

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR
ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENE-ALGER

FACULTE DE MATHEMATIQUES



THÈSE

Présentée pour l'obtention du diplôme de DOCTORAT

EN : MATHÉMATIQUES

Spécialité : Probabilités et Statistique

Par : GUESSOUM ZOHRA

Sujet :

RÉGRESSION NON PARAMÉTRIQUE DANS LES
MODÈLES CENSURÉS

soutenue publiquement le 21 Avril 2009, devant le jury composé de :

M.	K.	BOUKHETALA	Professeur	U.S.T.H.B	Président
M.	E.	OULD SAÏD	MCF-HDR	Univ. Littoral, France	Directeur de thèse
M.	W.	STUTE	Professeur	Univ. Giessen, Allemagne	Examinateur
Mlle.	H.	GUERBYENNE	Maître de Conférence	U.S.T.H.B	Examinatrice
M.	A.	LAKSACI	Maître de Conférence	Univ. Sidi Bel-Abbès	Examinateur
Mlle.	O.	SADKI	Maître de Conférence	U.S.T.H.B	Examinatrice

Résumé

Dans cette thèse, nous nous intéressons à l'estimation non paramétrique de la fonction de régression. Dans un premier temps, nous rappelons certains résultats établis sur le comportement asymptotique de quelques estimateurs non paramétriques de la fonction de régression, dans le cas où les données sont indépendantes et complètement observées, puis dans le cas où les données sont censurées à droite.

Dans un second temps, nous donnons nos résultats en considérant une suite de variables aléatoires $\{Y_i, i \geq 1\}$ indépendantes et identiquement distribuées (i.i.d.), de fonction de répartition H , censurée à droite par une suite $\{C_i, i \geq 1\}$ supposée i.i.d. et indépendante des v.a. $\{Y_i, i \geq 1\}$. Nous introduisons la covariable X et estimons la fonction de régression $r(x) = \mathbb{E}(Y|X = x)$. Nous donnons la vitesse de convergence uniforme presque sûre (sur un compact) de notre estimateur ainsi que la normalité asymptotique de l'estimateur convenablement normalisé. Nous obtenons des intervalles de confiance en utilisant la méthode du "plug-in". Une étude sur données simulées montre la qualité de l'estimateur ainsi que la qualité de la variable normalisée.

Mots clés : *Régression non paramétrique, Données censurées, Estimateur de Kaplan-Meier, Estimateur à noyau, Normalité asymptotique, Vitesse de convergence.*

Abstract

In this thesis, we are interested with the nonparametric estimate of the regression function. Firstly, we point out certain results established on the asymptotic behavior of some nonparametric estimators of the regression function, when the data are completely observed, and independent and identically distributed (i.i.d.). Then, we go on with the case where the data are right randomly censored.

Secondly, we give our results by considering a sequence of random variables $\{Y_i, i \geq 1\}$ independent and identically distributed, with a distribution function H , censored on the right by a sequence $\{C_i, i \geq 1\}$ supposed i.i.d. and independent from the interest r.v. $\{Y_i, i \geq 1\}$. We introduce the covariable X and estimate the regression function $r(x) = \mathbb{E}(Y|X = x)$. We give the rate of almost sure uniform convergence (over a compact set) of our estimator as well as the asymptotic normality of the suitably standardized estimator. We obtain confidence intervals by using the method of plug-in. A simulation study shows the quality of the estimator as well as the quality of the standardized variable.

Key words : *Nonparametric regression, Censored data, Kaplan-Meier estimator, Kernel estimator, Asymptotic normality, Rate of convergence.*

Remerciements

Avant toutes choses, je voudrais exprimer mes remerciements les plus sincères et les plus chaleureux à mon directeur de thèse, Monsieur le Professeur Elias Ould Saïd, pour avoir d'abord eu la patience de chercher un thème qui me motive et ensuite de m'avoir proposée ce sujet, de m'avoir guidée et encouragée, d'avoir lu et relu (combien de fois!) chaque étape de ce travail en m'aidant de ses commentaires et de son optimisme : MERCI.

Je remercie vivement Monsieur le Professeur Kamal Boukhetala de me faire l'honneur de présider ce jury, malgré ses nombreuses responsabilités.

Je suis très honoré que Monsieur le Professeur Winfried Stute ait accepté d'examiner ce travail bien qu'il ne soit pas rédigé dans sa langue et malgré ses énormes occupations et responsabilités. Qu'il trouve ici l'expression de ma sincère reconnaissance.

Mes remerciements sincères à Madame Hafida Guerbyenne qui a accepté sans hésitation d'examiner ce travail bien qu'elle soit très occupée.

J'adresse mes vifs remerciements à Mademoiselle Ourida Sadki, qui a répondu présente à chacune de mes sollicitations et qui a accepté d'examiner ce travail.

Je remercie chaleureusement Monsieur Ali Laksaci pour l'intérêt qu'il porte à mon travail en acceptant de faire partie du jury.

Je tiens à remercier également mes amis, Ouarda et Kader pour m'avoir entraînée dans cette belle aventure, en essayant de trouver (et en trouvant) toujours des réponses à mes questions, Dalila pour m'avoir écoutée me plaindre puis me reconforter, Fawzia et tous ceux qui m'ont encouragée dans cette voie .

Ma reconnaissance va aussi à tous les membres du laboratoire M.S.T.D. et aux collègues qui m'ont aidée de près ou de loin dans mon travail.

Je ne remercierai jamais assez mes chers parents et mes soeurs, qui ont toujours cru en moi, m'ont épaulée et combien de fois aidée par leurs encouragements ou par leur disponibilité; je n'en serai pas là sans eux.

*A mon père, à ma mère
et à ma p'tite Sara.*

Table des matières

1	Introduction	2
2	Régression non paramétrique : Cas de données complètes	8
2.1	Méthodes de régression	8
2.1.1	Méthode du noyau	9
2.1.1.1	L'estimateur à noyau de la densité	9
2.1.2	Quelques estimateurs à noyau de la fonction de régression :	11
2.1.2.1	L'estimateur de Nadaraya-Watson	11
2.1.2.2	L'estimateur de Gasser et Müller	12
2.1.2.3	L'estimateur des k-plus proches voisins (k-NN)	13
2.1.2.4	L'estimation par polynômes locaux	14
2.2	Propriétés asymptotiques de l'estimateur de Nadaraya-Watson	17
2.2.1	Consistance	17
2.2.1.1	Calcul de la variance :	17
2.2.1.2	Calcul du biais :	20
2.2.2	Optimalité asymptotique et choix de la fenêtre	23
2.2.2.1	Critère de selection local : MSE	23
2.2.2.2	Critère de selection global : MISE	24
2.2.2.3	La validation croisée	25
2.2.3	La normalité asymptotique	27
3	Régression non paramétrique : Cas de données censurées	31
3.1	Durées de vie	31
3.1.1	Caractérisation de la loi d'une durée de vie	31
3.1.2	Les trois types de censure	32
3.1.2.1	Censure de type I : fixée	32
3.1.2.2	Censure de type II : Attente (jusqu'au r-ième décès)	32

3.1.2.3	Censure de type III : aléatoire	32
3.2	Estimation de la fonction de survie	33
3.2.1	Estimateur de Nelson-Aalen	34
3.2.2	Estimateur de Kaplan-Meier	35
3.2.2.1	Propriétés asymptotiques de l'EKM	36
3.3	La régression non paramétrique pour des données censurées	39
3.3.1	Estimation de la régression dans le cas où (X, Y) est indépendant de C .	41
3.3.2	Estimation de la régression dans le cas où Y et C sont indépendantes conditionnellement à X	45
4	Vitesse de convergence et normalité asymptotique de l'estimateur de Nadaraya-Watson dans le modèle censuré	49
4.1	Introduction	49
4.2	Construction of the estimator of $r(\cdot)$	51
4.3	Assumptions and main results	53
4.4	Simulation study	55
4.4.1	Consistency	55
4.4.2	Asymptotic normality	57
4.5	Proofs	59
5	Conclusion et Perspectives	67
5.1	Conclusion	67
5.2	Perspectives	68
	Bibliographie	69

Chapitre 1

Introduction

Lorsque l'on souhaite décrire l'influence d'une variable quantitative sur un événement, ou le lien entre une variable explicative X et la variable réponse Y , on utilise des modèles dits de régression. Ayant observé X , la valeur moyenne de Y est donnée par une fonction de régression : c'est cette fonction qui nous renseigne sur le type de dépendance qu'il y a entre ces deux variables.

Considérons n réalisations indépendantes $(X_i, Y_i)_{i=1, \dots, n}$ du couple de variables aléatoires réelles (X, Y) , la relation de régression est modélisée par :

$$Y_i = r(X_i) + \varepsilon_i \quad i = 1, \dots, n \quad (1.1)$$

où $r(\cdot)$ est la fonction de régression inconnue et ε_i sont les observations d'erreurs. La représentation graphique de X_i et Y_i n'est pas suffisante pour établir une relation interprétable ; le but de l'analyse de régression est alors d'approximer la moyenne conditionnelle $r(X) := \mathbb{E}[Y|X]$, ce qui peut-être fait essentiellement de deux manières :

- L'approche paramétrique qui suppose que $r(X)$ est une fonction ayant une forme "connue" contenant des paramètres inconnus. Par exemple, la régression linéaire où l'on suppose que $r(X_i) = aX_i + b$ avec a et b les paramètres inconnus. On sait qu'à partir de ce type de modèle, on dispose d'une mesure synthétique du lien qui lie Y à X , d'écart-type et d'intervalles de confiance associés aux paramètres et aux valeurs prévues de la variable Y et de tests simples (inférence) à mettre en oeuvre sur la valeur des paramètres du modèle.

- L'approche non paramétrique qui consiste à ne donner aucune forme spécifique à $r(X)$. Le principal avantage (ou inconvénient, suivant le point de vue adopté) de cette approche c'est qu'elle ne nécessite aucune hypothèse à priori sur la forme du lien entre X et Y .

La question de savoir, quelle approche prendre dans une analyse de régression, a fait l'objet d'une rude bataille entre Pearson et Fisher au début du vingtième siècle. Fisher soutenait que l'approche non paramétrique donnait généralement une efficacité plus faible pour les estimateurs alors que Pearson était plus intéressé par la question spécifique, c-à-d déterminer la forme de la fonction de régression.

Une autre approche est de combiner les deux précédentes en une méthode appelée semi paramétrique. Cependant, l'approche non paramétrique (que l'on utilisera pour nos résultats) est plus intéressante pour différentes raisons : d'abord elle permet de faire des prévisions sans référence à un quelconque modèle paramétrique ; de plus, elle fournit un outil pour détecter les fausses observations en étudiant l'influence des points aberrants, enfin elle constitue une méthode de substitution pour les valeurs manquantes.

L'inférence statistique sur la fonction $r(\cdot)$ dépend des hypothèses qu'il est possible de faire sur cette fonction. Un modèle de régression paramétrique suppose que la forme de $r(\cdot)$ est connue à quelques paramètres près. Plus précisément, on suppose qu'il y'a un vecteur de paramètres $\beta = (\beta_0, \dots, \beta_p)^t \in \mathbb{R}^{p+1}$ et une fonction connue $r(\cdot, \beta)$ telle que $r(\cdot) = r(\cdot, \beta)$. Il est alors clair que pour un modèle paramétrique l'inférence sur $r(\cdot)$ est équivalente à l'inférence sur β .

Un modèle paramétrique peut être linéaire ou non linéaire, par exemple $r(x) = \beta_0 + \beta_1 x$ est un modèle linéaire alors que $r(x) = \beta_0 + \beta_1 \exp\{-\beta_2 x^{\beta_3}\}$ est un modèle non linéaire. En utilisant une méthode d'estimation appropriée, par exemple les moindres carrés, on peut se servir des données pour estimer les paramètres et donc la fonction $r(\cdot)$.

Le deuxième aspect de la régression est celui des méthodes non paramétriques. Ce sont des méthodes (qu'on discutera plus en détails dans le chapitre deux) permettant une grande flexibilité dans l'estimation de $r(\cdot)$, et en particulier ne nécessitant aucune condition paramétrique sur $r(\cdot)$. En général les hypothèses faites sur la fonction de régression sont du type différentiabilité, ou différentiable à dérivée seconde intégrable etc...

Quand elles sont appropriées, les méthodes paramétriques ont des avantages certains et de bonnes propriétés d'efficacité. Elles sont néanmoins déconseillées voire incorrectes lorsqu'on a peu d'informations sur la fonction $r(\cdot)$, et l'utilisation d'un cadre paramétrique peut s'avérer dangereux dans le sens où on obtient une inférence trompeuse sur la fonction de régression. Gasser *et al.* (1984) ont utilisé les méthodes non paramétriques pour montrer que certains modèles paramétriques appliqués dans l'étude de l'évolution de la taille d'une population, ont négligé certaines caractéristiques importantes dans cette étude.

Généralement les estimateurs non paramétriques sont moins efficaces que les estimateurs paramétriques lorsque le modèle paramétrique est valide. Pour la plupart des estimateurs paramétriques, la fonction de risque ou l'erreur quadratique moyenne décroît vers 0 avec un taux de

l'ordre de n^{-1} , alors que les estimateurs non paramétriques ont un taux de l'ordre de $n^{-\delta}$ avec $0 < \delta < 1$. A titre d'exemple si $r(\cdot)$ est deux fois différentiable le taux est de l'ordre de $n^{-4/5}$. Historiquement, le principe des régressions non paramétriques remonte au 19ème siècle selon Cleveland et Loader (1996), toutefois les premiers travaux modernes sur ce sujet datent des années cinquantes : l'estimation de fonctions de densité par des méthodes d'opérateurs à noyaux (kernel) avec les travaux fondateurs de Rosenblatt (1956) qui a proposé un estimateur de la densité et Parzen (1962) qui a contribué à la théorie asymptotique dans le cas univarié ; ceux ci ont été étendus à la régression par la méthode du noyau (régression avec lissage par un opérateur noyau), dans deux papiers fondateurs publiés la même année : Nadaraya (1964) et Watson (1964). Par la suite est venue la régression locale polynomiale, Cleveland et Devlin (1988), qui généralise ces méthodes.

L'estimation non paramétrique et en particulier l'estimation de la régression, constitue un champ important de la statistique depuis deux ou trois décennies ; pour s'en convaincre, il suffit de consulter les revues bibliographiques de Collomb (1981) et (1985) qui, dès le début des années 80 font déjà état de développements nombreux et variés sur ce thème, puis l'ouvrage collectif de Schimek (2000) qui dresse un peu le bilan actuel des diverses connaissances en la matière. D'autres ouvrages généraux sur ce thème incluent ceux de Bosq et Lecoutre (1987). Ce champs de recherche est potentiellement porteur à la fois de développements théoriques intéressants et de multiples possibilités d'applications.

Notre principal objectif est de faire de la régression sur des durées de vie, c'est à dire sur des variables aléatoires représentant la durée de temps jusqu'à l'apparition d'un certain événement qui peut représenter, en médecine, la durée de survie d'un patient, en industrie, la durée séparant deux pannes successives, en économie la durée de chômage etc.... Si on note Y cette v.a., qu'on est sur des données complètement observées (c'est-à-dire non censurées ou non tronquées) et qu'on dispose d'un échantillon $\{Y_i, 1 \leq i \leq n\}$ de taille n de la v.a. Y , où les Y_i sont indépendants et identiquement distribués (i.i.d.) de densité commune f , l'estimateur de la densité obtenu par Rosenblatt et Parzen qu'on note $\hat{f}_{PR,n}$ est ce qu'on appelle un estimateur à noyau c'est à dire un estimateur basé sur une fonction noyau qu'on note K et un paramètre de lissage appelé fenêtre et noté h , comme on le verra au chapitre deux. Si le choix du noyau K est réputé comme peu influent sur l'estimateur, il n'en est pas de même pour la fenêtre h . En effet en minimisant l'écart quadratique moyen intégré (MISE) entre la densité f de la v.a. Y et son estimateur $\hat{f}_{PR,n}$

$$MISE(h) = \mathbb{E} \int \left(f(x) - \hat{f}_{PR,n}(x) \right)^2 dx$$

et en approximant le MISE, Silverman (1986) a montré que

$$h_{opt} = \frac{c_1^{-2/5} c_2^{1/5} c_3^{-1/5}}{n^{1/5}}$$

où $c_1 = \int x^2 K(x) dx$, $c_2 = \int K^2(x) dx$ et $c_3 = \int (f''(x))^2 dx$. Pour ce h optimal, la fonction à noyau qui minimise le MISE est le noyau d'Epanechnikov défini par

$$K(x) = \frac{3(1 - \frac{x^2}{5})}{4\sqrt{5}} \mathbb{1}_{(|x| < \sqrt{5})}$$

où $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A .

Si f est gaussienne et si le noyau K est gaussien, Silverman (1986) a montré que

$$h_{opt} = 1.059 \hat{\sigma} n^{-1/5}$$

où $\hat{\sigma}$ est un estimateur robuste de l'écart-type de la loi en question. En pratique on prend $h_{opt} = 0.9 \min(\hat{\sigma}, \frac{\hat{q}_3 - \hat{q}_1}{1.349}) n^{-1/5}$ où $\hat{q}_3 - \hat{q}_1$ désigne l'écart interquartile empirique et on prend 0.9 au lieu de 1.059 pour réduire le risque de surlissage. Finalement la sélection de la fenêtre optimale dépend de la connaissance de f !

Une autre méthode de sélection de la fenêtre est la méthode par validation croisée (qu'on verra au chapitre deux), qui consiste à minimiser, pour la densité, l'erreur quadratique intégrée (ISE) définie par

$$ISE(h) = \int \left(\hat{f}_{PR,n}(x) - f(x) \right)^2 dx. \quad (1.2)$$

Les techniques d'estimation de la fonction de régression sont similaires à celles développées pour l'estimation de la densité, nous commençons donc par rappeler quelques références sur cette dernière. De nombreux auteurs se sont intéressés à l'estimation de la densité ; sans être exhaustif, on peut citer Devroye et Penrod (1984) qui ont établis la convergence uniforme de l'estimateur des k -voisins les plus proches de la densité, Giné et Guillou (2002) qui ont montré la convergence uniforme de l'estimateur à noyau de la densité en donnant une vitesse de convergence de l'ordre de $O\left(\left[\frac{|\log h_n|}{nh_n}\right]^{1/2}\right)$ en utilisant la technique des classes de *Vapnik - Červonenkis* (les VC classes) mais en imposant plus de conditions sur la fenêtre h_n dont $|\log h_n| / \log \log n$ tendant vers l'infini, ainsi que $h_n = ch_{2n}$. Leurs résultats améliorent et étendent au cas multivarié les résultats de Silverman (1978). Giné *et al.* (2004) ont étendu à \mathbb{R}^d les travaux de Stute (1984) sur le comportement asymptotique de l'estimateur de la densité sur des compacts de \mathbb{R}^d . Leur

résultat a ensuite été repris par Dony et Einmahl (2006) qui l'a étendu en le montrant uniformément par rapport à h . Dans le cas censurés Zhang (1998) a montré la convergence uniforme presque sûre de l'estimateur à noyau de la densité en utilisant la théorie des martingales et a donné une vitesse de convergence de l'ordre de $O(|\log h_n|/nh_n)$.

Pour ce qui est de l'estimation de la fonction de régression dans le cas de données complètes, qui fait l'objet du chapitre deux, Stone (1977) fut le premier à établir un résultat de consistance faible de l'estimateur de la fonction de régression, en étudiant l'estimateur du plus proche voisin. Depuis, Devroye (1978) a établi la convergence uniforme presque sûre de l'estimateur de Nadaraya-Watson sous des conditions plus faibles que celles imposées par Nadaraya (1970), entre autres, il ne suppose pas la continuité du noyau K . Devroye et Wagner (1980) ainsi que Spiegelman et Sacks (1980) ont établi, indépendamment, la consistance pour une certaine classe d'estimateurs à noyaux pour lesquels ils donnent une vitesse de convergence de l'ordre de $O(n^{-2/2+d})$ où d représente la dimension du vecteur de covariables X , Gordon et Olshen (1980) et Breiman *et al.* (1983) la consistance pour les estimateurs par partitionnement. Stone (1980, 1982) a donné le taux de convergence optimal pour la régression non paramétrique.

Le chapitre deux est donc consacré à la régression non paramétrique dans le cas où les données sont complètes et i.i.d. Nous commençons le chapitre par l'estimation de la densité pour laquelle nous rappelons les différentes étapes d'estimation avant d'arriver à l'estimation par la méthode du noyau. Nous rappelons ensuite les définitions d'un certain nombre d'estimateurs de la fonction de régression dont l'estimateur de Nadaraya-Watson, celui de Gasser-Muller, les k -voisins les plus proches et les polynômes locaux, en donnant pour quelques un d'entre eux certaines propriétés de consistance et en insistant sur l'estimateur de Nadaraya-Watson (qui est un cas particulier de l'estimateur par polynômes locaux), pour lequel nous montrons la consistance au sens de la moyenne quadratique en calculant explicitement le biais et la variance asymptotiques. Nous étudions aussi l'optimalité asymptotique en choisissant la fenêtre h_n optimale par les critères de l'erreur quadratique moyenne (MSE) et de l'erreur quadratique moyenne intégrée (MISE). Nous remarquons, en utilisant ces critères, que le choix de la fenêtre dépend de la fonction de régression et de la densité qui sont inconnues, d'où la recherche d'une alternative à ces procédures par la méthode de la validation croisée. Nous clôturons ce chapitre par la normalité asymptotique dont la première démonstration est due à Schuster (1972), qui a étendu le résultat de Nadaraya (1964) et nous rappelons un théorème de normalité dont la preuve est due à Bierens (1987) légèrement différente de celles qu'on retrouve habituellement dans la littérature (cf Schuster 1972 et Lemdani et Ould-Saïd 2006)

Dans le chapitre trois nous abordons la régression dans le cas de données censurées et nous rappelons rapidement quelques définitions en particulier l'estimateur de Kaplan-Meier pour la

fonction de survie, pour lequel nous donnons les propriétés asymptotiques c'est-à-dire la consistance et la normalité asymptotique. Nous étudions par la suite l'estimation de la fonction de régression dans deux modèles différents : le premier modèle suppose l'indépendance entre la variable de censure C et le couple (Y, X) formé par la variables durée de vie et la covariable. Sous ce modèle et sous certaines conditions sur les fonctions de survie de Y et de C respectivement, nous donnons des résultats de consistance au sens de la moyenne quadratique pour deux estimateurs particuliers, l'estimateur de Nadaraya-Watson et celui des k -voisins les plus proches. Les mêmes résultats ont été établis pour le deuxième modèle, où on suppose que Y et C sont indépendantes conditionnellement à X .

Le quatrième chapitre est consacré à la présentation de nos résultats qui est la contribution principale de cette thèse. Nous étudions l'estimateur de Nadaraya-Watson dans le cas d'un modèle censuré et dans lequel nous supposons l'indépendance entre le couple (Y, X) et la variable C , nous établissons un résultat de convergence uniforme presque sûre sur des compacts avec une vitesse de convergence de l'ordre de $O\left(\max\left\{\frac{\log n}{nh_n}, h_n^2\right\}\right)$, en utilisant des inégalités de type exponentielle, plus précisément l'inégalité de Bernstein. Nous faisons remarquer que si nous choisissons h_n de l'ordre de $O\left[\left(\frac{\log n}{n}\right)^{1/5}\right]$ nous obtenons la vitesse optimale de convergence de l'ordre de $O\left[\left(\frac{\log n}{n}\right)^{2/5}\right]$, optimalité au sens du critère de la convergence uniforme de l'estimation de la densité (cf. Stute 1982), puis nous établissons la normalité asymptotique de cet estimateur, et donnons des intervalles de confiance pour la fonction de régression $r(x)$. Des simulations viennent illustrer nos résultats et prouvent leurs performances à échantillon fini. Nous faisons remarquer aussi que la vitesse de convergence, que nous avons calculé en utilisant la technique des VC classes, est de l'ordre de $O\left(\max\left\{\frac{\log n}{nh_n^2}, h_n^2\right\}\right)$ et est donc moins bonne que celle que nous obtenons avec Bernstein. Les résultats sont présentés dans la version originale c'est-à-dire celle à paraître dans la revue *Statistics and Decision*.

Le dernier chapitre présente les conclusions de notre travail en récapitulant ce que nous avons présenté et en donnant quelques perspectives et extensions possibles.

