leq T_Y (x \in \mathbb{R}^d)\},$$

et

$$\mathcal{F}_n^+ = \{\{(x, y) \in \mathbb{R}^d \times \mathbb{R} : f(x) \geq y\}, f \in \mathcal{F}_n\}$$

l'ensemble des graphes des fonctions dans $\mathcal{F}_n$ et dont $\mathcal{V}_{\mathcal{F}_n^+}$ dénote la dimension V.C. (voir l'appendice pour un rappel sur ces notions).
Nous sommes maintenant en mesure d'énoncer le résultat suivant

Théorème 3 *Kohler et al. [31]*

Soit $\mathcal{F}_n$ une famille de fonctions $f : \mathbb{R}^d \to \mathbb{R}$ satisfaisant

$$\frac{\mathcal{V}_{\mathcal{F}_n^+}}{n} \underset{n \to \infty}{\to} 0, \tag{2.8}$$

où

$$\inf_{f\in\mathcal{B}_n^*\mathcal{F}_n}\sup_{x\in[-A,A]^d}|f(x)-g(x)|\underset{n\to\infty}{\to}0, \tag{2.9}$$

pour tout $A\in\mathbb{R}^+$ *et pour toute fonction continue* $g:\mathbb{R}^d\to\mathbb{R}$, *s'annulant en dehors de* $[-A,A]^d$ *et bornée par* T_Y.

Sous H_D, *on a*

$$\int_{\mathbb{R}^d}|r_{n,1}(x)-r(x)|^2\,\mu(dx)\underset{n\to\infty}{\to}0\ p.s.$$

2.3.2 Spline de lissage

De la même manière, Kohler et al. [31] définissent l'estimateur spline de lissage de la fonction de régression en posant d'abord

$$r_{n,2}(x)=\underset{f\in\mathcal{W}^p(\mathbb{R}^d)}{\arg\,min}\left(\frac{1}{n}\sum_{i=1}^{n}\delta_i\frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)}+\lambda_n\mathbf{J}_p^2(f)\right),$$

où le terme de pénalité $\lambda_n\mathbf{J}_p^2(f)$ est donné par la relation (1.13) au chapitre précédent.
Finalement, nous posons

$$r_{n,2}(x)=T_{[0,M_n]}(\tilde{r}_{n,2}(x)),$$

Le théorème suivant concerne la convergence de cet estimateur.

Théorème 4 *Kohler et al. [31]*

Soit $p\in\mathbb{N}$ *avec* $2p>d$, *et pour* $n\in\mathbb{N}$, *soit* $\lambda_n>0$ *tel que*

$$\lambda_n\underset{n\to\infty}{\longrightarrow}0 \tag{2.10}$$

et

$$n\lambda_n\underset{n\to\infty}{\to}\infty. \tag{2.11}$$

Sous l' hypothèse H_1 *et si* $|X|$ *est presque sûrement bornée, alors*

$$\int_{\mathbb{R}^d} |r_{n,2}(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\to} 0 \quad p.s. \tag{2.12}$$

L'estimation de la fonction de régression par les méthodes des moindres carrés et spline de lissage pour des données censurées à droite représente la plateforme pour l'estimation dans le cas de censure mixte, objet de notre contribution dans cette thèse et qui sera présentée aux chapitres suivants.

Cette partie a fait l'objet d'une publication dans la revue : Statistics and Probability letters.

3 Estimation de la fonction de régression par la méthode des moindres carrés dans un modèle de censure mixte

Notre but est d'estimer la fonction de régression $r(x) = \mathbf{E}(Y|X = x)$ dans le cas où la variable réponse est censurée selon le modèle I de Patilea et Rolin [37]. Rappelons que cela veut dire qu' on ne peut qu'observer un échantillon tiré de $(X, Z = \max(\min(Y, R), L), A)$, où R et L sont des variables de censure et A est l'indicateur de censure qui désigne laquelle des variables latentes Y, R où L est réellement observée. Dans ce contexte Messaci [33] a introduit des estimateurs à poids (à noyau, à partitions et des plus proches voisins) de la fonction de régression $r(x)$ convergeant vers la valeur optimale presque sûrement.
Dans l'esprit de continuité de ce travail, nous proposons un estimateur par la méthode des moindres carrés. Sous l'hypothèse de continuité de la répartition de L sur $]0, +\infty[$, que le couple (L, R) est indépendant de (X, Y) et enfin en utilisant quelques conditions concernant les supports des variables Y, R et L, nous démontrons la "consistence forte" de notre estimateur.
Cette investigation étend le travail de Kohler et al. [31] concernant l'estimateur par la méthode des moindres carrés de la fonction de régression proposé pour Y censurée seulement à droite.

3.1 Principe de l'estimation et hypothèses

Notons par X un vecteur aléatoire de $\mathbb{R}^d$ et soit Y une variable aléatoire bornée et positive. Soient R et L deux variables aléatoires positives de censure. Rappelons que pour toute variable aléatoire V, nous notons F_V (resp. S_V) sa fonction de répartition (resp. sa fonction de survie) et

$$T_V = \sup\{t : F_V(t) < 1\} \text{ et } I_V = \inf\{t : F_V(t) \neq 0\}$$

désignent les points terminaux du support de la variable V.

Nous nous proposons d'estimer $r(x) = \mathbf{E}(Y|X = x)$ à partir de l'échantillon formé des observations i.i.d $\mathcal{D}_n = \{X_i,\ Z_i = \max(\min(Y_i, R_i), L_i),\ A_i\}$ de même loi que (X, Z, A) où $Z = \max(\min(Y, R), L)$ et

$$A = \begin{cases} 0 & \text{si} \quad L < Y < R \\ 1 & \text{si} \quad L < R \leq Y \\ 2 & \text{si} \quad \min(Y, R) \leq L \end{cases}.$$

Supposons que les variables Y, R et L vérifient les hypothèses suivantes

H_1 : Y, R et L sont indépendantes.

H_2 : (L, R) est indépendant de (X, Y).

H_3 : $\exists T < T_R$ et $I > I_L$ tel que,

$$\forall n \in \mathbb{N}, \forall i (1 \leq i \leq n) : A_i = 0 \Rightarrow I \leq Z_i \leq T \text{p.s.},$$

H_4 : F_L est continue sur $]0, \infty[$,

H_5 : $T_R \leq T_Y \leq T_L < \infty$ et $I_Y \leq I_L < I_R$.

H_1 est une hypothèse inhérente au modèle de Patilea. Une hypothèse du même type que H_2 est aussi considérée dans le cas de la seule censure à droite. L'hypothèse H_3 nous semble acceptable du fait que $I_L < Z_i < T_R$ lorsque $A_i = 0$. l'hypothèse H_5, quant à elle, assure en particulier, que le modèle est identifiable.
Soit h une application de $\mathbb{R}^d \times \mathbb{R} \to \mathbb{R}$. En reprenant l'idée introduite dans Messaci [33], nous proposons comme "estimateur" sans biais de $\mathbf{E}\{h(X, Y)\}$

la quantité

$$\frac{1}{n}\sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{h(X_i, Z_i)}{S_R(Z_i)F_L(Z_i)}. \tag{3.1}$$

En effet, sous les hypothèses H_1, H_2 et H_4 et pour tout B dans $\sigma(X, Y)$ (tribu engendrée par le couple (X, Y)) il existe un borélien C tel que $B = (X, Y)^{-1}(C)$. L'indépendance de (X, Y) et (L, R) permet d'ecrire

$$\begin{aligned}
\int_B (1_{A=0})\, dP &= \int_B (1_{L<Y<R})\, dP \\
&= \int_{C\times\mathbb{R}_+^2} (1_{l<y<r})\, dP_{(X,Y,L,R)} \\
&= \int_{C\times\mathbb{R}_+^2} (1_{l<y<r})\, dP_{(X,Y)} \otimes dP_{(L,R)}.
\end{aligned}$$

Maintenant par le théorème de Fubini et l'indépendance de R et L, nous obtenons

$$\begin{aligned}
\int_B (1_{A=0})\, dP &= \int_C \left(\int_{\mathbb{R}_+^2} (1_{l<y<r})\, dP_{(L,R)} \right) dP_{(X,Y)} \\
&= \int_C \left(\int_{\mathbb{R}+} (1_{l<y})\, dP_L \times \int_{\mathbb{R}+} (1_{y<r})\, dP_R \right) dP_{(X,Y)} \\
&= \int_C (F_L(y)S_R(y))\, dP_{(X,Y)} \\
&= \int_B F_L((Y_1)S_R(Y_1) dP,
\end{aligned}$$

car F est continue. De plus, $F_L(Y)S_R(Y)$ étant clairement mesurable par rapport à $\sigma(X, Y)$, le résultat suivant en découle

$$E\left(1_{\{A=0\}} \middle| X, Y\right) = S_R(Y)F_L(Y). \tag{3.2}$$

Nous pouvons donc en déduire que

$$\begin{aligned}
E\left(\frac{1_{\{A=0\}}h(X, Z)}{S_R(Z)F_L(Z)}\right) &= E\left(E\left(\frac{1_{\{A=0\}}h(X, Y)}{S_R(Z)F_L(Z)}\right)\middle|(X, Y)\right) \\
&= E\left(\frac{h(X, Y)}{S_R(Y)F_L(Y)} E\left(1_{\{A=0\}}\middle|(X, Y)\right)\right) \\
&= E\left(h(X, Y)\right).
\end{aligned}$$

Autrement dit

$$E\left(1_{\{A=0\}}\frac{h(X,Z)}{S_R(Z)F_L(Z)}\right) = E\left(h(X,Y)\right). \tag{3.3}$$

Le problème qui se pose dans la formule (3.1) est que les fonctions S_R et F_L sont en général inconnues, nous allons les remplacer respectivement par leurs estimateurs $\hat{S}_n$ et $\hat{F}_n$, dont nous rappelons les expressions à la section suivante.

3.1.1 Estimation de S_R et F_L

Soit $Z_j^{'}(1 \leq j \leq M)$ les valeurs distinctes de Z_i rangées dans l'ordre croissant. Pour $k \in \{0,1,2\}$, posons

$$D_{kj} = \sum_{i=1}^{n} 1_{\{Z_i = Z_j^{'}, A_i = k\}}, \quad N_j = \sum_{i=1}^{n} 1_{\{Z_i \leq Z_j^{'}\}},$$

et

$$U_{j-1} = n \prod_{j \leq l \leq M} \left\{1 - \frac{D_{2l}}{N_l}\right\}.$$

Alors, Patilea et Rolin [37] proposent d'estimer S_R par

$$\hat{S}_n(Z_j^{'}) = \prod_{1 \leq l \leq j} \left\{1 - \frac{D_{1l}}{U_{l-1} - N_{l-1}}\right\}. \tag{3.4}$$

Et en inversant le temps dans l'estimateur de Kaplan-Meir, nous pouvons déduire l'estimateur $\hat{F}_n$ de F_L (cas de la censure à gauche) donné par

$$\hat{F}_n(Z_j^{'}) = \prod_{j < l \leq M} \left\{1 - \frac{1_{\{A_j = 2\}}}{l}\right\}. \tag{3.5}$$

Rappelons qu'en vertu des hypothèses H_1 et H_5, il a été démontré dans Patilea et Rolin [37] que

$$\sup_{t \in \mathbb{R}^+} \left|\hat{S}_n(t) - S_R(t)\right| \underset{n \longrightarrow \infty}{\longrightarrow} 0 \text{ p.s.} \tag{3.6}$$

$$\sup_{t \in \mathbb{R}^+} \left|\hat{F}_n(t) - F_L(t)\right| \underset{n \longrightarrow \infty}{\longrightarrow} 0 \text{ p.s.} \tag{3.7}$$

Remarquons que l'hypothèse H_3 implique que

$$S_R(T) > 0 \text{ et } F_L(I) > 0. \tag{3.8}$$

Et en vertu des équations (3.6)–(3.8), nous déduisons que pour n assez grand

$$\hat{S}_n(T) > 0 \text{ et } \hat{F}_n(I) > 0 \text{ p.s.} \tag{3.9}$$

3.1.2 Construction de l'estimateur

Rappelons que lorsque Y est complètement observée, l'estimateur de la fonction de régression par la méthode des moindres carrés, obtenu par minimisation du risque empirique L_2, est donnée par

$$\arg\min_{f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^{n}|f(X_i)-Y_i|^2,$$

où $\mathcal{F}_n$ est une classe de fonctions qui dépend de la taille de l'échantillon n. Ainsi, dans notre contexte, d'après la relation(3.3) et après estimation de S_R et F_L, l'estimateur par la méthode des moindres carrés de $r(x) = \mathbf{E}(Y|X=x)$ est, en premier temps, donné par

$$\widetilde{r}_n = \arg\min_{f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}\left(\frac{0}{0}:=0\right), \tag{3.10}$$

$\mathcal{F}_n$ est une certaine famille de fonctions qui sera clarifiée au théorème 5.
En vertu du lemme 2 dans le chapitre 2, nous voyons que $\hat{S}_n(Z_i)$ ne s'annule pas dans l'expression de $\widetilde{r}_n$ si $A_i = 0$. Il est aussi facile de vérifier que $\hat{F}_n(Z_i)$ ne s'annule pas non plus si $A_i = 0$.

Mais du fait que Y est bornée, nous allons imposer à notre estimateur de l'être aussi, à cette fin réintroduisons la notation de l'application de troncature suivante

Pour $0 \le t < \infty$ et $x \in \mathbb{R}$, définissons

$$T_{[0,t]}(x) = \begin{cases} t & \text{si } x > t \\ x & \text{si } 0 \le x \le t \\ 0 & \text{si } x < 0, \end{cases} \tag{3.11}$$

et pour $f : \mathbb{R}^d \to \mathbb{R}$, définisons $(T_{[0,t]}f)(x) := T_{[0,t]}(f(x))$.
Nous pouvons aussi réutiliser le fait que cette application vérifie la relation suivante.

$$\forall b > a, \quad |\mathbb{T}_{[0,b]}(x) - \mathbb{T}_{[0,a]}(x)| \le (b-a). \tag{3.12}$$

Y étant bornée et du fait que pour $M_n = \max_{1\leq i\leq n} Z_i$ avec $M_n \underset{n\to\infty}{\to} T_L$ p.s., nous proposons finalement comme estimateur de $r(x)$

$$r_n(x) = \mathbb{T}_{[0,M_n]}(\tilde{r}_n(x)),$$

3.2 Résultat

Nous introduisons le même type d'ensembles utilisés dans le modèle de censure à droite, que nous adaptons à notre cas comme suit

$$\mathcal{B}_n^*\mathcal{F}_n = \{f \in \mathcal{F}_n : 0 \leq f(x) \leq T_L(x \in \mathbb{R}^d)\}.$$

$$\mathcal{F}_n^*\mathcal{F}_n = \{g : \mathbb{R}^d \to \mathbb{R}^+ / \exists f \in \mathcal{F}_n, \forall x \in \mathbb{R}^d : g(x) = \mathbb{T}_{[0,T_L]}(f(x))\}.$$

Le résultat suivant concerne la convergence de l'estimateur r_n proposé. Le lecteur peut se référer à l'appendice pour quelques définitions et résultats, sur la théorie de Vapnik-Chervonenkis, utilisés dans ce travail.

Théorème 5 *Sous les hypothèses H_1–H_5, on a*

$$\int_{\mathbb{R}^d} |r_n(x) - r(x)|^2 \, \mu(dx) \underset{n\to\infty}{\to} 0 \ p.s.$$

Pour un choix de la famille $\mathcal{F}_n$ de fonctions $f : \mathbb{R}^d \to \mathbb{R}$ satisfaisant

$$\frac{\mathcal{V}_{\mathcal{F}_n^+}}{n} \underset{n\to\infty}{\to} 0, \tag{3.13}$$

où $\mathcal{V}_{\mathcal{F}_n^+}$ dénote la dimension V.C de l'ensemble des graphes des fonctions dans $\mathcal{F}_n$ et

$$\inf_{f\in\mathcal{B}_n^*\mathcal{F}_n} \sup_{x\in[-A,A]^d} |f(x) - g(x)| \underset{n\to\infty}{\to} 0, \tag{3.14}$$

pour tout $A \in \mathbb{R}^+$ et pour toute fonction continue $g : \mathbb{R}^d \to \mathbb{R}$, qui s'annule en dehors de $[-A, A]^d$ et qui est bornée par T_L.

