REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ DJILLALI LYABES
FACULTÉ DES SCIENCES

**THÈSE DE DOCTORAT 3ÈME CYCLE**
**EN : MATHÉMATIQUES**
**Spécialité : PROBABILITÉ ET STATISTIQUES**

**Présenter Par**

*Sihem SEMMAR*

**Titre**

# Sur l'estimation récursive de la fonction de densitée conditionnelle pour des données censurées

Soutenu publiquement le 30/11/2016, devant le jury composé de :

| | | | | |
|---|---|---|---|---|
| M | **A. LAKSACI** | Professeur | Univ. Sidi Bel-Abbès | Encadreur |
| M | **S. KHARDANI** | Maître de Conférence | ENI de Monastir. Tunisie | Co-Encadreur |
| M | **A. GHERIBALLAH** | Professeur | Univ. Sidi Bel-Abbès | Président |
| M | **M. ATTOUCH** | Professeur | Univ. Sidi Bel-Abbès | Examinateur |
| M | **F. MADANI** | Maître de Conférence | Univ. Saida | Examinateur |

À la mémoire de ma grand mére.

**Remerciements**

Tout au long de mon parcours universitaire, j'ai eu la chance de rencontrer des personnes formidables, sur qui j'ai pu compter et qui m'ont aidé à réaliser ce travail. Leur soutien m'était d'une importance capitale. Je tiens à les saluer et les remercier.

Je voudrais en tout premier lieu, exprimer ma profonde reconnaissance à mes directeurs de thèse les professeur ***Ali Laksaci*** et ***Salah Khardani*** pour la confiance qu'il m'ont témoignée en acceptant la direction de cette thèse aussi pour leur encadrement et leurs encouragements durant toute la période de la réalisation de ce travail malgré toutes leurs occupations. pour le temps et la patience que vous m'avez accordés tout au long de ces années en me fournissant d'excellentes conditions logistiques. Je vous dis Merci

Je tiens à remercier vivement le Professeur ***Abdelkader Gheriballah*** , qui m'a fait l'honneur de présider le jury de cette thèse, malgré ses nombreuses responsabilités.

Je suis trés honorée que Monsieur le Professeur ***Mohammed Kadi Attouch*** à accepté d'être examinateur de mon travail. Je le remercie également pour la confiance qu'il m'a témoignée tout au long de ces années et pour tous ses conseils et remarques constructives.

Je veux exprimer ma reconnaisance à Monsieur ***Fethi Madani*** qui a accepté sans hésitation d'examiner ce travail bien qu'il soit occupé. Je lui adresse mes sentiments les plus respectueux.

Un grand merci au Membre du laboratoire de Statistique et Processus Stochastiques de l'université de Sidi Bel Abbes, ainsi les enseignants du département de Probabilité et Statistiques.

Bien sûre je n'aurais pu atteindre cet objectif sans l'aide continue de mes enseignants, qui tout au long de mon apprentissage, m'ont transmis la passion du savoir et de la science et sûrement aussi le bonheur de partager ses connaissances grâce à l'enseignement. Je cite en particulier mes premiers enseignants, ma mère et mon père qui ont su croire en moi et qui m'ont apporté toute leur aide quand j'en ai eu besoin, tous les remerciements du monde ne suffiraient pas. Ce mémoire de thèse leur est dédié.

Les mots me manquent pour remercier, à sa juste valeur mon époux, pour ses encouragements et son soutien perpétuel et réconfortant, Je ne saurai passer sous silence l'apport inestimable des autres membres de ma famille qui m'ont soutenue, de près ou de loin durant mes études doctorales. A mon très cher frère *Mohammed Abderrahmane* et à ma chère soeur *Fatima*. Merci ma famille, je n'aurais rien fait de tout cela sans votre amour.

Je souhaite que cette thèse soit digne de toutes les personnes qui ont rendu ce travail possible par leur présence, leur encouragment, leur amitié et leur amour.

## Résumé

Dans cette thése, nous nous intéressons à la prévision non paramétriques récursifs dans les modèles de censure données incomplètes (censurées). Plus précisément, nous nous intéressons à la fonction densité conditionnelle pour lesquelles nous construisons des estimateurs et étudions le comportement asymptotique.

Dans la première partie, nous considérons une suite de v.a. $\{T_i, i \geq 1\}$ indépendante et identiquement distribuée (iid), de densité $g$, censurée à droite par une suite aléatoire $\{C_i, i \geq 1\}$ supposée iid et indépendante de $\{T_i, i \geq 1\}$. Nous établissons la convergence et la normalité asymptotique d'un estimateur à noyau récursif de la densité conditionnelle, et on va démontrer que cette propriété asymptotique est très utile dans des nombreuses analyses statistiques telle la prévision des intervalles de confiance. Une étude sur des données simulées de taille finie à étè réalisée.