Chapitre 2

Régression non paramétrique : Cas de données complètes

2.1 Méthodes de régression

La méthode la plus communément utilisée pour étudier la relation entre deux variables est la régression linéaire simple, qui suppose un modèle de la forme

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$$

où les erreurs aléatoires ε_i sont non corrélées, de moyennes nulles et de variances σ^2 indépendantes de X . Cette méthode paramétrique possède l'avantage d'être facile à interpréter et, lorsque les postulats sur les résidus ε_i sont vérifiés, permet de faire des tests d'hypothèses statistiques formels sur les paramètres. Par contre lorsque la linéarité de la relation est mise en défaut, on préfère choisir un modèle plus flexible qui reflète mieux la relation entre X et Y . On utilise le modèle de régression non paramétrique

$$Y_i = m(X_i) + \varepsilon_i$$

où ε_i représente la variation de Y_i autour de X_i et obéit aux mêmes hypothèses que dans le cas linéaire et à part certaines conditions de continuités et de lissages, il n'y a habituellement aucune contrainte associée à $m(X)$. On cherchera, dans une famille fixée de fonctions, quelle est celle pour laquelle les Y sont les plus proches de $m(X)$. Cette proximité se mesure en général par un risque utilisant l'erreur quadratique moyenne (MSE en anglais), et on essaiera alors de déterminer la fonction $m^*(X_i)$ qui rendra cette erreur la plus petite possible, c'est à dire à trouver une fonction $m^*(X_i)$ qui minimise l'erreur quadratique moyenne :

$$\mathbb{E} |m^*(X) - Y|^2 = \min_m \mathbb{E} |m(X) - Y|^2.$$

Il est connu que ce minimum est donné par l'espérance conditionnelle :

$$r(x) := \mathbb{E}(Y | X = x), \quad (2.1)$$

en effet, déterminer $m^*(x)$ revient à calculer $\arg \min_m \mathbb{E}(m(X) - Y)^2 | X = x)$. En différenciant l'espérance $\mathbb{E}(m(X) - Y)^2 | X = x)$ par rapport à $m(\cdot)$, en égalant le résultat à 0, et finalement en isolant $m(\cdot)$, on obtient :

$$\begin{aligned} \frac{\partial}{\partial f} \mathbb{E}(m(X) - Y)^2 | X = x) &= 2\mathbb{E}(m(X) - Y) | X = x) \\ &= 2m(X) - \mathbb{E}(Y | X = x) \\ &= 0 \end{aligned}$$

ce qui implique que

$$m^*(x) = \mathbb{E}(Y | X = x) =: r(x).$$

Le fait que la dérivée seconde, qui est égale à 2, soit positive nous permet de conclure que c'est bien un minimum.

2.1.1 Méthode du noyau

Les estimateurs du type noyau introduits indépendamment par Nadaraya (1964) et Watson (1964) sont une des techniques les plus populaires d'estimation pour des modèles de régression du type (2.1). L'estimateur obtenu est en fait une moyenne pondérée des valeurs y_i pour les i dont le point x_i est proche du point d'estimation. La première étape consiste à choisir un point d'estimation x_0 , c'est à dire une valeur de x pour laquelle on veut estimer $r(x_0)$. Ensuite une fonction de noyau symétrique autour de 0 et unimodale doit être choisie, qui doit être maximale en 0 et à support compact (à l'exception du noyau gaussien). L'estimation par les méthodes dites à noyau a d'abord commencé avec l'estimation de la densité. L'étude du comportement asymptotique de ces estimateurs ainsi que les différentes techniques utilisées ont été étendus aux estimateurs de la fonction de régression. Nous commencerons donc par donner quelques résultats sur l'estimation non paramétrique de la densité.

2.1.1.1 L'estimateur à noyau de la densité

Soit X_1, X_2, \dots, X_n un n -échantillon d'une variable aléatoire X de densité f . On répartit les observations dans k classes en choisissant la borne inférieure de la première classe, et on obtient

l'estimateur par histogramme de la densité :

$$\hat{f}(x) = \frac{1}{n} \frac{\#d'observations\ dans\ la\ classe\ contenant\ x}{longueur\ de\ la\ classe}. \quad (2.2)$$

Si on choisit des classes de même longueur h , (2.2) devient :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \mathbb{1}_{\{z_i < x < z_{j+1}\}}$$

avec $x \in [z_i, z_{i+1}]$ la i ème classe. Il est évident que cet estimateur est sensible à la fois au nombre de classes choisies mais aussi à la borne inférieure de la 1ère classe, en particulier en déplaçant légèrement l'extrémité d'une classe on peut modifier substantiellement $\hat{f}(x)$ en un point x qui passerait d'une classe à fort effectif à une classe à faible effectif. Pour s'affranchir de ces contraintes, une première modification de $\hat{f}(x)$ a été faite et on obtient l'estimateur de la "fenêtre mobile" :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n \mathbb{1}_{\{x - \frac{h}{2} < X_i < x + \frac{h}{2}\}}. \quad (2.3)$$

Soit $K_0(x)$ la fonction définie par :

$$K_0(x) = \mathbb{1}_{\{|x| < \frac{1}{2}\}}$$

alors (2.3) s'écrit :

$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^n K_0\left(\frac{x - X_i}{h}\right)$$

K_0 n'est rien d'autre que la densité de la loi uniforme sur $[0, 1]$, son rôle est de séparer les X_i de l'échantillon en ne gardant que ceux qui sont proches de x . Pour remédier au caractère trop brutal de cette dichotomie, Parzen et Rosenblatt ont eu l'idée de remplacer K_0 par une densité de probabilité K continue et bornée, tendant assez vite vers zéro à l'infini : l'idée instinctive est celle d'une loi Gaussienne centrée réduite. On voit alors que si on substitue K_0 par une fonction de ce type, on obtiendra une somme où les valeurs correspondantes à des X_i loin de x sont infinitésimales, les autres de plus en plus fortes avec la proximité de x et X_i : on maintient l'effet initial cherché mais en le "lissant". On définit ainsi une classe générale d'estimateurs de f : l'estimateur de la densité de Parzen-Rosenblatt, en posant :

$$\hat{f}_{PR}(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right). \quad (2.4)$$

Considérée comme fonction de x , \hat{f}_{PR} est elle-même une densité continue bornée, tendant vers 0 à l'infini et donc plus ressemblant à f qu'un histogramme ou une fenêtre mobile! $K(x)$ est

ce qu'on appelle une fonction à "noyau", et h_n est appelée fenêtre (bandwidth en anglais) et c'est en fait une suite de nombres indexés par n strictement positifs et tendant vers 0 lorsque n tend vers l'infini. La fonction K définie de \mathbb{R} dans \mathbb{R} sera supposée mesurable et satisfaisant certaines hypothèses basiques parmi celles énoncées ci-dessous :

$$(K1) \quad K \text{ est bornée,}$$

$$(K2) \quad \lim_{t \rightarrow +\infty} |t| K(t) = 0,$$

$$(K3) \quad K(\cdot) \in L_1(\mathbb{R}),$$

$$(K4) \quad \int K(t) dt = 1.$$

Le tableau ci dessous donne un aperçu de certaines fonctions à noyau pouvant être employées :

noyau gaussien	$K(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},$
Epanechnikov	$K(x) = \frac{3(1-\frac{x^2}{5})}{4\sqrt{5}} I(x < \sqrt{5}),$
noyau triangulaire	$K(x) = (1 - x) I(x < 1),$
noyau rectangulaire	$K(x) = \frac{1}{2} I(x < 1).$

2.1.2 Quelques estimateurs à noyau de la fonction de régression :

2.1.2.1 L'estimateur de Nadaraya-Watson

Notons que la fonction $r(x) = \mathbb{E}(Y | X = x)$ peut s'écrire sous la forme

$$r(x) = \frac{\int y f_{X,Y}(x, y) dy}{f(x)} =: \frac{\bar{r}_1(x)}{f(x)}. \quad (2.5)$$

Un estimateur à noyau de la densité conjointe $\hat{f}_{X,Y}$ peut être obtenu en suivant les mêmes étapes que pour $\hat{f}_{PR}(x)$, et on a alors :

$$\hat{f}_{X,Y}(x, y) = \frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right) K\left(\frac{y - Y_i}{h_n}\right) \quad (2.6)$$

et on obtient l'estimateur de Nadaraya- Watson (1964) (NW)

$$\hat{r}_{NW;n}(x) := \frac{\frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} =: \frac{\hat{r}_{1,n}(x)}{\hat{f}_{PR}(x)} \quad (2.7)$$

en effet si on remplace la densité conjointe $f_{X,Y}$ par son estimateur défini dans (2.6) dans le numérateur de (2.5) on a :

$$\begin{aligned} \int y \hat{f}_{X,Y}(x,y) dy &= \int \frac{1}{nh_n^2} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) y K\left(\frac{y-Y_i}{h_n}\right) dy \\ &= \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) \int \frac{1}{h_n} y K\left(\frac{y-Y_i}{h_n}\right) dy. \end{aligned}$$

Si on suppose de plus que le noyau K est symétrique on montre en faisant le changement de variables $u = \frac{y-Y_i}{h_n}$ que $\int \frac{1}{h_n} y K\left(\frac{y-Y_i}{h_n}\right) dy = Y_i$, d'où l'expression (2.7). Le noyau K détermine la forme du voisinage autour du point x alors que la fenêtre h_n contrôle la taille de ce voisinage, c'est à dire le nombre d'observations prises pour effectuer la moyenne locale : il est alors naturel que la fenêtre h_n soit prépondérante pour la consistance de l'estimateur $\hat{r}_{NW;n}$. En fait, h_n détermine le degré de lissage de l'estimateur $\hat{r}_{NW;n}$. Supposons que l'estimateur soit seulement évalué aux points d'observations $\{X_i, i = 1, \dots, n\}$, alors, lorsque K est à support compact, on obtient :

$$\lim_{h_n \rightarrow 0} \hat{r}_{NW;n}(X_i) = \frac{K(0)Y_i}{K(0)} = Y_i$$

plus précisément on a :

$$\lim_{h_n \rightarrow 0} \hat{r}_{NW;n}(x) = \begin{cases} Y_i & \text{si } x = X_i \quad 1 \leq i \leq n \\ 0 & \text{ailleurs} \end{cases}$$

ce qui veut dire que lorsque h_n tend vers 0 l'estimateur $\hat{r}_{NW;n}$ a tendance à reproduire les données, la courbe obtenue est proche d'une interpolation des points $\{(X_i, Y_i), i = 1, \dots, n\}$: c'est un phénomène de sous-lissage (la variance de l'estimateur est trop grande). Inversement,

$$\lim_{h_n \rightarrow +\infty} \hat{r}_{NW;n}(x) = \frac{\sum_{i=1}^n K(0)Y_i}{\sum_{i=1}^n K(0)} = \frac{1}{n} \sum_{i=1}^n Y_i$$

on voit alors que l'estimateur $\hat{r}_{NW;n}$ ne dépend plus de x : lorsque h_n tend vers l'infini, on a un phénomène de sur-lissage (le biais est trop grand). Ces deux comportements nous conduisent à penser que le choix de h_n est en fait un arbitrage biais/variance, que nous étudierons plus en détails dans le comportement asymptotique de l'estimateur NW.

2.1.2.2 L'estimateur de Gasser et Müller

Le dénominateur aléatoire de l'estimateur de Nadaraya-Watson est un inconvénient majeur, notamment pour l'étude des dérivées de cet estimateur. Dans le cadre du dispositif expérimental

où les X_i sont ordonnés, Gasser et Müller (1979), ont proposé l'estimateur suivant

$$\hat{r}_{GM;n}(x) := \sum_{i=1}^n \left(\int_{s_{i-1}}^{s_i} Y_i K\left(\frac{x-u}{h_n}\right) du \right)$$

où $s_i = \frac{X_i + X_{i+1}}{2}$, $X_0 = -\infty$ et $X_{n+1} = +\infty$.

L'estimateur de Gasser et Müller est une modification d'une version antérieure développée par Priestley et Chao (1972). Pour une étude complète de l'estimateur de Gasser et Müller nous citons l'ouvrage de Müller (1988).

2.1.2.3 L'estimateur des k-plus proches voisins (k-NN)

L'estimateur des k -plus proches voisins (en anglais k th nearest neighbor, k -NN) est une moyenne pondérée dans un voisinage qui varie : ce voisinage est défini à partir des variables X choisies parmi les k plus proche de x . Cet estimateur a été introduit par Loftsgaarden et Quesenberry (1965) dans un problème d'estimation de la densité et a été utilisé par Cover et Hart (1967) pour faire de la classification. En régression, L'estimateur k -NN est défini par

$$\hat{r}_{k;n}(x) := \frac{1}{n} \sum_{i=1}^n W_{k,i}(x) Y_i$$

où les poids $(W_{ki}(x))_{i=1,\dots,n}$ sont définis à partir de l'ensemble

$J_x = \{i / X_i \text{ est l'une des } k \text{ observations les plus proches de } x\}$, et on a

$$W_{k,i}(x) = \begin{cases} \frac{n}{k} & \text{si } i \in J_x \\ 0 & \text{sinon.} \end{cases}$$

Pour fixer les idées, on peut considérer l'exemple suivant :

soit $(X_i, Y_i)_{i=1,\dots,5}$ définis par $\{(1, 5), (7, 12), (3, 1), (2, 0), (5, 4)\}$. Calculons l'estimateur des 3-plus proches voisins pour $x = 4$ (on a donc $k = 3$). Les 3 observations les plus proches de x sont les 3 dernières d'où $J_x = J_4 = \{3, 4, 5\}$ ce qui nous permet de construire d'abord les poids :

$W_{3,1}(4) = W_{3,2}(4) = 0$; $W_{3,3}(4) = \frac{5}{3} = W_{3,4}(4) = W_{3,5}(4)$; et ensuite l'estimateur $\hat{r}_{3;5}(4) = \frac{1}{5} \sum_{i=1}^n W_{k,i}(x) Y_i = \frac{5}{3}$.

Le paramètre k contrôle le degré de lissage de la courbe estimée : il joue un rôle similaire à celui de h_n dans l'estimation par noyau. Si k est grand, le k -NN estimateur est alors égal à la moyenne des Y_i . Si $k = 1$, le 1-NN estimateur reproduit une des observations Y_i associée à l'observation X_i la plus proche de x . Ceci conduit à poser $k = k_n$ qui doit tendre vers l'infini pour avoir une variance réduite, et en même temps le biais doit être faible, ce qui est réalisé si

le voisinage autour de x tend asymptotiquement vers 0, c'est-à-dire si $\frac{k_n}{n}$ tend vers 0 : ces deux conditions constituant un arbitrage biais-variance, qui asymptotiquement conduit à prendre $k \simeq n^{4/5}$ grâce à la proposition suivante

Proposition 2.1. (Lai 1977).

Soit $k \rightarrow +\infty$ et $\frac{k}{n} \rightarrow 0$ lorsque $n \rightarrow +\infty$. Le biais et la variance de l'estimateur k -NN, $\hat{r}_{k;n}(x)$ sont donnés par

$$\begin{aligned}\mathbb{E}(\hat{r}_{k;n}(x)) - r(x) &\simeq \frac{1}{2x(f(x))^3} [(r''f + 2r'f')(x)] \left(\frac{k}{n}\right)^2 \\ \text{Var}(\hat{r}_{k;n}(x)) &\simeq \frac{\sigma^2(x)}{k}.\end{aligned}$$

En effet le choix de $k = O(n^{4/5})$ implique que l'erreur quadratique moyenne (MSE) converge vers 0 avec un taux de l'ordre de $\frac{1}{k} = O(n^{-4/5})$.

2.1.2.4 L'estimation par polynômes locaux

L'estimation de la fonction de régression par la méthode des polynômes locaux (LP) est une généralisation de l'estimateur de Nadaraya-Watson. L'idée maitresse de cette approche est de considérer le problème de la régression sous l'angle des moindres carrés (M.C.). Nous rappelons que la fonction de régression est elle-même solution d'un problème des M.C. Si on reprend l'estimateur de Nadaraya-Watson, alors on a d'après (2.7)

$$\hat{r}_{1,n}(x) - \hat{r}_{NW;n}(x) \hat{f}_{PR}(x) = 0.$$

L'estimateur de la régression $\hat{r}_{NW;n}$ peut donc être regardé comme la solution du problème des moindres carrés (M.C.) pondérés suivant

$$\arg \min_{\theta \in \mathbb{R}} \sum_{i=1}^n (Y_i - \theta)^2 K\left(\frac{x - X_i}{h_n}\right)$$

c'est à dire que l'estimateur $\hat{r}_{NW;n}$ est obtenu par une approximation des M.C. localement constante. Le principe de l'estimation localement polynomiale consiste en l'ajustement local d'un polynôme de degré p aux données $(X_i, Y_i)_{i=1, \dots, n}$. Pour $p \in \mathbb{N}$, nous cherchons à ajuster le polynôme

$$a_0 + a_1(\cdot - x) + a_2(\cdot - x)^2 + \dots + a_p(\cdot - x)^p$$

aux données $(X_i, Y_i)_{i=1, \dots, n}$ par la méthode des Moindres Carrés. Pour cela, on doit supposer l'existence de la $(p+1)$ ème dérivée de la fonction de régression $r(\cdot)$ au point x pour pouvoir

faire un développement de Taylor et on a alors

$$\begin{aligned} r(u) &\approx r(x) + r'(x)(u-x) + \frac{r''(x)}{2}(u-x)^2 + \dots + \frac{r^{(p)}(x)}{p!}(u-x)^p \\ &\approx \sum_{j=0}^p \frac{r^{(j)}(x)}{j!}(u-x)^j =: \sum_{j=0}^p a_j(u-x)^j \end{aligned} \quad (2.8)$$

pour u appartenant à un voisinage de x .

Pour ajuster localement le polynôme (2.8) aux données $(X_i, Y_i)_{i=1, \dots, n}$ par la méthode des M.C. pondérés par $K(\frac{\cdot-x}{h_n})$ on doit minimiser par rapport au vecteur $a = (a_0, a_1, \dots, a_p)^t \in \mathbb{R}^{p+1}$ la quantité

$$\sum_{i=1}^n \left\{ Y_i - \sum_{j=0}^p a_j (X_i - x)^j \right\}^2 K \left(\frac{X_i - x}{h_n} \right). \quad (2.9)$$

Comme pour l'estimateur de Nadaraya-Watson, les paramètres K et h_n déterminent la forme et la taille du voisinage autour de x . Soit $\hat{a} = (\hat{a}_0, \hat{a}_1, \dots, \hat{a}_p)^t \in \mathbb{R}^{p+1}$ le vecteur qui minimise (2.9). D'après (2.8), la dérivée k -ème, $r^{(k)}(x)$, peut-être estimée par $\hat{a}_k k!$ pour $k = 0, \dots, p$, d'où la définition

Définition 2.1. *La statistique*

$$\hat{r}_{LP;n,p}^{(k)}(x) := \hat{a}_k k! \quad 0 \leq k \leq p \quad (2.10)$$

est l'estimateur localement polynômial d'ordre p de la dérivée k -ème de la régression $r(x)$, c-à-d $r^{(k)}(x)$.

Remarque 2.1. *Lorsque $k = p = 0$, on retrouve évidemment l'estimateur de Nadaraya-Watson.*

Définition 2.2. *L'estimateur localement polynômial d'ordre 1 de la fonction de régression $r(x)$ est appelé estimateur localement linéaire.*

Si on prend $p = 1$ dans (2.9) et $k = 0$ dans (2.10), on a l'expression de cet estimateur

$$\hat{r}_{LL;n}(x) := \frac{\hat{m}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{m}_{n,1}(x)\hat{f}_{n,1}(x)}{\hat{f}_{n,0}(x)\hat{f}_{n,2}(x) - \hat{f}_{n,1}(x)\hat{f}_{n,1}(x)}$$

où

$$\begin{aligned} \hat{f}_{n,j}(x) &: = \frac{1}{nh_n} \sum_{i=1}^n \left(\frac{X_i - x}{h_n} \right)^j K \left(\frac{X_i - x}{h_n} \right) & j = 0, 1, 2 \\ \hat{m}_{n,j}(x) &: = \frac{1}{nh_n} \sum_{i=1}^n Y_i \left(\frac{X_i - x}{h_n} \right)^j K \left(\frac{X_i - x}{h_n} \right) & j = 0, 1. \end{aligned}$$

Fan (1992) a montré que l'estimateur localement linéaire $\hat{r}_{LL;n}(x)$ a un meilleur biais que l'estimateur de Nadaraya-Watson et une meilleure variance que l'estimateur de Gasser-Müller. D'après Fan (1993), il est le meilleur estimateur sur la classe des fonctions de régression à dérivées seconde bornée, parmi tous les estimateurs linéaires.

Biais et variance asymptotiques de l'estimateur LP

Les moments de K et K^2 sont dénotés par

$$\mu_j(K) = \int_{\mathbb{R}} u^j K(u) du \quad \text{et} \quad \mu_j(K^2) = \int_{\mathbb{R}} u^j K^2(u) du$$

soient :

$$\begin{aligned} N_p &= (\mu_{j+l-2}(K))_{1 \leq j, l \leq p+1} \in \mathcal{M}_{p+1}(\mathbb{R}) \\ M_{k,p} &= N_p \text{ où on remplace la } (k+1)\text{ème colonne par le vecteur } (1, u, \dots, u^p)^t \\ K_{k,p}(u) &= k! \frac{|M_{k,p}|}{|N_p|} K(u) \\ v(x) &= \text{Var}(Y|X = x) \end{aligned}$$

où $\mathcal{M}_{p+1}(\mathbb{R})$ désigne l'ensemble des matrices carrées d'ordre $p+1$ à valeurs dans \mathbb{R} .

Le biais et la variance conditionnels de l'estimateur par polynômes locaux sont donnés par le théorème.

Théorème 2.1. (*Ruppert and Wand 1994*).

Supposons que $f(x) > 0$, que les fonctions $f(\cdot)$, $r^{(p+2)}(\cdot)$ et $v(\cdot)$ soient continues dans un voisinage de x , que la fonction noyau $K(\cdot)$ soit bornée, à support compact et que les moments d'ordres impairs soient nuls, alors si $h_n \rightarrow 0$ et $nh_n^{2p+1} \rightarrow +\infty$ lorsque $n \rightarrow +\infty$, on a

- pour $p - k$ impair

$$\mathbb{E} \left(\hat{r}_{LP;n,p}^{(k)}(x) - r^{(k)}(x) | X_1, \dots, X_n \right) = \left\{ \int u^{p+1} K_{k,p}(u) du \right\} \left\{ \frac{r^{(p+1)}(x)}{(p+1)!} \right\} h_n^{p-k+1} + o(h_n^{p-k+1})$$

- et pour $p - k$ pair

$$\begin{aligned} \mathbb{E} \left(\hat{r}_{LP;n,p}^{(k)}(x) - r^{(k)}(x) | X_1, \dots, X_n \right) &= \left[\left\{ \int u^{p+1} K_{k,p}(u) du \right\} \left\{ \frac{r^{(p+2)}(x)}{(p+2)!} \right\} \right. \\ &\quad \left. + \left\{ \int u^{p+2} K_{k,p}(u) du - r \int u^{p+1} K_{k-1,p}(u) du \right\} \left\{ \frac{r^{(p+1)}(x) f'(x)}{(p+1)! f(x)} \right\} \right] h_n^{p-k+2} \\ &\quad + o(h_n^{p-k+2}). \end{aligned}$$

- Dans tous les cas

$$\text{Var} \left(\hat{r}_{LP;n,p}^{(k)}(x) | X_1, \dots, X_n \right) = \{K_{k,p}^2(u)du\} \left\{ \frac{1}{nh_n^{2k+1}} \frac{v(x)}{f(x)} \right\} (1 + o(1)).$$

où le terme $o(1)$ tend vers 0 lorsque h_n tend vers 0.

2.2 Propriétés asymptotiques de l'estimateur de Nadaraya-Watson

L'estimateur à noyau de la régression est donc dépendant du choix des deux paramètres que sont la fenêtre h_n et le noyau K . Nous verrons par la suite que le paramètre crucial est la fenêtre pour obtenir de bonnes propriétés asymptotiques. Toutefois le noyau ne doit pas être négligé, il permet de réduire le biais de l'estimateur en s'appuyant sur des propriétés de régularité de la courbe de régression.

2.2.1 Consistance

Dans ce qui suit nous allons déterminer les conditions sur la fenêtre et le noyau, nécessaires à la consistance de l'estimateur $\hat{r}_{NW;n}$. Nous utiliserons pour cela la décomposition biais-variance suivante :

$$\mathbb{E} [(\hat{r}_{NW;n}(x) - r(x))^2] = \text{Var} [\hat{r}_{NW;n}] + [\mathbb{E} (\hat{r}_{NW;n}(x)) - r(x)]^2. \quad (2.11)$$

On dénote par $\xrightarrow{L_2}$ (resp. $\xrightarrow{\mathbb{P}}$) la convergence en norme L_2 (resp. en probabilité). Lorsque (2.11) tend vers 0, on déduit que $\hat{r}_{NW;n} \xrightarrow{L_2} r(x)$ ce qui implique que $\hat{r}_{NW;n} \xrightarrow{\mathbb{P}} r(x)$.