Un exemple de classe de fonctions vérifiant les conditions de ce théorème, que nous reprenons de Kohler et al. [31], est le suivant

Exemple 1 *Soient* $M \in \mathbb{N}^*$, $A_n \in \mathbb{R}$ *et* $K_n \in \mathbb{N}$ *tels que*

$$\frac{K_n^d}{n} \underset{n\longrightarrow\infty}{\longrightarrow} 0, \tag{3.15}$$

$$A_n \underset{n\longrightarrow\infty}{\longrightarrow} \infty \text{ et } \frac{A_n}{K_n} \underset{n\longrightarrow\infty}{\longrightarrow} 0. \tag{3.16}$$

Soit $\pi_n = \{B_{n,1}, B_{n,2}, \ldots, B_{n,K_n^d}\}$ *la partition de* $[-A_n, A_n]^d$ *en cubes de côté* K_n^d. *Soit* $\mathcal{F}_n$ *l'ensemble des fonctions polynomiales par morceaux de degré* M *(ou moins, en chaque coordonnée) par rapport à* π_n.
$\mathcal{F}_n$ *est un espace vectoriel de dimension* $K_n^d.(M+1)^d$. *L'application du lemme 8 de l'appendice permet de déduire que* $V_{\mathcal{F}_n^+} \leq K_n^d.(M+1)^d$. *La condition (3.13) du théorème est vérifiée par l'hypothèse (3.15). Pour montrer que la condition (3.14)du théorème est vérifier, considérons une fonction continue* g *vérifiant* $0 \leq g(x) \leq T_L$ *avec* $x \in \mathbb{R}^d$ *et* $g(x) = 0$ *pour tous* $x \in \mathbb{R}^d \setminus [-A_n, A_n]^d$. *Pour* $x_i \in B_{n,i}$ $(i = 1, \ldots, K_n^d)$ *définissons* $f_g \in \mathcal{F}_n$ *par*

$$f_g(x) = \sum_i g(x_i) 1_{B_{n,i}}(x), \ (x \in \mathbb{R}^d).$$

Pour des points arbitraires $u_1, u_2 \in B_{n,i}$ $(i = 1, \ldots, K_n^d)$, *nous avons* $||u_1 - u_2|| \leq \sqrt{d}\frac{2A_n}{K_n}$. *De plus, pour un* x *donné, il existe un cube* $B_{n,i}$ *vérifiant*

$f_g(x) = g(x_i)$. *Il en résulte que*

$$\begin{aligned}
& \inf_{\substack{f\in\mathcal{F}_n \\ ,0\le f(x)\le T_L, x\in\mathbb{R}^d}} \sup_{x\in[-A_n,A_n]^d} |f(x)-g(x)| \\
\le\ & \sup_{x\in[-A_n,A_n]^d} |f_g(x)-g(x)| \\
\le\ & \max_{i=1,\ldots,K_n^d} \sup_{u_1,u_2\in B_{n,i}} |g(u_1)-g(u_2)| \\
\le\ & \sup_{||u_1-u_2||\le\sqrt{d}.(2A_n/K_n)} |g(u_1)-g(u_2)| \\
\longrightarrow\ & 0,
\end{aligned}$$

et cela par l'hypothèse (3.16) et par la continuité uniforme de g.

Introduisons la quantité $\bar{r}_n(x) = \mathbb{T}_{[0,T_L]}(\tilde{r}_n(x))$. En premier lieu, le lemme suivant permet de voir que le théorème est prouvé si et seulement si

$$\int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\to} 0 \text{ p.s.}$$

Lemme 3 *Sous les hypothèses* H_1*–*H_5*, nous avons*

$$\int_{\mathbb{R}^d} |r_n(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\to} 0 \ p.s \ \Leftrightarrow \int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\to} 0 \ p.s.$$

Preuve 3 *En effet, d'après la relation (4.7), on a*

$$|r_n(x) - \bar{r}_n(x)| \le (T_L - M_n),$$

ce qui implique que

$$\int_{\mathbb{R}^d} |r_n(x) - \bar{r}_n(x)|^2 \le (T_L - M_n)^2 \to 0 \ p.s.$$

puisque $\lim_{n\to+\infty} M_n = T_L$ *p.s. en vertu de* H_5.

Ce résultat nous permet de baser la démonstration de notre théorème sur l'inégalité suivante

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) \qquad (3.17)\\
\leq\ & 2 \sup_{f \in \mathcal{F}_n^* \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|f(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X) - Y|^2 \right| \\
+\ & \inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \int_{\mathbb{R}^d} |f(x) - r(x)|^2 \mu(dx),
\end{aligned}$$

que nous commençons par montrer.

– D'un côté, nous avons

$$\begin{aligned}
&\int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) \\
=\ & \left\{ \mathbf{E}\left(|\bar{r}_n(X) - Y|^2 \,|\mathcal{D}_n \right) - \inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \mathbf{E}\,|f(X) - Y|^2 \right\} \\
+\ & \left\{ \inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \mathbf{E}\,|f(X) - Y|^2 - \mathbf{E}\,|r(X) - Y|^2 \right\}.
\end{aligned}$$

– De plus, la fonction de régression vérifie

$$\inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \mathbf{E}\,|f(X) - Y|^2 - \mathbf{E}\,|r(X) - Y|^2 = \inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \int_{\mathbb{R}^d} |f(x) - r(x)|^2 \mu(dx).$$

– D'un autre coté

$$\begin{aligned}
&\mathbf{E}\left(|\bar{r}_n(X) - Y|^2 \,|\mathcal{D}_n \right) - \inf_{f \in \mathcal{B}_n^* \mathcal{F}_n} \mathbf{E}\,|f(X) - Y|^2 \\
=\ & \sup_{f \in \mathcal{B}_n^* \mathcal{F}_n} \left\{ \mathbf{E}\left(|\bar{r}_n(X) - Y|^2 \,|\mathcal{D}_n \right) - \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \right. \\
&+ \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \\
&+ \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|f(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \\
&\left. + \frac{1}{n} \sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|f(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X) - Y|^2 \right\} \leq \sum_{i=1}^{4} Q_{n,i},
\end{aligned}$$

où les $Q_{n,i}$ sont explicités ci dessous.

– Du fait que $\tilde{r} \in \mathcal{F}_n^*\mathcal{F}_n$,$\bar{r}_n \in \mathcal{F}_n^*\mathcal{F}_n$ et $\mathcal{B}_n^*\mathcal{F}_n \subset \mathcal{F}_n^*\mathcal{F}_n$, il est clair que

$$\begin{aligned} Q_{n,1} = & \sup_{f\in\mathcal{B}_n^*\mathcal{F}_n} \left\{ \mathbf{E}\left(|\bar{r}_n(X)-Y|^2 |\mathcal{D}_n\right) - \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|\bar{r}_n(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \right\} \\ \leq & \sup_{f\in\mathcal{F}_n^*\mathcal{F}_n} \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X)-Y|^2 \right|, \end{aligned}$$

et

$$\begin{aligned} Q_{n,4} = & \sup_{f\in\mathcal{B}_n^*\mathcal{F}_n} \left\{ \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X)-Y|^2 \right| \right\} \\ \leq & \sup_{f\in\mathcal{F}_n^*\mathcal{F}_n} \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X)-Y|^2 \right|. \end{aligned}$$

– Puisque $\bar{r}_n(X_i) \leq T_L$ et $Z_i \leq T_L$ p.s., nous obtenons

$$1_{\{A_i=0\}}|\tilde{r}_n(X_i)-Z_i| \geq 1_{\{A_i=0\}}|\bar{r}_n(X_i)-Z_i|,$$

ce qui implique que

$$Q_{n,2} = \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|\bar{r}_n(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|\tilde{r}_n(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \leq 0$$

– En vertu de la définition de $\tilde{r}_n$ (voir(3.10)), il est évident que

$$Q_{n,3} = \sup_{f\in\mathcal{B}_n^*\mathcal{F}_n} \left\{ \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|\tilde{r}_n(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \right\} \leq 0.$$

L'inégalité (3.17) est donc démontrée. Il reste à prouver que les deux termes du second membre de l'équation tendent vers zéro presque sûrement quand $n \to +\infty$, pour cela nous allons procéder en deux étapes.

– Dans la première étape, nous montrons que

$$\lim_{n\to\infty} \sup_{f\in\mathcal{F}_n^*\mathcal{F}_n} \left| \frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}} \frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \mathbf{E}\,|f(X)-Y|^2 \right| = 0 \text{ p.s.} \tag{3.18}$$

A cette fin, utilisons les inégalités suivantes

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}-\mathbf{E}\,|f(X)-Y|^2\right|\\
\leq&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}-\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)\hat{F}_n(Z_i)}\right|\\
+&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)\hat{F}_n(Z_i)}-\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}\right|\\
+&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}-E\,|f(X)-Y|^2\right|.
\end{aligned}$$

- Comme $f\in\mathcal{F}_n^*\mathcal{F}_n$ implique que $0\leq f(x)\leq T_L$, nous obtenons en vertu des relations (3.6)- - (3.8)

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}-\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)\hat{F}_n(Z_i)}\right|\\
\leq&\frac{T_L^2}{\hat{S}_n(T)S_R(T)\hat{F}_n(I)}\sup_{t\in\mathbb{R}^+}\left|\hat{S}_n(t)-S_R(t)\right|\underset{n\to\infty}{\to}0,\ \text{p.s.}
\end{aligned}\tag{3.19}$$

et on a

$$\begin{aligned}
&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)\hat{F}_n(Z_i)}-\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}\right|\\
\leq&\frac{T_L^2}{F_L(I)S_R(T)\hat{F}_n(I)}\sup_{t\in\mathbb{R}^+}\left|\hat{F}_n(t)-F_L(t)\right|\underset{n\to\infty}{\to}0\ \text{p.s.}
\end{aligned}\tag{3.20}$$

- Introduisons les notations suivantes

$$V=(X,Z,1_A),V_1=(X_1,Z_1,1_{A_1}),\ldots,V_n=(X_n,Z_n,1_{A_n}),$$

n vecteurs aléatoires i.i.d de même répartition que V.
Posons

$$\begin{aligned}
\mathcal{H}_n=\big\{&h:\mathbb{R}^d\times[0,T_L]\times\{0,1\}\to\mathbb{R}^+:\exists f\in\mathcal{F}_n^*\mathcal{F}_n\text{ tel que}\\
&h(x,z,1_A)=\frac{1_A\,|f(x)-z|^2}{S_R(z)F_L(z)}\\
&\text{pour tout }\ (x,z,1_A)\in\mathbb{R}^d\times[0,T_L]\times\{0,1\}\big\}.
\end{aligned}$$

Les fonctions de $\mathcal{H}_n$ sont positives et bornées par $\frac{T_L^2}{S_R(T)F_L(I)}$, et

$$\begin{aligned}\mathbf{E}h(V) =&\mathbf{E}\left(\frac{1_A\left|f(X)-Z\right|^2}{S_R(Z)F_L(Z)}\right)\\ =&\mathbf{E}\left[\mathbf{E}\left(\frac{1_A\left|f(X)-Z\right|^2}{S_R(Z)F_L(Z)}\mid X,Y\right)\right]\\ =&\mathbf{E}\left(\left|f(X)-Z\right|^2\right),\end{aligned}$$

sous H_1, H_2 et H_4. De plus

$$\begin{aligned}&\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}}\frac{\left|f(X_i)-Z_i\right|^2}{S_R(Z_i)F_L(Z_i)}-\mathbf{E}\left|f(X)-Y\right|^2\right|\\ =&\sup_{f\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^n h(V)-\mathbf{E}h(V)\right|.\end{aligned}$$

Pour tout $h_1,h_2\in\mathcal{H}_n$, soient f_1,f_2 leurs fonctions correspondantes dans $\mathcal{F}_n^*\mathcal{F}_n$, alors

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^n\left|h_1(V_i)-h_2\left(V_i\right)\right|\\ &\frac{1}{n}\sum_{i=1}^n\left|1_{\{A_i=0\}}\frac{\left|f_1(X_i)-Z_i\right|^2}{S_R(Z_i)F_L(Z_i)}-1_{\{A_i=0\}}\frac{\left|f_2(X_i)-Z_i\right|^2}{S_R(Z_i)F_L(Z_i)}\right|\\ \leq\ &\frac{1}{S_R(T)F_L(I)}\frac{1}{n}\sum_{i=1}^n\left|\left(f_1(X_i)+f_2(X_i)-2Z_i\right)\left(f_1(X_i)-f_2(X_i)\right)\right|\\ \leq\ &\frac{2T_L}{S_R(T)F_L(I)}\frac{1}{n}\sum_{i=1}^n\left|f_1(X_i)-f_2(X_i)\right|,\end{aligned}$$

ce qui implique que

$$\mathcal{N}\left(\varepsilon,\mathcal{H}_n,V_1^n\right)\leq\mathcal{N}\left(\varepsilon\frac{S_R(T)F_L(I)}{2T_L},\mathcal{F}_n^*\mathcal{F}_n,X_1^n\right),$$

où $\mathcal{N}\left(\varepsilon,\mathcal{F}_n,Z_1^n\right)$ dénote le nombre recouvrant (se référer à l'appendice).
Par application du théorème 7 donné à l'appendice, nous obte-

nons pour tout $\delta > 0$

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{f\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^{n}h(V_i)-\mathbf{E}h(V)\right|>\delta\right\}\\
\leq\ & 8\mathbf{E}\left\{\mathcal{N}\left(\delta\frac{S_R(T)F_L(I)}{16T_L},\mathcal{F}_n^*\mathcal{F}_n,X_1^n\right)\right\}\exp\left(-\frac{n\delta^2S_R^2(T)F_L^2(I)}{128T_L^4}\right),
\end{aligned}$$

qui est, d'après le lemme 7 de l'appendice et le fait que $\log x \leq x$ est borné pour un δ assez petit par

$$16\left(-\frac{64eT_L}{\delta\left(S_R(T)F_L(I)\right)^4}\right)^{2\mathcal{V}_{\mathcal{F}_n^*\mathcal{F}_n^+}}\exp\left(-\frac{n\delta^2S_R^2(T)F_L^2(I)}{128T_L^4}\right).$$

La relation $\mathcal{V}_{\mathcal{F}_n^*\mathcal{F}_n^+}\leq\mathcal{V}_{\mathcal{F}_n^+}$ combinée avec la condition (3.13) du théorème, permet d'appliquer le lemme de Borel-Cantelli, pour arriver à

$$\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}-\mathbf{E}\left|f(X)-Y\right|^2\right|\underset{n\to\infty}{\to}0\text{ p.s.}$$

Cette dernière avec les équations (3.19) et (3.20) implique la formule (3.18).

– Dans la deuxième étape, nous montrons que

$$\inf_{f\in\mathcal{B}_n^*\mathcal{F}_n}\int_{\mathbb{R}^d}\left|f(x)-r(x)\right|^2\mu(dx)\underset{n\to\infty}{\to}0\text{ p.s.}\tag{3.21}$$

Soit $\mathrm{C}^0(\mathbb{R}^d)$ l'ensemble de toutes les fonctions continues à support compact. Comme $\mathrm{C}^0(\mathbb{R}^d)$ est dense dans L^2, pour tout $\varepsilon > 0$ il existe une fonction h de $\mathrm{C}^0(\mathbb{R}^d)$ vérifiant

$$\int_{\mathbb{R}^d}\left|h(x)-r(x)\right|^2\mu(dx)\leq\varepsilon\text{ p.s.}\tag{3.22}$$

Choisissons $A > 0$ tel que $h(x) = 0,$ si $x \notin [-A, +A]^d$ et

$$\mu\left([-A,+A]^d\right)^c\leq\frac{\varepsilon}{T_L^2}.\tag{3.23}$$

Définissons $\bar{h}(x) = T_{[0,T_L]}(h(x))$, et comme $\sup(\bar{h}(x), r(x)) \leq T_L$ pour tout $x \in \mathbb{R}^d$, nous obtenons de (3.22)

$$\int_{\mathbb{R}^d}\left|\bar{h}(x)-r(x)\right|^2\mu(dx)\leq\int_{\mathbb{R}^d}\left|h(x)-r(x)\right|^2\mu(dx)\leq\varepsilon\tag{3.24}$$

– Maintenant et en tenant compte des relations (3.23) et (3.24), nous déduisons que

$$\inf_{f\in\mathcal{B}_n^*\mathcal{F}_n}\int_{\mathbb{R}^d}|f(x)-r(x)|^2\,\mu(dx)\leq 2\inf_{f\in\mathcal{B}_n^*\mathcal{F}_n}\sup_{x\in[-A,+A]^d}|f(x)-\bar{h}(x)|+3\varepsilon,$$

qui prouve (3.21) grâce à la relation(3.14) du théorème 5 (Remarquons que $\bar{h}\in \mathrm{C}^0(\mathbb{R}^d)$).

Ainsi s'achève la démonstration du théorème de convergence de notre estimateur, résultat qui sera illustré par une étude de simulation plus loin.

Cette partie à fait l'objet d'un article soumis.