Dans la deuxième partie, nous traitons le cas où la suite $\{T_i, i \geq 1\}$ est supposée fortement mélangeante, alors que les $\{C_i, i \geq 1\}$ sont iid. Nous construisons un estimateur à noyau récursif de la densité conditionnelle dont nous établissons la convergence presque sûre et la normalité asymptotyque.

**Mots clefs** : *$\alpha$-mélange, convergence forte, données censurées, estimateurs de Kaplan-Meier, estimateur à noyau,estimateurs récursifs, densitée conditionnelle, mode conditionnel, normalité asymptotique.*

**Abstract**

In this work, we focus on the recursive nonparametric prevision in censorship models, we are interested in the problem of conditional density functions, for independent and dependent data, for which we construct estimators and study the asymptotic behavior.
In the first part, we consider an independent and identically distributed (iid) sequence random variables (rvs) $\{T_i, i \geq 1\}$ with density $g$. This sequence is right-censored by another iid sequence of rvs $\{C_i, i \geq 1\}$ which is supposed to be independent of $\{T_i, i \geq 1\}$. We establish the consistency and asymptotic normality of a recursive kernel estimator of the conditional density, and we will demonstrate that this asymptotic property is very useful in many statistical analyzes such the prevision of confidence intervals.
In a second part we deal with the case where the sequence $\{T_i, i \geq 1\}$ is supposed to be stationary and strongly mixing whereas the $\{C_i, i \geq 1\}$ are iid, We build a recursive kernel estimator of the conditional density in which we establish the almost sure convergence and asymptotyque normality.



**Keywords** : *$\alpha$-mixing, asymptotic normality, conditional density, strong consistency, censored data, Kaplan-Meier estimator, Kernel estimator, recursive estimator.*

# Table des matières

# Chapitre 1

# Introduction

On considère $n$ observations d'une série chronologique $x_1, x_2, \cdots, x_n$, d'où on voudrait prévoir $x_{n+h}$ (h est l'horizon de prévision)pour cela on utilise une méthode non paramétriques. La particularité de la statistique non-paramétrique est que le paramètre inconnu qu'on cherche à détecter, à estimer ou à classifier n'est pas supposé d'appartenir à une famille indicée par un petit nombre de paramètres réels. En général, dans la théorie non-paramétrique on suppose que le nombre de paramètres qui décrivent la loi des observations est une fonction croissant du nombre d'observations, ou encore que le nombre de paramètres est infini.
Ce chapitre est divisé en trois sections. Nous commençons par une étude bibliographique à la premier section. En suite nous présentons le plan de notre thèse dans le deuxième paragraphe. Une brève présentation des résultats obtenus est donné dans la dernière section.

## 1.1 Context bibliographique

Le travail développé dans ce manuscrit de thèse se situe l'intersection entre les thématiques importantes de la Statistique, à savoir l'estimation non paramétrique (avec des méthodes à noyau), la théorie des données incomplètes (censures et troncatures) et la recursivité . Pour la partie d'estimation non paramétrique, nous nous sommes intéressés plus particulièrement à l'estimation non paramétrique de la densité conditionnelle (qui représentent une bonne alternative de regréssion) conditionnels et non conditionnels. Nous nous sommes placé, dans un premier temps, dans le cas où on dispose de réalisations de variables aléatoires indépendantes et identiquement distribuées (i.i.d.). Ensuite nous nous sommes intéressés au cas d'observations vérifiant une hypothèse de dépendance (mélange). En ce qui concerne la thématique des données incomplètes, on s'intéresse à des variables d'intérêts censurées à droite. Pour des raisons d'optimisation du temps de calcul et la nature des données étudiés, nous avons choisi d'étudier des estimateurs récursifs.
L'estimation non-paramétrique de la densité et la densité conditionnelle est un sujet qui a donné lieu à un grand nombre de travaux. Son champ d'application est très vaste et couvre divers domaines, comme l'analyse de la régression, des séries chronologiques et la théorie de la fiabilité. Par exemple Singh utilise des résultats sur l'estimation de la densité, en particu-

lier pour estimer l'information de Fisher. Pour une revue des différentes méthodes on pourra consulter Wegman (1972), Fukunaga (1972) et Tarter et Koonmal (1976)).