2.2.1.1 Calcul de la variance :

Pour calculer la variance de l'estimateur $\hat{r}_{NW;n}$ ainsi que son expression asymptotique, nous supposons que le noyau K vérifie les hypothèses (K1-K4).

Posons :

$$\bar{\sigma}^2(x) := \text{Var}(Y|X=x) = \frac{1}{f(x)} \int y^2 f_{X,Y}(x,y) dy - [r(x)]^2$$

et $\kappa =: \int K^2(t)dt$, lorsque ces deux expressions sont bien définies. Nous avons besoin aussi du lemme de Bochner que nous rappelons ci-dessous (voir Tsybakov (2003, p. 11)).

Lemme 2.1. (Bochner). Supposons que g soit une fonction bornée sur \mathbb{R} , continue dans un voisinage du point $x_0 \in \mathbb{R}$ et que Q soit une fonction sur \mathbb{R} telle que $\int |Q(t)| dt < +\infty$, alors

$$\lim_{h \rightarrow 0} \frac{1}{h} \int Q\left(\frac{z-x_0}{h}\right) g(z) dz = g(x_0) \int Q(t) dt.$$

Preuve

Posons $t = \frac{z-x_0}{h}$, alors pour tout $h > 0$

$$\begin{aligned} \left| \frac{1}{h} \int Q\left(\frac{z-x_0}{h}\right) g(z) dz - g(x_0) \int Q(t) dt \right| &= \left| \int [g(x_0 + th) - g(x_0)] Q(t) dt \right| \\ &\leq \sup_{|t| < h^{-\frac{1}{2}}} |g(x_0 + th) - g(x_0)| \int |Q(t)| dt \\ &\quad + \int_{|t| \geq h^{-\frac{1}{2}}} |g(x_0 + th) - g(x_0)| |Q(t)| dt \\ &\leq \sup_{|s| < h^{\frac{1}{2}}} |g(x_0 + s) - g(x_0)| \int |Q(t)| dt \\ &\quad + 2 \sup_t |g(t)| \int_{|t| \geq h^{-\frac{1}{2}}} |Q(t)| dt. \end{aligned}$$

On obtient le résultat en faisant tendre h vers 0.

Nous sommes maintenant en mesure d'évaluer la variance de l'estimateur de Nadaraya-Watson grâce à la proposition suivante

Proposition 2.2. On suppose que $\mathbb{E}(Y^2) < +\infty$. A chaque point de continuité des fonctions $r(x)$, $f(x)$ et $\bar{\sigma}^2(x)$ tels que $f(x) > 0$, on a

$$\text{Var} [\hat{r}_{NW;n}(x)] = \frac{1}{nh_n} \left(\frac{\bar{\sigma}^2(x)}{f(x)} \kappa \right) (1 + o(1)) \quad (2.12)$$

où le terme $o(1)$ tend vers 0 lorsque h_n tend vers 0.

Preuve

En utilisant le lemme de Bochner, on obtient

$$\begin{aligned} \text{Var} [\hat{f}_{PR;n}(x)] &= \frac{1}{nh_n^2} \left[\mathbb{E} \left(K^2\left(\frac{x-X}{h_n}\right) \right) - \mathbb{E}^2 \left(K\left(\frac{x-X}{h_n}\right) \right) \right] \\ &= \frac{1}{nh_n} \left[\int K^2(t) f(x - th_n) dt - h_n^2 \left(\int K(t) f(x - th_n) dt \right)^2 \right] \\ &= \frac{1}{nh_n} f(x) \kappa (1 + o(1)) \end{aligned} \quad (2.13)$$

lorsque h_n tend vers 0. De plus, nous avons

$$\begin{aligned} \text{Var} [\hat{r}_{1,n}(x)] &= \frac{1}{nh_n^2} \left[\mathbb{E} \left(Y^2 K^2 \left(\frac{x - X_i}{h_n} \right) \right) - \mathbb{E}^2 \left(Y K \left(\frac{x - X_i}{h_n} \right) \right) \right] \\ &= \frac{1}{nh_n} \left[\int K^2(t) \tilde{r}_2(x - th_n) dt - h_n^2 \left(\int K(t) \bar{r}_1(x - th_n) dt \right)^2 \right] \\ &= \frac{1}{nh_n} \tilde{r}_2(x) \kappa (1 + o(1)) \end{aligned}$$

où $\bar{r}_1(x)$ est défini dans 2.5 et

$$\tilde{r}_2(x) = \int y^2 f_{X,Y}(x, y) dy.$$

De même la covariance de $\hat{f}_{PR}(x)$ et $\hat{r}_{1,n}(x)$ est donnée par

$$\mathbb{E} \left[\hat{f}_{PR}(x) - \mathbb{E} \left(\hat{f}_{PR}(x) \right) \right] \left[\hat{r}_{1,n}(x) - \mathbb{E} \left(\hat{r}_{1,n}(x) \right) \right] = \frac{1}{nh_n} \bar{r}_1(x) \kappa (1 + o(1)). \quad (2.14)$$

Si on considère le vecteur $\theta_n(x)$ défini par $\theta_n(x) = \begin{pmatrix} \hat{f}_{PR}(x) \\ \hat{r}_{1,n}(x) \end{pmatrix}$ et $\Sigma_{\theta_n(x)}$ sa matrice de variance covariance, alors lorsque h_n tend vers 0,

$$\Sigma_{\theta_n(x)} = \frac{1}{nh_n} \begin{pmatrix} f(x) & \bar{r}_1(x) \\ \bar{r}_1(x) & \tilde{r}_2(x) \end{pmatrix} \kappa (1 + o(1)).$$

En remarquant que,

$$\begin{pmatrix} -\frac{\bar{r}_1(x)}{(f(x))^2} & \frac{1}{f(x)} \end{pmatrix} \begin{pmatrix} f(x) & \bar{r}_1(x) \\ \bar{r}_1(x) & \tilde{r}_2(x) \end{pmatrix} \begin{pmatrix} -\frac{\bar{r}_1(x)}{(f(x))^2} \\ \frac{1}{f(x)} \end{pmatrix} = \frac{\tilde{r}_2(x)}{(f(x))^2} - \frac{(\bar{r}_1(x))^2}{(f(x))^3}.$$

On obtient alors

$$\begin{aligned} \text{Var} [\hat{r}_{NW;n}(x)] &= \frac{1}{nh_n} \left(\frac{\tilde{r}_2(x)}{(f(x))^2} - \frac{(\bar{r}_1(x))^2}{(f(x))^3} \right) \kappa (1 + o(1)) \\ &= \frac{1}{nh_n} \left(\frac{\bar{\sigma}^2(x)}{f(x)} \kappa \right) (1 + o(1)). \end{aligned}$$

On en conclut que si h_n tend vers 0 et nh_n tend vers $+\infty$, lorsque n tend vers $+\infty$, la variance tend vers 0.

Remarque 2.2. Dans le calcul des variances et covariances précédentes, nous retrouvons invariablement le terme $\kappa := \int K^2(t) dt$. Un choix approprié du noyau K assure la finitude de cette intégrale. On peut prendre par exemple $K(\cdot)$ à variation bornée sur \mathbb{R} et à support compact, suppositions qui implique évidemment les hypothèses $K(1-3)$. Si on veut obtenir une variance minimale, on peut minimiser κ suivant K dans une certaine classe de noyaux fixée, on peut citer l'article de Gasser, Müller et Mammitzsch (1985), pour une expression explicite des noyaux de variance minimale.

2.2.1.2 Calcul du biais :

Le calcul du biais repose essentiellement sur des développements de Taylor, ce qui nous conduit à poser certaines conditions de régularités sur les fonctions $r(\cdot)$ et $f(\cdot)$ qui détermineront l'ordre du biais asymptotique en fonction du paramètre de lissage h_n . Afin de faciliter les calculs et du fait que l'estimateur $\hat{r}_{NW;n}(\cdot)$ est lui-même sous forme d'un rapport aléatoire, nous allons introduire un terme de centrage sous forme de rapport, défini par $\tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) := \frac{\mathbb{E}(\hat{r}_{1;n}(x))}{\mathbb{E}(\hat{f}_{PR}(x))}$. Pour évaluer le biais de l'estimateur, nous commençons d'abord par donner une proposition due à Nadaraya (1989 p.116-117).

Proposition 2.3. (Nadaraya 1989).

Si Y est bornée et si nh_n tend vers $+\infty$, lorsque n tend vers $+\infty$, on a

$$\mathbb{E}(\hat{r}_{NW;n}(x)) = \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) + O((nh_n)^{-1}).$$

Si $\mathbb{E}(Y^2) < +\infty$ et si nh_n^2 tend vers $+\infty$, lorsque n tend vers $+\infty$, on a

$$\mathbb{E}(\hat{r}_{NW;n}(x)) = \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) + O((\sqrt{n}h_n)^{-1}).$$

Preuve

Remarquons que

$$\begin{aligned} \frac{1}{\hat{f}_{PR}(x)} &= \frac{1}{\mathbb{E}(\hat{f}_{PR}(x))} - \frac{\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x))}{\mathbb{E}^2(\hat{f}_{PR}(x))} \\ &\quad + \frac{\left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x))\right)^2}{\hat{f}_{PR}(x)\mathbb{E}^2(\hat{f}_{PR}(x))}. \end{aligned} \tag{2.15}$$

Multiplions l'expression (2.15) par $\hat{r}_{1;n}(x)$ et passons à l'espérance, on obtient

$$\begin{aligned} \mathbb{E}(\hat{r}_{NW;n}(x)) &= \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) - \frac{\mathbb{E}\left\{\hat{r}_{1;n}(x) \left[\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x))\right]\right\}}{\mathbb{E}^2(\hat{f}_{PR}(x))} \\ &\quad + \mathbb{E}\left\{\frac{\hat{r}_{1;n}(x) \left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x))\right)^2}{\hat{f}_{PR}(x)\mathbb{E}^2(\hat{f}_{PR}(x))}\right\} \\ &=: \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) + \frac{a_n(x) + b_n(x)}{\mathbb{E}^2(\hat{f}_{PR}(x))}, \end{aligned}$$

où

$$\begin{aligned} a_n(x) &:= \mathbb{E} \left[\hat{r}_{1,n}(x) \hat{f}_{PR}(x) \right] - \mathbb{E}(\hat{r}_{1,n}(x)) \mathbb{E}(\hat{f}_{PR}(x)) \\ b_n(x) &:= \mathbb{E} \left\{ \frac{\hat{r}_{1,n}(x)}{\hat{f}_{PR}(x)} \left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x)) \right)^2 \right\}. \end{aligned}$$

On a clairement d'après l'expression (2.14) que

$$a_n(x) = \frac{1}{nh_n} \bar{r}_1(x) \kappa + o(1) = O\left(\frac{1}{nh_n}\right). \quad (2.16)$$

En ce qui concerne $b_n(x)$, lorsque Y est bornée il existe une constante $M > 0$ telle que $|Y| \leq M$. Alors l'estimateur $\hat{r}_{NW;n}(x)$ est lui même borné, en effet

$$\hat{r}_{NW;n}(x) = \frac{\hat{r}_{1,n}(x)}{\hat{f}_{PR}(x)} = \frac{\frac{1}{nh_n} \sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \leq M$$

d'où d'après l'expression (2.13)

$$b_n(x) \leq M \mathbb{E} \left\{ \left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x)) \right)^2 \right\} = M \left\{ \frac{1}{nh_n} f(x) \kappa + o(1) \right\} = O\left(\frac{1}{nh_n}\right). \quad (2.17)$$

On a alors, grâce a (2.16) et (2.17) le résultat de la proposition pour Y bornée.

Lorsque $\mathbb{E}(Y^2) < +\infty$, en utilisant l'inégalité de Cauchy-Schwartz on obtient

$$\begin{aligned} b_n(x) &\leq \mathbb{E} \left\{ \max_{1 \leq i \leq n} |Y_i| \left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x)) \right)^2 \right\} \\ &\leq \left[\mathbb{E} \left(\sum_{i=1}^n Y_i^2 \right) \right]^{1/2} \mathbb{E} \left[\left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x)) \right)^4 \right]^{1/2} \\ &\leq \sqrt{n} [\mathbb{E}(Y_i^2)]^{1/2} \mathbb{E} \left[\left(\hat{f}_{PR}(x) - \mathbb{E}(\hat{f}_{PR}(x)) \right)^4 \right]^{1/2} \\ &\leq \sqrt{n} [\mathbb{E}(Y_i^2)]^{1/2} O\left(\frac{1}{nh_n}\right) = O\left(\frac{1}{\sqrt{nh_n}}\right). \end{aligned} \quad (2.18)$$

D'après (2.16) et (2.18), on a le résultat de la seconde partie de la proposition.

Ceci nous permet d'énoncer le biais asymptotique de l'estimateur $\hat{r}_{NW;n}(\cdot)$ lorsque Y est bornée

Proposition 2.4. *On suppose que $r(x)$ et $f(x)$ sont de classe \mathcal{C}^2 sur \mathbb{R} et que le noyau K est d'ordre 2 c-à-d*

$$\int K(t) dt = 1 \quad , \quad \int tK(t) dt = 0 \quad , \quad \int t^2 K(t) dt < +\infty$$

alors lorsque h_n tend vers 0 et nh_n tend vers $+\infty$ on a

$$\mathbb{E}(\hat{r}_{NW;n}(x)) - r(x) = \frac{h_n^2}{2} \left\{ \left(r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right) \int t^2 K(t) dt \right\} + o(1).$$

Preuve

On a

$$\begin{aligned} \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) - r(x) &= \frac{\mathbb{E}(\hat{r}_{1;n}(x))}{\mathbb{E}(\hat{f}_{PR}(x))} - r(x) \\ &= \frac{1}{\frac{1}{h_n} \mathbb{E}\left(K\left(\frac{x-X}{h_n}\right)\right)} \int \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) \bar{r}_1(u) du - r(x) \\ &= \frac{1}{\frac{1}{h_n} \mathbb{E}\left(K\left(\frac{x-X}{h_n}\right)\right)} \left\{ \int \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) \bar{r}_1(u) du - \bar{r}_1(x) + \bar{r}_1(x) \right. \\ &\quad \left. - r(x) \int \frac{1}{h_n} K\left(\frac{x-u}{h_n}\right) f(u) du \right\} \\ &= \frac{1}{\frac{1}{h_n} \mathbb{E}\left(K\left(\frac{x-X}{h_n}\right)\right)} \left\{ \int K(t) [\bar{r}_1(x - th_n) - \bar{r}_1(x)] dt \right. \\ &\quad \left. - r(x) \int K(t) [f(x - th_n) - f(x)] dt \right\} \end{aligned}$$

un développement de Taylor à l'ordre 1 pour $\frac{1}{h_n} \mathbb{E}\left(K\left(\frac{x-X}{h_n}\right)\right)$ et à l'ordre 2 pour $\bar{r}_1(\cdot)$ et $f(\cdot)$ donne

$$\begin{aligned} \tilde{\mathbb{E}}(\hat{r}_{NW;n}(x)) - r(x) &= \frac{h_n^2}{2} \frac{1}{f(x)} [\bar{r}_1''(x) - r(x)f''(x)] \int t^2 K(t) dt + o(1) \\ &= \frac{h_n^2}{2} \left\{ r''(x) + 2r(x) \frac{f'(x)}{f(x)} \right\} + o(1). \end{aligned} \tag{2.19}$$

La proposition précédente et l'expression (2.19) achèvent la démonstration.

Remarque 2.3. On voit bien que dans le terme du biais asymptotique, apparaissent les dérivées de $r(\cdot)$ et $f(\cdot)$: l'estimateur de Nadaraya-Watson souffre donc d'un biais élevé dans la région où la dérivée seconde de la vraie fonction de régression est grande, ou lorsque la quantité $\frac{f'(x)}{f(x)}$ est grande. En comparaison, sous des hypothèses similaires, l'estimateur de Gasser-Müller a un meilleur biais. En effet, on montre que

$$\mathbb{E}(\hat{r}_{GM;n}(x)) - r(x) = \frac{h_n^2}{2} r''(x) \int t^2 K(t) dt + o(1)$$

et on voit que le biais ne dépend ni de f ni de ses dérivées. Plusieurs techniques ont été développées afin de réduire le biais de $\hat{r}_{NW;n}$, on peut citer les articles de Müller et Song (1993), Linton et Nielsen (1994), Mammen et Marron (1997), Müller (1997), Choi, Hall et Rousson (2000) et Hall et Müller (2003), pour une exposition des différentes méthodologie existantes.

2.2.2 Optimalité asymptotique et choix de la fenêtre

On a vu que par construction même, les estimateurs à noyau $\hat{r}_{NW;n}$ dépendent de deux paramètres : le noyau K et la largeur de la fenêtre h_n . Dans la pratique on aura besoin de décider quels choix effectuer pour ces deux paramètres. Au vu de la section précédente, il est clair que le paramètre h_n , en contrôlant la régularité de la fonction estimée, joue un rôle essentiel. Remarquons aussi que la consistance de l'estimateur de NW est basée sur les conditions nécessaires et suffisantes sur la fenêtre h_n

$$h_n \rightarrow 0 \text{ et } nh_n \rightarrow +\infty \text{ lorsque } n \rightarrow +\infty.$$

On se propose à présent de déterminer la fenêtre optimale dans le cas de la régression un peu plus en détails que dans la cas de la densité (voir chapitre 1). Pour cela nous allons utiliser deux procédures : la première sera locale pour obtenir des résultats ponctuels via le critère de l'erreur quadratique moyenne (MSE), et la seconde pour obtenir des résultats uniformes via le critère de l'erreur quadratique intégrée moyenne (MISE)

2.2.2.1 Critère de selection local : MSE

Sous les hypothèses de la proposition 2.4 nous avons

$$\begin{aligned} \mathbb{E}(\hat{r}_{NW;n}(x)) - r(x) &= \frac{h_n^2}{2} \left\{ \left(r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right) \int t^2 K(t) dt \right\} + o(1) \\ &= : \frac{h_n^2}{2} [\bar{b}(x)] + o(1). \end{aligned} \quad (2.20)$$

De même, sous les hypothèses de la proposition 2.2 nous avons

$$\begin{aligned} Var[\hat{r}_{NW;n}(x)] &= \frac{1}{nh_n} \left(\frac{\bar{\sigma}^2(x)}{f(x)} \kappa \right) (1 + o(1)) \\ &= : \frac{1}{nh_n} (\bar{v}(x))^2 (1 + o(1)). \end{aligned} \quad (2.21)$$

Le comportement asymptotique du risque quadratique de l'estimateur de NW au point x est donné par le théorème.

Théorème 2.2. *Sous les hypothèses des propositions 2.2 et 2.4 nous obtenons*

$$\begin{aligned} MSE(h_n) & : = \mathbb{E} [(\hat{r}_{NW; n}(x) - r(x))^2] \\ & = \left\{ \frac{h_n^4}{4} (\bar{b}(x))^2 + \frac{1}{nh_n} (\bar{v}(x))^2 \right\} (1 + o(1)). \end{aligned}$$

La preuve résulte des expressions (2.20) et (2.21) ainsi que de la décomposition en biais-variance

$$MSE(h_n) = [\mathbb{E} (\hat{r}_{NW; n}(x) - r(x))^2 + Var [\hat{r}_{NW; n}(x)] .$$

Nous obtenons alors l'expression asymptotique de l'erreur quadratique moyenne (AMSE)

$$AMSE(h_n) = \frac{h_n^4}{4} (\bar{b}(x))^2 + \frac{1}{nh_n} (\bar{v}(x))^2. \quad (2.22)$$

Notons qu'en réalité le risque quadratique asymptotique (2.22) dépend non seulement de h_n mais aussi de K . Nous supposons dans un premier temps, le noyau K fixé. La fenêtre optimale au sens du AMSE est obtenu en minimisant suivant h_n la quantité (2.22), c'est à dire

$$h_n^{opt} = \arg \min_{h_n} AMSE(h_n).$$

Lorsque $\bar{b}(x) \neq 0$, nous obtenons

$$h_{MSE,n}^{opt} = \frac{1}{n^{1/5}} \left\{ \frac{\frac{\bar{\sigma}^2(x) \kappa}{f(x)}}{\left(r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right)^2 \left(\int t^2 K(t) dt \right)^2} \right\}^{1/5}$$

et

$$\min_{h_n} AMSE(h_n) = \frac{5}{4} \left\{ \frac{\bar{\sigma}^2(x)}{f(x)} \right\}^{4/5} \left| r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right|^{2/5} \kappa^{4/5} \left(\int t^2 K(t) dt \right)^{2/5} n^{-4/5}.$$

2.2.2.2 Critère de selection global : MISE

On s'intéresse à l'estimation de la fonction de régression sur un intervalle $I \subseteq \mathbb{R}$, et au risque global de l'estimateur NW. On introduit pour cela l'erreur quadratique intégrée moyenne ou MISE

$$\begin{aligned} MISE(h_n) & := \mathbb{E} \left[\int_I (\hat{r}_{NW; n}(x) - r(x))^2 dx \right] \\ & = \int_I MSE(h_n) dx \\ & = \int_I \mathbb{E} [(\hat{r}_{NW; n}(x) - r(x))^2] dx \end{aligned}$$

d'après le théorème de Tonelli-Fubini. Nous avons le théorème suivant

Théorème 2.3. *Sous les hypothèses des propositions 2.2 et 2.4, nous avons*

$$MISE(h_n) = \left[\int_I \left\{ \frac{h_n^4}{4} (\bar{b}(x))^2 + \frac{1}{nh_n} (\bar{v}(x))^2 \right\} dx \right] (1 + o(1)). \quad (2.23)$$

La fenêtre optimale, au sens du critère asymptotique de l'erreur quadratique intégrée moyenne ou AMISE, est donnée par

$$h_{MISE,n}^{opt} = \frac{1}{n^{1/5}} \left\{ \frac{\int_I \bar{v}^2(x) dx}{\int_I \left[\left(r''(x) + 2r'(x) \frac{f'(x)}{f(x)} \right) \int t^2 K(t) dt \right]^2 dx} \right\}^{1/5}.$$

Remarque 2.4. *On voit bien que quelque soit le critère choisi, AMSE ou AMISE, la fenêtre optimale dépend de paramètres inconnus comme la fonction de régression et la densité de X , et n'est donc pas utilisable en pratique. Il y'a plusieurs alternatives à cette procédure, la première est celle dite du plug-in qui consiste à remplacer les fonctions inconnues par leurs estimées. Une autre alternative est la méthode par validation croisée que nous détaillons maintenant.*

2.2.2.3 La validation croisée

Nous avons observé dans les paragraphes précédents que l'efficacité de l'estimateur NW est liée au paramètre de lissage, la fenêtre h_n . Il faut choisir convenablement la fenêtre, afin d'équilibrer le terme de variance et le terme du biais de manière à ce qu'ils ne dépendent pas de quantités inconnues. Afin de construire des estimateurs qui minimisent l'erreur quadratique indépendamment des fonctions inconnues, il faut utiliser d'autres méthodes dont la plus commune est appelée la procédure par validation croisée. Cette technique est due à Rudemo (1982) qui l'a appliqué sur l'estimation par histogramme et par noyau de la densité et l'idée principale de cette méthode est celle de la minimisation, par rapport à h_n d'un estimateur du risque quadratique moyen $MSE(h_n)$. En nous inspirant de ce qui a été fait dans Tsybakov (2003) pour les densités, nous avons

$$\begin{aligned} MSE(h_n) &= \mathbb{E} (\hat{r}_{NW; n}(x) - r(x))^2 \\ &= \mathbb{E} (\hat{r}_{NW; n}(x))^2 - 2\mathbb{E} (\hat{r}_{NW; n}(x)r(x)) + \mathbb{E} (r(x))^2. \end{aligned} \quad (2.24)$$

Comme le dernier terme dans (2.24) ne dépend pas de h_n , minimiser $MSE(h_n)$ revient à minimiser

$$J(h_n) := [\mathbb{E}(\hat{r}_{NW; n}(x))^2 - 2\mathbb{E}(\hat{r}_{NW; n}(x)r(x))] . \quad (2.25)$$

Remarquons que

$$\mathbb{E}[\hat{r}_{NW; n}(x)r(x)] = \mathbb{E}[\hat{r}_{NW; n}(x)\mathbb{E}[Y/X = x]] = \mathbb{E}[Y \hat{r}_{NW; n}(X)]$$

dont un estimateur est donné par

$$\hat{E} := \frac{1}{n} \sum_{i=1}^n Y_i \hat{r}_{n,-i}(X_i)$$

où

$$\hat{r}_{n,-i}(x) =: \frac{\sum_{j \neq i} Y_j K\left(\frac{x-X_j}{h_n}\right)}{\sum_{j \neq i} K\left(\frac{x-X_j}{h_n}\right)} .$$

L'estimateur $\hat{r}_{n,-i}(x)$ est communément appelé dans la littérature, estimateur *leave-one-out*. C'est en fait l'estimateur de Nadaraya-Watson construit avec les $(n-1)$ couples aléatoires $\{(X_1, Y_1); \dots; (X_{i-1}, Y_{i-1}); (X_{i+1}, Y_{i+1}); \dots; (X_n, Y_n)\}$. De même il est possible d'approximer le premier terme de (2.25) par

$$\frac{1}{n} \sum_{i=1}^n \hat{r}_{n,-i}^2(X_i)$$

Finalement, il paraît raisonnable de choisir h_n qui minimise la quantité

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

qui est égale à

$$\frac{1}{n} \sum_{i=1}^n [Y_i - \hat{r}_{n,-i}(X_i)]^2 - \frac{1}{n} \sum_{i=1}^n Y_i^2 . \quad (2.26)$$

Le deuxième terme de (2.26) ne dépend pas de h_n et n'intervient donc pas dans la minimisation. Le critère de sélection de la fenêtre h_n se réduit à

Choisir \hat{h}_n qui minimise

$$CV(h_n) := \frac{1}{n} \sum_{i=1}^n [Y_i - \hat{r}_{n,-i}(X_i)]^2 . \quad (2.27)$$

Cette méthode de sélection de la fenêtre est appelée procédure par validation croisée. Les références principales sur cette méthode sont Hall (1984), Härdle et Marron (1985) et Härdle et Kelly (1987).