4 Estimation spline de lissage de la fonction de régression dans un modèle de censure mixte

Afin de rendre ce chapitre indépendant du précédent, nous rappelons toutes les notions nécessaires. Nous nous intéressons à l'estimation de la fonction de régression $r(x)$ qui, comme nous l'avons déjà mentionné, représente la moyenne conditionnelle de la variable d'intérêt Y sachant une valeur de la variable explicative X, pour Y ne pouvant être complètement observée. Plus précisément, nous ne disposons que d'un échantillon de $(X, Z = \max(\min(Y, R), L), A)$ où R et L sont des variables de censure et A est une variable qui nous indique laquelle des variables Y, R ou L est réellement observée. C'est toujours le modèle 1 de censure étudié dans Patilea et Rolin (cf [37]), dans le cadre duquel Messaci (cf [33]) a proposé des estimateurs à poids de $r(x)$ alors que Kebabi et al. (cf [26]) ont défini des estimations des moindres carrés que nous avons présentés dans le chapitre précédent. A présent, notre intérêt se porte sur l'introduction de l'estimateur spline de lissage et la preuve de sa consistence forte. Ces travaux, effectués dans un cadre de censure mixte, étendent les résultats obtenus par Kohler et al. (cf [31]) pour Y censurée seulement à droite.

4.1 Notations et hypothèses

La covariable X est un vecteur aléatoire de $\mathbb{R}^d$, la variable réponse Y et les variables de censure R et L sont supposées bornées et positives. Rap-

pelons aussi que pour toute v.a. V, nous notons par F_V sa fonction de répartition, S_V sa fonction de survie ($S_V = 1-F_V$), $I_V = inf\{t : F_V(t) \neq 0\}$ et $T_V = sup\{t : F_V(t) < 1\}$ sont les points terminaux du support de V.

Notre but est d'estimer $r(x) = \mathbf{E}[Y|X = x]$ à partir de l'échantillon formé des observations i.i.d. $\mathcal{D}_n = \{X_i, Z_i, A_i \ ; \ \ 1 \leq i \leq n\}$ de même loi que $(X, Z = \max(\min(Y, R), L), A)$ où

$$A = \begin{cases} 0 & \text{si} \quad L < Y < R \\ 1 & \text{si} \quad L < R \leq Y \\ 2 & \text{si} \quad \min(Y, R) \leq L \end{cases} .$$

Nous allons travailler dans des conditions similaires au cas des moindres carrés, c'est à dire sous l'hypothèse englobant les conditions suivantes

H_1 : Y, R et L sont indépendantes.

H_2 : (L, R) est indépendant de (X, Y).

H_3 : $\exists T < T_R$ et $I > I_L$ tel que,

$$\forall n \in \mathbb{N}, \forall i (1 \leq i \leq n) : A_i = 0 \Rightarrow I \leq Z_i \leq T \text{p.s.},$$

H_4 : F_L est continue sur $]0, \infty[$,

H_5' : $T_R \leq T_Y$ et $I_Y \leq I_L < I_R$.

Soulignons le fait que toutes ces hypothèses ont été imposées pour l'étude d'estimateurs à poids dans Messaci (cf [33]) et que $H_1 - H_4$ ont été déja utilisées au chapitre précédent alors que H_5' est un peu moins restrictive que H_5 du chapitre précédent.

4.2 Construction de l'estimateur

Quand la variable aléatoire réelle Y est complètement observée, l'estimateur spline de lissage est obtenu en ajoutant un terme de pénalité au

risque empirique L_2 ; donné par

$$\underset{f\in\mathcal{W}^p(\mathbb{R}^d)}{\arg\, min}\left(\frac{1}{n}\sum_{i=1}^{n}|f(X_i)-Y_i|^2+\lambda_n J_p^2(f)\right), \tag{4.1}$$

où $p\in\mathbb{N}$, $d<2p$ et

• $\mathcal{W}^p(\mathbb{R}^d)$ est l'espace de Sobolev de toutes les fonctions de $\mathbb{R}^d$ dans $\mathbb{R}$ dont les dérivées d'ordre total p, sont dans $L_2(\mathbb{R}^d)$,

•
$$\mathbf{J}_p^2(f)=\sum_{\substack{\alpha_1,...,\alpha_d\in\mathbb{N}\\ \alpha_1+...+\alpha_d=p}}\frac{p!}{\alpha_1!...\alpha_d!}\int_{\mathbb{R}^d}\left|\frac{\partial^p f}{\partial {x_1}^{\alpha_1}...\partial {x_d}^{\alpha_d}}(x)\right|^2dx,$$

• $\lambda_n>0$ est un paramètre de l'estimation.

Soit h une application de $\mathbb{R}^d\times\mathbb{R}$ dans $\mathbb{R}$ et du fait que sous les hypothèses H_1, H_2 et H_4

$$\mathbf{E}\left(\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{h(X_i,Z_i)}{S_R(Z_i)F_L(Z_i)}\right)=\mathbf{E}h(X,Y),$$

nous estimons $\mathbf{E}h(X,Y)$ par

$$\frac{1}{n}\sum_{i=1}^{n}1_{\{A_i=0\}}\frac{h(X_i,Z_i)}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}, \tag{4.2}$$

où $\hat{S}_n$ et $\hat{F}_n$ sont des estimateurs fortement consistants de S_R et F_L respectivement, leurs expressions sont données ci dessous.
Soit $(Z'_j)_{1\le j\le M}$ $(M\le n)$ les valeurs distinctes de Z_i rangées dans l'ordre croissant.
Posons
$D_{kj}=\sum_{1\le i\le n}1_{\{Z_i=Z'_j,A_i=k\}}$ $(k\in\{0,1,2\})$ et $N_j=\sum_{1\le i\le n}1_{\{Z_i\le Z'_j\}}$.
Patilea et Rolin (cf [37]) ont proposé d'estimer S_R par

$$\hat{S}_n(t)=\prod_{j/Z'_j\le t}\left\{1-\frac{D_{1j}}{U_{j-1}-N_{j-1}}\right\}\quad\text{où}\quad U_{j-1}=n\prod_{j\le l\le M_n}\left\{1-\frac{D_{2l}}{N_l}\right\}.$$

$\hat{F}_n$ est l'estimateur de F_L (cas de la censure à gauche) obtenu de l'estimateur de Kaplan-Meier en inversant le temps, il est donc donné par

$$\hat{F}_n(t)=\prod_{j/Z'_j>t}\left\{1-\frac{1_{\{A_j=2\}}}{j}\right\}.$$

Sous les hypothèses H_1– H_5', Patilea et Rolin [37] ont montré que

$$\lim_{n\to\infty} \sup_{t\in\mathbb{R}^+} |\hat{S}_n(t) - S_R(t)| = 0. \quad \text{p.s.} \tag{4.3}$$

D'autre part, l'hypothèse H_5' implique que

$$\lim_{n\to\infty} \sup_{I_L<t} |\hat{F}_n(t) - F_L(t)| = 0 \quad \text{p.s.} \tag{4.4}$$

Remarquons que l'hypothèse H_3 entraîne

$$S_R(T) > 0 \quad \text{et} \quad F_L(I) > 0. \tag{4.5}$$

Des relations (4.3), (4.4) et (4.5), nous déduisons que pour n assez grand

$$\hat{S}_n(T) > 0 \quad \text{et} \quad \hat{F}_n(I) > 0 \quad \text{p.s.} \tag{4.6}$$

En tenant compte des relations (4.1) et (4.2), nous proposons d'estimer en premier lieu $r(x)$ par

$$\tilde{r}_n(x) = \underset{f\in\mathcal{W}^p(\mathbb{R}^d)}{\arg\,min} \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|f(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} + \lambda_n \mathbf{J}_p^2(f) \right).$$

Nous ferons encore usage de la variable $\mathbb{T}_{[0,t]}(x)$ donnée par l'équation (3.11). Remarquons que

$$\forall b > a, \left| \mathbb{T}_{[0,b]}(x) - \mathbb{T}_{[0,a]}(x) \right| \leq (b-a). \tag{4.7}$$

Nous proposons finalement d'estimer $r(x)$ par

$$r_n(x) = \mathbb{T}_{[0,M_n]}(\tilde{r}_n(x)),$$

où $M_n := \max\{Z_1, \ldots, Z_n\}$.

4.3 Résultat et preuve

En adaptant la preuve du théorème 3 de Kohler et al. (cf [31]) à notre contexte, nous pouvons montrer le résultat suivant.

Théorème 6 *Soit $p \in \mathbb{N}$ avec $2p > d$, et pour $n \in \mathbb{N}$, soit $\lambda_n > 0$ tel que*

$$\lambda_n \underset{n\to\infty}{\longrightarrow} 0, \tag{4.8}$$

et

$$n\lambda_n \underset{n\to\infty}{\to} \infty. \tag{4.9}$$

Sous les hypothèses $H_1 - H_5'$ et si $|X|$ est presque sûrement bornée, alors

$$\int_{\mathbb{R}^d} |r_n(x) - r(x)|^2 \mu(dx) \underset{n\to\infty}{\to} 0 \quad p.s. \tag{4.10}$$

Introduisons la quantité $\bar{r}_n(x) = \mathbb{T}_{[0,T_R\vee T_L]}(\tilde{r}_n(x))$, qui joue un rôle central dans la démonstration du fait que (4.10) équivaut à

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) = 0 \quad \text{p.s.} \tag{4.11}$$

En effet, il suffit de voir que

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |r_n(x) - \bar{r}_n(x)|^2 \mu(dx) = 0 \quad p.s.$$

Par application de la formule (3.12),

$$|r_n(x) - \bar{r}_n(x)| \leq (T_R \vee T_L) - M_n,$$

ce qui montre que

$$\int_{\mathbb{R}^d} |r_n(x) - \bar{r}_n(x)|^2 \mu(dx) \leq [(T_R \vee T_L) - M_n]^2 \to 0 \text{ p.s.}$$

Grâce à l'hypothèse H_5' de laquelle nous déduisons que $\lim_{n\to\infty} M_n = T_R \vee T_L$ p.s.
Afin de prouver (4.11), nous avons besoin de donner le lemme suivant.

Lemme 4 *Sous les hypothèses imposées au théorème 6, nous avons*

$$\mathbf{E}\left(|\bar{r}_n(X) - Y|^2 | \mathcal{D}_n\right) - \frac{1}{n}\sum_{i=1}^{n} 1_{\{A_i=0\}} \frac{|\bar{r}_n(X_i) - Z_i|^2}{S_R(Z_i)F_L(Z_i)} \underset{n\to\infty}{\longrightarrow} 0 \quad p.\ s.$$

si $X \in [0,1]^d$ p.s.

Preuve 4 *du lemme*

Puisque $0 \in \mathcal{W}^p(\mathbb{R}^d)$ *et par définition de* $\tilde{r}_n$*, il vient*

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1_{\{A_i=0\}}|\tilde{r}_n(X_i)-Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}+\lambda_n \boldsymbol{J}_p^2(\tilde{r}_n)$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\frac{1_{\{A_i=0\}}Z_i^2}{S_R(Z_i)F_L(Z_i)}+\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1_{\{A_i=0\}}Z_i^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}-\frac{1_{\{A_i=0\}}Z_i^2}{\hat{S}_n(Z_i)F_L(Z_i)}\right|$$

$$+\frac{1}{n}\sum_{i=1}^{n}\left|\frac{1_{\{A_i=0\}}Z_i^2}{\hat{S}_n(Z_i)F_L(Z_i)}-\frac{1_{\{A_i=0\}}Z_i^2}{S_R(Z_i)F_L(Z_i)}\right|$$

$$\leq \frac{1}{n}\sum_{i=1}^{n}\frac{1_{\{A_i=0\}}Z_i^2}{S_R(Z_i)F_L(Z_i)}+\frac{T^2}{\hat{F}_n(I)\hat{S}_n(T)F_L(I)}\sup_{t>I_L}|F_L(t)-\hat{F}_n(t)|$$

$$+\frac{T^2}{F_L(I)\hat{S}_n(T)S_R(T)}\sup_{t\in\mathbb{R}^+}|S_R(t)-\hat{S}_n(t)| \underset{n\to\infty}{\to} \mathbf{E}\left\{\frac{1_A Z^2}{F_L(Z)S_R(Z)}\right\} \quad p.s. \quad ,$$

par application de (4.3), (4.4), (4.5), (4.6) et de la loi forte des grands nombres. Nous obtenons donc, pour n suffisamment grand, sous la condition (4.9) du théorème

$$\boldsymbol{J}_p^2(\tilde{r}_n) \leq \frac{2}{\lambda_n}\mathbf{E}\left\{\frac{1_A Z^2}{F_L(Z)S_R(Z)}\right\} = n\frac{2\mathbf{E}\{\frac{1_A Z^2}{F_L(Z)S_R(Z)}\}}{n\lambda_n} \leq n,$$

ce qui entraîne que

$\bar{r}_n \in \mathcal{F}_n^*\mathcal{F}_n := \{g : \mathbb{R}^d \to \mathbb{R}^+ / \exists f \in \mathcal{F}_n, \forall x \in \mathbb{R}^d : g(x) = \mathbb{T}_{T_R \vee T_L}(f(x))\}$,

avec $\mathcal{F}_n = \left\{f : f \in \mathcal{W}^p(\mathbb{R}^d) \quad / \quad \boldsymbol{J}_p^2(f) \leq n\right\}$.

Par ailleurs, introduisons les notations suivantes : $V = (X, Z, 1_A)$, $V_i = (X_i, Z_i, 1_{A_i})_{1\leq i\leq n}$ *sont n vecteurs i.i.d. de même loi que* V*, et*

$$\mathcal{H}_n = \Bigg\{h : \mathbb{R}^d \times [0, T_L \vee T_R] \times \{0,1\} \to \mathbb{R}^+ : \exists f \in \mathcal{F}_n^*\mathcal{F}_n \text{ telle que}$$
$$\forall (x, z, 1_A) \in \mathbb{R}^d \times [0, T_L \vee T_R] \times \{0,1\},\ h(x, z, 1_A) = \frac{1_A|f(x)-z|^2}{S_R(z)F_L(z)}\Bigg\}.$$

Si $h \in \mathcal{H}_n$, *alors* $|h| \leq \frac{(T_L \vee T_R)^2}{S_R(T)F_L(I)}$ *et en vertu de* H_1

$$\begin{aligned}\mathbf{E}h(V) &= \mathbf{E}\left(\frac{1_A|f(X)-Z|^2}{S_R(Z)F_L(Z)}\right)\\ &= \mathbf{E}\left[\mathbf{E}\left(\frac{1_A|f(X)-Z|^2}{S_R(Z)F_L(Z)}\middle|X,Y\right)\right]\\ &= \mathbf{E}\left(|f(X)-Y|^2\right).\end{aligned}$$

De plus

$$\begin{aligned}\sup_{f\in\mathcal{F}_n^*\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^n 1_{\{A_i=0\}}\frac{|f(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)} - \mathbf{E}|f(X)-Y|^2\right| \\ = \sup_{f\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^n h(V_i) - \mathbf{E}h(V_i)\right|.\end{aligned}$$

Pour h_1 *et* h_2 *dans* $\mathcal{H}_n$, *soient* f_1 *et* f_2 *les fonctions leur correspondant respectivement dans* $\mathcal{F}_n^*\mathcal{F}_n$, *il s'ensuit que*

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n |h_1(V_i)-h_2(V_i)| &= \frac{1}{n}\sum_{i=1}^n\left|1_{\{A_i=0\}}\frac{|f_1(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)} - 1_{\{A_i=0\}}\frac{|f_2(X_i)-Z_i|^2}{S_R(Z_i)F_L(Z_i)}\right|\\ &\leq \frac{1}{S_R(T)F_L(I)}\frac{1}{n}\sum_{i=1}^n\left|\left(f_1(X_i)+f_2(X_i)-2Z_i\right)\left(f_1(X_i)-f_2(X_i)\right)\right|\\ &\leq \frac{2(T_R\vee T_L)}{S_R(T)F_L(I)}\frac{1}{n}\sum_{i=1}^n\left|f_1(X_i)-f_2(X_i)\right|,\end{aligned}$$

qui implique que

$$\mathcal{N}\left(\epsilon,\mathcal{H}_n,(V_1,\ldots,V_n)\right) \leq \mathcal{N}\left(\epsilon\frac{S_R(T)F_L(I)}{2(T_L\vee T_R)},\mathcal{F}_n^*\mathcal{F}_n,(X_1,\ldots,X_n)\right),$$

où $\mathcal{N}(\epsilon,\mathcal{G},(W_1,\ldots,W_n))$ *dénote le nombre recouvrant (cf l'annexe).*

Par application du théorème 7 de l'annexe, nous obtenons pour tout $\delta > 0$

$$\begin{aligned}&\mathbf{P}\left\{\sup_{f\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^n h(V_i)-\mathbf{E}h(V)\right|>\delta\right\}\\ \leq& 8\mathbf{E}\left\{\mathcal{N}\left(\frac{\delta S_R(T)F_L(I)}{16(T_L\vee T_R)},\mathcal{F}_n^*\mathcal{F}_n,(X_1,\ldots,X_n)\right)\right\}\ \exp\left(\frac{-n\delta^2S_R^2(T)F_L^2(I)}{128(T_L\vee T_R)^4}\right).\end{aligned}$$

Le théorème 8 de l'appendice montre alors l'existence des constantes c_1, c_2 et c_3, indépendantes de n, et telles que

$$\begin{aligned}&\mathbf{P}\left\{\sup_{h\in\mathcal{H}_n}\left|\frac{1}{n}\sum_{i=1}^{n}h(V_i)-\mathbf{E}h(V)\right|>\delta\right\}\\&\leq\left(c_1\frac{16n(T_L\vee T_R)^2}{\delta S_R(T)F_L(I)}\right)^{c_2\left(\frac{16\sqrt{n}(T_L\vee T_R)}{\delta F_L(I)S_R(T)}\right)^{\frac{d}{p}}+c_3}\exp\left(-\frac{n\delta^2F_L^2(I)S_R^2(T)}{128(T_L\vee T_R)^4}\right),\end{aligned}$$

qui est le terme d'une série convergente puisque $2p > d$. Le résultat du lemme (4) n'est alors qu'une conséquence directe du lemme de Borel-Cantelli.