les principales méthodes non-paramétriques pour l'estimation de la densité sont la méthode du noyau introduite par Rosenblatt (1956) et Parzen (1962), la méthode des séries orthogonales étudiée entre autre par Schwartz (1967) et Watson (1964) et la méthode de l'histogramme introduite par Graunt puis developpée par Scott, Tran (1994), Carbon et Tran (1996). Parmi l'ensemble de ces estimateurs, l'un des plus utilisés reste l'estimateur à noyau défini par :

$$f_n^{\mathrm{PR}}(x) := \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right), \forall x \in \mathbb{R}^d,$$

où $K$ est un noyau défini dans $\mathbb{R}^d$, borné et intégrable par rapport à la mesure de Lebesgue et $h_n$ une suite réelle appelée paramètre de lissage, tendant vers zéro à l'infini. D'autres conditions complémentaires sur $K$ et $h_n$ sont nécessaires pour l'étude de cet estimateur. L'étude de $f_n^{\mathrm{PR}}(x)$ a donné lieu à une vaste littérature statistique, pour une représentation globale des résultats obtenus sur cet estimateur, notamment dans le cadre de données dépendantes, nous renvoyons aux livres de Prakasa-Rao (1983) , Bosq (1998) , Bosq, Lecoutre (1987), Bosq et Blanke(2007).

De nombreux auteurs se sont intéressés à l'estimation de la densité, on peut citer Devroye et Penrod (1984) qui ont établi 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. 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 $\mathbb{R}^d$. Dans le cas des variables dépendantes, Cai et Roussas (1992) montrent la convergence simple de cet estimateur à noyau .
En ce qui concerne l'estimateur de la fonction densité conditionnelle, Roussas (1968) fut le premier à établir ses propriétés asymptotiques pour des données markoviennes, ainsi que sa convergence en probabilité. Youndjé (1993) s'est intéressé à l'étude de la densité conditionnelle pour des données complètes indépendantes. Laksasi et Yousfate (2002) ont établi, pour un processus markovien stationnaire, la convergence en norme $\mathcal{L}^p$ pour l'estimateur à noyau de la densité conditionnelle. Pour Hall et al., (2004) la densité conditionnelle joue un rôle clef en statistique appliquée et particulièrement en économie, pour Fan and Yim (2004), une densité conditionnelle offre le résumé le plus informatif de la relation entre variable dépendante et indépendante, enfin Efromovich, (2007) dit que la densité conditionnelle de la variable dépendante sachant le prédicteur décrit l'association ultime entre le prédicteur et la variable dépendante.

La récursivité peut s'avérer cruciale lorsque l'on cherche à inférer sur des phénomènes qui évoluent dans le temps et qui nécessitent une mise à jour constante des estimations effectuées. L'intérêt des méthodes récursives est de prendre en compte l'arrivée temporelle des informations et d'affiner ainsi au fil du temps les estimations. L'idée est d'utiliser les estimations calculées sur la base de données initiales et de les remettre à jour en tenant uniquement compte des nouvelles données arrivant dans la base. Le gain en terme de temps de calcul peut être très intéressant et les applications d'une telle approche sont nombreuses. En outre, les estimateurs récursifs peuvent s'avérer préférables aux versions non récursives du fait de leur plus faible variance asymptotique.

Concernant la métode d'estimation récursive du noyau, une premiére forme a été introduite par Wolverton et Wagner (1969). De nombreuses variantes récursives ont également été proposées et étudiées depuis. Deheuvels (1973) s'est intéressé à l'estimation sequentiel de la densité, Aussi Wegman et Davies (1979) étudient l'estimation récursif d'une densité de probabilité, leur idée consiste à partager le paramétre de lissage de l'estimation à noyau habituel en deux puissance. Pour des résultats récents, nous renvoyons à Mokkadem et al. (2006) et Amiri et al. (2014).
Derniérement deux approches principales ont été développées dans le papier de Benziadi et al. (2014), le premier est basé sur l'estimation récursive à double noyau de la fonction de répartition conditionnelle et le second est obtenu en utilisant l'approche robuste
Les algorithmes stochastiques de recherche du zéro $z^*$ d'une fonction inconnue $h : R \longrightarrow R$ sont construits de la façon suivante : (i) $Z_0 \in \mathbb{R}$, (ii) : on définit récursivement la suite $(Z_n)$ en posant

$$Z_n = Z_{n-1} + \beta_n W_n$$

où $W_n$ est une observation de la fonction $h$ au point $Z_{n-1}$ et où le pas $(\beta_n)$ est une suite de réels positifs qui tend vers zéro. Soit $(X_1, ..., X_n)$ un échantillon de la loi d'une variable aléatoire $X$ de densité de probabilité $f$ . Pour construire un estimateur de $f$ en un point $x$ par la méthode des algorithmes stochastiques, on définit un algorithme de recherche du zéro de la fonction $h : y \to f(x) - y$. On procède donc de la façon suivante :
i) on choisit $f(x) \in R$,
ii) Pour tout $n \succeq 1$, on pose