2.2.3 La normalité asymptotique

La première démonstration de la normalité asymptotique de l'estimateur de Nadaraya-Watson est due à Schuster (1972), qui a étendu le résultat de Nadaraya (1964) dans lequel celui-ci avait montré la normalité asymptotique de

$$\sqrt{nh_n} (\hat{r}_{NW;n}(x) - \mathbb{E}(\hat{r}_{NW;n}(x)))$$

sous la condition que Y soit borné et que $nh_n^2 \rightarrow +\infty$. On peut citer aussi le théorème 1.3 et 1.4 p 117,120 de Nadaraya (1989) et le théorème 4.2.1 p 99 de Härdle (1990), qui proposent une démonstration basée sur la décomposition suivante

$$\hat{r}_{NW;n}(x) - r(x) \approx \frac{\hat{r}_{1,n}(x) - r(x)\hat{f}_{PR}(x)}{f(x)} \quad (2.28)$$

équivalence qui provient du fait que l'estimateur à noyau a la même distribution limite que le terme de droite dans l'expression (2.28). Pour notre part nous avons choisi de donner un théorème de normalité en reprenant le résultat de Bierens (1987) qui en donne une preuve légèrement différente de celles qu'on peut trouver dans la littérature (cf schuster 1972 et Johnstone 1979). D'après les expressions (2.7), (2.4) et (1.1), nous remarquons que

$$\begin{aligned} (\hat{r}_{NW;n}(x) - r(x))\hat{f}_{PR}(x) &= \frac{1}{nh_n} \sum_{i=1}^n \varepsilon_i K\left(\frac{x - X_i}{h_n}\right) \\ &+ \frac{1}{nh_n} \sum_{i=1}^n \left\{ (r(X_i) - r(x)) K\left(\frac{x - X_i}{h_n}\right) - \mathbb{E} \left[(r(X_i) - r(x)) K\left(\frac{x - X_i}{h_n}\right) \right] \right\} \\ &+ \frac{1}{nh_n} \sum_{i=1}^n \mathbb{E} \left[(r(X_i) - r(x)) K\left(\frac{x - X_j}{h_n}\right) \right] \\ &= : \hat{Q}_1(x) + \hat{Q}_2(x) + \hat{Q}_3(x) \end{aligned} \quad (2.29)$$

Pour établir la normalité asymptotique de l'estimateur de Nadaraya Watson, nous allons étudier le comportement des processus \hat{Q}_j , $j = 1, 2, 3$.

Pour $f(x) > 0$ posons $\sigma_\varepsilon^2(x) = \mathbb{E}(\varepsilon_i^2 / X_i = x)$ la variance conditionnelle de ε et pour $p > 0$,

si $\mathbb{E}(|\varepsilon_i|^p) < +\infty$ posons $\sigma_\varepsilon^p(x) = \mathbb{E}|\varepsilon_i|^p / X_i = x$. Quelques hypothèses nécessaires à l'établissement de la normalité asymptotique sont données ci-dessous

(H1) Pour $p > 0$, $\sigma_\varepsilon^p(x)f(x)$ est continue et uniformément bornée sur \mathbb{R} .

(H2) Il existe $\delta > 0$ tel que $\sigma_\varepsilon^{2+\delta}(x)f(x)$ est uniformément bornée sur \mathbb{R} .

(H3) Les fonctions $r^2(x)f(x)$ et $\sigma_\varepsilon^2(x)f(x)$ sont continues et uniformément bornées.

(H4) Les fonctions $f(x)$ et $r(x)f(x)$ ainsi que leurs dérivées premières et secondes sont continues et uniformément bornées.

(H5) $\int tK(t)dt = 0$ et $\int t^2K(t)dt < +\infty$.

Lemme 2.2. *Sous l'hypothèse (H1) nous avons, lorsque $n \rightarrow +\infty$,*

$$\sqrt{nh_n}\hat{Q}_1(x) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \sigma_\varepsilon^2(x)f(x) \int K^2(t)dt\right).$$

Preuve

Posons $v_{n,i}(x) = \varepsilon_i \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sqrt{h_n}}$, alors comme

$$\sqrt{nh_n}\hat{Q}_1(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n v_{n,i}(x)$$

il suffit de montrer que $(v_{n,i}(x))$ satisfait la condition de Lyapounov pour le théorème de la limite centrale. Nous avons d'une part que, $\mathbb{E}(v_{n,i}(x)) = 0$ et

$$\begin{aligned} \mathbb{E}(v_{n,i}^2(x)) &= \mathbb{E}\left(\frac{\varepsilon_i^2 K^2\left(\frac{x-X_i}{h_n}\right)}{h_n}\right) = \frac{1}{h_n} \mathbb{E}\left[\mathbb{E}\left(\varepsilon_i^2 K^2\left(\frac{x-X_i}{h_n}\right) / X_i\right)\right] \\ &= \frac{1}{h_n} \mathbb{E}\left[K^2\left(\frac{x-X_i}{h_n}\right) \mathbb{E}(\varepsilon_i^2 / X_i)\right] = \frac{1}{h_n} \mathbb{E}\left[K^2\left(\frac{x-X_i}{h_n}\right) \sigma_\varepsilon^2(X_i)\right] \\ &= \int \sigma_\varepsilon^2(x - th_n) K^2(t) f(x - th_n) dt \\ &\xrightarrow{n \rightarrow +\infty} \sigma_\varepsilon^2(x) f(x) \int K^2(t) dt \end{aligned}$$

d'autre part, pour $\delta > 0$, nous avons

$$\begin{aligned} \sum_{i=1}^n \mathbb{E} \left(\frac{v_{n,i}(x)}{\sqrt{n}} \right)^{2+\delta} &= \frac{1}{h_n} \left(\frac{1}{\sqrt{nh_n}} \right)^\delta \mathbb{E} \varepsilon_i^{2+\delta} K^{2+\delta} \left(\frac{x - X_i}{h_n} \right) \\ &= \left(\frac{1}{\sqrt{nh_n}} \right)^\delta \int \sigma_\varepsilon^{2+\delta}(x - th_n) K^{2+\delta}(t) f(x - th_n) dt = O \left(\frac{1}{\sqrt{nh_n}} \right)^\delta \\ &\quad \xrightarrow[n \rightarrow +\infty]{} 0 \end{aligned}$$

ce qui implique que

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n v_{n,i}(x) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \sigma_\varepsilon^2(x) f(x) \int K^2(t) dt \right).$$

Ceci achève la preuve du lemme.

Lemme 2.3. *Sous les hypothèses (H3-H4), nous avons*

$$\lim_{n \rightarrow +\infty} \mathbb{E} \left(\sqrt{nh_n} \hat{Q}_2(x) \right)^2 = 0.$$

Preuve

Nous avons

$$\begin{aligned} \mathbb{E} \left(\sqrt{nh_n} \hat{Q}_2(x) \right)^2 &= \int [r(x - th_n) - r(x)]^2 K^2(t) f(x - th_n) dt \\ &\quad - h_n \left\{ \int [r(x - th_n) - r(x)] K(t) f(x - th_n) dt \right\}^2 \\ &\quad \xrightarrow[n \rightarrow +\infty]{} 0 \end{aligned}$$

car $r^2(x)f(x)$ et $r(x)f(x)$ sont uniformément bornées.

Lemme 2.4. *Sous les hypothèses (H1-H5) on a*

$$\lim_{n \rightarrow +\infty} \frac{1}{h_n^2} \hat{Q}_3(x) < +\infty.$$

Preuve

En faisant un développement de Taylors à l'ordre deux pour chacune des fonctions $r(x -$

$th_n)f(x - th_n)$ et $f(x - th_n)$ au voisinage de x , nous obtenons

$$\begin{aligned}
 \hat{Q}_3(x) &= \int [r(x - th_n) - r(x)] K(t) f(x - th_n) dt \\
 &= \int [r(x - th_n) f(x - th_n) - r(x) f(x - th_n)] K(t) dt \\
 &= \int [r(x - th_n) f(x - th_n) - r(x) f(x)] K(t) dt - r(x) \int [f(x - th_n) - f(x)] K(t) dt \\
 &\approx -h_n \int t \frac{d}{dx} (r(x) f(x)) K(t) dt + \frac{h_n^2}{2} \int t^2 \frac{d^2}{dx^2} (r(x) f(x)) K(t) dt \\
 &\quad + r(x) h_n \int t \frac{d}{dx} f(x) K(t) dt - \frac{h_n^2}{2} r(x) \int t^2 \frac{d^2}{dx^2} f(x) K(t) dt.
 \end{aligned}$$

Posons

$$\tilde{b}(x) = \frac{1}{2} \frac{d^2}{dx^2} (r(x) f(x)) \int t^2 K(t) dt - \frac{1}{2} r(x) \frac{d^2}{dx^2} f(x) \int t^2 K(t) dt < +\infty$$

alors

$$\lim_{n \rightarrow +\infty} \frac{1}{h_n^2} \hat{Q}_3(x) = \tilde{b}(x)$$

ceci achève la preuve du lemme.

Nous sommes maintenant en mesure d'énoncer le théorème de normalité

Théorème 2.4. *Sous les hypothèses (H1-H5) et pour $f(x) > 0$ nous avons*

Si $\lim_{n \rightarrow +\infty} h_n^2 \sqrt{nh_n} = \lambda$ avec $0 \leq \lambda < +\infty$ alors

$$\sqrt{nh_n} (\hat{r}_{NW; n}(x) - r(x)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(\frac{\lambda \tilde{b}(x)}{f(x)}, \frac{\sigma_\varepsilon^2(x) \int K^2(t) dt}{f(x)} \right).$$

La preuve est une application directe des trois lemmes précédents.

Remarque 2.5. *Le cas $\lambda = 0$ donne évidemment le résultat classique*

$$\sqrt{nh_n} (\hat{r}_{NW; n}(x) - r(x)) \xrightarrow{\mathcal{L}} \mathcal{N} \left(0, \frac{\sigma_\varepsilon^2(x) \int K^2(t) dt}{f(x)} \right).$$

Remarque 2.6. *on peut aisément étendre ce résultat de normalité au cas où $X \in \mathbb{R}^d$. On pourra se référer à Bierens (1987) pour des preuves détaillées.*

Chapitre 3

Régression non paramétrique : Cas de données censurées

3.1 Durées de vie

Dans cette section, nous rappelons ce qu'est un modèle censuré, un modèle de survie ainsi que certains résultats de régression pour ce type de modèles.

Un modèle de survie est un modèle basé sur des durées de vie. Le terme de durée de vie est employé de manière générale pour désigner le temps qui s'écoule jusqu'à la survenue d'un événement particulier qui n'est pas forcément la mort : par exemple, le délai de rémission d'une maladie (temps écoulé entre le début du traitement et la rechute), le délai de guérison (temps qui sépare le diagnostic de la guérison) ou encore le délai de panne (temps entre la panne d'une machine et sa réparation) etc...

Des censures apparaissent souvent dans les modèles de survie : une censure se produit lorsque la durée de vie n'est pas observée pour une raison ou une autre, cela peut-être dû à une perte de vue du patient, à la fin de la période d'analyse, à une limitation de la précision d'un instrument de mesure ; ce qu'on observe alors c'est l'instant qui découle de ces raisons.

3.1.1 Caractérisation de la loi d'une durée de vie

Soit Y une variable aléatoire positive représentant la durée de vie d'un individu ou d'un sujet prenant part à une étude expérimentale (médicale, fiabilité etc...). Si Y est absolument continue, alors sa loi de probabilité peut-être définie par l'une des cinq fonctions suivantes :

- la fonction de répartition (qu'on notera f.d.r.) $H(t) = \mathbb{P}(Y \leq t)$,

- la densité de probabilité $h(t) = \frac{dH(t)}{dt}$,
- la fonction de survie $\bar{H}(t) = S(t) = 1 - H(t)$,
- le taux de hasard ou le risque instantané $\lambda(t) = \lim_{dt \rightarrow 0} \frac{\mathbb{P}(t < Y \leq t+dt / Y > t)}{dt} = \frac{h(t)}{S(t)}$,
- le taux de hasard cumulé $\Lambda(t) = \int_0^t \lambda(s) ds = -\log S(t)$.

Le taux de hasard caractérise la probabilité de mourir dans un petit intervalle de temps après l'instant t conditionnellement au fait d'avoir survécu jusqu'à l'instant t . On peut déduire la fonction de survie à partir du taux de hasard cumulé par la relation

$$\begin{aligned} S(t) &= \exp(-\Lambda(t)) = \exp\left(-\int_0^t \lambda(s) ds\right) \\ &= \exp\left(-\int_0^t \frac{dS(s)}{S(s)}\right). \end{aligned} \quad (3.1)$$

3.1.2 Les trois types de censure

3.1.2.1 Censure de type I : fixée

Au lieu d'observer les variables Y_1, Y_2, \dots, Y_n qui nous intéressent, on n'observe Y_i que lorsqu'elle est inférieure à une durée fixée C , sinon on sait seulement que Y_i est supérieur à C . On observe donc une variable T_i telle que $T_i = \min(Y_i, C)$.

3.1.2.2 Censure de type II : Attente (jusqu'au r -ième décès)

On observe les durées de vie de n patients jusqu'à ce que r d'entre eux soient décédés et d'arrêter l'étude à ce moment. Si on ordonne les durées Y_1, Y_2, \dots, Y_n , on obtient les statistiques d'ordre $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$. La date de la censure est alors $Y_{(r)}$ et on observe $T_{(1)} = Y_{(1)}, T_{(2)} = Y_{(2)}, \dots, T_{(r)} = Y_{(r)}, T_{(r+1)} = Y_{(r)}, \dots, T_{(n)} = Y_{(r)}$.

3.1.2.3 Censure de type III : aléatoire

A chaque patient i on lui associe non seulement son temps de survie Y_i mais aussi son temps de censure C_i . Ce temps de censure sera propre à chaque patient et peut-être dû à plusieurs raisons dont :

- Perte de vue : le patient ne revient plus dans l'étude.

- Arrêt du traitement : il peut y avoir des effets secondaires si désastreux pour le patient qu'on est obligé d'arrêter le traitement.

- Fin de l'étude : l'étude se termine alors que certains patients sont toujours vivants (heureusement !)

Ce qu'on observe alors c'est le couple (T_i, δ_i) où $T_i = \min(Y_i, C_i)$ et $\delta_i = \mathbb{I}_{\{Y_i \leq C_i\}}$ (l'indicatrice de non censure).

3.2 Estimation de la fonction de survie

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé. Soit Y_1, \dots, Y_n, \dots une suite de variables aléatoires (v.a.) positives indépendantes et identiquement distribuées (i.i.d.) de fonction de répartition (f.d.r.) commune H . Soit C_1, \dots, C_n, \dots une suite de v.a. de censure, positives, i.i.d. et de f.d.r. G . Généralement et conformément aux études biomédicales, les v.a. C_i sont supposées être indépendantes des Y_i . Soit $(T_i, \delta_i)_{i=1, \dots, n}$ l'échantillon réellement observé.

Posons $\bar{H}(t) = \mathbb{P}(Y > t)$, $\bar{G}(t) = \mathbb{P}(C > t)$, alors les T_i sont de f.d.r. L donnée par $\bar{L}(t) = \mathbb{P}(T > t) = \bar{H}(t)\bar{G}(t)$. Posons aussi $\tau_F = \sup \{y / \bar{F}(y) > 0\}$ où F est une f.d.r.

Définissons un processus ponctuel "d'événement" par

$$N_i(t) = \mathbb{I}_{\{T_i \leq t, \delta_i = 1\}}.$$

On note par $\bar{T}(t)$ le nombre de sujets présents jusqu'à t

$$\bar{T}(t) = \sum_{i=1}^n T_i(t) := \sum_{i=1}^n \mathbb{I}_{\{T_i > t\}}$$

et par $\bar{N}(t)$ le nombre d'observations non censurées inférieures ou égales à t

$$\bar{N}(t) = \sum_{i=1}^n N_i(t).$$

3.2.1 Estimateur de Nelson-Aalen

Définition 3.1. *L'estimateur de Nelson-Aalen est défini par*

$$\hat{\Lambda}_{NA}(t) = \int_0^t \frac{d\bar{N}(s)}{\bar{T}(s)} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{T_{(i)} \leq t\}}}{1 - \bar{N}(T_{(i)}^-)} \quad (3.2)$$

où

$$d\bar{N}(s) = \bar{N}(s) - \bar{N}(s^-).$$

Cet estimateur de Λ est fondé sur la remarque suivante

$$\begin{aligned} \Lambda(s + ds) - \Lambda(s) &\simeq \lambda(s)ds \\ &= \mathbb{P}(\text{événement ait lieu dans } (s, s + ds) / \text{à risque en } s). \end{aligned}$$

Il est alors naturel d'estimer cette quantité par $(\bar{N}(s + ds) - \bar{N}(s)) / \bar{T}(s)$. En sommant ces quantités sur les intervalles de $[0, t]$ et en faisant tendre ces intervalles vers 0, de telle sorte que chacun ne contienne qu'un seul événement, on obtient (3.2). Le comportement asymptotique de l'estimateur de Nelson-Aalen est résumé dans le théorème suivant

Théorème 3.1. *(Shorack et Wellner 1986 p. 304). Pour tout τ fixé, $\tau < \tau_H$,*

$$\sup_{0 \leq t \leq \tau} \left| \hat{\Lambda}_{NA}(t) - \Lambda(t) \right| \rightarrow 0 \text{ p.s. } \text{ quand } n \rightarrow +\infty.$$

Définissons le processus suivant, pour tout $0 \leq t \leq \tau_F$

$$Z_n(t) = \sqrt{n} \left(\hat{\Lambda}_{NA}(t) - \Lambda(t) \right).$$

Nous avons le résultat de normalité suivant

Théorème 3.2. *(Breslow et Cowley (1974)). Si $\tau < +\infty$ et si H et G sont continues, alors $Z_n \xrightarrow{\mathcal{L}} Z$ quand $n \rightarrow +\infty$, où Z est un processus gaussien centré tel que*

$$\text{cov}(Z(s), Z(t)) = \int_0^t \frac{dH(s)}{(1 - H(s))(1 - L(s))}.$$

3.2.2 Estimateur de Kaplan-Meier

Cet estimateur (que l'on notera EKM) est aussi appelé estimateur PL (Product Limit) car il s'obtient comme la limite d'un produit. En effet pour $t_1 < t$

$$\begin{aligned}\bar{H}(t) &= \frac{\mathbb{P}(Y > t, Y > t_1)}{\mathbb{P}(Y > t_1)} \mathbb{P}(Y > t_1) \\ &= \mathbb{P}(Y > t / Y > t_1) \bar{H}(t_1),\end{aligned}$$

pour $t_2 < t_1$ on a $\bar{H}(t_1) = \mathbb{P}(Y > t_1 / Y > t_2) \bar{H}(t_2)$ et ainsi de suite. Si l'on choisit pour les dates où l'on conditionne celles où il s'est produit un événement (mort ou censure), on estime seulement des quantités de la forme

$$\mathbb{P}(Y > T_{(i)} / Y > T_{(i-1)}) =: p_i,$$

p_i est la probabilité de survivre pendant l'intervalle $I_i =]T_{(i-1)}, T_{(i)}]$ quand on est vivant au début de cet intervalle. Soit R_i le nombre de sujets qui sont vivants (donc "à risque" de mourir) juste avant l'instant $T_{(i)}$. Soit M_i le nombre de morts à l'instant $T_{(i)}$ et soit $q_i = 1 - p_i =$ *Probabilité de mourir pendant l'intervalle I_i*

sachant qu'on était vivant au début de cet intervalle. Alors l'estimateur naturel de q_i est

$$\hat{q}_i = \frac{M_i}{R_i} = \frac{\text{nombre de morts à } T_{(i)}}{\text{nombre de sujets à risque}}.$$

Supposons qu'il n'y ait pas d'ex-aequo (c-à-d des temps de mort identiques pour plusieurs sujets). Si $\delta_i = 1$ c'est qu'il y a eu un mort en $T_{(i)}$ et donc $M_i = 1$, si $\delta_i = 0$ c'est qu'il y'a eu une censure en $T_{(i)}$ et donc $M_i = 0$. On a alors

$$\hat{p}_i = \begin{cases} 1 - \frac{1}{R_i} & \text{en cas de mort en } T_{(i)} \\ 1 & \text{en cas de censure.} \end{cases}$$

Il est clair que $R_i = n - i + 1$. On obtient finalement l'EKM pour la fonction de survie de la variable durée de vie Y

$$\hat{S}_{KM} =: \bar{H}_n(t) =: \begin{cases} \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n-i+1}\right)^{\delta_{(i)}} & \text{si } t < T_{(n)}, \\ 0 & \text{si } t \geq T_{(n)} \end{cases}$$

et donc on a aussi l'EKM pour la fonction de survie de la variable de censure C

$$\bar{G}_n(t) =: \begin{cases} \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n-i+1}\right)^{1-\delta_{(i)}} & \text{si } t < T_{(n)}, \\ 0 & \text{si } t \geq T_{(n)} \end{cases}$$

où $(T_{(i)}, \delta_{(i)})_{i=1, \dots, n}$ sont tels que $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$ et les $\delta_{(i)}$ sont les indicatrices concomitantes.

Remarque 3.1. *L'estimateur de Kaplan Meier peut aussi se mettre sous la forme suivante*

$$\bar{H}_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{\delta_{(i)}}{n-i+1}\right)^{\mathbf{1}_{\{T_{(i)} \leq t\}}} & \text{si } t < T_{(n)}, \\ 0 & \text{si } t \geq T_{(n)} \end{cases}$$

et

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbf{1}_{\{T_{(i)} \leq t\}}} & \text{si } t < T_{(n)}, \\ 0 & \text{si } t \geq T_{(n)}. \end{cases}$$

3.2.2.1 Propriétés asymptotiques de l'EKM

Consistance

Peterson (1977) a établi la consistance de l'EKM aussi bien pour la fonction de survie $\bar{H}(t)$ que pour celle de la censure $\bar{G}(t)$. Posons

$$\bar{H}^*(t) = \mathbb{P}(Y > t, \delta = 1) = \mathbb{E}(\mathbb{I}_{\{Y > t, \delta = 1\}}) \quad \text{et} \quad \bar{G}^*(t) = \mathbb{P}(C > t, \delta = 0) = \mathbb{E}(\mathbb{I}_{\{C > t, \delta = 0\}}),$$

fonctions qui peuvent être estimées par les moyennes empiriques

$$\bar{H}_n^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Y_i > t, \delta_i = 1\}} \quad \text{et} \quad \bar{G}_n^*(t) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{C_i > t, \delta_i = 0\}}.$$

Théorème 3.3. (Peterson(1977)). *Supposons que \bar{H} et \bar{G} n'aient aucun saut en commun, alors pour tout $t < \tau_L$*

$$\begin{aligned} \bar{H}_n(t) &\xrightarrow{p.s.} H(t) \quad \text{lorsque } n \rightarrow +\infty \\ \bar{G}_n(t) &\xrightarrow{p.s.} G(t) \quad \text{lorsque } n \rightarrow +\infty. \end{aligned}$$

Preuve Nous établissons la preuve pour le processus $\bar{G}_n(t)$, le cas $\bar{H}_n(t)$ est similaire. Pour une fonction S on note par $D(S)$ l'ensemble des points de sauts de S . Cox (1972) a montré que pour toute fonction de survie S ,

$$S(t) = \left\{ \exp \left(\int_{[0,t] \setminus D(S)} \frac{dS(s)}{S(s)} \right) \right\} \prod_{\substack{s \in D(S) \\ s \leq t}} \frac{S(s)}{S(s^-)}$$

où $S(s^-) = \lim_{u \nearrow s} S(u)$.