Nous sommes maintenant en mesure de donner la preuve du théorème.

Preuve 5 *nous rappelons qu'il suffit de montrer que*

$$\int_{\mathbb{R}^d}|\bar{r}_n(x)-r(x)|^2\mu(dx)\to 0\quad p.s.\quad (n\to\infty).$$

Pour $\epsilon > 0$ arbitraire, soit $g_\epsilon : \mathbb{R}^d \to \mathbb{R}$ une application bornée, indéfiniment différentiable, à support compact et satisfaisant

$$\int_{\mathbb{R}^d}|r(x)-g_\epsilon(x)|^2\mu(dx)<\epsilon\quad et\quad \boldsymbol{J}_p^2(g_\epsilon)<\infty. \tag{4.12}$$

Effectuons la décomposition suivante

$$\begin{aligned}
\int_{\mathbb{R}^d} |\bar{r}_n(x) - r(x)|^2 \mu(dx) =& \mathbf{E}(|\bar{r}_n(X) - Y|^2 | \mathcal{D}_n) - \mathbf{E}|r(X) - Y|^2 \\
=& \mathbf{E}(|\bar{r}_n(X) - Y|^2 | \mathcal{D}_n) - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) F_L(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i) \hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i) F_L(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i) F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} \\
&+ \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |g_\epsilon(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} - \mathbf{E}|g_\epsilon(X) - Y|^2 \\
&+ \mathbf{E}|g_\epsilon(X) - Y|^2 - \mathbf{E}|r(X) - Y|^2 \\
:=& \sum_{i=1}^{9} Q_{n,i}.
\end{aligned}$$

– *En raison du lemme (4)*

$$Q_{n,1} = \mathbf{E}(|\bar{r}_n(X) - Y|^2 | \mathcal{D}_n) - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} \to 0.$$

– *Par application de (4.3), il vient*

$$\begin{aligned}
|Q_{n,2}| &= \left| \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{S_R(Z_i) F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}} |\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i) F_L(Z_i)} \right| \\
&\leq \frac{(T_L \vee T_R)^2}{S_R(T) F_L(I) \hat{S}_n(T)} \sup_{t \in \mathbb{R}^+} |\hat{S}_n(t) - S_R(t)| \underset{n\to\infty}{\to} 0 \quad p.s..
\end{aligned}$$

La preuve que

$$Q_{n,3} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \underset{n\to\infty}{\to} 0 \ p.s.$$

est identique à ce qui a été fait dans $Q_{n,3}$, mais en utilisant (4.4) au lieu de (4.3).

– *Comme*

$$\bar{r}_n(X_i) = T_{[0,T_R\vee T_L]}(\tilde{r}_n(X_i))$$

et que $Z_i \leq T_R \vee T_L$, il vient que

$$Q_{n,4} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|\bar{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \leq 0.$$

– *Par définition de $\tilde{r}_n$ et de g_ϵ, nous déduisons que*

$$Q_{n,5} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|\tilde{r}_n(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \leq \lambda_n \boldsymbol{J}_p^2(g_\epsilon),$$

qui tend vers zéro en raison de la condition (4.8) du théorème et de la relation (4.12).

– *Du fait que g_ϵ est bornée, que $Z_i \leq T_L \vee T_R$ combinés avec (4.3) ou (4.4), nous obtenons :*

$$Q_{n,6} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i)F_L(Z_i)} \underset{n\to\infty}{\longrightarrow} 0 \ p.s.$$

et

$$Q_{n,7} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{\hat{S}_n(Z_i)F_L(Z_i)} - \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{S_R(Z_i)F_L(Z_i)} \underset{n\to\infty}{\longrightarrow} 0 \ p.s.$$

– *L'application de la loi forte des grands nombres conduit à*

$$Q_{n,8} = \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}|g_\epsilon(X_i) - Z_i|^2}{S_R(Z_i)F_L(Z_i)} - \mathbf{E}|g_\epsilon(X) - Y|^2 \underset{n\to\infty}{\longrightarrow} 0.$$

– *Finalement, en vertu de (4.12)*

$$Q_{n,9} = \mathbf{E}|g_\epsilon(X) - Y|^2 - \mathbf{E}|r(X) - Y|^2 \leq \epsilon.$$

En faisant tendre ϵ vers 0, le résultat visé s'ensuit.

5 Simulation

Notre étude de simulation concerne quelques exemples de calcul des estimateurs de la fonction de régression introduits et ceux en vue de les comparés aux vrais fonctions de régression aux deux chapitres précédents dont nous gardons les notations. A cette fin, nous allons procéder en trois étapes, les deux premières se rapportant chacune à une méthode et la dernière concerne la comparaison entre elles.

5.1 Estimateur des moindres carrés

Rappelons que l'estimateur des moindres carrés en présence de censure mixte, s'écrit pour $M_n = \max_{1\leq i\leq n} Z_i$

$$r_n(x) = \mathbb{T}_{[0,M_n]}(\tilde{r}_n(x)),$$

où

$$\widetilde{r}_n = \arg\min_{f\in\mathcal{F}_n} \frac{1}{n}\sum_{i=1}^{n} \frac{1_{\{A_i=0\}}}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)} \left|f(X_i) - Z_i\right|^2 \left(\frac{0}{0} := 0\right). \tag{5.1}$$

Pour notre étude de simulation nous choisissons comme famille de fonctions $\mathcal{F}_n$ la classe des polynômes de degré inférieur ou égal à trois. Une fois nos échantillons générés, nous commençons par calculer l'estimateur $\hat{S}_n$ de Patiléa et Rolin [37] et celui de Kaplan et Meier [25] $\hat{F}_n$. Puis la méthode utilisée est de chercher dans chaque classe de polynômes de degré fixé (1, 2 et 3) celui qui minimise la quantité donnée dans (5.1). Finalement, nous prenons le meilleur des trois.

Ainsi, nous allons considérer $w_i = \frac{1_{\{A_i=0\}}}{\hat{S}_n(Z_i)\hat{F}_n(Z_i)}$ comme étant des poids qui dépendent de la $i^{-\text{ème}}$ donnée, c'est-à-dire que nous allons réaliser une minimisation au sens des moindres carrés avec poids. Pour cela, nous avons fait

appel à une commande de "Matlab" noté "lscov" qui permet de résoudre le système d'équations que nous obtenons en minimisant (5.1), avec une pondération, pour nos différents cas.
Dans toute la suite nous considérons trois tailles d'échantillon $n = 50, 100, 500$.

5.1.1 Modèle linéaire

Prenons $Y = 2X + 1 + \varepsilon$ où X suit la loi uniforme sur l'intervalle $[0, 1]$ et $\varepsilon \simeq \mathcal{N}(0, 0.25)$. Les variables de censure sont $R \simeq \exp(5.5)$ et $L \simeq \exp(0.45)$. Nous obtenons les graphes représentant la courbe théorique r et les polynômes de degré 1 noté $(P1)$, 2 noté $(P2)$ et 3 noté $(P3)$.

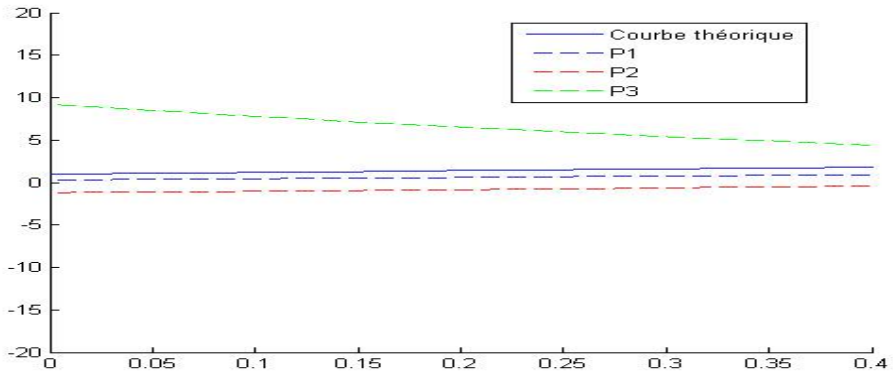

Figure 5.1: $r(x) = 2x + 1$, avec $n = 50$ et un taux de censure à gauche et à droite de 22% et 12% respectivement.

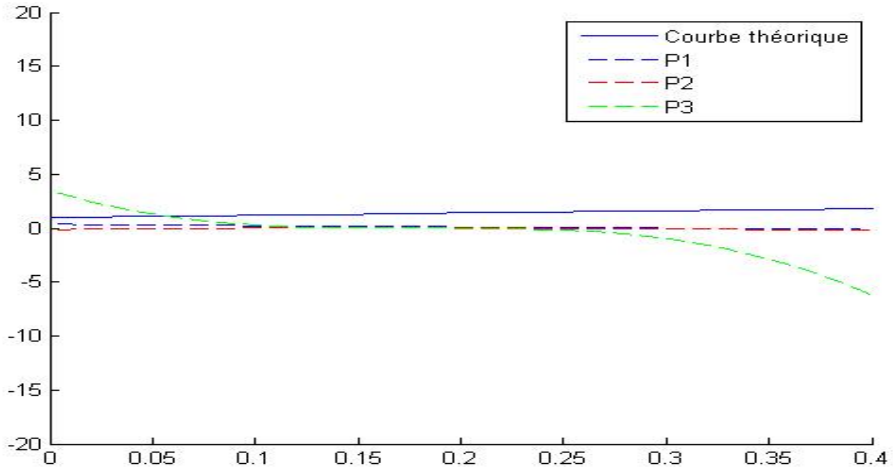

FIGURE 5.2: $r(x) = 2x + 1$, avec $n = 100$ et un taux de censure à gauche et à droite de 10% et 20% respectivement.

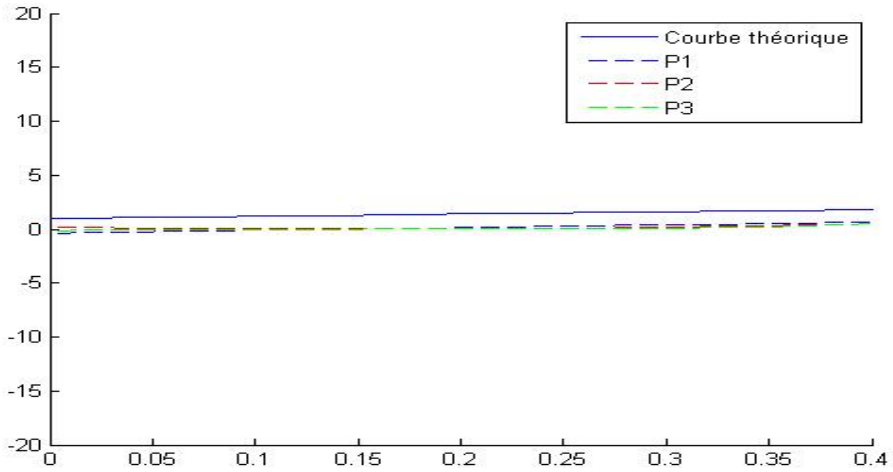

FIGURE 5.3: $r(x) = 2x + 1$, avec $n = 500$ et un taux de censure à gauche et à droite de 16.2% et 16% respectivement.

n	erreur1	erreur2	erreur3
50	1.9519	123.3321	13204
100	5.9262	11.1666	13584
200	1.2422	1.4079	31.7794

Table 5.1: Tableau des erreurs d'approximation.

Le tableau 5.1 résume, pour les différentes tailles des échantillons les moyennes des écarts au carré entre la courbe théorique et les polynômes $P1, P2$ et $P3$ notés par erreur1, erreur2 et erreur3 respectivement.
D'après la figure (5.3), les trois polynômes $P1, P2$ et $P3$ semblent s'approcher de r. Afin de les départager, nous utilisons le tableau 5.1 qui montre que le polynôme de degré 1 est le meilleur et l'erreur d'approximation devient de plus en plus petite au fur et à mesure que la taille de l'échantillon augmente.

5.1.2 Modèle cosinusoïdale

Prenons $Y = cos(2X + 1) + \varepsilon$ où X suit la loi uniforme sur l'intervalle $[5, 6]$ et $\varepsilon \simeq \mathcal{N}(0, 0.25)$. Les variables de censure sont $R \simeq \exp(4)$ et $L \simeq \exp(0.07)$.

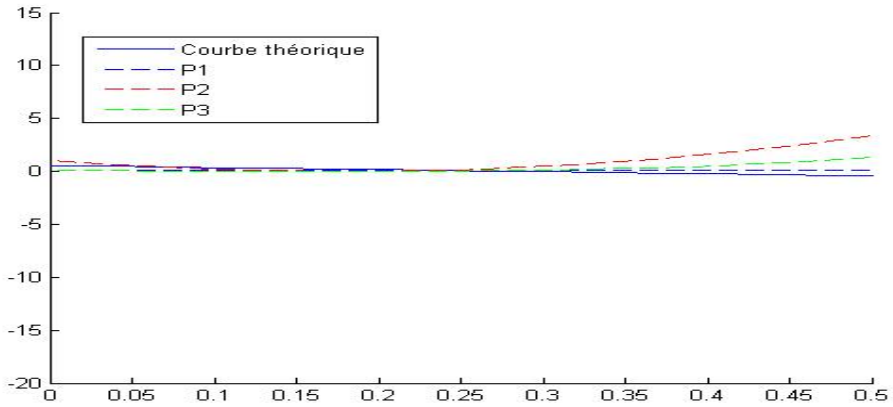

Figure 5.4: $r(x) = \cos(2x + 1)$, avec n=50 et un taux de censure à gauche et à droite de 12% et 18% respectivement.

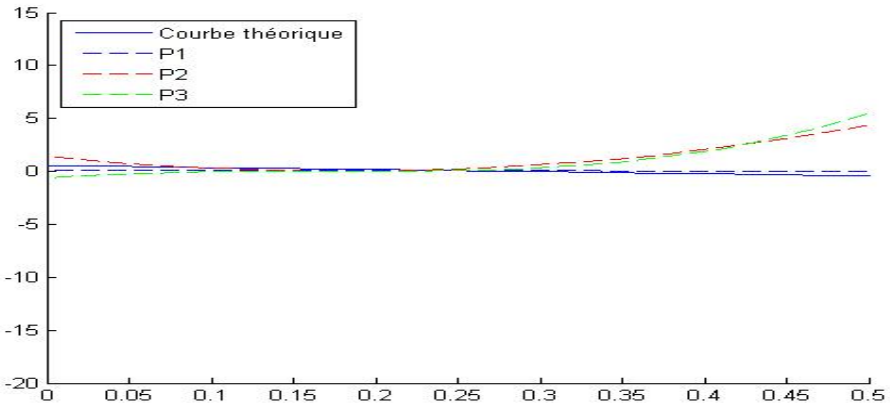

FIGURE 5.5: $r(x) = \cos(2x+1)$, avec n=100 et un taux de censure à gauche et à droite de 15.6% et 13.8% respectivement.

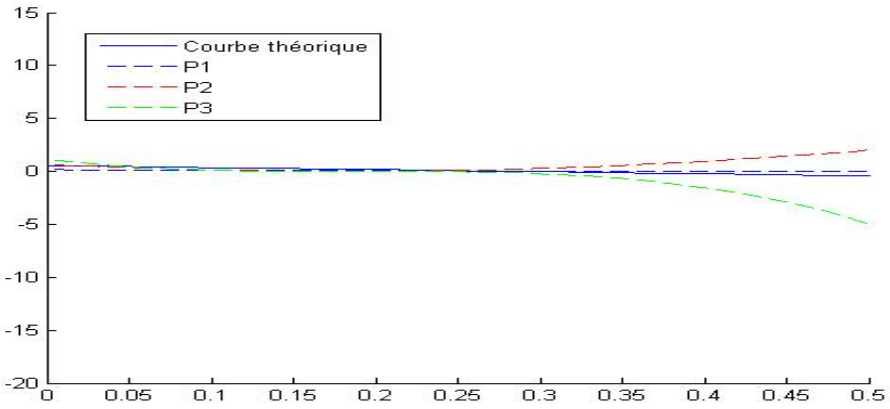

FIGURE 5.6: $r(x) = \cos(2x+1)$, avec n=500 et un taux de censure à gauche et à droite de 19% et 14% respectivement.

n	erreur1	erreur2	erreur3
50	0.4237	92.4620	39.8786
100	0.2937	154.3758	939.4889
500	0.2408	154.3758	966.3336

Table 5.2: Tableau des erreurs d'approximation.