$$f_n(x) = f_{n-1}(x) + \beta_n W_n(x)$$

où $W_n(x)$ doit être une observation de la fonction $h$ au point $f_{n-1}(x)$. Soient K un noyau (i.e. une fonction telle que $(\int_{\mathbb{R}} K(x)dx = 1)$ et $(h_n)$ une fenêtre (i.e. une suite déterministe positive qui tend vers zéro) ; $f(x)$ peut être estimée par $Z_n(x) = h_n^{-1} K\left(h_n^{-1}(x - X_n)\right)$, ce qui mène à poser $W_n(x) = Z_n(x) - f_{n-1}(x)$.
Un estimateur récursif de la densité $f$ au point $x$ s'écrit alors sous la forme

$$f_n(x) = (1 - \beta_n) f_{n-1}(x) + \beta_n h_n^{-1} K\left(h_n^{-1}(x - X_n)\right) \tag{1.1}$$

La relation (1.1) définit toute une classe d'estimateurs récursifs à noyau d'une densité de probabilité. Notons que si l'on pose $\beta_n = \frac{1}{n}$, alors l'estimateur $f_n$ défini par l'algorithme (1.1) se réecrit sous la forme

$$f_n(x) = \frac{1}{n} \sum_{k=1}^{n} K\left(h_k^{-1}(x - X_k)\right) \tag{1.2}$$

dans ce cas, $f_n$ est l'estimateur récursif introduit par Wolverton et Wagner (1969). D'autre part, dans le cas où on pose $\beta_n = \frac{h_n}{\sum_{k=1}^n h_k}$, l'estimateur $f_n$ défini par l'algorithme se réécrit sous la forme

$$f_n(x) = \frac{1}{\sum_{k=1}^n h_k} \sum_{k=1}^n K\left(h_k^{-1}(x - X_k)\right) \tag{1.3}$$

$f_n$ est alors l'estimateur récursif introduit par Deheuvels (1973) et étudié par Duflo (1997). Les deux estimateurs définis par (11.1) et (21.1) peuvent etre vu sous la forme suivante

$$f_n^\ell(x) := \frac{1}{\sum_{i=1}^n h_i^{(1-\ell)}} \sum_{i=1}^n \frac{1}{h_i^\ell} K\left(\frac{x - X_i}{h_i}\right), \ x \in \mathbb{R} \tag{1.4}$$

pour $\ell = 0, 1$ La question qui se pose naturellement est de savoir quel est le choix optimal du pas. Slaoui (2010) a réalisé plusieurs travaux sur l'étude du choix optimal du pas de l'algorithme en explicitant le biais et la variance de l'estimateur $f_n$ . Finalement, il considère le point de vue de l'estimation par intervalles de confiance en donnant la vitesse de convergence presque sûre de l'estimateur $f_n$.

Dans le cas iid, Davies (1973), Deheuvels (1973,1974), Roussas (1992), Wegman, Davies (1973) et Wertz (1985 étudient la famille $f_n^{\mathrm{H}}(x)$ et les cas $\ell = 0$, et $\ell = 1$. En particulier, en dimension $d = 1$, Deheuvels (1973,1974) établit la convergence en moyenne quadratique de la famille $f_n^{\mathrm{H}}(x)$ et donne des conditions nécessaires et suffisantes pour sa convergence presque sûre. Roussas (1992), Wegman et Davies (1979) établissent les vitesses de convergence presque sûre exactes dans les cas $\ell = 0$ et $\ell = 1$. Aussi, Isogai (1984) établit sous certaines conditions, la normalité asymptotique pour $\ell = 1$ dans le cas iid.

Dans le cas dépendant, seuls les cas $\ell = 1/2$ et $\ell = 1$ ont été étudiés dans la littérature. Les résultats sur la convergence en moyenne quadratique et la normalité asymptotique pour $\ell = 1$ sont établis par Masry (1986), pour des processus stationnaires fortement mélangeants. La vitesse de convergence presque sûre ponctuelle pour $\ell = 0$ et $\ell = 1$ est étudiée par Takahata (1980), Masry et Györfi (1987), d'abord sous des conditions de $\rho$-mélangeance, ensuite pour $\ell = 1$, par Masry (1987), pour des processus fortement mélangeants. Un résultat uniforme est également obtenu dans le cas $\ell = 1$ par Tran (1989), sous des conditions de forte mélangeance. La normalité asymptotique pour $\ell = 1$ est également examinée par Lian et Baek (2004) pour des suites de variables négativement associées. Les approches utilisées dans ces travaux, notamment pour la convergence en moyenne quadratique et la normalité asymptotique, ne se généralisent pas aisément, en dimension supérieure pour des valeurs plus petites de $\ell$, alors qu'en particulier, le cas $\ell = 0$ est intéressant du fait de la faible variance de l'estimateur.