Donc

$$\bar{G}(t) = \left\{ \exp \left(\int_{[0,t] \setminus D(\bar{G})} \frac{d\bar{G}(s)}{\bar{G}(s)} \right) \right\} \prod_{\substack{s \in D(\bar{G}) \\ s \leq t}} \frac{\bar{G}(s)}{\bar{G}(s^-)} \quad (3.3)$$

$$\bar{G}_n(t) = \prod_{\substack{s \in D(\bar{G}_n) \\ s \leq t}} \frac{\bar{G}_n(s)}{\bar{G}_n(s^-)} \quad (\text{car } \bar{G}_n(\cdot) \text{ est une fonction de survie discrète}).$$

D'autre part on a

$$\begin{aligned} \bar{G}^*(t) &= \mathbb{P}(C > t, \delta = 0) = \mathbb{P}(C > t, Y > C) \\ &= \mathbb{E}(\mathbb{P}(C > t, \delta = 0 / C)) = \mathbb{E}(\mathbb{E}(\mathbb{I}_{\{C > t\}} \mathbb{I}_{\{Y > C\}} / C)) \\ &= \mathbb{E}[\mathbb{I}_{\{C > t\}} \mathbb{E}(\mathbb{I}_{\{Y > C\}})] \quad (\text{car } Y \text{ est indépendant de } C) \\ &= \mathbb{E}[\mathbb{I}_{\{C > t\}} \bar{H}(C)] = - \int_t^{+\infty} \bar{H}(s) d\bar{G}(s); \end{aligned} \quad (3.4)$$

Notons que

$$\bar{G}^*(t) + \bar{H}^*(t) = \mathbb{P}(T > t) = \bar{L}(t) = \bar{G}(t) \bar{H}(t). \quad (3.5)$$

Définissons maintenant, pour toutes fonctions S_1, S_2 la fonction Φ par

$$\Phi(S_1, S_2, t) = \left\{ \exp \left(\int_{[0,t] \setminus D(S_1)} \frac{dS_1(s)}{S_1(s) + S_2(s)} \right) \right\} \prod_{\substack{s \in D(S_1) \\ s \leq t}} \frac{S_1(s) + S_2(s)}{S_1(s^-) + S_2(s^-)}.$$

Nous allons montrer d'abord que $\bar{G}(t) = \Phi(\bar{G}^*, \bar{H}^*, t)$, la convergence de \bar{G}_n vers \bar{G} découlera de la continuité de la fonction Φ au sens de la norme sup. D'après l'égalité (3.4), \bar{G}^* et \bar{G} ont les mêmes points de sauts, de plus nous avons

$$\begin{aligned} \frac{d\bar{G}^*(s)}{\bar{G}^*(s) + \bar{H}^*(s)} &= \frac{d\bar{G}^*(s)}{\bar{G}(s) \bar{H}(s)} \quad (\text{d'après (3.5)}) \\ &= \frac{\bar{H}(s) d\bar{G}(s)}{\bar{G}(s) \bar{H}(s)} \quad (\text{d'après (3.4)}) \\ &= \frac{d\bar{G}(s)}{\bar{G}(s)} \end{aligned}$$

et

$$\frac{\bar{H}^*(s) + \bar{G}^*(s)}{\bar{H}^*(s^-) + \bar{G}^*(s^-)} = \frac{\bar{H}(s)\bar{G}(s)}{\bar{G}(s^-)\bar{H}(s^-)} \quad (\text{d'après (3.5)}).$$

Si s est un point de saut pour la fonction $\bar{G}(\cdot)$, alors comme par hypothèse, \bar{H} et \bar{G} n'ont pas les mêmes points de sauts donc s ne peut pas être un point de saut pour \bar{H} , d'où $\bar{H}(s^-) = \bar{H}(s)$ et nous avons

$$\frac{\bar{H}^*(s) + \bar{G}^*(s)}{\bar{H}^*(s^-) + \bar{G}^*(s^-)} = \frac{\bar{G}(s)}{\bar{G}(s^-)}$$

et donc $\Phi(\bar{G}^*, \bar{H}^*, t) = \bar{G}(t)$ d'après (3.3). D'un autre côté, nous avons

$$\bar{H}_n^*(s) + \bar{G}_n^*(s) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{T_i > t\}} =: \bar{L}_n(t) \quad (3.6)$$

et

$$\begin{aligned} \bar{G}_n(t)\bar{H}_n(t) &= \prod_{T_{(i)} \leq t} \left(1 - \frac{1}{n-i+1}\right)^{1-\delta_i} \left(1 - \frac{1}{n-i+1}\right)^{\delta_i} = \prod_{T_{(i)} \leq t} \left(\frac{n-i}{n-i+1}\right) \\ &= \frac{n-1}{n} \frac{n-2}{n-1} \cdots \frac{n - \sum_{i=1}^n \mathbb{I}_{\{T_i \leq t\}}}{n - \sum_{i=1}^n \mathbb{I}_{\{T_i \leq t\}} + 1} = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{T_i > t\}} =: \bar{L}_n(t). \end{aligned} \quad (3.7)$$

Nous obtenons alors d'après (3.6) et (3.7)

$$\frac{\bar{H}_n^*(s) + \bar{G}_n^*(s)}{\bar{H}_n^*(s^-) + \bar{G}_n^*(s^-)} = \frac{\bar{G}_n(s)\bar{H}_n(s)}{\bar{G}_n(s^-)\bar{H}_n(s^-)} = \frac{\bar{G}_n(s)}{\bar{G}_n(s^-)}$$

d'où $\Phi(\bar{G}_n^*, \bar{H}_n^*, t) = \bar{G}_n(t)$ pour $t \leq T_{(n)}$. Par le théorème de Glivenko-Cantelli on a que $\bar{G}_n^*(t)$ et $\bar{H}_n^*(t)$ convergent presque sûrement vers $\bar{G}^*(t)$ et $\bar{H}^*(t)$ (resp.). La continuité de Φ assure que $\Phi(\bar{G}_n^*, \bar{H}_n^*, t)$ converge presque sûrement vers $\Phi(\bar{G}^*, \bar{H}^*, t)$, c-à-d que $\bar{G}_n(t) \xrightarrow{p.s.} \bar{G}(t)$.

Normalité asymptotique

Posons $L_1(t) = \mathbb{P}(T \leq t, \delta = 1) = \mathbb{P}(T_i \leq \min(t, C_i)) = \int_0^t (1 - G(s)) dH(s)$,

et $L_0(t) = \mathbb{P}(T \leq t, \delta = 0)$. La variance asymptotique de $\bar{H}_n(t)$ est donnée par

$$V(t) = \int_0^t \frac{d\bar{L}_1(s)}{(1 - L(s^-))^2} = \int_0^t \frac{dH(s)}{(1 - H(t^-))^2 (1 - G(s^-))}.$$

L'estimateur empirique de $V(t)$ est obtenu à l'aide de la formule d'approximation de Greenwood

$$V_n(t) = \int_0^t \frac{d\bar{N}(s)}{(1 - \bar{T}(s^-))^2} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{T_{(i)} \leq t\}}}{(1 - \bar{N}(T_{(i)}^-))^2}.$$

$V_n(t)$ est un estimateur consistant de $V(t)$; Klein (1991) montra que V_n est plus performant que d'autres estimateurs de la variance de \bar{H}_n .

Définissons le processus suivant, pour tout $0 \leq t \leq \tau_F$

$$U_n(t) = \sqrt{n} (\bar{H}_n(t) - H(t)).$$

Nous avons le résultat de normalité suivant

Théorème 3.4. (Breslow et Cowley (1974)). Si $\tau < +\infty$ et si H et G sont continues, alors $U_n \xrightarrow{\mathcal{L}} U$ quand $n \rightarrow +\infty$ où U est un processus gaussien centré tel que

$$\text{cov}(U(s), U(t)) = (1 - H(s))(1 - H(t)) \int_0^s \frac{dL(y)}{(1 - H(y))^2}.$$

Le premier type d'approximation du processus $(\bar{H}_n - H)$ par une moyenne de v.a. i.i.d. fut obtenu par Burke et al (1981,1986) avec un reste de l'ordre de $O\left(\frac{\log^2 n}{n^{1/2}}\right)$. Ce résultat est basé sur les travaux de Komlos et al (1975) sur l'approximation des processus empiriques. Un second type d'approximation fut établi par Lo et Singh (1986) avec un taux de $O\left(\left(\frac{\log n}{n}\right)^{3/4}\right)$. Ce taux a été ensuite amélioré à $O\left(\frac{\log n}{n}\right)$ par Lo et al (1989).

3.3 La régression non paramétrique pour des données censurées

Ce type de modèle a été largement étudié dans la littérature, on peut citer Gonzalez-Manteiga et Cadarso-Suarez (1991), Stute (1993, 1999), Cai (2003), Cai et Ould-Said (2003) etc...

Lorsque les données $(Y_i)_{i=1, \dots, n}$ sont i.i.d. et complètement disponibles, l'estimateur usuel de la moyenne des Y_i est la moyenne arithmétique définie par

$\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$. Si par contre on est sur un modèle censuré (T_i, δ_i) avec une distribution G de la censure connue, alors un estimateur non biaisé de $\mathbb{E}(Y)$ est donné par

$$\hat{Y} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i T_i}{G(T_i)}. \quad (3.8)$$

En effet, si on suppose l'indépendance entre Y et C , alors d'après les propriétés de l'espérance conditionnelle on a

$$\begin{aligned}\mathbb{E}(\hat{Y} | X) &= \mathbb{E}\left(\frac{\delta T}{\bar{G}(T)} | X\right) = \mathbb{E}\left(\frac{\mathbb{I}_{\{Y \leq C\}}}{\min}(Y, C) \bar{G}(\min(Y, C)) | X\right) \\ &= \mathbb{E}\left(\mathbb{I}_{\{Y \leq C\}} \frac{Y}{\bar{G}(Y)} | X\right) = \mathbb{E}\left[\mathbb{E}\left(\mathbb{I}_{\{Y \leq C\}} \frac{Y}{\bar{G}(Y)} | X, Y\right) | X\right] \\ &= \mathbb{E}\left(\bar{G}(\cdot) \frac{Y}{\bar{G}(Y)} | X\right) = \mathbb{E}(Y | X) = r(X).\end{aligned}$$

Lorsque \bar{G} est inconnue, on le remplace dans (3.8) par l'EKM $\bar{G}_n(\cdot)$ et on obtient ce qu'on

appelle la moyenne de Kaplan-Meier définie par

$$M_n := \frac{1}{n} \sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}_n(T_i)}.$$

Sursula, Tsai et Van Rysin (1984) ont montré que

$$M_n = \int_0^{+\infty} \bar{H}_n(t) dt = \int_0^{\tau_H} \bar{H}_n(t) dt$$

comme H_n converge presque sûrement vers H alors d'après le théorème de convergence dominée, M_n converge presque sûrement vers $\mathbb{E}(Y)$.

Pour la variable Y représentant une durée de vie on considère la covariable X et la variable de censure C . Dans le modèle censuré à droite ce qu'on observe donc c'est le triplet $(X_i, T_i, \delta_i)_{i=1, \dots, n}$. La loi du vecteur aléatoire (X, T, δ) n'identifie pas la loi conditionnelle de Y sachant X . Le problème de l'identifiabilité peut-être résolu en imposant des conditions supplémentaires sur le vecteur (X, T, δ) . Deux cas sont envisagés

- Cas où (X, Y) est indépendant de C .
- Cas où Y et C sont indépendants conditionnellement sachant X .

Nous allons établir la consistance de l'estimateur de la fonction de régression dans chacun de ces deux cas, ce qui va impliquer que la loi du vecteur (X, T, δ) identifie la loi du vecteur (X, Y) .

3.3.1 Estimation de la régression dans le cas où (X, Y) est indépendant de C

Le modèle où on suppose l'indépendance entre la censure C et le couple (X, Y) est assez plausible dans la réalité : C'est un modèle qu'on peut retrouver dans le cas où la censure est indépendante des caractéristiques de l'individu. On peut donner l'exemple d'une compagnie d'assurance qui offre des contrats d'assurances maladie à ses clients : soit Y_i la durée de vie de chaque client pour lequel la compagnie dispose d'un dossier médical où plusieurs facteurs sont enregistrés et on note par X le vecteur de covariables représentant ces facteurs. Il arrive que certains clients rompent leurs contrat d'assurance (par exemple par manque d'argent, parce qu'ils changent d'assureur, etc...), et donc la durée de vie dont dispose l'assureur pour ces clients c'est leurs temps de censures C_i (date à laquelle le client n'est plus dans le fichier de la compagnie). On est donc clairement en présence d'une indépendance entre C et (X, Y) .

Dans toute la suite de cette section nous supposerons les conditions suivantes vérifiées :

(C1) (X, Y) est indépendant de C .

(C2) \bar{H} et \bar{G} n'ont aucun saut en commun.

(C3) $\tau_H < +\infty$

(C4) \bar{G} est continue en τ_H et $\bar{G}(\tau_H) > 0$.

Remarque 3.2. La condition $\bar{G}(\tau_H) > 0$ implique que $\tau_H < \tau_G$ et on a alors $\tau_H = \tau_L$ par définition même de τ_L .

Définition 3.2. On appelle estimateur par moyenne locale pour un modèle censuré, un estimateur de la forme

$$\tilde{m}_n(x) = \sum_{i=1}^n W_{ni}(x) \frac{\delta_i T_i}{\bar{G}_n(T_i)}, \quad (3.9)$$

où les poids $W_{ni}(x)$ sont positifs ou nuls et vérifient $\sum_{i=1}^n W_{ni}(x) \leq 1$.

Les exemples les plus répandus sont les poids associés aux estimateurs à noyaux, aux k plus proches voisins et au partitionnement.

Les résultats qui vont suivre sont des résultats de consistance forte, c'est à dire de convergence

en moyenne quadratique, pour deux estimateurs particuliers, à savoir l'estimateur de Nadaraya-Watson et l'estimateur des k points les plus proches. Nous énoncerons pour chacun d'eux un théorème de consistance forte et nous donnerons une preuve pour l'estimateur de Nadaraya-Watson, on pourra se référer au livre de Györfi et al (2002) pour des preuves complètes.

Nous commençons par donner la définition d'un noyau régulier.

Définition 3.3. *Le noyau K est dit régulier s'il est positif et s'il existe un intervalle $I_{0,l}$ centré en 0, de longueur $2l$ et une constante $a > 0$ tels que*

$$a\mathbb{1}_{\{x \in I_{0,l}\}} \leq K(x) \leq 1$$

et

$$\int \sup_{u \in x+I_{0,l}} K(u) du < \infty.$$

Théorème 3.5. (Köhler–Mathé–Pinter 2002). *Sous les conditions (C1-C4) , s'il existe une constante M telle que $|Y| < M$, si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$, nous avons*

$$\int (\hat{r}_{NW;n}(x) - r(x))^2 f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Où $\hat{r}_{NW;n}(x)$ est défini à partir de (3.9) en prenant $W_{ni}(x) = \frac{K(\frac{x-X_i}{h_n})}{\sum_{i=1}^n K(\frac{x-X_i}{h_n})}$.

La preuve de ce théorème utilise les trois lemmes suivants qu'on trouve dans Györfi et al (2002). Posons

$$\tilde{r}_{1,n}(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i}{G(T_i)} K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)} \tag{3.10}$$

$$\bar{r}_{1,n}(x) = \frac{\sum_{i=1}^n \frac{\delta_i T_i}{G_n(T_i)} K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)}. \tag{3.11}$$

Lemme 3.1. (Györfi et al. 2002, lemme 23.6 p. 479). *Si le noyau K est régulier, alors il existe une constante $\rho = \rho(K)$ tel que, pour tout $x \in \mathbb{R}$ et $h_n > 0$,*

$$\int \frac{K\left(\frac{x-X_i}{h_n}\right)}{\int \left(K\left(\frac{x-X_i}{h_n}\right)\right) f(x) dx} f(x) dx \leq \rho. \tag{3.12}$$

Lemme 3.2. (Györfi et al. 2002, lemme 23.9 p. 483). Pour n assez grand et s'il existe une constante M telle que $|Y| < M$, nous avons

$$\mathbb{P} \left\{ \int |r(x) - \tilde{r}_{1,n}(x)| f(x) dx > \varepsilon \right\} \leq \exp(-n\varepsilon^2 | (8M\rho)^2). \quad (3.13)$$

Lemme 3.3. (Györfi et al. 2002, lemme 26.1 p. 549). Sous les conditions (C3-C4), nous avons

$$\frac{1}{n} \sum_{j=1}^n \delta_j T_j \left| \frac{1}{\bar{G}(T_j)} - \frac{1}{\bar{G}_n(T_j)} \right| \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty. \quad (3.14)$$

Preuve du Théorème 3.5

Remarquons que

$$\int (\hat{r}_{NW;n}(x) - r(x))^2 f(x) dx \leq cst \int |\hat{r}_{NW;n}(x) - r(x)| f(x) dx$$

il suffit donc, pour établir la résultat, de montrer que

$$\int |\hat{r}_{NW;n}(x) - r(x)| f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Or

$$\begin{aligned} |\hat{r}_{NW;n}(x) - r(x)| &\leq |\hat{r}_{NW;n}(x) - \bar{r}_{1,n}(x)| + |\bar{r}_{1,n}(x) - \tilde{r}_{1,n}(x)| + |\tilde{r}_{1,n}(x) - r(x)| \\ &\leq I_{n,1}(x) + I_{n,2}(x) + I_{n,3}(x). \end{aligned}$$

D'après (3.13) nous avons

$$\int I_{n,3}(x) f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

D'autre part d'après (3.12) nous avons

$$\int \frac{K\left(\frac{x-X_i}{h_n}\right)}{\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)} f(x) dx \leq \rho. \quad (3.15)$$

Donc

$$\begin{aligned} \int I_{n,2}(x) f(x) dx &\leq \frac{1}{n} \sum_{i=1}^n \left| \frac{\delta_i T_i}{\bar{G}(T_i)} - \frac{\delta_i T_i}{\bar{G}_n(T_i)} \right| \int \frac{K\left(\frac{x-X_i}{h_n}\right)}{\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)} f(x) dx \\ &\leq \rho \frac{1}{n} \sum_{i=1}^n \left| \frac{\delta_i T_i}{\bar{G}(T_i)} - \frac{\delta_i T_i}{\bar{G}_n(T_i)} \right| \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty \end{aligned}$$

d'après (3.14). De plus

$$\begin{aligned} I_{n,1}(x) &= |\hat{r}_{NW;n}(x) - \bar{r}_{1,n}(x)| \leq |\hat{r}_{NW;n}(x)| \left| 1 - \frac{\bar{r}_{1,n}(x)}{\hat{r}_{NW;n}(x)} \right| \\ &\leq |\hat{r}_{NW;n}(x)| \left| 1 - \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)} \right| \end{aligned}$$

et comme $|\hat{r}_{NW;n}(x)| \leq \frac{\tau_H}{G(\tau_H)} =: c$, alors d'après (3.13)

$$\begin{aligned} \int I_{n,1}(x)f(x)dx &\leq c \int \left| 1 - \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)}{n\mathbb{E}\left(K\left(\frac{x-X_i}{h_n}\right)\right)} \right| f(x)dx \\ &\longrightarrow 0 \quad p.s. \end{aligned}$$

Ceci achève la démonstration.

Théorème 3.6. (Köhler-Mathé-Pinter 2002). *Sous les conditions (C1-C4), s'il existe une constante M telle que $|Y| < M$, si $k_n \rightarrow \infty$ et $k_n/n \rightarrow 0$, nous avons*

$$\int (\hat{r}_{k-NN;n}(x) - r(x))^2 f(x)dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Où $\hat{r}_{k-NN;n}(x)$ est défini à partir de (3.9) en prenant $W_{ni}(x) = \frac{1}{k_n} \mathbb{I}_{J_x}$ et où J_x représente l'ensemble

$$J_x = \{i \mid X_i \text{ est l'une des } k \text{ observations les plus proches de } x\},$$

La preuve de ce théorème est basée sur un résultat de Pinter (2001) qui établie la consistance pour l'estimateur par moyenne locale $\tilde{m}_n(x)$ défini en (3.9) sous les conditions suivantes sur $W_{ni}(x)$

$$\lim_{n \rightarrow \infty} \int \sum_{i=1}^n (W_{ni}(x)Y_i - r(x))^2 f(x)dx = 0 \quad p.s. \quad (3.16)$$

et

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n \left| \frac{\delta_i T_i}{G(T_i)} - \frac{\delta_i T_i}{G_n(T_i)} \right| \int W_{ni}(x)f(x)dx = 0 \quad p.s. \quad (3.17)$$

On montre par la suite que les poids $W_{ni}(x) = \frac{1}{k_n} \mathbb{I}_{J_x}$, de l'estimateur des k -voisins les plus proches vérifient les conditions (3.16) et (3.17).

3.3.2 Estimation de la régression dans le cas où Y et C sont indépendantes conditionnellement à X

Pour ce type de modèle, l'estimation de la régression a été étudié par plusieurs auteurs, on peut citer Beran (1981) qui en a donné une première approche complète, en introduisant une classe d'estimateurs non paramétriques pour estimer la distribution de survie conditionnelle en présence de censure à droite. Dabrowska (1987, 1989) a ensuite donné quelques résultats de consistance pour cet estimateur. Une alternative à l'approche de Beran a été donnée par Horvath (1981) qui a proposé d'estimer la distribution de survie conditionnelle en intégrant un estimateur de la densité de survie conditionnelle.

Posons

$$\begin{aligned}\bar{H}(y | x) &= \mathbb{P}(Y > y | X = x) \\ \bar{G}(y | x) &= \mathbb{P}(C > y | X = x) \\ \bar{L}(y | x) &= \mathbb{P}(T > y | X = x) = \bar{H}(y | x)\bar{G}(y | x)\end{aligned}$$

et pour toute f.d.r. conditionnelle F

$$\tau_F(x) = \sup\{y : \bar{F}(y | x) > 0\}.$$

Deux problèmes se posent alors : le premier est l'estimation des distributions conditionnelles $\bar{H}(y | x)$ et $\bar{G}(y | x)$ et le second est d'estimer la fonction de régression $r(x) = \mathbb{E}(Y | X = x)$. Pour résoudre ces deux problèmes nous définissons d'abord les estimateurs par moyenne locale de $\bar{H}(y | x)$ et $\bar{G}(y | x)$, nous donnons un théorème de consistance forte pour ces estimateurs, dû à Beran (1981) et repris par Györfi et al (2002) puis nous donnons des résultats de consistance forte pour les deux estimateurs de la régression précédemment étudiés, à savoir l'estimateur de Nadaraya-Watson et l'estimateur des k -voisins les plus proches.

Définition 3.4. *Les estimateurs par moyenne locale des fonctions $\bar{H}(y | x)$ et $\bar{G}(y | x)$ sont définis par*

$$\bar{H}_n(y | x) = \prod_{\substack{T_{(i)} \leq y \\ \sum_{j=i}^n W_{n(j)}(x) > 0}} \left(\frac{\sum_{j=i+1}^n W_{n(j)}(x)}{\sum_{j=i}^n W_{n(j)}(x)} \right)^{\delta_{(i)}} \mathbb{I}_{\{y \leq T_{(n)}\}}$$

et

$$\bar{G}_n(y | x) = \prod_{\substack{T_{(i)} \leq y \\ \sum_{j=i}^n W_{n(j)}(x) > 0}} \left(\frac{\sum_{j=i+1}^n W_{n(j)}(x)}{\sum_{j=i}^n W_{n(j)}(x)} \right)^{1-\delta_{(i)}} \mathbb{I}_{\{y \leq T_{(n)}\}},$$

où $(T_{(i)}, \delta_{(i)})_{i=1, \dots, n}$ sont tels que $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$ et les $\delta_{(i)}$ sont les indicatrices concomitantes. Les $W_{ni}(x)$ sont des des fonctions de poids dépendant de x , vérifiant $W_{ni}(x) \geq 0$ et $\sum_{i=1}^n W_{ni}(x) \leq 1$, et tels que $W_{n(i)}(x)$ sont les poids ordonnés associés a l'ordre des $T_{(i)}$.

Remarque 3.3. Les estimateurs par partitionnement, par noyau et les k voisins les plus proches sont des exemples d'estimateurs par moyenne locale.

Remarque 3.4. Si on prend $W_{ni}(x) = \frac{1}{n}$ pour tout i alors

$$\bar{H}_n(y | x) = \prod_{T_{(i)} \leq y} \left(\frac{\sum_{j=i+1}^n W_{n(j)}(x)}{\sum_{j=i}^n W_{n(j)}(x)} \right)^{\delta_{(i)}} \mathbb{I}_{\{y \leq T_{(n)}\}} = \prod_{T_{(i)} \leq y} \left(\frac{n-i}{n-i+1} \right)^{\delta_{(i)}} \mathbb{I}_{\{y \leq T_{(n)}\}}$$

on retrouve l'estimateur de Kaplan-Meier de $\bar{H}_n(y)$.