D'après les figures 5.4, 5.5 et 5.6, les polynômes $P1$ et $P2$ semblent convenables dès $n = 50$. Le tableau 5.2 des erreurs permet de les départager. Notre choix se porte donc sur le polynôme $P1$.

5.1.3 Modèle exponentielle

Les figures (5.6, 5.7 et 5.8) sont le résultat de l'étude du modèle $Y = \exp(2X + 1) + \epsilon$ où ϵ et les variables latentes X, R, L suivent les mêmes lois qu'au modèle cosinusoïdale. Ici aussi le choix se porte sur le polynôme $P1$.

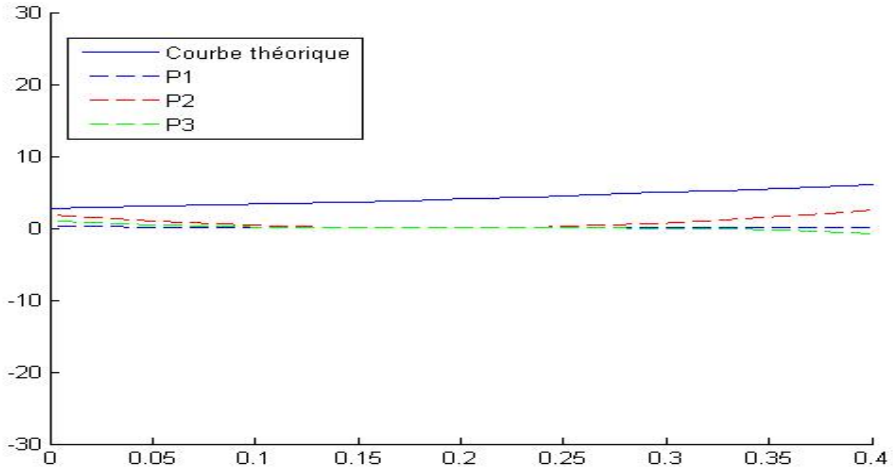

Figure 5.7: $r(x) = \exp(2x+1)$, avec $n = 50$ et un taux de censure à gauche et à droite de 12% et 16% respectivement.

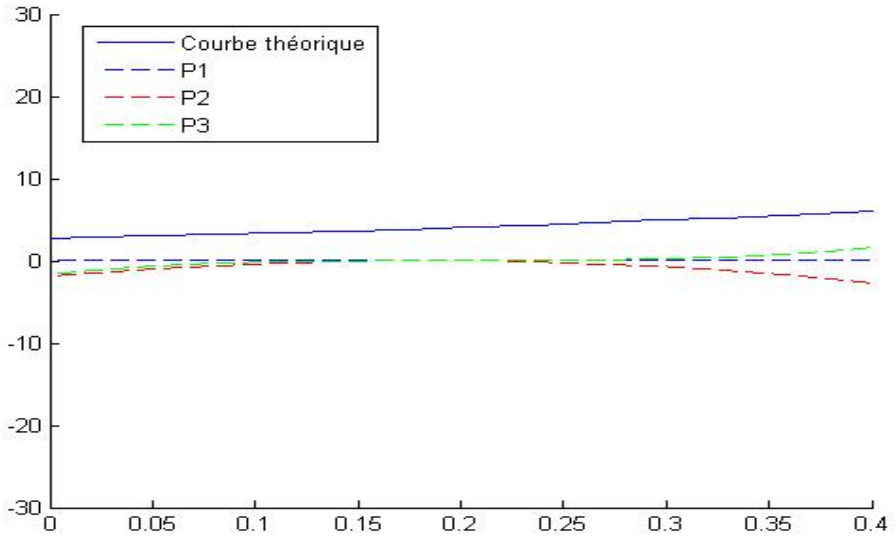

FIGURE 5.8: $r(x) = \exp(2x + 1)$, avec $n = 100$ et un taux de censure à gauche et à droite de 15% et 17%.

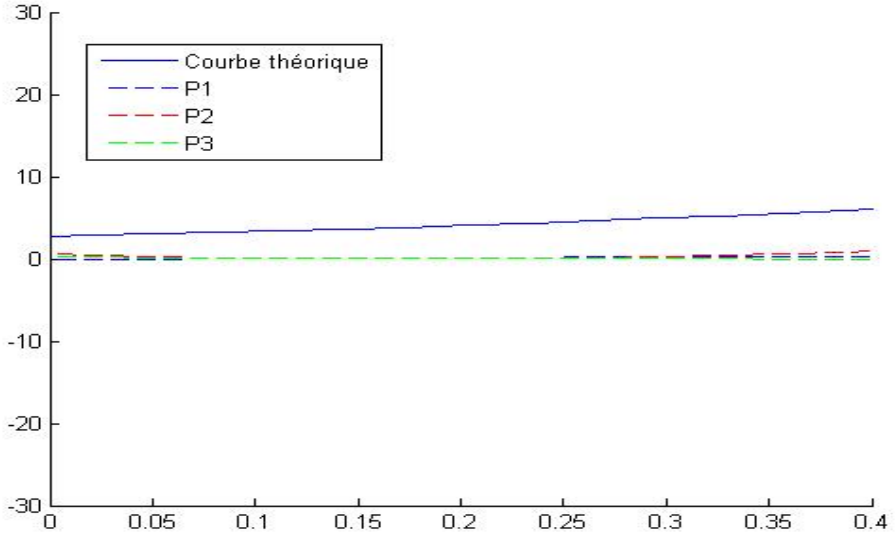

FIGURE 5.9: $r(x) = \exp(2x + 1)$, avec $n = 500$ et un taux de censure à gauche et à droite de 16% et 18% respectivement.

n	erreur1	erreur2	erreur3
50	1.8445	3.8175	368.49
100	1.6849	46.542	2589.4
200	1.6055	32.656	96.925

Table 5.3: Tableau des erreurs d'approximation.

A travers tous les exemples précédents nous concluons que notre estimateur des moindres carrés est assez performant dès la taille $n = 50$ et s'améliore lorsque la taille augmente. Faisons remarquer que les taux de censure sont assez significatifs puisqu'ils sont de l'ordre de 30%.

5.2 Estimateur spline de lissage

Dans cette partie nous allons traiter les mêmes modèles que dans la section précédente (en vue d'une comparaison ultérieure entre les deux méthodes d'estimation). Rappelons que notre estimateur spline de lissage pour le cas unidimensionnel est donné sous la forme

$$r_n(.) = \underset{f \in C^2(\mathbb{R})}{\arg\, min} \left\{ \frac{1}{n} \sum w_i |f(X_i) - Y_i|^2 + J_n(f) \right\},$$

où le terme de pénalité est donnée par $J_n(f) \geq 0$ et $J_n(f) = \lambda_n \int_{-\infty}^{+\infty} |f^{(2)}(x)|^2 dx$. Ici l'utilisation d'une commande interne du logiciel de calcul "Matlab", noté "csaps" nous permet d'obtenir les graphes suivants.

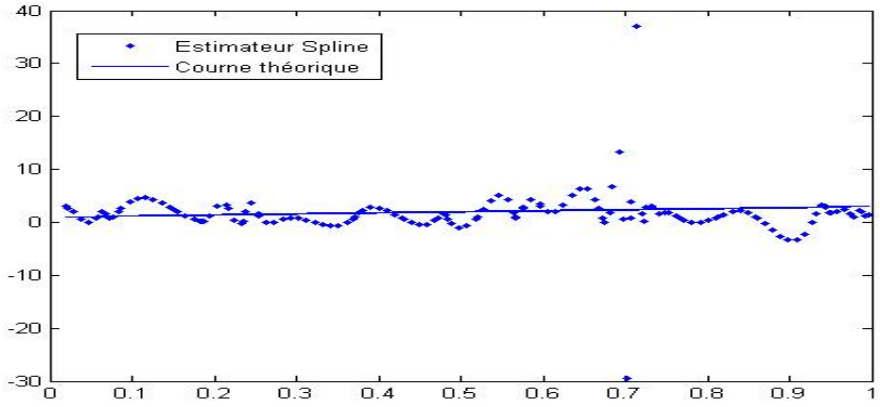

FIGURE 5.10: $r(x) = 2x + 1$, $n = 50$ et avec un taux de censure à gauche et à droite de 14% et 26% respectivement.

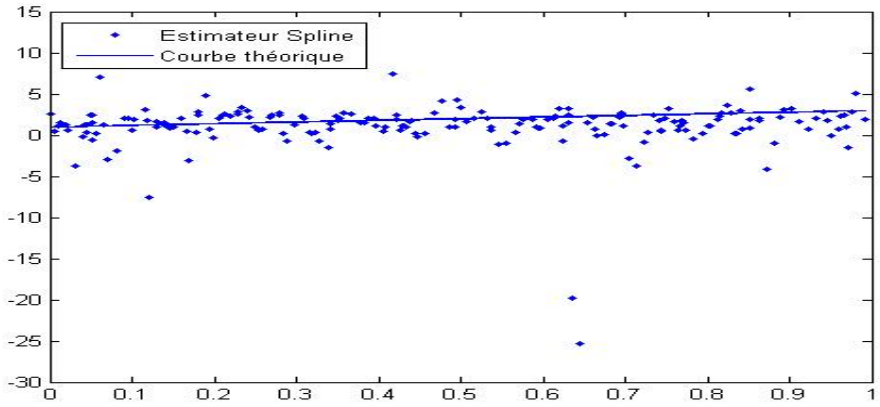

FIGURE 5.11: $r(x) = 2x + 1$, $n = 100$ et avec un taux de censure à gauche et à droite de 15% et de 17% respectivement.

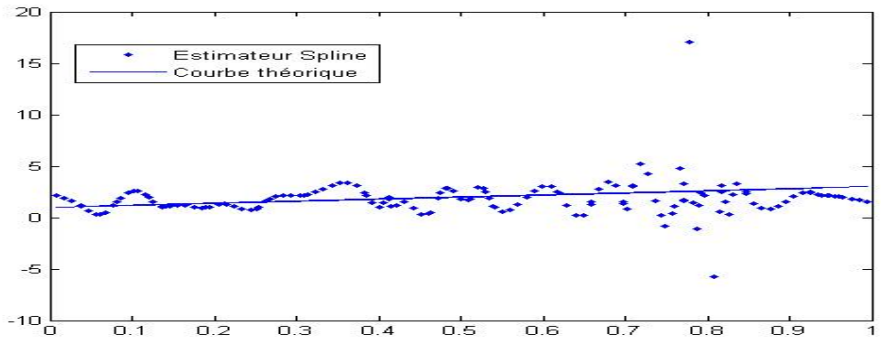

Figure 5.12: $r(x) = 2x + 1$, $n = 500$ et avec un taux de censure à gauche et à droite de 8% et 26% respectivement.

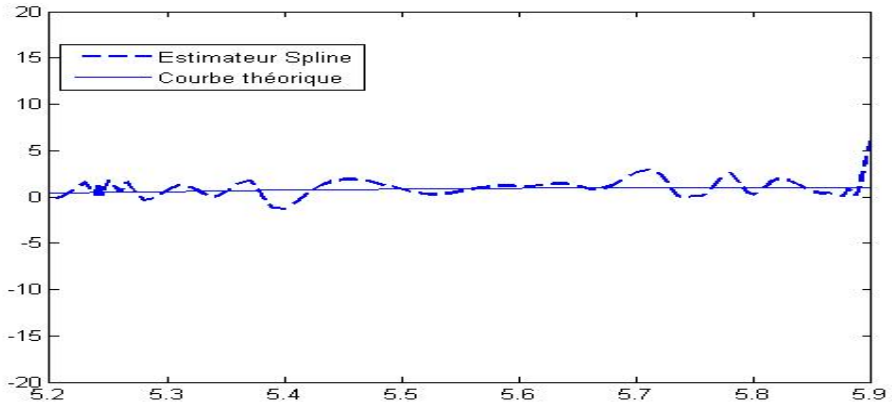

Figure 5.13: $r(x) = \cos(2x + 1)$, avec $n = 50$ et un taux de censure à gauche et à droite de 18% et 14% respectivement.

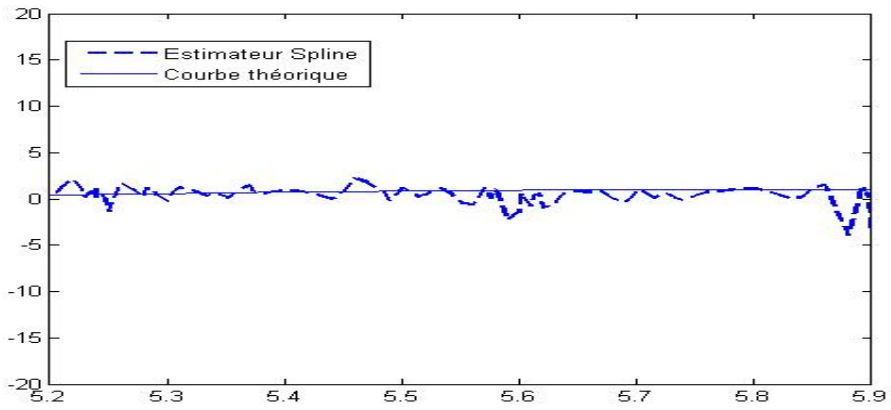

FIGURE 5.14: $r(x) = \cos(2x + 1)$, avec $n = 100$ et un taux de censure à gauche et à droite de 15% et 16% respectivement.

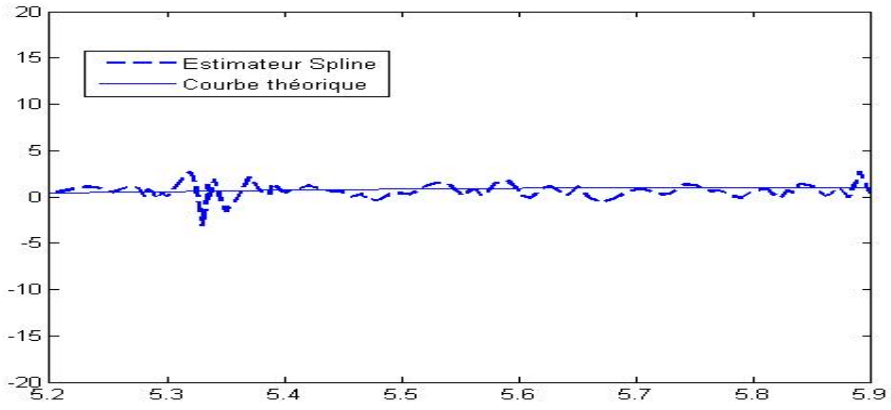

FIGURE 5.15: $r(x) = \cos(2x + 1)$, avec $n = 500$ et un taux de censure à gauche et à droite de 17.4% et 13% respectivement.

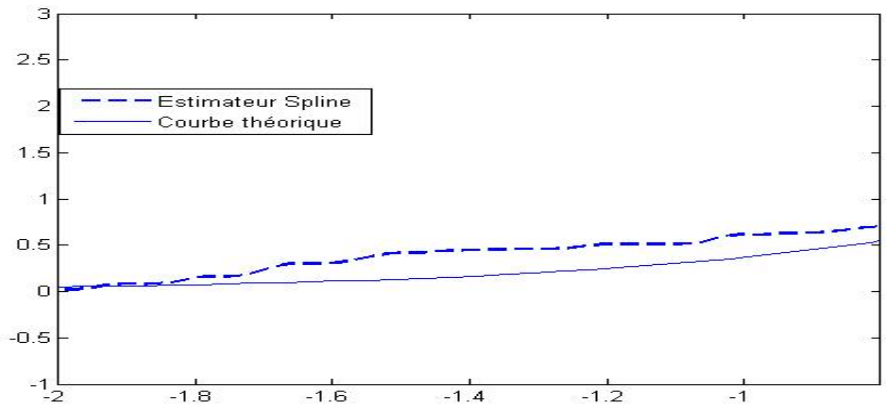

Figure 5.16: $r(x) = \exp(2x + 1)$, $n = 50$ et un taux de censure à gauche et à droite de 10% et 24% respectivement.

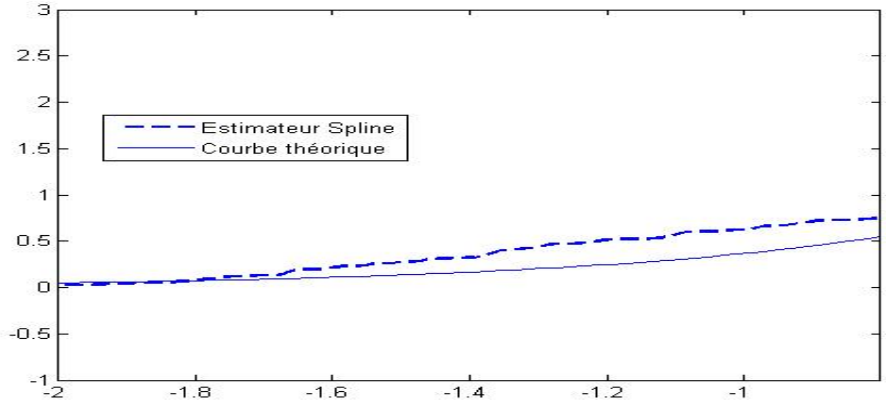

Figure 5.17: $r(x) = \exp(2x + 1)$, $n = 100$ et un taux de censure à gauche et à droite de 17% et 18% respectivement.

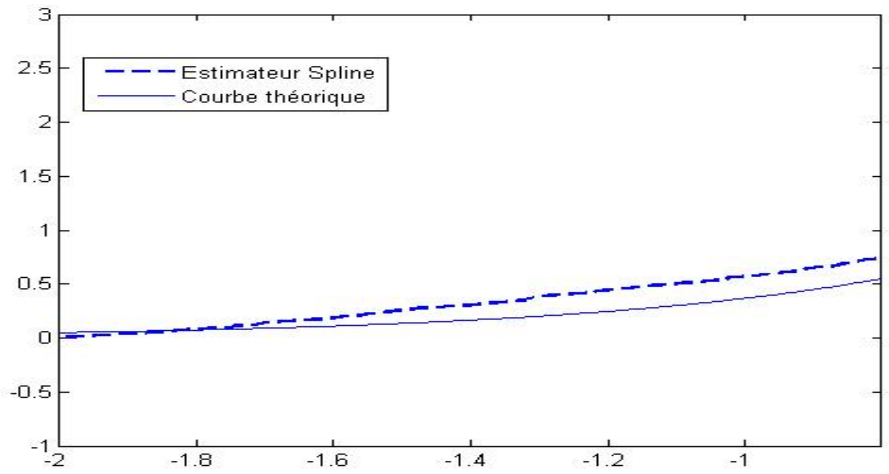

FIGURE 5.18: $r(x) = \exp(2x + 1)$, $n = 500$ et un taux de censure à gauche et à droite de 16.6% et 16.8% respectivement.

Ces différents graphes permettent de conclure à la performance de l'estimateur spline de lissage pour un taux de censure avoisinant 30%.

5.3 Comparaison des deux modèles

Comme annoncé plus haut, nous allons comparer nos deux méthodes d'estimation à travers les modèles précédemment étudiés en les représentant sur les mêmes graphes.

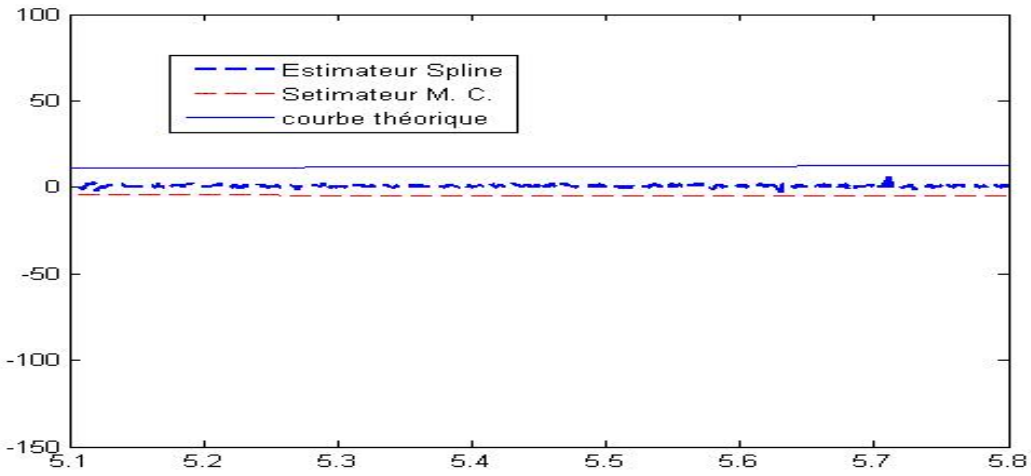

Figure 5.19: $r(x) = 2x + 1$, $n = 500$ et un taux de censure à gauche et à droite de 17% et 13% respectivement.

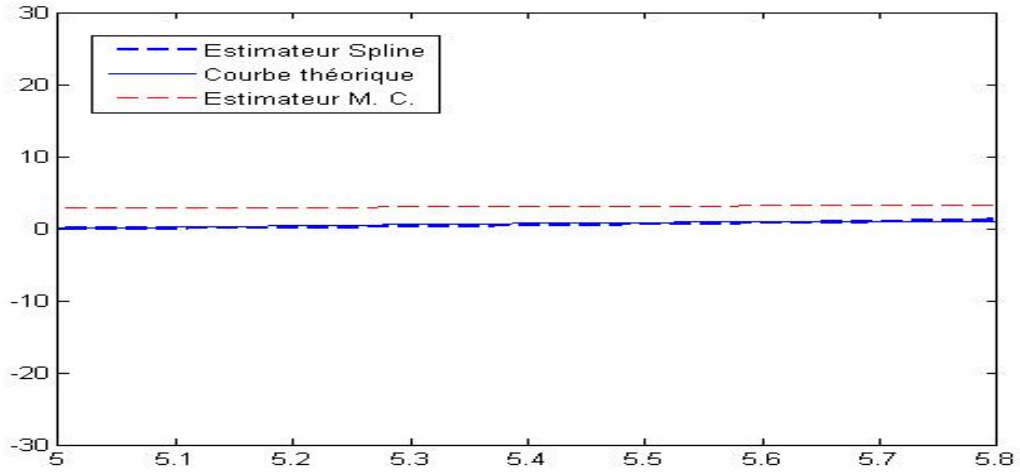

Figure 5.20: $r(x) = \cos(2x + 1)$, $n = 500$ et un taux de censure à gauche et à droite de 16% et 15.2% respectivement.

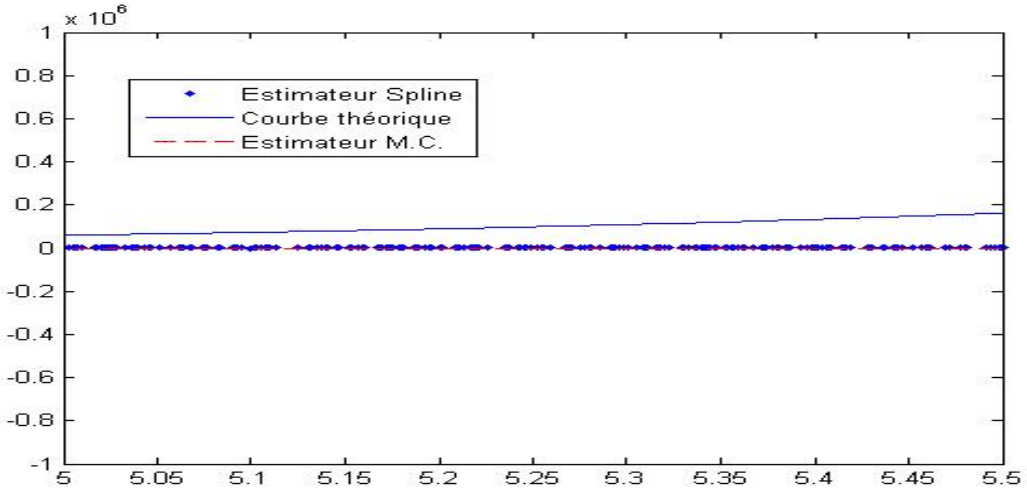

FIGURE 5.21: $r(x) = \exp(2x + 1)$, $n = 500$ et un taux de censure à gauche et à droite de 13.6% et 16.8% respectivement.

Les figures (5.19, 5.20 et 5.21), nous permettent de conclure que l'estimateur des moindres carrés, malgré son efficacité, est moins adhérant à la courbe théorique que l'estimateur spline de lissage et cela pour les trois modèles considérés.

Perspectives

Pour compléter cette thèse nous présentons une liste de quelques points susceptibles de faire l'objet de futurs travaux.

- ✓ Les hypothèses concernant les liens entres les supports des variables latentes pouvant paraître restrictives, il nous semble possible de s'en passer mais en restreignant l'ensemble de convergence des estimateurs.
- ✓ Étude des taux de convergence de nos estimateurs.
- ✓ Étude de l'estimateur de la fonction de régression par la méthode des B-spline dans un modèle de censure mixte ainsi que les taux de convergence.
- ✓ Étude de la convergence presque complète de nos estimateurs.
- ✓ Étude d'autres modèles de censure mixte.

Appendice

Dans les chapitres [3] et [4], nous sommes amenés à montrer que des quantités de la forme

$$\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1,\ldots,n} f(Z_i) - \mathbf{E}f(Z_i)\right|$$

tendent vers zéro.
Cet appendice nous fournit des conditions suffisantes pour obtenir ces résultats. Il est basé sur la théorie de Vapnik-Chervonenkis (cf [9], [21] et [39] pour plus de détails).

Définition 1 (nombre recouvrant) *Soient $\epsilon > 0$ et $\mathcal{F}$ une famille de fonctions $f : \mathbb{R}^d \to \mathbb{R}$.*

- *Toute famille finie de fonctions $f_1, \ldots, f_N : \mathbb{R}^d \to \mathbb{R}$ telle que pour tout $f \in \mathcal{F}$, il existe un $k = k(f) \in 1, 2, ..., N$ vérifiant*

$$||f - f_k||_{L_p} < \epsilon,$$

  *est appelé un $\epsilon-$**recouvrement** ($\epsilon-$**cover**) de $\mathcal{F}$ (par rapport à la norme de L_p).*
- *Soit $\mathcal{N}(\epsilon, \mathcal{F}, ||.||_{L_p})$ le cardinal du plus petit $\epsilon-$**recouvrement** ($\epsilon-$**covering number**) de $\mathcal{F}$. On pose $\mathcal{N}(\epsilon, \mathcal{F}, ||.||_{L_p}) = \infty$ s'il n'existe pas de $\epsilon-$**recouvrement** de $\mathcal{F}$. Alors $\mathcal{N}(\epsilon, \mathcal{F}, ||.||_{L_p})$ est dit un $\epsilon-$ **nombre recouvrant** de $\mathcal{F}$ (par rapport la norme de L_p).*

Nous allons maintenant définir le nombre englobant qui est en relation directe avec le nombre recouvrant comme on le verra dans la suite.

Définition 2 (nombre englobant) *Soient $\epsilon > 0$, $\mathcal{F}$ un ensemble de fonctions de $\mathbb{R}^d \to \mathbb{R}$ et $f_1, \ldots, f_N \in \mathcal{F}$ une famille finie de fonctions vérifiant pour tous $1 \leq j \leq k \leq N$*

$$||f_j - f_k||_{L_p} \geq \epsilon \tag{5.2}$$

est dite $\epsilon -$ **emballage** *($\epsilon -$* **packing***) de $\mathcal{F}$.*

Soit $\mathcal{M}(\epsilon, \mathcal{F}, ||.||_{L_p})$ le plus grand cardinal de toutes les familles vérifiant (5.2). On pose $\mathcal{M}(\epsilon, \mathcal{F}, ||.||_{L_p}) = \infty$ s'il existe une telle famille pour tout $N \in \mathbb{N}$. $\mathcal{M}(\epsilon, \mathcal{F}, ||.||_{L_p})$ est dit un $\epsilon-$ **nombre englobant** *($\epsilon-$* **packing nomber***) de $\mathcal{F}$ (par rapport à la norme de L_p).*

Soit $z_1^n = (z_1, \ldots, z_n)$ un point fixe de $\mathbb{R}^d$, soit μ_n la mesure empirique correspondante, c'est-à-dire,

$$\mu_n(A) = \frac{1}{n} \sum_{i=1}^{n} 1_A(z_i) \text{ avec } A \subseteq \mathbb{R}^d.$$

Dans le cas où l'espace L_p est muni de la mesure μ_n, alors

$\mathcal{N}(\epsilon, \mathcal{F}, ||.||_{L_p})$ est noté par $\mathcal{N}_p(\epsilon, \mathcal{F}, z_1^n)$(où encore $\mathcal{N}(\epsilon, \mathcal{F}, z_1^n)$ pour $p = 1$).

$\mathcal{M}(\epsilon, \mathcal{F}, ||.||_{L_p})$ est noté par $\mathcal{M}_p(\epsilon, \mathcal{F}, z_1^n)$.

En d'autres termes

$\mathcal{N}_p(\epsilon, \mathcal{F}, z_1^n)$ est le plus petit nombre $N \in \mathbb{N}$ pour qui il existe les fonctions $f_1, \ldots, f_N : \mathbb{R}^d \to \mathbb{R}$ avec la propriété que pour tout $f \in \mathcal{F}$ il y a un

$k = k(f) \in 1, \ldots, N$ *tel que*

$$\left(\frac{1}{n} \sum_{i=1}^{n} |f(z_i) - f_k(z_i)|^p \right)^{\frac{1}{p}} < \epsilon,$$

$\mathcal{M}_p(\epsilon, \mathcal{F}, z_1^n)$ *est le plus grand nombre* $N \in \mathbb{N}$ *pour qui il existe les fonctions* $f_1, \ldots, f_N \in \mathcal{F}$ *avec*

$$\left(\frac{1}{n} \sum_{i=1}^{n} |f_j(z_i) - f_k(z_i)|^p \right)^{\frac{1}{p}} \geq \epsilon,$$

pour tous $1 \leq j \leq k \leq N$.

Le lemme suivant donne une relation qui lie ces deux nombres.

Lemme 5 *Soient* $\mathcal{F}$ *une classe de fonction sur* $\mathbb{R}^d$, $p \geq 1$ *et* $\epsilon > 0$. *Alors*

$$\mathcal{M}_p(2\epsilon, \mathcal{F}, z_1^n) \leq \mathcal{N}_p(\epsilon, \mathcal{F}, z_1^n) \leq \mathcal{M}_p(\epsilon, \mathcal{F}, z_1^n),$$

pour tous $z_1, \ldots, z_n \in \mathbb{R}^d$.

Preuve 6 *Soit* $\{f_1, \ldots, f_l\}$ *un* $2\epsilon-$**emballage** *de* $\mathcal{F}$. *Alors, tout ensemble*

$$U_\epsilon(f) = \{h : \mathbb{R}^d \to \mathbb{R} : ||h - f||_{L_p} < \epsilon\},$$

peut contenir, au plus, un des f_i. *Ceci prouve la première inégalité.*
Pour la deuxième inégalité, on suppose que $\mathcal{M}_p(\epsilon, \mathcal{F}, z_1^n) < \infty$*(autrement la démonstration est triviale). Soit* $\{g_1, \ldots, g_l\}$ *un* $\epsilon-$**emballage** *de* $\mathcal{F}$, *de cardinal maximum* $l = \mathcal{M}(\epsilon, \mathcal{F}, ||.||_{L_p})$. *Soit* $h \in \mathcal{F}$, *alors,* $\{h, g_1, \ldots, g_l\}$ *est un sous-ensemble de* $\mathcal{F}$ *de cardinal*$= l + 1$, *donc il ne peut pas être un* $\epsilon-$**emballage** *de* $\mathcal{F}$. *Ainsi, il existe* $j \in \{1, \ldots, l\}$ *tel que*

$$||h - g_j||_{L_p} < \epsilon.$$

Ceci prouve que c'est un $\epsilon-$**recouvrement** *de* $\mathcal{F}$*, par rapport à la norme* L_p*, donc*

$$\mathcal{N}_p(\epsilon, \mathcal{F}, z_1^n) \leq \mathcal{M}_p(\epsilon, \mathcal{F}, z_1^n).$$

Définition 3 (dimension VC) *Soit* $\mathcal{D}$ *une classe de sous ensembles de* $\mathbb{R}^d$ *et* $F \subseteq \mathbb{R}^d$*. On dit que* $\mathcal{D}$ *"brise" (shatters)* F *si pour tout sous ensemble* E *de* F *il existe* $D \in \mathcal{D}$ *tel que* $E = F \cap D$*. La dimension VC de* $\mathcal{D}$*, noté* $V_{\mathcal{D}}$*, est le plus grand entier* k *pour lequel il existe un ensemble de cardinal* k *brisé par* $\mathcal{D}$*.*

Le nombre recouvrant peut être utilisé pour borner certaines probabilités, comme le montre le résultat suivant.