Supposons que

- (D1) Y et C indépendants conditionnellement à X .
- (D2) \bar{H} et \bar{G} n'ont aucun saut en commun.
- (D3) $\tau_H(x) < \tau_G(x)$ pour tout x .

Théorème 3.7. (Györfi et al. 2002 p. 556). Sous les conditions (D1-D3) et si les poids $W_{ni}(x)$ sont tels que l'estimateur de la regression par moyenne locale converge uniformément presque sûrement pour Y borné, alors pour $y < \tau_L(x)$ et pour $n \rightarrow \infty$, nous avons

$$\begin{aligned} \bar{H}_n(y | x) &\rightarrow \bar{H}(y | x) \quad p.s. \\ \bar{G}_n(y | x) &\rightarrow \bar{G}(y | x) \quad p.s. \end{aligned}$$

La preuve est analogue en tout points à celle du théorème 3.3.

Nous allons maintenant donner un estimateur de la fonction de régression $r(x)$ à partir de l'estimateur de la fonction de survie conditionnelle $\bar{H}_n(y | x)$.

Posons

$$r_n(x) = \int_0^\infty \bar{H}_n(y | x) dy = \int_0^{T(n)} \bar{H}_n(y | x) dy. \quad (3.18)$$

Théorème 3.8. (Pinter 2001) *Sous les conditions du théorème 3.7, nous avons*

$$\int (r_n(x) - r(x))^2 f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Preuve Comme Y est borné par τ_H , alors $r(x) \leq \tau_H$ et donc

$$|r(x) - r_n(x)| \leq \tau_H$$

la condition (D3) implique que $\tau_H(x) = \tau_L(x)$ et donc

$$\begin{aligned} \int (r_n(x) - r(x))^2 f(x) dx &\leq \tau_H \int |r(x) - r_n(x)| f(x) dx \\ &\leq \tau_H \int \left| \int_0^\infty (\bar{H}(y | x) - \bar{H}_n(y | x)) \right| f(x) dx \\ &\leq \tau_H \int \left| \int_0^{\tau_H(x)} (\bar{H}(y | x) - \bar{H}_n(y | x)) \right| f(x) dx \end{aligned}$$

qui tend vers 0 d'après le théorème précédent et le théorème de la convergence dominée.

Pinter (2001) a montré que l'estimateur (3.18) s'écrit aussi

$$r_n(x) = \sum_{i=1}^n W_{ni}(x) \frac{\delta_i T_i}{\bar{G}_n(T_i | X_i)} \quad (3.19)$$

où les poids $W_{ni}(x)$ sont ceux définissant $\bar{H}_n(y | x)$.

Les conditions sur les poids $W_{ni}(x)$ dans le théorème 3.7 sont satisfaites pour les estimateurs à noyaux et pour les k -voisins les plus proches et nous avons comme conséquences de ces résultats, les corollaires suivants, pour l'estimateur de Nadaraya-Watson et l'estimateur des k -voisins les plus proches définis à partir de (3.19) (c'est-à-dire avec

$$W_{ni}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \text{ et } W_{ni}(x) = \frac{1}{k_n} \mathbb{I}_{J_x} \text{ resp.}).$$

Corollaire 3.1. *Sous les conditions (D1-D3) et si $h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ alors*

$$\int (\hat{r}_{NW;n}(x) - r(x))^2 f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Corollaire 3.2. *Sous les conditions (D1-D3) et si $k_n/\log n \rightarrow \infty$ et $k_n/n \rightarrow 0$ alors*

$$\int (\hat{r}_{k-NN;n}(x) - r(x))^2 f(x) dx \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty.$$

Chapitre 4

Vitesse de convergence et normalité asymptotique de l'estimateur de Nadaraya-Watson dans le modèle censuré

Dans ce qui suit, nous étudions le comportement de l'estimateur de Nadaraya-Watson de la fonction de régression dans le cas d'un modèle censuré à droite. Nous établissons la convergence uniforme presque sûre sur des compacts et donnons la vitesse de convergence pour cet estimateur. Nous prouvons par la suite la normalité asymptotique de l'estimateur considéré, et illustrons nos résultats théoriques par des simulations pour différentes fonctions de régression. Nous présentons nos résultats dans leur version originale.

4.1 Introduction

Let $(Y_n)_{n \geq 1}$ be a sequence of independent and identically distributed (iid) random variables (rv), with a common unknown absolutely continuous distribution function (df) H . In survival analysis, the random variables can be lifetimes of patients under study. In reality it is not possible to observe the survival time of all patients, and often some of them are still alive at the end of the study, or withdraw, or die from other causes than those addressed by the study. In those cases, we observe another rv C called censoring. Then the observable rv is the minimum of the survival time and the censoring time. Assuming that $(C_n)_{n \geq 1}$ is a sequence of iid censoring random variables with a common unknown continuous distribution function G , we then observe the pairs $(T_i, \delta_i)_{i \geq 1}$ where $T_i = \min(Y_i, C_i)$ and $\delta_i = \mathbb{I}_{\{Y_i \leq C_i\}}$ (the indicator of nocensoring). Now let X be a real-valued random covariable with density $f(\cdot)$. In regression analysis the

purpose is that of estimating Y given X , ie : finding a function $m^*(X)$ which achieves the minimum of the mean squared error :

$$\mathbb{E} |m^*(X) - Y|^2 = \min_m \mathbb{E} |m(X) - Y|^2.$$

It is well known that the solution of this problem is the regression function $r(x)$ defined by

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

In statistics literature, many papers deal with estimating the regression function. In the case of complete data, the first result concerning the existence of weakly universally consistent estimates was established by Stone (1977) who considered the nearest neighbor estimates. Later, this result was extended to other types of estimates like kernel, partitioning, least square, smoothing spline, local polynomial by Lugosi and Zeger (1995), Györfi and Walk (1996, 1997), Walk (1997), Köhler (1997, 1999), and Köhler and Krzyzak (2001). In the case of randomly right-censored data, under the linear model,

$$Y = \Upsilon X + \varepsilon,$$

where Υ is an unknown real-valued parameter of interest and the error ε has zero mean and is uncorrelated with X , the purpose is to estimate the linear regression parameter Υ by parametric methods. Far from being exhaustive, we can quote Miller (1977), Buckley and James (1979), Koul, Suarla and Van Rysin (1981) for parametric models and Gonzalez-Manteiga and Cadarso-Suarez (1991), Stute (1993, 1999), Cai (2003) and Cai and Ould-Saïd (2003) for a nonparametric approach. In the case of non linear regression,

$$Y = r(X) + \varepsilon,$$

where $r(X)$ is a non linear function, the first fully non parametric approach in the censoring case was given by Beran (1981), who introduced a class of nonparametric regression estimators for the conditional survival function in the presence of right-censoring. He proved some consistency results of these estimates, which were extended by Dabrowska (1987, 1989) : both of them assumed Y and C to be conditionally independent given X . In order to get a consistent estimate of the regression function $r(x)$, the estimator of the conditional survival function could be used, though the computations may be difficult practically. Carbonez *et al.* (1995) introduced a nonparametric estimator of $r(x)$ -the so-called partitioning estimate- which generalizes the notion of partitioning estimate and proved its strong consistency. Köhler *et al.* (2002) gave a simpler proof in the randomly right-censoring case for kernel, nearest neighbor, least squares and penalized least squares estimates. Both of Carbonez *et al.* (1995) and Köhler *et al.* (2002),

assumed (X, Y) and C to be independent. This condition is plausible whenever the censoring is independent of the characteristics of the patients under study. Furthermore, as Y is a lifetime, it can be assumed to be bounded.

Our goal is to establish strong uniform convergence with a rate for the kernel regression estimate under random censorship and state its asymptotic normality. We point out that Köhler *et al.* (2002) did not obtain the asymptotic normality result. In the case of the conditional quantile estimate for the same model Ould-Saïd (2006a) established strong consistency with rates while Ould Saïd and Lemdani (2006b) proved the strong consistency and the asymptotic normality of the kernel estimator of $r(x)$ for a left-truncated model and gave a rate of convergence for the estimate.

The paper is organized as follows : in Section 2 we define a kernel conditional regression estimator for the random censorship model . In Section 3 we give the assumptions and the main results with an application to build a confidence band. Some simulations are drawn to grant further support to our theoretical results for the consistency (in different cases) as well as the asymptotic normality in Section 4. The proofs, together with auxiliary results are collected in Section 5.

4.2 Construction of the estimator of $r(\cdot)$

We suppose that the joint density of (X, Y) denoted by $f_{X,Y}$ exists, then $r(x)$ can be written as

$$r(x) = \frac{\int_{\mathbb{R}} y f_{X,Y}(x, y) dy}{f(x)} =: \frac{\bar{r}_1(x)}{f(x)}. \quad (4.1)$$

In the case of no censoring, it is well known that a linear estimate of the conditional regression is given by

$$r_n(x) = \sum_{i=1}^n W_{in}(x) Y_i,$$

where $W_{in}(x) = W_{in}(x, X_1, X_2, \dots, X_n)$ are weights that depend on x and X_1, X_2, \dots, X_n . We recall that the well-known kernel estimator is obtained by choosing the following weights

$$W_{in}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)},$$

where K is a kernel function on \mathbb{R} and (h_n) a non negative sequence which goes to zero as n goes to infinity. (In the case where the denominator of $W_{ni}(x)$ is zero, we use the convention $0/0 = 0$.)

In the censoring model, following Carbonez *et al.* (1995) and Köhler *et al.* (2002), we define $\tilde{r}_n(x)$ as an estimate of $r(x)$ by

$$\tilde{r}_n(x) = \sum_{i=1}^n W_{in}(x) \frac{\delta_i T_i}{\bar{G}(T_i)} =: \frac{\tilde{r}_{1,n}(x)}{f_n(x)},$$

where

$$\tilde{r}_{1,n}(x) = \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}(T_i)} K\left(\frac{x - X_i}{h_n}\right),$$

$f_n(\cdot)$ is the well-known kernel estimator of the marginal density $f(\cdot)$ given by

$$f_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right), \quad (4.2)$$

and $\bar{G} = 1 - G$ is the survival function of the censoring rv C .

In practice, G is usually unknown and we replace it by the corresponding Kaplan-Meier (1958) estimator G_n defined by

$$1 - G_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1 - \delta_{(i)}}{n - i + 1}\right)^{\mathbf{1}_{\{T_{(i)} \leq t\}}}, & \text{if } t < T_{(n)}, \\ 0, & \text{if } t \geq T_{(n)} \end{cases}$$

where $T_{(1)} \leq T_{(2)} \leq \dots \leq T_{(n)}$ are the order statistics of the T_i and $\delta_{(i)}$ is the indicator of nocensoring associated to $T_{(i)}$.

Hence a feasible estimator of $r(x)$ is given by

$$r_n(x) = \sum_{i=1}^n W_{in}(x) \frac{\delta_i T_i}{\bar{G}_n(T_i)} := \frac{r_{1,n}(x)}{f_n(x)}, \quad (4.3)$$

where

$$r_{1,n}(x) = \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}_n(T_i)} K\left(\frac{x - X_i}{h_n}\right). \quad (4.4)$$

The estimator (4.3) has already been used by Köhler *et al.* (2002) with a slight difference at the endpoint in the Kaplan-Meier estimator.

Let $\tau_H = \sup\{y, \bar{H}(y) > 0\}$ and $\tau_G = \sup\{y, \bar{G}(y) > 0\}$ be a upper endpoints of \bar{H} and \bar{G} , respectively. We assume that $\tau_H < \infty$, $\bar{G}(\tau_H) > 0$ (this implies that $\tau_H \leq \tau_G$) and that C and (X, T) are independent.

Remark 4.1. *This last condition can be weakened to the standard well-known condition, that T and C are conditionally independent given $X = x$. In this case, we propose the following estimator for the regression function $r(x)$*

$$\widehat{r}_n(x) = \sum_{i=1}^n W_{in}(x) \frac{\delta_i T_i}{\bar{G}_n(T_i | X_i)} := \frac{\widehat{r}_{1,n}(x)}{f_n(x)}$$

where $\bar{G}_n(T_i | X_i)$ is Beran's estimator of the survival conditional distribution of the censored rv C given X . As far as we know and contrarily to the unconditional case (see Deheuvels and Einmahl (2000)) no rate of convergence for this estimate has so far been stated. We think this issue has to be addressed if we aim to get rates of convergence.

4.3 Assumptions and main results

Let \mathcal{C} be a compact set of \mathbb{R} which is included in $\mathcal{C}_0 = \{x \in \mathbb{R} / f(x) > 0\}$. We will make use of the following assumptions .

A1. The bandwidth h_n satisfies : $\lim_{n \rightarrow +\infty} h_n = 0$, $\lim_{n \rightarrow +\infty} \frac{nh_n}{\log n} = +\infty$ and $\log \log n = o(\frac{1}{h_n^\mu})$ where $0 < \mu < 1$.

A2. The kernel K is bounded, symmetric and has compact support. It is also Hölderian of order $\gamma > 0$.

Furthermore $\int_{\mathbb{R}} t |K(t)| dt < +\infty$ and $\int_{\mathbb{R}} t^2 K(t) dt < +\infty$ hold.

A3. The function $\bar{r}_1(x)$ defined in (4.1) is twice differentiable and $\sup_{x \in \mathcal{C}} |\bar{r}_1''(x)| < +\infty$.

A4. The marginal density $f(\cdot)$ is twice differentiable and satisfies a Lipschitz condition. Furthermore $f(x) > \Gamma$ for all $x \in \mathcal{C}$ and $\Gamma > 0$.

A5. The integral defined by $\int_{\mathbb{R}} \frac{y^2}{\bar{G}(y)} f_{X,Y}(x, y) dy =: \bar{r}_2(x)$ is twice differentiable and $\sup_{x \in \mathcal{C}} |\bar{r}_2''(x)| < +\infty$. In addition the following assumptions are used for the asymptotic normality result.

B1. The bandwidth h_n satisfies $\lim_{n \rightarrow +\infty} nh_n^5 = 0$.

B2. There exists $\beta > 2$ such that $\iint y^\beta f_{X,Y}(x, y) dy dx < +\infty$.

Remark 4.2. *Assumptions **A1**, **A2**, and **A5** are very common in functional estimation. **A1**, **A2** and **A3** are needed to study the bias term of $\widehat{r}_{1,n}(x)$, which is the kernel estimate of $\bar{r}_1(x)$. Selecting μ between 0 and 1 in Assumption **A1** is linked to Lemma 4 whereas Lemma 1 allows to take μ between 0 and 4. Assumption **A4** is used to conclude the convergence of the kernel density estimator. Assumptions **B1-B2** are needed for get asymptotic normality.*

The following theorem gives the rate of the uniform almost sure convergence of the kernel regression estimator.

Theorem 4.1. *Under Assumptions **A1-A5**, we have for $n \rightarrow \infty$:*

$$\sup_{x \in \mathcal{C}} |r_n(x) - r(x)| = O \left(\max \left\{ \sqrt{\frac{\log n}{nh_n}}, h_n^2 \right\} \right) \quad a.s.$$

Remark 4.3. *If we choose $h_n = O \left(\left(\frac{\log n}{n} \right)^{\frac{1}{5}} \right)$ then Theorem 4.1 becomes*

$$\sup_{x \in \mathcal{C}} |r_n(x) - r(x)| = O \left[\left(\frac{\log n}{n} \right)^{\frac{2}{5}} \right],$$

which is the optimal rate in the sense of the uniform convergence criterion for density estimation (see Stute, 1982).

For the asymptotic normality result we need some notations.

Let

$$\Sigma(x) = \begin{pmatrix} \bar{r}_2(x) & \bar{r}_1(x) \\ \bar{r}_1(x) & f(x) \end{pmatrix}$$

where $\bar{r}_1(x)$ is defined in (4.1) and $\bar{r}_2(x)$ in **A5**.

Theorem 4.2. *Under Assumptions **A1-A5** and **B1-B2***

$$\sqrt{nh_n} (r_n(x) - r(x)) \xrightarrow{\mathcal{D}} N(0, \sigma^2(x)),$$

where

$$\sigma^2(x) = \kappa \frac{\bar{r}_2(x) f^2(x) - \bar{r}_1^2(x) f(x)}{f^4(x)},$$

$\kappa = \int_{\mathbb{R}} K^2(t) dt$ and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution.

Remark 4.4. *From Cauchy-Schwartz's inequality it is clear that $\bar{r}_2(x) f(x) - \bar{r}_1^2(x) > 0$ (the equality does not hold since the mapping $y \mapsto y$ is not constant) and then $\Sigma(x)$ is positive definite as soon as $f(x) > 0$.*

Confidence interval

A plug-in estimate $\sigma_n(x)$ of the asymptotic variance can easily be obtained using (4.4), (4.2) and

$$r_{2,n}(x) = \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i T_i^2}{G_n^2(T_i)} K\left(\frac{x - X_i}{h_n}\right), \quad (4.5)$$

which are estimates of $\bar{r}_1(\cdot)$, $f(\cdot)$ and $\bar{r}_2(\cdot)$, respectively.

This yields the following confidence interval of asymptotic level $1 - \zeta$ of $r(x)$

$$\left[r_n(x) - t_{1-\zeta/2} \frac{\sigma_n(x)}{\sqrt{nh_n}}, r_n(x) + t_{1-\zeta/2} \frac{\sigma_n(x)}{\sqrt{nh_n}} \right]$$

where $t_{1-\zeta/2}$ denotes the $1 - \zeta/2$ quantile of the standard normal distribution.

Remark 4.5. *Our result is an extension of Schuster's (1972) result to the censorship model and it completes works of Carbonez et al. (1995) and Köhler et al. (2002) where no asymptotic normality result is given.*

Remark 4.6. *In the proof of Theorem 4.1, we use Bernstein's inequality which is a point-wise exponential inequality and we use a standard idea about covering a compact set by a finite number of intervals to get the uniformity. Another interesting idea consists in using Vapnik-Cervonenkis' theory through Pollard's inequality (see Devroye et al. (1996) Theorem 29.1) together with bounds on the covering number, which gives the uniformity straightforward. However, the choice of the kernels in the last case must satisfy Condition K1 of Giné and Guillou (2002) which is more restrictive than the Hölder continuity. In this case, one may only obtain a rate of convergence of order $\sqrt{\frac{\log n}{nh_n^2}}$, which is slower than under our assumptions.*

4.4 Simulation study

This section is divided into two parts : the first one shows the behavior of our estimate for some particular conditional regression functions, whereas the second deals with asymptotic normality.

4.4.1 Consistency

First we consider the linear regression model : $Y_i = \Upsilon X_i + \Upsilon_0 + b\varepsilon_i$, for which we simulate X_i and ε_i as two independent iid sequences distributed as $\mathcal{N}(0, 1)$ and where b is a constant appropriately chosen (here we take $b = 0.2$). We also simulate n iid rv $C_i \rightsquigarrow \mathcal{N}(0, 1)$, and then take $T_i = \min(Y_i, C_i)$ with indicative rv $\delta_i = \mathbb{1}_{\{Y_i \leq C_i\}}$.

We calculate our estimator based on the observed data $(X_i, T_i, \delta_i,)$ $i = 1, \dots, n$, by choosing a Gaussian kernel K . In this simulation, the optimal bandwidth choice is much more crucial than the choice of the kernel.

It is well known that in nonparametric estimation, optimality (in the MSE sense) is not seriously affected by the choice of the kernel K but is affected by that of the bandwidth h_n . Here we use the optimal bandwidth obtained by minimizing the supremum norm between the estimated and the true values over a set of the known bandwidths values $h_n \in [0.1, 0.5]$. Note that, outside of this interval, the values give a bad estimator. The supremum is taken over a compact set $\mathcal{C} = [0, 1]$ for the linear case.

Obviously the bandwidth choice satisfies Assumptions **A1** and **B1**. We recall that an automatically optimal choice can be made by using cross validation method. We notice that the quality of fit increases with n (Figure 4.1).

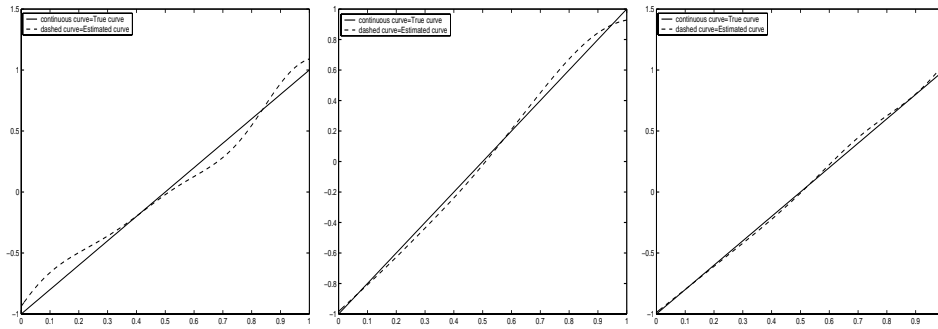


FIG. 4.1 – $r(x) = 2x - 1$, with $n=100, 500$ and 1000 respectively.

To conclude this part, we consider the case of non linear regression by choosing the following three models :

$$\begin{aligned}
 Y_i &= \sin\left(\frac{3}{2}X_i\right) + \varepsilon_i && \text{sine case} \\
 Y_i &= \exp(X_i - 0.2) + \varepsilon_i && \text{exponential case} \\
 Y_i &= \frac{3}{2}X_i^2 - \frac{1}{2} + \varepsilon_i && \text{parabolic case,}
 \end{aligned}$$

with $\mathcal{C} = [-3, 1]$ for sine case and $\mathcal{C} = [-2, 2]$ for the other cases.

Figures 4.2 and 4.3 show that the quality of fit for the non linear model is as good as for the linear model. Note that for the exponential model, we take $C_i \rightsquigarrow \text{Exp}(2.5)$ because of a strong censoring in the normal case and then a bad quality of fit.

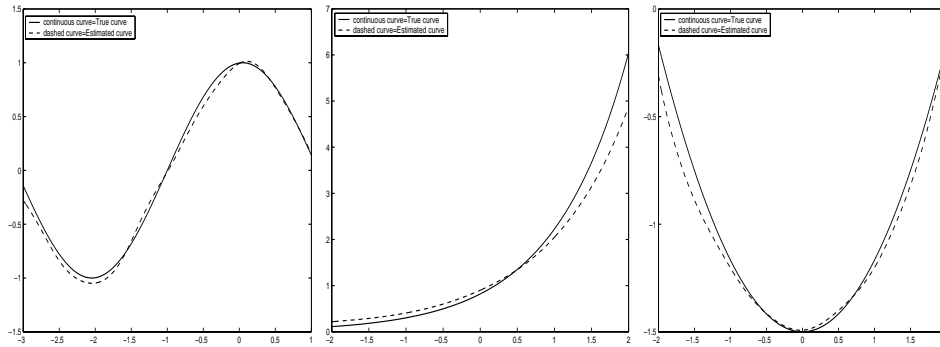


FIG. 4.2 – Sine, Exponential and Parabolic case with $n = 500$.

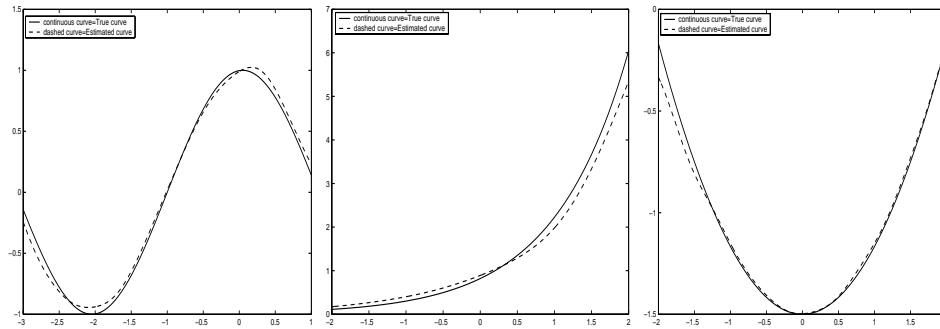


FIG. 4.3 – Sine, exponential and parabolic case with $n = 1000$.

4.4.2 Asymptotic normality

In this subsection we address the asymptotic normality issue for large sample. We compare the shape of the estimated density (with normalized deviations) to that of the standard normal density in the case of the linear regression model

$$Y_i = 2X_i - 1 + \varepsilon_i, \quad i = 1, \dots, n.$$

Then $r(x) = 2x - 1$ and the data are like as in previous section. The steps are as follows : we estimate the regression function $r(x)$ by $r_n(x)$, and we calculate the normalized deviation between this estimate and the theoretical regression function, for $x = 0$, ie

$$\bar{r}_n = \frac{\sqrt{nh_n}}{\hat{\sigma}}(r_n(0) - r(0)) = \frac{\sqrt{nh_n}}{\hat{\sigma}}(r_n(0) + 1).$$

Using this scheme, we draw m independent n -samples. The bandwidth h_n is chosen as before and according to assumptions **A1** and **B1**. To estimate the density function of \bar{r}_n by the kernel method we choose $h'_m = Cm^{-1/5}$ (see e.g. Silverman 1986 p 45). In all cases, we take either $C = 0.9$ or $C = 1.5$. It is well known that this constant depends on unknown parameters

which can be estimated. As mentioned before, the bandwidth can be chosen by cross-validation method.

We compare the shape of the estimated density (with normalized deviation) to that of the standard normal density in the case of the linear regression model. The figures below give the density of \bar{r}_n , the histogram and the Q-Q-Plot against a standard normal distribution. We consider different values of n and B and give only the plots for $n = 200$, $m = 300$ (Figure 4.4) and $n = 500$, $m = 300$ (Figure 4.5). For $n = 200$ the P-value from the Jarque-Bera test is $p = 0.61$ and for $n = 500$ the P-value is around 0.84. We see clearly that our asymptotic law has a good behavior with respect to the standard normal law.

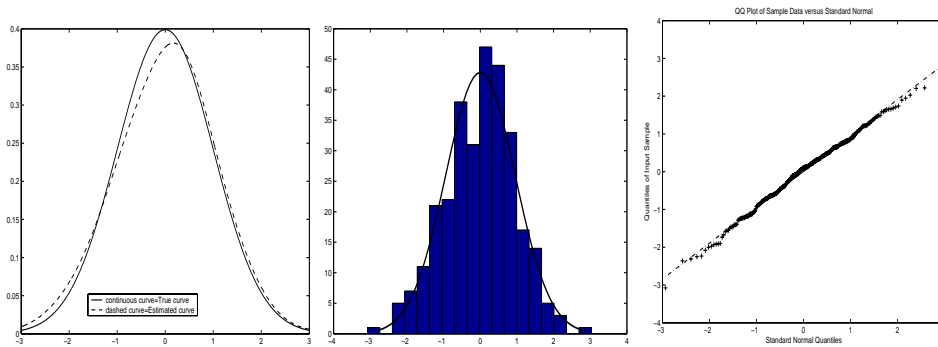


FIG. 4.4 – $n = 200$ and $m = 300$.

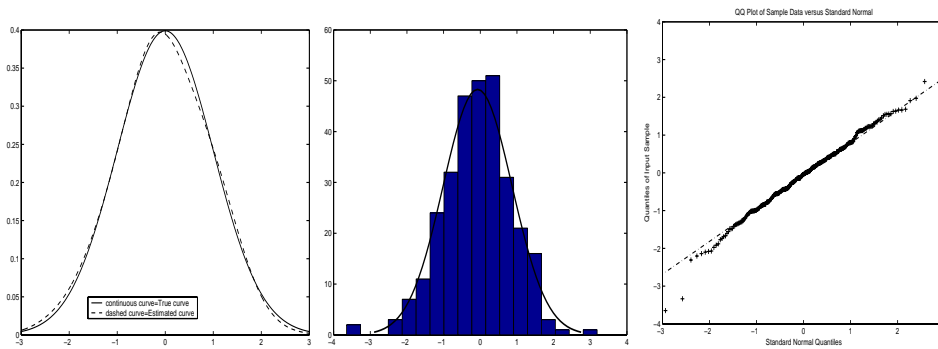


FIG. 4.5 – $n = 500$ and $m = 300$.

4.5 Proofs

We split the proof of Theorem 3.2 by using the following Lemmata.

Lemma 4.1. *Under Assumptions **A1** and **A2**, we have for $n \rightarrow \infty$*

$$\sup_{x \in \mathcal{C}} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| = O\left(\frac{1}{\sqrt{nh_n^\mu}}\right) \quad a.s.$$

Proof

We have :

$$\begin{aligned} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| &= \frac{1}{nh_n} \left| \sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}_n(T_i)} K\left(\frac{x - X_i}{h_n}\right) - \frac{\delta_i T_i}{\bar{G}(T_i)} K\left(\frac{x - X_i}{h_n}\right) \right| \\ &= \frac{1}{nh_n} \left| \sum_{i=1}^n \frac{\mathbb{1}_{\{Y_i < C_1\}} Y_i}{\bar{G}_n(Y_i)} K\left(\frac{x - X_i}{h_n}\right) - \frac{\mathbb{1}_{\{Y_i < C_1\}} Y_i}{\bar{G}(Y_i)} K\left(\frac{x - X_i}{h_n}\right) \right| \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n \left| Y_i K\left(\frac{x - X_i}{h_n}\right) \right| \left| \frac{1}{\bar{G}_n(Y_i)} - \frac{1}{\bar{G}(Y_i)} \right| \\ &\leq \frac{1}{nh_n} \sum_{i=1}^n \left| Y_i K\left(\frac{x - X_i}{h_n}\right) \right| \left| \frac{\bar{G}_n(Y_i) - \bar{G}(Y_i)}{\bar{G}_n(Y_i) \bar{G}(Y_i)} \right| \\ &\leq \frac{1}{\bar{G}_n(\tau_H) \bar{G}(\tau_H)} \sup_{t < \tau_H} (|\bar{G}_n(t) - \bar{G}(t)|) \frac{\tau_H}{nh_n} \sum_{i=1}^n \left| K\left(\frac{x - X_i}{h_n}\right) \right|. \end{aligned}$$

Then by using the strong law of large numbers and the law of iterated logarithm (Deheuvels and Einmahl 2000) :

$$\sup_{x \in \mathcal{C}} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| \leq \frac{\tau_H}{\bar{G}^2(\tau_H)} \mathbb{E} \left(\left| \frac{1}{h_n} K\left(\frac{x - X_1}{h_n}\right) \right| \right) \sqrt{\frac{\log \log n}{n}}.$$

Assumptions **A1** and **A2** conclude the proof.

Lemma 4.2. *Under Assumptions **A1**, **A2** and **A3**, we have, for $n \rightarrow \infty$*

$$\sup_{x \in \mathcal{C}} |\mathbb{E}(\tilde{r}_{1,n}(x)) - \bar{r}_1(x)| = O(h_n^2) \quad a.s.$$

Proof

We have

$$\begin{aligned} |\mathbb{E}(\tilde{r}_{1,n}(x)) - \bar{r}_1(x)| &= \left| \mathbb{E} \left(\frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i T_i}{\bar{G}(T_i)} K\left(\frac{x - X_i}{h_n}\right) \right) - \bar{r}_1(x) \right| \\ &= \left| \mathbb{E} \left(\frac{1}{h_n} \frac{\delta_1 T_1}{\bar{G}(T_1)} K\left(\frac{x - X_1}{h_n}\right) \right) - \bar{r}_1(x) \right|. \end{aligned} \quad (4.6)$$

Note that, on the one hand we have :

$$\begin{aligned}
\mathbb{E} \left(\frac{1}{h_n} \frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \right) &= \mathbb{E} \left(\mathbb{E} \left(\frac{1}{h_n} \frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \middle| X_1 \right) \right) \\
&= \mathbb{E} \left(\frac{1}{h_n} K \left(\frac{x - X_1}{h_n} \right) \mathbb{E} \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} \middle| X_1 \right) \right) \\
&= \int \frac{1}{h_n} K \left(\frac{x - u}{h_n} \right) \mathbb{E} \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} \middle| X_1 = u \right) f(u) du.
\end{aligned}$$

On the other hand

$$\begin{aligned}
\mathbb{E} \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} \middle| X_1 = u \right) &= \mathbb{E} \left(\frac{\mathbb{1}_{\{Y_1 < C_1\}} Y_1}{\bar{G}(Y_1)} \middle| X_1 = u \right) \\
&= \mathbb{E} \left(\frac{Y_1}{\bar{G}(Y_1)} \mathbb{E}(\mathbb{1}_{\{Y_1 < C_1\}} \middle| Y_1) \middle| X_1 = u \right) \\
&= \mathbb{E}(Y_1 \middle| X_1 = u) = r(u)
\end{aligned}$$

holds. So (4.6) becomes :

$$\begin{aligned}
\left| \mathbb{E} \left(\frac{1}{h_n} \frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \right) - \bar{r}_1(x) \right| &= \left| \int \frac{1}{h_n} K \left(\frac{x - u}{h_n} \right) r(u) f(u) du - \bar{r}_1(x) \right| \\
&= \left| \int \frac{1}{h_n} K \left(\frac{x - u}{h_n} \right) \bar{r}_1(u) du - \bar{r}_1(x) \right| \\
&= \left| \int K(t) [\bar{r}_1(x - h_n t) - \bar{r}_1(x)] dt \right|
\end{aligned}$$

A Taylor expansion around x gives

$$\bar{r}_1(x - h_n t) - \bar{r}_1(x) = -h_n t \bar{r}'_1(x) + \frac{h_n^2 t^2}{2} \bar{r}''_1(x_0)$$

where x_0 is between $x - h_n t$ and x . Then

$$\begin{aligned}
\sup_{x \in \mathcal{C}} |\mathbb{E}(\tilde{r}_{1,n}(x)) - \bar{r}_1(x)| &= \sup_{x \in \mathcal{C}} \left| \int K(t) [\bar{r}_1(x - h_n t) - \bar{r}_1(x)] dt \right| \\
&= \sup_{x \in \mathcal{C}} \left| \int K(t) \left[-h_n t \bar{r}'_1(x) + \frac{h_n^2 t^2}{2} \bar{r}''_1(x^*) \right] dt \right| \\
&\leq h_n \sup_{x \in \mathcal{C}} \left| \int t K(t) \bar{r}'_1(x) dt \right| + h_n^2 \sup_{x \in \mathcal{C}} \left| \int \frac{t^2}{2} K(t) \bar{r}''_1(x^*) dt \right| \tag{4.7}
\end{aligned}$$

The first term in (4.7) is equal to zero because of Assumption **A2**. The second term goes to zero for $n \rightarrow \infty$ from Assumptions **A1** and **A3**. This concludes the proof.

Lemma 4.3. *Under Assumptions **A1**, **A2** and **A5**, we have, for $n \rightarrow \infty$*

$$\sup_{x \in \mathcal{C}} |\tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x))| = O \left[\left(\sqrt{\frac{\log n}{nh_n}} \right) \right] \quad a.s.$$

Proof

Since \mathcal{C} is compact, then it can be covered by a finite number \mathfrak{s}_n of intervals having length h_n^η and centers at points $x_1, \dots, x_{\mathfrak{s}_n}$, respectively, where $\eta > \frac{\gamma+1}{\gamma}$.

Since \mathcal{C} is bounded, there exist a constant $M > 0$ such that $\mathfrak{s}_n \leq \frac{M}{h_n^\eta}$. Then we set, for $x \in \mathcal{C}$:

$$\Delta_i(x) = \frac{1}{nh_n} \frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(\frac{1}{nh_n} \frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \right).$$

It is obvious that

$$\sum_{i=1}^n \Delta_i(x) = \tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x)).$$

If we write $\tilde{\Delta}_i(x) = \Delta_i(x) - \Delta_i(x_k)$, we have clearly $|\Delta_i(x)| \leq |\tilde{\Delta}_i(x)| + |\Delta_i(x_k)|$, then

$$\begin{aligned} \sup_{x \in \mathcal{C}} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| &\leq \sup_{x \in \mathcal{C}} \left\{ \frac{1}{n} \sum_{i=1}^n \frac{\delta_i |T_i|}{\bar{G}(T_i)} \frac{1}{h_n} \left| K \left(\frac{x - X_i}{h_n} \right) - K \left(\frac{x_k - X_i}{h_n} \right) \right| \right\} \\ &\quad + \sup_{x \in \mathcal{C}} \left\{ \mathbb{E} \left(\frac{\delta_1 |T_1|}{\bar{G}(T_1)} \frac{1}{h_n} \left| K \left(\frac{x - X_1}{h_n} \right) - K \left(\frac{x_k - X_1}{h_n} \right) \right| \right) \right\} \end{aligned}$$

holds.

Since K is Hölderian of order γ then :

$$\sup_{x \in \mathcal{C}} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| \leq \sup_{x \in \mathcal{C}} \left(\frac{2\mathbb{E}(|Y_1|)}{\bar{G}(\tau_H)} \frac{1}{h_n} \left| \frac{x - x_k}{h_n} \right|^\gamma \right) \leq \frac{\mathbb{E}(|Y_1|)}{\bar{G}(\tau_H)} \frac{h_n^{\eta\gamma}}{h_n^{\gamma+1}}.$$

So

$$\sup_{x \in \mathcal{C}} \left| \sum_{i=1}^n \tilde{\Delta}_i(x) \right| = o(1) \quad a.s.$$

On the other hand, for all $\epsilon > 0$, we have

$$\mathbb{P} \left\{ \max_{k=1, \dots, \mathfrak{s}_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} \leq \sum_{i=1}^{\mathfrak{s}_n} \mathbb{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\}. \quad (4.8)$$

Then let $U_i = nh_n \Delta_i(x_k)$. We have $\mathbb{E}(U_i) = 0$ and

$|U_i| \leq 2\tau_H \bar{K} v =: M_1$, where \bar{K} is the upper bound of K and $v = \frac{1}{\bar{G}(\tau_H)}$. In order to apply

Bernstein's inequality (see Ferraty and Vieu 2006, corollary A.9, p 235), we calculate $S^2 = \text{Var}(U_i)$, so,

$$\begin{aligned} S^2 &= \mathbb{E} \left(\frac{\delta_1^2 T_1^2}{\bar{G}^2 T_1} K^2 \left(\frac{x - X_1}{h_n} \right) \right) - \mathbb{E}^2 \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \right) \\ &= \int \int y^2 K^2 \left(\frac{x - u}{h_n} \right) f_{X,Y}(u, y) du dy - \left(\int \int y K \left(\frac{x - u}{h_n} \right) f_{X,Y}(u, y) du dy \right)^2 \\ &= \int K^2 \left(\frac{x - u}{h_n} \right) \bar{r}_2(u) du - \left(\int K \left(\frac{x - u}{h_n} \right) \bar{r}_1(u) du \right)^2 \end{aligned}$$

By a Taylor expansion of order one, under Assumptions **A2** and **A5** we get $S^2 = O(h_n)$. Then applying Bernstein's inequality, we have

$$\begin{aligned} \mathbb{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} &= P \left(\left| \sum_{i=1}^n U_i \right| > \epsilon h_n n \right) \leq 2 \exp \left\{ -\frac{\epsilon^2 h_n n}{2(c + \epsilon M_1)} \right\} \\ &=: A \end{aligned}$$

where c is a generic constant.

So (4.8) becomes

$$\begin{aligned} \mathbb{P} \left\{ \max_{k=1, \dots, s_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} &\leq \mathfrak{s}_n A \leq M h_n^{-\beta} A \\ &= M (n h_n)^{-\beta} n^{\beta - \frac{\epsilon^2}{2(c + \epsilon M_1)} n h_n / \log n} \end{aligned} \quad (4.9)$$

From Assumption **A1**, $\frac{n h_n}{\log n} \rightarrow \infty$ follows, which means that the last term in (4.9) is the general term of a convergent series. Then by Borel-Cantelli's Lemma the first term in (4.9) goes to zero almost surely. Now if we replace ϵ by $\epsilon_0 \sqrt{\frac{\log n}{n h_n}}$ we can choose ϵ_0 so that the term $n^{\beta - \frac{\epsilon_0^2}{2(c + \epsilon_0 M_1)}}$ is the general term of a convergent serie and then the result holds.

Next we proceed to prove Theorem 4.1.

Proof of Theorem 4.1 :

Observe that

$$\begin{aligned} \sup_{x \in \mathcal{C}} |r_n(x) - r(x)| &\leq \sup_{x \in \mathcal{C}} \left| \frac{r_{1,n}(x)}{f_n(x)} - \frac{\tilde{r}_{1,n}(x)}{f_n(x)} \right| + \left| \frac{\tilde{r}_{1,n}(x)}{f_n(x)} - \frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{f_n(x)} \right| \\ &\quad + \left| \frac{\mathbb{E}(\tilde{r}_{1,n}(x))}{f_n(x)} - \frac{\bar{r}_1(x)}{f_n(x)} \right| + \left| \frac{\bar{r}_1(x)}{f_n(x)} - \frac{\bar{r}_1(x)}{f(x)} \right| \\ &\leq \frac{1}{\inf_{x \in \mathcal{C}} f_n(x)} \left\{ \sup_{x \in \mathcal{C}} |r_{1,n}(x) - \tilde{r}_{1,n}(x)| + \sup_{x \in \mathcal{C}} |\tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x))| \right. \\ &\quad \left. + \sup_{x \in \mathcal{C}} |\mathbb{E}(\tilde{r}_{1,n}(x) - \bar{r}_1(x))| + \sup_{x \in \mathcal{C}} (|\bar{r}_1(x)| \Gamma^{-1}) \sup_{x \in \mathcal{C}} |f(x) - f_n(x)| \right\} \end{aligned} \quad (4.10)$$

The kernel estimator $f_n(x)$ is almost surely bounded away from 0 by the second part of Assumption **A4**. Now writing

$$\sup_{x \in \mathcal{C}} |f(x) - f_n(x)| \leq \sup_{x \in \mathcal{C}} |f_n(x) - \mathbb{E}(f_n(x))| + \sup_{x \in \mathcal{C}} |\mathbb{E}(f_n(x)) - f(x)|$$

By Assumptions **A2** and **A4** and a Taylor expansion we get

$$\sup_{x \in \mathcal{C}} |\mathbb{E}(f_n(x)) - f(x)| = O(h_n^2). \quad (4.11)$$

Using assumptions **A1** and **A2**, and following step by step the proof of Lemma 4.3 with $\overline{G}(Ti) = 1$ and $\delta_i = 1$ we get

$$\sup_{x \in \mathcal{C}} |f_n(x) - \mathbb{E}(f_n(x))| = O\left(\sqrt{\frac{\log n}{nh_n}}\right). \quad (4.12)$$

Then (4.10), (4.11), (4.12) and Lemmas 4.1, 4.2 and 4.3 conclude the proof.

For the proof of Theorem 4.2, we need some auxiliary results.

Note that

$$\begin{aligned} r_{1,n}(x) - \bar{r}_1(x) &= (r_{1,n}(x) - \tilde{r}_{1,n}(x)) + (\tilde{r}_{1,n}(x) - \mathbb{E}(\tilde{r}_{1,n}(x))) \\ &+ (\mathbb{E}(\tilde{r}_{1,n}(x)) - \bar{r}_1(x)) \\ &=: \Lambda_{1,n}(x) + \Lambda_{2,n}(x) + \Lambda_{3,n}(x) \end{aligned} \quad (4.13)$$

and

$$\begin{aligned} (f_n(x)) - f(x) &= (f_n(x) - \mathbb{E}(f_n(x))) + (\mathbb{E}(f_n(x)) - f(x)) \\ &=: \Gamma_{1,n}(x) + \Gamma_{2,n}(x). \end{aligned} \quad (4.14)$$

Lemma 4.4. *Under the assumptions of Lemma 4.1 and **B1**, both $\sqrt{nh_n}\Lambda_{1,n}(x)$ and $\sqrt{nh_n}\Gamma_{2,n}(x)$ are $o_{\mathbb{P}}(1)$ as $n \rightarrow \infty$, for any x .*

Proof

From Lemma 4.1

we have :

$$\begin{aligned} \sqrt{nh_n}\Lambda_{1,n}(x) &= \sqrt{nh_n}(r_{1,n}(x) - \tilde{r}_{1,n}(x)) = \sqrt{nh_n}O\left(\frac{1}{\sqrt{nh_n^\mu}}\right) \\ &= O(\sqrt{h_n^{1-\mu}}) = o_{\mathbb{P}}(1). \end{aligned}$$

From Formula (4.11) and Assumption **B1** we have

$$\begin{aligned}\sqrt{nh_n}\Gamma_{2,n}(x) &= \sqrt{nh_n}(\mathbb{E}(f_n(x)) - f(x)) = \sqrt{nh_n}O(h_n^2) \\ &= O(\sqrt{nh_n^5}) = o_{\mathbb{P}}(1).\end{aligned}$$

Lemma 4.5. *Under the assumptions of Lemma 4.2 and **B1**, we have that $\sqrt{nh_n}\Lambda_{3,n}(x)$ is $o_{\mathbb{P}}(1)$ as $n \rightarrow \infty$, for any x .*

Proof

From Lemma 4.2 and Assumption **B1** we have

$$\begin{aligned}\sqrt{nh_n}\Lambda_{3,n}(x) &= \sqrt{nh_n}(\mathbb{E}(\tilde{r}_{1,n}(x)) - \bar{r}_1(x)) = \sqrt{nh_n}O(h_n^2) \\ &= O(\sqrt{nh_n^5}) = o_{\mathbb{P}}(1).\end{aligned}$$

Lemma 4.6. *Under Assumptions **A2-A5** and **B2**, we have*

$$\sqrt{nh_n}(\Lambda_{2,n}(x), \Gamma_{1,n}(x)) \xrightarrow{D} N(0, \kappa\Sigma(x)).$$

Proof

Let us calculate variance and covariance of $\sqrt{nh_n}\Lambda_{2,n}(x)$ and $\sqrt{nh_n}\Gamma_{1,n}(x)$. A Taylor expansion and Assumptions **A2** and **A5** give

$$\begin{aligned}Var\left(\sqrt{nh_n}\Lambda_{2,n}(x)\right) &= \frac{1}{h_n}Var\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}K\left(\frac{x - X_1}{h_n}\right)\right) \\ &= \frac{1}{h_n}\left\{\mathbb{E}\left(\frac{\delta_1^2 T_1^2}{\bar{G}^2(T_1)}K^2\left(\frac{x - X_1}{h_n}\right)\right) - \mathbb{E}^2\left(\frac{\delta_1 T_1}{\bar{G}(T_1)}K\left(\frac{x - X_1}{h_n}\right)\right)\right\} \\ &= \bar{r}_2(x)\kappa + o(1)\end{aligned}\tag{4.15}$$

where $\kappa = \int_{\mathbb{R}} K^2(t)dt$. In the same way, by **A2** and **A4** we have

$$\begin{aligned}Var\left(\sqrt{nh_n}\Gamma_{1,n}(x)\right) &= \frac{1}{h_n}Var\left(K\left(\frac{x - X_1}{h_n}\right)\right) \\ &= \frac{1}{h_n}\left\{\mathbb{E}\left(K^2\left(\frac{x - X_1}{h_n}\right)\right) - \mathbb{E}^2\left(K\left(\frac{x - X_1}{h_n}\right)\right)\right\} \\ &= f(x)\kappa + o(1).\end{aligned}\tag{4.16}$$

In addition with **A3**

$$\begin{aligned}
\text{Cov}\sqrt{nh_n} \left(\Lambda_{2,n}(x), \sqrt{nh_n}\Gamma_{1,n}(x) \right) &= \frac{1}{h_n} \left\{ \mathbb{E} \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} K^2 \left(\frac{x - X_1}{h_n} \right) \right) \right. \\
&\quad \left. - \mathbb{E} \left(\frac{\delta_1 T_1}{\bar{G}(T_1)} K \left(\frac{x - X_1}{h_n} \right) \right) \mathbb{E} \left(K \left(\frac{x - X_1}{h_n} \right) \right) \right\} \\
&= \bar{r}_1(x) \kappa + \mathbf{o}(1).
\end{aligned} \tag{4.17}$$

Now we have to prove that all linear combinations of $\sqrt{nh_n}\Lambda_{2,n}(x)$ and $\sqrt{nh_n}\Gamma_{1,n}(x)$ are asymptotically Gaussian. So for a given pair of real numbers

$$\begin{aligned}
\omega_n(x) &= a_1 \sqrt{nh_n} \Lambda_{2,n}(x) + a_2 \sqrt{nh_n} \Gamma_{1,n}(x) \\
&= \sum_{i=1}^n \left\{ a_1 \frac{1}{\sqrt{nh_n}} \left[\frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(\frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) \right) \right] \right. \\
&\quad \left. + a_2 \frac{1}{\sqrt{nh_n}} \left[K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{x - X_i}{h_n} \right) \right) \right] \right\} \\
&=: \sum_{i=1}^n \omega_{ni}(x).
\end{aligned}$$

Let $\tau_{ni}^\beta(x) = \mathbb{E} |\omega_{ni}(x)|^\beta$ for $\beta > 2$. Then by Jensen's inequality (see Shorack and Wellner 1986, p. 843)

$$\begin{aligned}
\tau_{ni}^\beta(x) &= \mathbb{E} \left| a_1 \frac{1}{\sqrt{nh_n}} \left\{ \frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(\frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) \right) \right\} \right. \\
&\quad \left. + a_2 \frac{1}{\sqrt{nh_n}} \left\{ K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{x - X_i}{h_n} \right) \right) \right\} \right|^\beta \\
&\leq 2^{\beta-1} \left(\frac{h_n}{n} \right)^{\frac{\beta}{2}} \left\{ a_1^\beta \mathbb{E} \left| \frac{1}{h_n} \frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) \right|^\beta \right. \\
&\quad \left. + a_2^\beta \mathbb{E} \left| \frac{1}{h_n} K \left(\frac{x - X_i}{h_n} \right) \right|^\beta \right\}.
\end{aligned} \tag{4.18}$$

Since

$$\mathbb{E} \left(\frac{\delta_i T_i}{\bar{G}(T_i)} K \left(\frac{x - X_i}{h_n} \right) \right) = O(h_n) \text{ and } \mathbb{E} \left(K \left(\frac{x - X_i}{h_n} \right) \right) = O(h_n),$$

then from Assumption **B2**, the sum in inequality (4.18) is bounded, so

$$\tau_n^\beta(x) =: \sum_{i=1}^n \tau_{ni}^\beta(x) = O(n^{1-\beta/2} h_n^{\beta/2}),$$

which is $o(1)$, since $1 - \beta/2 < 0$.