Théorème 7 *[cf [21]] Soit* $\mathcal{F}$ *une famille de fonctions* $f : \mathbb{R}^d \to [0, B]$ *et soit* $Z_1^n = (Z_1, \ldots, Z_n)$ *une suite de vecteurs aléatoires i. i. d. à valeurs dans* $\mathbb{R}^d$*. Alors pour tout* $\epsilon > 0$

$$\begin{aligned}\mathbf{P}&\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \mathbf{E}f(Z_i)\right| > \epsilon\right\} \\ &\leq 8\mathbf{E}\left\{\mathcal{N}\left(\frac{\epsilon}{8}, \mathcal{F}, Z_1^n\right)\right\}\exp\left(\frac{-n\epsilon^2}{128B^2}\right).\end{aligned}$$

Preuve 7 *La démonstration se fera en quatre étapes.*

- *Étape 1. Introduction d'un échantillon intermédiaire.*

  Remplaçons l'espérance de $f(Z_i)$ *à l'intérieur de la probabilité par la moyenne empirique basée sur un échantillon* $\acute{Z}_1^n = (\acute{Z}_1, \acute{Z}_2, \ldots, \acute{Z}_n)$ *qui est une suite de vecteurs aléatoires i. i. d. de même loi que* Z*,*

indépendante de Z_1^n. *Soit* $f \in \mathcal{F}_n$ *tel que*

$$\left| \frac{1}{n} \sum_{i=1}^{n} f(Z_i) - \mathbf{E} f(Z_i) \right| > \epsilon,$$

si une telle fonction existe. Sinon, soit f^* *une fonction arbitraire dans* $\mathcal{F}_n$. *Notez que* f^* *dépend de* Z_1^n *et que* $\mathbf{E}\{f^*(Z)|Z_1^n\}$ *est l'espérance de* $f^*(Z)$ *sachant* Z. *L'application de l'inégalité de Chebyshev et le fait que* f^* *est borné par* B *donnent*

$$\begin{aligned} &\mathbf{P}\left\{ \left| \mathbf{E}\{f^*(Z)|Z_1^n\} - \frac{1}{n} \sum_{i=1}^{n} f^*(\acute{Z}_i) \right| > \frac{\epsilon}{2} \middle| Z_1^n \right\} \\ &\quad \leq \frac{\mathbf{Var}\{f^*(Z)|Z_1^n\}}{n(\frac{\epsilon}{2})^2} \leq \frac{\frac{B^2}{4}}{n\frac{\epsilon^2}{2}} = \frac{B^2}{n\epsilon^2}, \end{aligned}$$

ce qui implique que

$$\begin{aligned} \mathbf{Var}\{f^*(Z)|Z_1^n\} = \mathbf{Var}&\left\{ (f^*(Z) - \frac{B}{2}) \mid Z_1^n \right\} \\ \leq &\mathbf{E}\left\{ \left| f^*(Z) - \frac{B}{2} \right|^2 |Z_1^n \right\} \\ \leq &\frac{B^2}{4}. \end{aligned}$$

Ainsi, pour $n \geq \frac{2B^2}{\epsilon^2}$, *nous obtenons*

$$\mathbf{P}\left\{ \left| \mathbf{E}\{f^*(Z)|Z_1^n\} - \frac{1}{n} \sum_{i=1}^{n} f^*(\acute{Z}_i) \right| > \frac{\epsilon}{2} |Z_1^n \right\} \geq \frac{1}{2}. \qquad (5.3)$$

Donc,

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{f\in\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}f(Z_i)-\frac{1}{n}\sum_{i=1}^{n}f(\acute{Z}_i)\right|>\frac{\epsilon}{2}\right\}\\
&\geq\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}f^*(Z_i)-\frac{1}{n}\sum_{i=1}^{n}f^*(\acute{Z}_i)\right|>\frac{\epsilon}{2}\right\}\\
&\geq\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}f^*(Z_i)-\mathbf{E}\{f^*(Z)|Z_1^n\}\right|>\epsilon\,,\right.\\
&\left.\left|\frac{1}{n}\sum_{i=1}^{n}f^*(\acute{Z}_i)-\mathbf{E}\{f^*(Z)|Z_1^n\}\right|\leq\frac{\epsilon}{2}\right\}.
\end{aligned}$$

La dernière probabilité peut être déterminée en calculant, en premier lieu, la probabilité correspondante sachant Z_1^n, puis en passant par le calcul de l'espérance par rapport à Z_1^n. Si

$$\left|\frac{1}{n}\sum_{i=1}^{n}f^*(Z_i)-\mathbf{E}\{f^*(Z)|Z_1^n\}\right|>\epsilon,$$

alors la probabilité précédente sachant Z_1^n est égale à

$$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n}f^*(\acute{Z}_i)-\mathbf{E}\{f^*(Z)|Z_1^n\}\right|\leq\frac{\epsilon}{2}|Z_1^n\right\}.$$

Sinon elle est nulle.

Ce qui précède implique que

$$
\begin{aligned}
&\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f^*(Z_i) - \mathbf{E}\{f^*(Z)|Z_1^n\}\right| > \epsilon\ ,\right.\\
&\left.\left|\frac{1}{n}\sum_{i=1}^{n} f^*(\acute{Z}_i) - \mathbf{E}\{f^*(Z)|Z_1^n\}\right| \le \frac{\epsilon}{2}\right\}\\
&= \mathbf{E}\left\{1_{\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f^*(Z_i) - \mathbf{E}\{f^*(Z)|Z_1^n\}\right| > \epsilon\right\}}\right.\\
&\left.\times\ \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f^*(\acute{Z}_i) - \mathbf{E}\{f^*(Z)|Z_1^n\}\right| \le \frac{\epsilon}{2}\middle| Z_1^n\right\}\right\}\\
&\ge \frac{1}{2}\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} f^*(Z_i) - \mathbf{E}\{f^*(Z)|Z_1^n\}\right| > \epsilon\right\}\\
&= \frac{1}{2}\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \mathbf{E}\{f(Z)\}\right| > \epsilon\right\},
\end{aligned}
$$

où, la dernière inégalité vient de (5.3). Ainsi nous obtenons pour $n \ge \frac{2B^2}{\epsilon^2}$,

$$
\begin{aligned}
&\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \mathbf{E}\{f(Z)\}\right| > \epsilon\right\}\\
&\le 2\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(Z_i) - \frac{1}{n}\sum_{i=1}^{n} f(\acute{Z}_i)\right| > \frac{\epsilon}{2}\right\}.
\end{aligned}
$$

- *Étape 2. Introduction de variables signes.*

  Soient $U_1, \ldots, U_n$ *des variables indépendantes qui suivant la loi uniforme sur* $\{-1, +1\}$ *et indépendantes de* $Z_1, \ldots, Z_n, \acute{Z}_1, \ldots, \acute{Z}_n$. *La répartition jointe de* $Z_1^n, \acute{Z}_1^n$, *n'est pas affectée si un changement aléatoire*

se fait sur leurs composantes correspondantes. Donc

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(Z_i)-\frac{1}{n}\sum_{i=1}^{n}f(\acute{Z}_i)\right|>\frac{\epsilon}{2}\right\}\\
&=\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}U_i\left[f(Z_i)-\frac{1}{n}\sum_{i=1}^{n}f(\acute{Z}_i)\right]\right|>\frac{\epsilon}{2}\right\}\\
&\leq\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}U_if(Z_i)\right|>\frac{\epsilon}{4}\right\}+\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}U_if(\acute{Z}_i)\right|>\frac{\epsilon}{4}\right\}\\
&=2\mathbf{P}\left\{\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}U_if(Z_i)\right|>\frac{\epsilon}{4}\right\}.
\end{aligned}$$

- *Étape 3. Conditionnement et introduction du recouvrement.*

  Calculons l'espérance conditionnelle de cette dernière probabilité par rapport à Z_1^n, c'est équivalent à considérer

$$\mathbf{P}\left\{\exists f\in\mathcal{F}:\left|\frac{1}{n}\sum_{i=1}^{n}U_if(Z_i)\right|>\frac{\epsilon}{4}\right\},\tag{5.4}$$

  en fixant $z_1,\ldots,z_n\in\mathbb{R}^d$.

  *Soit $\mathcal{F}_{\frac{\epsilon}{8}}$ un L_1 $\frac{\epsilon}{8}-$***recouvrement** *de $\mathcal{F}$ en z_1^n et soit $f\in\mathcal{F}$. Alors, il existe $\bar{f}\in\mathcal{F}_{\frac{\epsilon}{8}}$ $(0\leq\bar{f}\leq B)$ tel que*

$$\frac{1}{n}\sum_{i=1}^{n}|f(z_i)-\bar{f}(z_i)|<\frac{\epsilon}{8}.\tag{5.5}$$

  Ce qui implique

$$\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n}U_if(Z_i)\right|&=\left|\frac{1}{n}\sum_{i=1}^{n}U_i\bar{f}(Z_i)+\frac{1}{n}\sum_{i=1}^{n}U_i\left[f(Z_i)-\bar{f}(z_i)\right]\right|\\
&\leq\left|\frac{1}{n}\sum_{i=1}^{n}U_i\bar{f}(Z_i)\right|+\frac{1}{n}\sum_{i=1}^{n}U_i\left[f(Z_i)-\bar{f}(z_i)\right]\\
&<\left|\frac{1}{n}\sum_{i=1}^{n}U_i\bar{f}(Z_i)\right|+\frac{\epsilon}{8}.
\end{aligned}$$

En utilisant cette affirmation, nous pouvons borner la probabilité donnée par (5.4) comme suit

$$\mathbf{P}\left\{\exists f \in \mathcal{F}_{\frac{\epsilon}{8}} : \left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| + \frac{\epsilon}{8} > \frac{\epsilon}{4}\right\} \leq \left|\mathcal{F}_{\frac{\epsilon}{8}}\right| \max_{f\in\mathcal{F}_{\frac{\epsilon}{8}}} \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| > \frac{\epsilon}{8}\right\}.$$

En choisissant $\mathcal{F}_{\frac{\epsilon}{8}}$ de taille minimal, nous obtenons

$$\mathbf{P}\left\{\exists f \in \mathcal{F} : \left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| > \frac{\epsilon}{4}\right\} \leq \mathcal{N}\left(\varepsilon, \mathcal{F}, z^n\right) \max_{f\in\mathcal{F}_{\frac{\epsilon}{8}}} \mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| > \frac{\epsilon}{8}\right\}.$$

- *Étape 4. Application de l'inégalité de Hoeffding.*

  Commençons par la rappeler pour des variables H_i indépendantes à valeurs dans $[a_i, b_i] \subset \mathbb{R}$, alors pour tout $\epsilon > 0$

  $$\mathbf{P}\left\{\sum H_i - \sum \mathbf{E}(H_i) \geq \epsilon\right\} \leq \left(\frac{-2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right).$$

  Bornons la dernière probabilité obtenue à l'étape 3, c'est-à-dire

  $$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| > \frac{\epsilon}{8}\right\},$$

  où $z_1, \ldots, z_n \in \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ et $0 \leq f \leq B$.

  Puisque $U_1 f(z_1), \ldots, U_n f(z_n)$ sont des variables aléatoires indépendantes avec

  $$-B \leq U_i f(z_i) \leq B, (i = 1, \ldots, n),$$

  l'inégalité de Hoeffding permet d'écrire

  $$\mathbf{P}\left\{\left|\frac{1}{n}\sum_{i=1}^{n} U_i f(Z_i)\right| > \frac{\epsilon}{8}\right\} \leq 2\exp\left(-\frac{2n(\frac{\epsilon}{8})^2}{(2B)^2}\right) \leq 2\exp\left(-\frac{n\epsilon^2}{128B^2}\right).$$

  Dans le cas où $n \geq \frac{2B^2}{\epsilon^2}$ alors le lemme est démontré après ces quatre étapes. Pour $n < \frac{2B^2}{\epsilon^2}$ la borne de la probabilité est triviale, du fait que le membre droit de l'équation est supérieur à un.

Dans la partie suivante nous allons, en premier lieu, reprendre le théorème 9.4 donné dans Györfi et al. [21] qui permet de borner le nombre englobant pour un p quelconque (cf [21] pour la preuve).

Lemme 6 (cf [21]) *Soit $\mathcal{F}$ une classe de fonctions $f : \mathbb{R}^d \to [0, B]$ vérifiant $V_{\mathcal{F}^+} \geq 2$. Alors pour tout $z_1^n \in \mathbb{R}^{n\times d}$ et tout $0 < \varepsilon < B/4$, nous avons*

$$\mathcal{M}_p\left(\varepsilon, \mathcal{F}, z_1^n\right) \leq 3\left(\frac{2eB^p}{\varepsilon^p}\log 3\left(\frac{3eB^p}{\varepsilon^p}\right)\right)^{V_{\mathcal{F}^+}},$$

où, $\mathcal{F}^+ := \{\{(z,t) \in \mathbb{R}^d \times \mathbb{R}; f(z) \geq t\}; f \in \mathcal{F}\}$ est l'ensemble de tous les graphes des fonctions de $\mathcal{F}$.

Le lemme (5) de l'appendice permet d'appliquer ce résultat au nombre recouvrant. En fait seul le cas $p = 1$ nous intéresse dans notre travail, nous énonçons donc, ci après, une version antérieure à la précédente.

Lemme 7 (cf [31], lemme 4) *Soit $\mathcal{F}$ une classe de fonctions $f : \mathbb{R}^d \to [-B, B]$. Alors pour tout $z^n \in \mathbb{R}^{n\times d}$ et tout $0 < \varepsilon < B$*

$$\mathcal{N}\left(\varepsilon, \mathcal{F}, z_1^n\right) \leq 2\left(\frac{4eB}{\varepsilon}\log 3\left(\frac{4eB}{\varepsilon}\right)\right)^{V_{\mathcal{F}^+}}.$$

La première version du lemme suivant, servant à borner la dimension VC, fut donné dans Steele (75) et Dudley (78), mais nous reprenons la preuve donnée dans Györfi et al. [21].

Lemme 8 ([21]) *Soit $\mathcal{F}$ un espace vectoriel de fonctions numériques définies sur $\mathbb{R}^d$, de dimension K. Alors la classe des ensembles $A = \{\{x \in \mathbb{R}^d : f(x) \geq 0\}, f \in \mathcal{F}\}$, admet une dimension V.C inférieure ou égale à K.*

Preuve 8 *Il suffit de montrer qu'aucun ensemble de cardinal $K + 1$, ne peut être brisé par des ensembles de la forme $\{x \in \mathbb{R}^d : f(x) \geq 0\}, f \in \mathcal{F}$.*

Choisissons d'abord un n-uplet $\{z_1, \ldots, z_n\}$ *de points distinct de* $\mathbb{R}^d$. *définissons, ensuite, l'application linéaire* $L : \mathcal{F} \to \mathbb{R}^{K+1}$ *par*

$$L(f) = (f(z_1), \ldots, f(z_{K+1}))^T, \text{ pour } f \in \mathcal{F}.$$

L'image $L\mathcal{F}$ *de* $\mathcal{F}$ *est un sous-espace (linéaire) de l'espace* $\mathbb{R}^{K+1}$, *donc sa dimension est inférieure ou égale à la dimension de* $\mathcal{F}$, *qui est au plus égale à* K. *Ainsi, il existe un vecteur non nul* $\gamma = (\gamma_1, \ldots, \gamma_{K+1})^T \in \mathbb{R}^{K+1}$, *qui est orthogonal à* $L\mathcal{F}$, *donc*

$$\gamma_1 f(z_1) + \ldots + \gamma_{K+1} f(z_{K+1}) = 0, \text{ pour tout } f \in \mathcal{F}. \tag{5.6}$$

En remplaçant γ *par* $-\gamma$ *si nécessaire, nous pouvons supposé qu'au moins une des* γ_i *est négative. L'équation (5.6) implique*

$$\sum_{i:\gamma_i \geq 0} \gamma_i f(z_i) = \sum_{i:\gamma_i < 0} (-\gamma_i) f(z_i), \text{ pour tout } f \in \mathcal{F}. \tag{5.7}$$

Supposons qu'il existe un $f \in \mathcal{F}$ *pour lequel* $\{z : f(z) \geq 0\}$, *correspond exactement à ses points* z_i *avec* $\gamma_i \geq 0$. *Pour ce* f, *le membre de gauche de l'équation (5.7) doit être positif alors que celui de droite est négatif. Ce qui constitue une contradiction et finalise la preuve.*

Le lemme qui va suivre donne une borne supérieure pour le nombre recouvrant

Théorème 8 (cf [30]) *Soit* $L, c > 0$ *et*

$$\mathcal{F} = \{\mathbb{T}_{[0,L]} f : f \in W^p(\mathbb{R}^d) \text{ et } J_p^2(f) \leq c\},$$

où $W^p(\mathbb{R}^d)$ *est l'espace de Sobolev contenant toutes les fonctions admettant des dérivées faibles d'ordre* p *et le terme de pénalité s'écrit sous la forme*

suivante

$$\lambda_n \boldsymbol{J}_p^2(f) = \lambda_n \sum_{\substack{\alpha_1,...,\alpha_d \in \mathbb{N} \\ \alpha_1+...+\alpha_d = p}} \frac{p!}{\alpha_1!...\alpha_d!} \int_{\mathbb{R}^d} \left| \frac{\partial^p f}{\partial {x_1}^{\alpha_1} ... \partial {x_d}^{\alpha_d}}(x) \right|^2 dx,$$

où $\lambda_n > 0$ *est un paramètre de l'estimation.*

Alors pour tout $0 < \epsilon < L$, *et* $z_1, \ldots, z_n \in [0,1]^d$

$$\mathcal{N}(\epsilon, \mathcal{F}, z_1^n) \leq \left(c_1 \frac{nL}{\epsilon} \right)^{c_2(\sqrt{c}/\epsilon)^{d/p}+c_3}$$

où c_1, c_2 *et* c_3 *sont des constantes positives ne dépendant que de* p *et* d.