From (4.15), (4.16) and (4.17) we have

$$\begin{aligned} \text{Var}(\omega_n(x)) &= a_1^2 \text{Var} \left(\sqrt{nh_n} \Lambda_{2,n}(x) \right) + a_2^2 \text{Var} \left(\sqrt{nh_n} \Gamma_{1,n}(x) \right) \\ &\quad + 2a_1 a_2 \text{Cov} \left(\sqrt{nh_n} \Lambda_{2,n}(x), \sqrt{nh_n} \Gamma_{1,n}(x) \right) \\ &= a_1^2 \bar{r}_2(x) \kappa + a_2^2 f(x) \kappa + 2a_1 a_2 \bar{r}_1(x) \kappa + o(1) \end{aligned}$$

and then $\text{Var}(\omega_n(x)) \xrightarrow{n \rightarrow \infty} \kappa a^t \Sigma(x) a$ where $a = (a_1, a_2)^t$. So, we deduce that $\lim_{n \rightarrow +\infty} (\text{Var}(\omega_n(x)))^{\beta/2}$ exists. Hence, from Lyapounov's Central Limit Theorem (see Pollard 1984, p. 51), we have the result since

$$\lim_{n \rightarrow +\infty} \frac{1}{(\text{Var}(\omega_n(x)))^{\beta/2}} \sum_{i=1}^n \mathbb{E} |\omega_{ni}(x)|^\beta = \lim_{n \rightarrow +\infty} \frac{1}{(\text{Var}(\omega_n(x)))^{\beta/2}} \tau_n^\beta(x) = 0.$$

Now we can prove theorem 4.2

Proof of Theorem 4.2 :

Notice if we define the mapping

$$\begin{aligned} \theta : \quad \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ (x, y) &\rightsquigarrow \theta(x, y) = \frac{x}{y}, \quad y \neq 0 \end{aligned}$$

then from formulae (4.3) and (4.1) it is clear that

$$r_n(x) - r(x) = \frac{r_{1,n}(x)}{f_n(x)} - \frac{\bar{r}_1(x)}{f(x)} = \theta(r_{1,n}(x), f_n(x)) - \theta(\bar{r}_1(x), f(x)).$$

We deduce from (4.13), (4.14), Lemmas 4.4, 4.5, 4.6 and by Mann-Wald's Theorem (see Rao 1965 p. 321) and using the delta method (see Van Der Vaart and Wellner (1996 p. 374) that

$$\sqrt{nh_n} (r_n(x) - r(x)) \xrightarrow{\mathcal{D}} N(0, \kappa \nabla \theta^t \Sigma(x) \nabla \theta)$$

where the gradient $\nabla \theta^t = \left(\frac{\partial \theta}{\partial x}, \frac{\partial \theta}{\partial y} \right)$ is evaluated at point $(\bar{r}_1(x), f(x))$. Then a simple calculation gives

$$\kappa \nabla \theta^t \Sigma(x) \nabla \theta = \kappa \frac{\bar{r}_2(x) f^2(x) - \bar{r}_1^2(x) f(x)}{f^4(x)} = \sigma^2(x).$$

Acknowledgments

The authors are grateful to an anonymous referee and the chief editor, whose careful reading gave them the opportunity to improve the quality of the presentation of the paper.

Chapitre 5

Conclusion et Perspectives

5.1 Conclusion

La statistique fonctionnelle constitue un champ de recherche d'actualité et diversifié à la fois par les aspects fondamentaux et par les différents domaines qu'elle recoupe. La littérature sur l'estimation non paramétrique que ce soit de la densité ou de la régression, est très riche, littérature à laquelle G. Collomb apporta une contribution déterminante comme en témoigne la compilation de ses travaux (Collomb (1983)). Nous nous en sommes tenu, dans cette thèse, à des résultats de convergence en norme L_2 et L_∞ . Nous renvoyons le lecteur vers Ioannides (1992) pour d'autres types de résultats en normes L_2 , Kim et Cox (2001) pour des développements plus récents.

Ce document est essentiellement axé sur l'estimation non paramétrique de la fonction de régression, pour laquelle nous avons donné quelques aspects particuliers qui tournent autour des méthodes à noyaux et nous avons mis en évidence le rôle du paramètre de lissage. L'estimateur de Nadaraya-Watson a été particulièrement étudié aussi bien dans le cas complet qu'en présence de censures.

L'introduction de cette thèse nous a permis de faire le point (sans être exhaustif) sur ce qui s'est déjà fait dans ce domaine. Nous avons ensuite rappelé les propriétés fondamentales de certains estimateurs de la fonction de régression en insistant sur l'estimateur de Nadaraya-Watson étant donné que nos résultats portent sur cet estimateur. Nous avons donc donné ses propriétés asymptotiques dans le cas d'un modèle complet, et nous avons donné quelques références sur le modèle censuré ce qui nous a permis de situer l'apport de notre travail, qui complète les travaux de Carbonez *et al* (1995) et Kohler *et al* (2002) où la vitesse de convergence n'a pas été obtenue ainsi que la normalité asymptotique. Des simulations ont illustré la performance de notre estimateur ainsi que sa normalité asymptotique. Les techniques utilisées dans notre

travail sont inspirées des travaux de Ould-Said (2006) pour ce qui de la vitesse de convergence et de Lemdani et Ould-Said (2006) pour la normalité asymptotique. Nous faisons remarquer que les résultats obtenus l'ont été pour une covariable X dans \mathbb{R} ; nous pouvons les étendre sans difficulté au cas \mathbb{R}^d , en utilisant dans le lemme 4.2 la technique des petites boules qui est une généralisation de la technique utilisée et nous obtenons une vitesse de convergence de l'ordre de $O\left(\max\left\{\sqrt{\frac{\log n}{nh_n^d}}, h_n^2\right\}\right)$.

5.2 Perspectives

Pour conclure nous pouvons soulever quelques pistes de recherche :

- Les résultats établis au chapitre quatre peuvent être étendus au cas où le modèle présente une certaine forme de dépendance, et nous pensons particulièrement au modèle d' α -mélange, pour lequel la littérature est très prolifique. Nous pensons, en particulier que la vitesse de convergence obtenue dans le cas indépendant peut-être maintenue au cas de l' α -mélange, en faisant un peu plus d'hypothèses sur le modèle.

- L'estimateur par moyenne locale (appelé aussi estimateur de Kaplan-Meier généralisé) introduit par Beran (1981) peut-être utilisé si on veut se mettre sur un modèle où les variables durée de vie et censure sont conditionnellement indépendantes de la covariable. A notre connaissance, la vitesse de convergence de cet estimateur n'a pas été établie comme c'est le cas dans le modèle indépendant.

- Le choix du paramètre de lissage a été fait dans le cas indépendant, il serait intéressant de l'étendre au cas dépendant.

- Des résultats en normes L^p , $p \geq 1$, en particulier en norme L^1 en (LAD, Least Absolute Error) qui permet d'obtenir des estimateurs robustes sont intéressants à établir.

- Il nous semble possible d'étendre nos résultats au cas de données fonctionnelles pour un modèle alpha mélangeant. Le cas de données complètes et indépendantes ayant été traité par Ferraty et Vieu (2006) et Ferraty, Mas et Vieu (2007).

Bibliographie

- [1] Beran, R. (1981). *Nonparametric regression with randomly censored survival data*, Technical Report university of California, Berkeley.
- [2] Bierens, H. J. (1987). *Kernel estimators of regression function*. Advances in Econometrics, Cambridge Univ. Press.
- [3] Bosq, D. et Lecoutre, J. P. (1987). *Théorie de l'estimation fonctionnelle*. Economica, Paris.
- [4] Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, J. C. (1983). *Tree structured methods for classification and regression*. Chapman et Hall.
- [5] Breslow, N. and Cowley, J. (1974). *A large sample study of the life table and product-limit estimates under random censorship*. Ann. Statist., 2, 437-443.
- [6] Buckley, J. and James, I. (1979). *Linear regression with censored data*. Biometrika, 66, 429-436.
- [7] Burke, M. D., Csörgo, S. and Horvarth, L. (1981). *Strong approximations of some biometric estimates under random censorship*. Z. Wahrsh. View. Gebiete, 56, 87-112.
- [8] Burke, M. D., Csörgo, S. and Horvarth, L. (1986). *A correction to an improvement of Strong approximations of some biometric estimates under random censorship*. Probab. Theory Related Fields, 79, 51-57.

-
- [9] Cai, Z. (2003). *Weighted local linear approach to censored nonparametric regression*. In Advances and Trends in Nonparametric Statistics, M.G. Akritas and D.M. Politis, 217-231.
- [10] Cai, Z. and Ould-Saïd, E. (2003). *Local M-estimation of regression function for time series data*. Statist. Probab. Lett., 65, 433-449.
- [11] Carbonez, A., Györfi, L. and Van der Meulin, E.C. (1995). *Partition-estimate of a regression function under random censoring*. Statist. Decisions, 13, 21-37.
- [12] Choi, C. E., Hall, P. and Rousson, V. (2000). *Data sharpening methods for bias reduction in nonparametric regression*. Ann. Statist., 28, 1339-1355.
- [13] Cleveland, W. S. and Loader, C. (1996). *Smoothing by local regression : principals and methods (with discussion)*. In W. Härdle & M. Shimek (eds.). Statistical Theory and Computational Aspects of Smoothing, 10-49, Heidelberg, Physica-Verlag.
- [14] Cleveland, W. S. and Devlin, S. J. (1988). *Locally weighted regression : an approach to regression analysis by local fitting*. J. Amer. Statist. Assoc., 83, 596-610.
- [15] Collomb, G. (1981). *Estimation nonparamétrique de la régression : revue bibliographique*. Int. Statist. Rev., 49, 75-93.
- [16] Collomb, G. (1983). *Méthodes Non Paramétriques en Régression, Analyse des Séries Temporelles, Prédiction et Discrimination*. Doctorat d'Etat, Toulouse 3.
- [17] Collomb, G. (1985). *Nonparametric regression : an up to date biblioraphy*. Statistics, 2, 297-307.
- [18] Cover, T. M. and Hart, P. E. (1967). *Nearest neighbor pattern classification*. IEEE Transactions of Information Theory, 13, 21-27.
- [19] Cox, D. R. (1972). *Regression models and life tables*. J. Royal. Statist. Soc., Serie B, 34, 187-202.

-
- [20] Dabrowska, D. M. (1987). *Nonparametric regression with censored survival data*. Scand. J. Statist., 14, 181-197.
- [21] Dabrowska, D.M. (1989). *Uniform consistency of the kernel conditionnal Kaplan-Meier estimate*. Ann. Statist., 17, 1157-1167.
- [22] Deheuvels, P. and Einmahl, J. H. J. (2000). *Functional limit laws for the increments of Kaplan-Meier product limit processes and applications*. Ann. Probab., 28, 1301-1335.
- [23] Devroye, L. P. (1978). *The uniform convergence of the Nadaraya-Watson regression function estimate*. Canad. J. of Statist., 6, 179-191.
- [24] Devroye, L. P., Györfi, L. and Lugosi, G. (1996). *A Probability Theory of Pattern Recognition*. Springer, New York.
- [25] Devroye, L. P. and Penrod, C. S. (1984). *The strong uniform convergence of multiplicative kernel estimates*. Canad. J. of Statist., 14, 211-219.
- [26] Devroye, L. P. and Wagner, T. J. (1980). *Distribution free consistency results in nonparametric discrimination and regression function estimation*. Ann. Statist., 8, 231-239.
- [27] Dony, J. and Einmahl, U. (2006). *Weighted uniform consistency of kernel density estimator with general bandwidth sequence*. Elect. J. Probab., 11, 844-859.
- [28] Fan, F. (1992). *Design-Adaptive nonparametric regression*. J. Amer. Statist. Assoc., 87, 998-1004.
- [29] Fan, F. (1993). *Local linear regression smoothers and their minimax efficiencies*. Ann. Statist., 21, 196-216.
- [30] Fan, F. and Gijbels, I. (1996). *Local polynomial modelling and its applications*. Monographs on Statist. and Appl. Probab., 66, Chapman and Hill, London.

-
- [31] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis : Theory and Practice*. Springer , New York.
- [32] Ferraty, F., Mass, A. and Vieu, P. (2007). *Nonparametric regression of functional data : Inference and practical aspects*. Aust. N. Z. J. Statist. 49, 267-286.
- [33] Gäscher, T. and Müller, H. G. (1979). *Kernel estimation of regression function*. In : *Smoothing Techniques for Curve Estimation*. Lecture Notes in Math., 75, 23-68, Springer Verlag, Berlin.
- [34] Gäscher, T. and Müller, H.G. (1984). *Estimating regression function and their derivatives by the kernel methods*. Scand.J. Statist., 11, 171-185.
- [35] Giné, E. and Guillou, A. (2002). *Rate of strong consistency for multivariate kernel density estimators*. Ann. Inst. H. Poincaré, 38, 907-921.
- [36] Giné, E., Koltchinskii V. and Zinn, I. 2004). *Weighted uniform consistency of kernel density estimators*. Ann. Probab., 32, 2570-2605.
- [37] Gonzalez-Manteiga, W. C. and Cadarso-Suarez M. (1991). *Linear regression with randomly right-censored data using prior nonparametric estimation*. In Nonparametric Functionnal Estimation and Related Topics (G. Roussas, ed) Kluwer Academic Publishers, Dordrecht, 315-328.
- [38] Gordon, L. and Olshen, R. A. (1980). *Consistent nonparametric regression from recursive partitioning schemes*. J. Multivariate Anal., 10, 611-627.
- [39] Guessoum, Z. and Ould-Saïd, E. (2009). *On nonparametric estimation of the regression function under random censorship model*. Statist. Decisions, à paraître.
- [40] Györfi, L. and Walk, H. (1996). *On the strong universal consistency of a serie type regression estimate*. Math. Meth. Statist., 5, 332-342.

-
- [41] Györfi, L. and Walk, H. (1997). *On the strong universal consistency of a recursive regression estimate by Pál Révész*. *Statist. Probab. Lett.*, 31, 177-183.
- [42] Györfi, L., Köhler, M., Krzyzak, A. and Walk, H. (2002). *A Distribution-Free Theory of Nonparametric Regression*. Springer series in statistics.
- [43] Hall, P. (1984). *Asymptotic properties of integrated square error and cross validation function*. *Z. Wahrsch. Verw. Gebiete*, 67, 175-195.
- [44] Hall, P. and Müller, H. G. (2003). *Order-preserving nonparametric regression with application to conditionnal distribution and quantile function estimation*. *J. Amer. Statist. Assoc.*, 98, 598-608.
- [45] Härdle, W. (1990). *Applied Nonparametric Regression*. Cambridge Univ. Press.
- [46] Härdle, W. and Marron, J. S. (1985). *Optimal bandwidth selection in nonparametric regression function estimation*. *Ann. Statist.*, 13, 1465-1481.
- [47] Härdle, W. and Kelly, G. (1987). *Nonparametric kernel regression estimation : optimal choice of bandwidth*. *Statistics*, 18, 21-35.
- [48] Horvath, J. L. (1981). *On nonparametric regression with randomly censored data*. In : *Proc. Third pannonian Symp. on Math. Statist.*, Visegrad, Hungary, 105-112.
- [49] Ioannides, D. (1992). *Integrated square error of nonparametric estimators of regression function : The fixed design*. *Statist. Probab. Lett.*, 15, 85-94.
- [50] Johnston, G. (1979). *Smooth nonparametric regression analysis*. Inst. of Statist. Mimeo Series, 1253, PhD, Dissertation, Univ. of N. Carolina and Chapell Hill.
- [51] Kaplan, E. L. and Meier, P. (1958). *Nonparametric estimation from incomplete observations*. *J. Amer. Statist. Assoc.*, 53, 457-481.

-
- [52] Kim, T. Y. and Cox, D. D. (1995). *Asymptotic behavior of some measures of accuracy in nonparametric curve estimation with dependent observations*. J. Multivariate. Anal., 53, 67-93.
- [53] Klein, J.P. (1991). *Small sample moments of some estimator of the variance of the Kaplan-Meier and Nelson-Aalen estimators*. Scand. J. Statist., 18, 333-340.
- [54] Köhler, M. (1997). *On the universal consistency of a least squares spline regression estimator*. Math. Methods Statist., 6, 349-364.
- [55] Köhler, M. (1999). *Universal consistency regression function estimation using hierarchical B-splines*. J. Multivariate Anal., 67, 138-164.
- [56] Köhler, M. and Krzyzak, A. (2001). *Nonparametric regression estimation using penalized least squares*. IEEE Trans. Inform. Theory, 47, 3054-3058.
- [57] Köhler, M., Måthé, K. and Pintér, M. (2002). *Prediction from randomly Right Censored Data*. J. Multivariate Anal., 80, 73-100.
- [58] Komlos, J., Major, P. and Tusnady, G. (1975). *Weak convergence and embedding limit theormes of probability theory*. Colloq. Math. Soc. Janos Bolyai, North Holland, 11, 149-165.
- [59] Koul, H., Susarla, V. and Van Rysin, J. (1981). *Regression analysis with randomly right-censored data*. Ann. Statist., 9, 1276-1288.
- [60] Lai, S. L. (1977). *Large sample properties of k-nearest neighbors procedures*. PhD dissertation, Dpt Math., UCLA, Los Angeles.
- [61] Lugosi, G. and Zeger, K. (1995). *Nonparametric estimation via empirical risk minimization*. IEEE Trans. Inform. Theory, 41, 677-687.
- [62] Linton, O. and Nielsen, J. P. (1994). *A multiplicative bias reduction method for nonparametric regression*. Statist. Probab. Lett., 19, 181-187.

-
- [63] Lo, H. S., Mack, Y. P. and Wang, J. L. (1989). *Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator*. Probab. Theory Related Fields, 80, 461-473.
- [64] Lo, H. S. and Singh, K. (1986). *The product limit estimator and bootstrap : some asymptotic representations*. Probab. Theory Related Fields, 71, 455-465.
- [65] Loftsgaarden, D. O. and Quesenberry, G. P. (1965). *A nonparametric estimate of a multivariate density function*. Ann. of Math. Statist., 39, 1049-1051.
- [66] Mammen, E. and Marron, J. S. (1997). *Mass centred kernel smoothers*. Biometrika, 84, 765-777.
- [67] Miller, R. G. (1977). *Developements in multiple comparison*. J. Amer. Statist. Assoc., 72, 779-788.
- [68] Müller, H. G. (1988). *Nonparametric regression analysis of longitudinal data*. Lecture Notes in Statistics, 46, Springer-Verlag, Berlin.
- [69] Müller, H. G. (1997). *Density adjusted kernel smoothers for random design nonparametric regression*. Statist. Probab. Lett., 36, 161-172.
- [70] Müller, H. G and Song, K. S. (1993). *Identity reproducing multivariate nonparametric regression*. J. Multivariate Anal., 46, 237-253.
- [71] Nadaraya, E. A. (1964). *On estimating regression*. Theor. Probab. Appl. 9, 141-142.
- [72] Nadaraya, E. A. (1970). *Remarks on nonparametric estimates for densities functions and regression curves*. Theory Probab. Appl., 15, 134-137.
- [73] Nadaraya, E. A. (1989). *Nonparametric estimation of probability densities and regression curves*. Kluwer, Dordrecht.

-
- [74] Ould-Saïd, E. (2006). *A strong uniform convergence rate of kernel conditional quantile estimator under random censorship*. Statist. Probab. Lett., 76, 579-58.
- [75] Ould-Saïd, E. and Lemdani, M. (2006). *Asymptotic properties of a nonparametric regression function estimator with randomly truncated data*. Ann. Instit. Statist. Math., 58, 357-378.
- [76] Parzen, E. (1962). *On estimation of a probability density function and mode*. Ann. Math. Statist., 33, 1065-1076.
- [77] Peterson, A. V. (1977). *Expressing the Kaplan-Meier Estimator as a function of subsurvival function*. J. Amer. statist. Assoc., 72, 854-858.
- [78] Pinter, M. (2001). *Consistency results in nonparametric estimation and classification*. PhD Thesis, Technical University of Budapest.
- [79] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer Verlag.
- [80] Priestley, M. B. and Chao, M. T. (1972). *Nonparametric function fitting*. J. Royal Statist. Soc., B 23, 395-436.
- [81] Rao. C. R. (1965). *Linear Statistical Inference and Its Applications*. Wiley.
- [82] Rudemo, M. (1982). *Empirical choice of histograms and kernel density estimators*. Scand. J. Statist., 9, 65-78.
- [83] Ruppert, D. and Wand, M. P. (1994). *Multivariate weighted least square regression*. Ann. Statist., 22, 1346-1370.
- [84] Rosenblatt, M. (1956). *Remarks on some nonparametric estimates of density function*. Ann. Math. Statist., 27, 832-837.
- [85] Schimek, M. G. (2000). *Smoothing and Regression : Approaches, computation and Application*. John Wiley & Son, New-York.

-
- [86] Schuster, E. F. (1972). *Joint Asymptotic Distribution of the estimated regression function at a finite number of distinct points*. Ann. Math. Statist., 43, 84-88.
- [87] Shorack G. R. and Wellner. J. A. (1986). *Empirical processes with application to statistics*. Wiley.
- [88] Silverman, B. W. (1978). *Weak and Strong uniform consistency of the kernel estimate of the density and its derivatives*. Ann. Statist., 6, 177-189.
- [89] Silverman, B. W. (1986). *Estimation for Statistics and Data Analysis*. Monographs on Statist. and Appl. Probab., Chapman and Hill.
- [90] Spiegelman, C. and Sacks, J. (1980). *Consistent window estimation in nonparametric regression*. Ann. Statist., 8, 240-246.
- [91] Stone, J. C. (1977). *Consistence nonparametric regression*. Ann. Statist., 5.4, 595-645.
- [92] Stone, J. C. (1980). *Optimal rates of convergence for nonparametric estimators*. Ann. Statist., 8, 240-246.
- [93] Stone, J. C. (1982). *Optimal global rates of convergence for nonparametric regression*. Ann. Statist., 10, 1040-1053.
- [94] Stute, W. (1982). *Law of logarithm for kernel density estimators*. Ann. Probab., 10, 414-422.
- [95] Stute, W. (1984). *The oscillation behavior of empirical processes. The multivariate case*. Ann. Probab., 12, 361-379.
- [96] Stute, W. (1993). *Consistent estimation under random censorship when covariables are present*. J. Multivariate Anal. 1, 89-103.
- [97] Stute, W. (1999). *Nonlinear censored regression*. Statist. Sinica, 9, 1089-1102, .
- [98] Sursula, V., Tsai, W. Y. and Van Rysin J. V. (1984). *A Buckley-James-Type estimator for the mean with censored data*. Biometrika, 71, 624-625.

- [99] Tsybakov, A. B. (2003). *Introduction à l'estimation non paramétrique*. Mathématiques et applications, 41, Springer.
- [100] Tukey, J. W. (1947). *Nonparametric estimation II. Statistically equivalent blocks and tolerance region*. Ann. of Math. Statist., 18, nř4, 529-539.
- [101] Van Der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical process*. Springer.
- [102] Walk, H. (1997). *Strong universal consistency of kernel and partitioning regression estimates*. Preprint 97-1, Mathematisches Institut A, Universität Stuttgart.
- [103] Watson, G. S. (1964). *Smooth regression analysis*. Sankhya, Series A, 26, 359-372.
- [104] Zhang, B. (1998). *A note on the strong uniform consistency of kernel density estimators under random censorship*. Sankhya, serie A, 2, 265-273.