Preuve 9 *Nous allons, en premier lieu, construire une partition rectangulaire du cube unitaire, puis nous utilisons la famille des fonctions polynomiales par morceaux sur cette partition pour obtenir une borne du nombre recouvrant.*

Soit $f \in W^p(\mathbb{R}^d)$ *avec* $J_p^2(f) \leq c$. *Soit* $\{A_1, \ldots, A_K\}$ *la partition de* $[0,1]^d$ *en rectangles de dimension* d *vérifiant*

1. $\int\limits_{A_i} \sum\limits_{\alpha_1+...+\alpha_d=p} \frac{p!}{\alpha_1!...\alpha_d!} \left| \frac{\partial^p f}{\partial x_1^{\alpha_1} ... \partial x_d^{\alpha_d}}(x) \right|^2 dx \leq c(\frac{\delta}{\sqrt{c}})^{d/p}$ *pour* $i = 1, \ldots, K$.
2. $\sup\limits_{x,z \in A_i} ||x - z||_\infty \leq (\frac{\delta}{\sqrt{c}})^{1/p}$ *pour* $i = 1, \ldots, K$.
3. $K \leq \left((\frac{\sqrt{c}}{\delta})^{1/p} + 1 \right)^d + (\frac{\sqrt{c}}{\delta})^{d/p}$.

Une telle partition existe, en effet divisons $[0,1]^d$ *en* $B_1, \ldots, B_{\tilde{K}}$ *cubes de même volume de longueur* $(\frac{\sqrt{c}}{\delta})^{1/p}$ *et tel que* $\tilde{K} \leq \left((\frac{\sqrt{c}}{\delta})^{1/p} + 1 \right)^d$. *Puis reprenons la partition sur chaque cube* B_i *pour obtenir des rectangles* $B_{i,1}, \ldots, B_{i,l_i}$

de dimension d vérifiant

$$\int_{B_{i,j}} \sum_{\alpha_1+\ldots+\alpha_d=p} \frac{p!}{\alpha_1!\ldots\alpha_d!} \left| \frac{\partial^p f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}(x) \right|^2 dx = c(\frac{\delta}{\sqrt{c}})^{d/p} \text{ pour } j = 1,\ldots,l_i-1,$$

et

$$\int_{B_{i,j}} \sum_{\alpha_1+\ldots+\alpha_d=p} \frac{p!}{\alpha_1!\ldots\alpha_d!} \left| \frac{\partial^p f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}(x) \right|^2 dx \leq c(\frac{\delta}{\sqrt{c}})^{d/p} \text{ pour } j = l_i.$$

Ce qui nous conduit à une partition de rectangles de taille

$$K = \sum_{i=1}^{\tilde{K}} l_i = \tilde{K} + \sum_{i=1}^{\tilde{K}} (l_i - 1),$$

et vérifiant 1) et 2). Comme

$$\begin{aligned} c \geq & \int_{\mathbb{R}^d} \sum_{\alpha_1+\ldots+\alpha_d=p} \frac{p!}{\alpha_1!\ldots\alpha_d!} \left| \frac{\partial^p f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}(x) \right|^2 dx \\ \geq & \sum_{i=1}^{\tilde{K}} \sum_{j=1}^{l_i-1} \int_{B_{i,j}} \sum_{\alpha_1+\ldots+\alpha_d=p} \frac{p!}{\alpha_1!\ldots\alpha_d!} \left| \frac{\partial^p f}{\partial x_1^{\alpha_1}\ldots\partial x_d^{\alpha_d}}(x) \right|^2 dx \\ = & \sum_{i=1}^{\tilde{K}} (l_i - 1) c (\frac{\delta}{\sqrt{c}})^{d/p}. \end{aligned}$$

Nous obtenons la dernière hypothèse qui est

$$K = \sum_{i=1}^{\tilde{K}} l_i = \tilde{K} + \sum_{i=1}^{\tilde{K}} (l_i - 1) \leq \left((\frac{\sqrt{c}}{\delta})^{1/p} + 1 \right)^d + (\frac{\sqrt{c}}{\delta})^{d/p}.$$

Dans la suite nous allons faire une approximation entre f et les polynômes de degré $p-1$ sur tout rectangle A_i. D'après l'intégrale unitaire de Sobolev, il existe un polynôme p_i de degré qui ne dépasse pas $p-1$ et une fonction

borné et infiniment différentiable $Q_\alpha(x,y)$ tel que pour tout $x \in A_i$

$$
\begin{aligned}
|f(x)-p_i(x)|^2 &= \left| \int_{A_i} \frac{1}{||x-z||_2^{d-p}} \sum_{\alpha_1+\ldots+\alpha_d=p} Q_\alpha(x,z) \left(\frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)(z) dz \right|^2 \\
&\leq \int_{A_i} ||x-z||_2^{2p-2d} dz \int_{A_i} \left| \sum_{\alpha_1+\ldots+\alpha_d=p} Q_\alpha(x,z) \left(\frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)(z) \right|^2 dz \\
&\leq \int_{A_i} ||x-z||_2^{2p-2d} dz \int_{A_i} \sum_{\alpha_1+\ldots+\alpha_d=p} |Q_\alpha(x,z)|^2 \sum_{\alpha_1+\ldots+\alpha_d=p} \left| \left(\frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)(z) \right|^2 d \\
&\leq \int_{A_i} ||x-z||_2^{2p-2d} dz \int_{A_i} \sum_{\alpha_1+\ldots+\alpha_d=p} \frac{p!}{\alpha_1! \ldots \alpha_d!} \left| \left(\frac{\partial^p f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \right)(z) \right|^2 dz.
\end{aligned}
$$

Utilisons 1) et 2) pour avoir

$$|f(x)-p_i(x)|^2 \leq (d(\delta/\sqrt{c})^{1/p})^{2p-2d} (\delta/\sqrt{c})^{d/p} c_0 c (\delta/\sqrt{c})^{d/p} = d^{2(p-d)} c_0 \delta^2.$$

Ce qui implique

$$\mathcal{N}((\sqrt{c_0} d^{(p-d)} + 1)\delta, \mathcal{F}, x_1^n) \leq \mathcal{N}(\delta, \mathbb{T}_L \mathcal{G}, x_1^n),$$

où $\mathbb{T}_L \mathcal{G} = \{\mathbb{T}_L g : g \in \mathcal{G}\}$ et $\mathcal{G}$ représente l'ensemble des polynômes de degré inférieur ou égale à $p-1$ avec le fait que la partition de $[0,1]^d$ sous forme de rectangle est constituée tout au plus de $K \leq (2^d+1)(\sqrt{c}/\delta)^{d/p} + 2^d$ rectangles.

Dans la dernière étape, nous allons borner le nombre de recouvrement de $\mathcal{G}$. Notons que la partition définie plus haut peut être obtenue par intersection avec les $2dK$ hyperplans tel que $\triangle_n \leq (n^d)^{2dK}$, où $\triangle_n$ est le nombre des partitions de $\{x_1, \ldots, x_n\}$ générée par intersection avec les hyperplans. La première proposition de Nobel [36] et le corolaire 29.2 de Devroye et al. (1996) [9] implique le résultat final.

Bibliographie

[1] A. Antoniodis, G. Grégoire, I. W. McKeague, Wavelet methods for curve estimation.*J. Am. Statist. Ass.*, **89**, (1994) 1340–1353.

[2] A. Antoniadis, Smoothing noisy data with traped coifletsseries. *Scand. J. Statist.*, **23**, (1996), 313–330.

[3] R. Beran, Nonparametric regression with randomly censored survival data. Technical Report, University of California, Berkeley (1981).

[4] N. Breslow et J. Crowley, A large sample study of the life table and product limit estimates under random censorship. *The Annals of statistics* **2**(3), (1974) 437–453.

[5] A. Carbonez, L. Györfi and E. C. Van der Meulen, Partitioningestimates of a regression function under random censoring. *Statistics and Decisions*, **13**, (1995) 21–37.

[6] D. M. Dabrowska, Nonparametric regression with censored data. *Scandinavian J. Statistics*, **14**, (1987) 181–197.

[7] D. M. Dabrowska, Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics*, **17**, (1989) 1157—1167.

[8] L. Devroye, Györfi, A. Krzyżak et G. Lugosi, On the strong universal consistency of nearest neighbor regression function estimates *Annals of Statistics*, **22** (1994) 1371—1385.

[9] L. Devroye, L.Györfi, G. Lugosi, *A probabilist theory of pattern Recognition*, Springer Verlag, (1996).

[10] N. R. Draper and H. Smith. *Applied Regression Analysis*, 2nd ed. Wiley, New York, (1981).

[11] D. Donoho et I. M. Johnstone, Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81** (1994) 425—455.

[12] D. Donoho, I. M. Johnstone, G. Kerkyacharian et D. Picard, Wavelet shrinkage : Asymptopia ? *Journal of the Royal Statistical Society*, Series B, **57** (1995) 301—369.

[13] D. Donoho, Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Statist. Ass.*, **90** 1200–1224.

[14] D. Donoho et I. M. Johnstone, Minimax estimation via wavelet shrinkage. *Annals of Statistics*, **26**(1998) 879—921.

[15] J. Duchon , Interpolation des fonctions de deux variables suivant le principe de la flexion des plaques minces. *R.A.I.R.O, Analyse Numérique* **10** (1976) 5–12.

[16] R. L. Eubank. *Nonparametric Regression and Spline Smoothing.* Marcel Dekker, New York, (1999).

[17] R. W. Farebrother. *Linear Least Squares Computations.* Marcel Dekker, New York, (1988).

[18] R. Gill, Large sample behaviour of the product limit estimator on the whole line. *The annals of statistics* 1**11**(1)(1983), 49–58.

[19] R. D. Gill, Lectures on survival analysis. In *Lectures on Probability Theory*, Bernard, P., editor, (1994) pages 115–241. Springer-Verlag.

[20] R. D. Gill and S. Johansen, A survey of the product-integration witn a view toward application in survival analysis, *Ann. Statist.* **18**, (1990) 1501-1555.

[21] L.Györfi, M. Kohler, A. Krzyżak, H. Walk, *A Distribution Free theory of Nonparametric Regression.* Springer-Verlag New York, Inc. (2002).

[22] D. J. Hart. *Nonparametric Smoothing and Lack-of-Fit Tests.* Springer-Verlag, New York, (1997).

[23] T. Hastie and R. J. Tibshirani. *Generalized Additive Models.* Chapman and Hall, London, U. K, (1990).

[24] D. Haussler, Decision theoretic generalizations of the PAC model for neural net and other learning applications, *Inform. Comput.* **100**, (1992) 78-150,.

[25] E. L. Kaplan, P. Meier, Nomparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.* **53**, (1958) 457-481.

[26] K. Kebabi, I.Laroussi, F. Messaci, Least squares estimators of the regression function with twice censored data, *Statist. Probab. Lett.***81**, (2011) 1588-1593.

[27] K. Kebabi et F. Messaci, Rate of the almost complete convergence of a kernel regression estimate with twice censored data. *Statist. Probab. Lett.*, **82** (11) (2012) 1908–1913.

[28] M. Kohler, On the universal consistency of a least squares spline regression estimator. *Math. Methods Statist.*, (1997) (6) 349–364.

[29] M. Kohler, Universally consistent regression function estimation using hierarchical b-splines. *J. Multivariate Anal.*, (1999) (67) 138–164.

[30] M. Kohler, A. Krzyżak, Nonparametric regression estimation using penalized least squares, *IEEE Trans. Inform. Theory.* **47**, (2001) 3054-3058.

[31] M. Kohler, K. Màthé, M. Pintér, Prediction from randomly right censored data, *J. Multivariate Anal.* **80**, (2002) 73-100.

[32] G. Lugisi et K. Zeger, Nonparametric estimation via empirical risk minimization. *IEEE Trans. Inform. Theory*, **41** (1995) 677–687.

[33] F. Messaci, Local averaging estimates of the regression function with twice censored data, *Statist. Probab. Lett.*, **80**, (2010) 1508-1511.

[34] D.Morales, L. Pardo, V. Quesada, Bayesian survival estimation for incomplete data when the life distribution is proportionally related to the censoring time distribution.*Comm. Statist. Theory Methods*, **20**, (1991) 831–850. MR1131189.

[35] E. A. Nadaraya, On estimating regression. *Theory of Probability and Its Applications*, **9** (1) (1964) 141–142.

[36] A. Nobel, Histogram Regression Estimation Using Data-dependent Partitions. *Ann. Statist.* **24**, 1084-1105.

[37] V. Patilea, J. M. Rolin, Product-limit estimators of the survival function with twice censored data, *Ann. Statist.* **34**, No 2, (2006) 925-938.

[38] A. V. Peterson. Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *J. Amer. Statist. Assoc.*, **72**, (1977) 854–858.

[39] D. Pollard, *Convergence of stochastic processes*, Springer Verlag, (1984).

[40] R. C. Rao. Linear Statistical Inference and Its Applications. Wiley, New York, 2nd edition.(1973).

[41] L. Schumaker. *Spline Functions : Basic Theory.* Wiley, New York (1981).

[42] G. A. F. Seber. *Linear Regression Analysis*. Wiley, New York (1977).

[43] B. W. Silverman, Weavelets in statisics : beyond the standard assumptions. *Phil. Trans. R. Soc. Lond.* A, **357**, 2459–2474.

[44] C. J. Stone, Consistent nonparametric regression. *Annals of Statistics*, **5** (1977) 595–645.

[45] W. Stute et J.-L. Wang, The strong law under random censorship. *The Annals of statistics* **21**(3) (1993) 1591–1607.

[46] B.W. Turnbull. The empirical distribution function with arbitrarily grouped, censored and truncated data. *Journal of the Royal Statistical Society*, **Series B, 38**, (1976), 290–295.

[47] V. N. Vapnik et A. Y. Chervonenkis, On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, **16** (1971) 264–280.

[48] V. N. Vapnik, *Estimation of Dependencies Based on Empirical Data*. Springer-Verlag, New York (1982).

[49] V. N. Vapnik, *Statistical Learning Theory*. Wiley, New York (1998).

[50] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, PA, (1990).

[51] H. Walk, Strong universal consistency of kernel and partitioning regression estimates. *Mathematisches Institut A.* Preprint 97-1, Universität Stittgart (1997).

[52] G. S. Watson, Smooth regression analysis. *Sankhya* Series A, **26** (1964) 359–372.

[53] B. B ; Winter ; A. Földes et L. Rejtő (1978), Glivenko-Cantelli theorems for the product limit estimate. *Problems of Control and Information Theory* **7**, 213–225.

تقدير دالة الانحدار بطريقة المربعات الصغرى و المربعات الصغرى المجازاة لمعطيات خاضعة لحجب مزدوج

ملخص
ضمن هذه الأطروحة سنهتم بطريقتين لتقدير دالة الانحدار دون وسائط .المسماة بطريقة المربعات الصغرى و المربعات الصغرى المجازاة.
مساهمتنا تكمن في تمديد هذه الطرق إلى متغير الإجابة خاضع إلى حجب مزدوج حيث بينا أن المقدرات المدروسة تتقارب شبه أكيد نحو القيمة المثلى.
في النهاية نقدم دراسة محاكاة الهدف منها التأكيد على جودة المقدرات المقدمة من جهة والمقارنة بينهما من جهة اخرى.
الكلمات المفتاحية : دالة الانحدار المربعات الصغرى ، المربعات الصغرى المجازة ،المعطيات الخاضعة لحجب مزدوج.

Estimation des moindres carrés et spline de lissage de la fonction de régression dans un modèle de censure

Résumé

Dans ce travail, nous nous intéressons à deux méthodes non paramétriques de la fonction de régression. A savoir la méthode des moindres carrés et celle des moindres carrés pénalisés. Notre apport se situe dans le fait que nous avons étendu ces méthodes au cas où la variable réponse est soumise à une censure mixte. Nous avons montré que les estimateurs introduits convergent presque sûrement vers la valeur optimale.

Mots clés: Fonction de régression. Moindres carrés. Moindres carrés pénalisés. Censure mixte.

Least squares and smoothing spline estimators of the regression function in a censored model

Summary

In this work, we are interested in two non-parametric methods of the regression function. Namely the method of least squares and the penalized least squares. Our contribution is in the fact that we extended these methods if the variable answer is subjected to a mixed censure. We showed that the introduced estimators almost surely converge towards the optimal value.

Key words: Function of regression. Least squares. Penalized least squares. Twice censure data.