Dans ce travail, nous nous intéressons essentiellement au problème de prévision dans les modèles de durées de vie à données incomplètes (censurées aléatoirement à droite). Nous considérons, pour cela, les cas non paramétriques conditionnel et non conditionnel. Nous proposons comme alternative à la méthode de régression la fonction densité et la fonction densité conditionnelle.

Au-delà des origines historiques de l'analyse statistique des durées de vie, démographiques et actuarielles, les trois grands domaines actuels de l'analyse des données de survie sont la fiabilité, la biostatistique et l'économie. Ils existe néanmoins *a priori* deux différences fondamentales entre les deux derniers et la fiabilité des systèmes. D'une part, au-delà de la détermination des caractéristiques de fiabilité d'un système physique à composantes multiples, il est important d'optimiser cette fiabilité en jouant sur l'architecture du système. Une telle approche d'optimisation pourrait être étendue aux études de survie médicale puisque, dans une certaine mesure, le corps humain est un système complexe et une greffe d'organe peut s'apparenter à un changement de pièce défectueuse.

Seconde différence principale : en biostatistique ou en économie, les individus sont rarement identiques entre eux, contrairement aux systèmes physiques. Un malade sera caractérisé par son âge, ses habitudes alimentaires, son environnement social...,etc... .Autre exemple, lors d'une étude pratique sur les durées de vie, en médecine, les chercheurs ne se contentent pas de relever uniquement la durée de vie observée. Les patients de l'échantillon bénéficieront d'un suivi médical qui permettra le relevé de nombreuses données tels les antécédents familiaux, le taux de diabète du sang, la densité de différents globules, etc.... Ces données ne sont généralement pas indépendantes de la durée de $T$, certaines d'entre elles pouvant être des circonstances aggravantes pour la maladie considérée. Un demandeur d'emploi possèdera des paramètres qui lui sont propres, tels que le diplôme le plus élevé obtenu, l'expérience professionnelle, la durée de sa situation en chômage, le sexe, ... Un certain nombre de variables exogènes seront à prendre en compte dans les modélisations.

Souvent les données de durées est de porter sur des variables aléatoires (v.a.) *positives.* Cela restreint *a priori* la classe des modèles paramétriques utilisables, encore que toute v.a. peut être transportée sur $\mathbb{R}^+$ par une transformation appropriée (par exemple exponentielle) et donc, pour des durées de vie, seule l'interprétation est spécifique.
Deuxième particularité : le calcul des probabilités identifie la loi de probabilité d'une v.a. par sa fonction de répartition, sa densité ou sa fonction caractéristique. L'interprétation d'une v.a. en termes de durée va permettre de définir d'autres notions associées, telles que la fonction de survie, le taux de hasard, la survie conditionnelle
ou la durée de vie moyenne restante.
Une troisième différence par rapport au modèle d'échantillonnage classique de la statistique est la présence de *données incomplètes.* Deux cas particuliers feront l'objet du paragraphe 1.5. Enfin, la présence de variables exogènes, qui ont essentiellement un caractère descriptif et qui vont intervenir dans l'écriture de la loi de la durée de vie est à noter $T$.

Quels modèles statistiques pourra-t-on utiliser pour analyser ces durées de survie ? Trois approches sont possibles : paramétrique, non paramétrique ou semi-paramétrique. Elles seront souvent complémentaires ou plus particulièrement adaptées à certaines données ou à certains types de problèmes.
Le modèle paramétrique spécifie l'appartenance de la vraie loi de notre variable d'intérêt $T$ à une classe $\mathcal{P}_0 = \{\mathbb{P}_\theta, \theta \in \Theta, \Theta \subset \mathbb{R}^d\}$. Il joue un rôle essentiel dans la théorie classique ; ayant l'avantage de permettre la modélisation avec un petit nombre de paramètres pouvant être estimés "facilement" à partir d'un nombre restreint de données observées. Ces

méthodes facilitent ensuite l'obtention d'intervalles de confiance et la construction de tests. L'inconvénient de cette pseudo-simplicité est la distorsion qui peut exister entre le "vrai" modèle du phénomène étudié et le modèle retenu.
La voie *non paramétrique* est susceptible de vaincre cette difficulté. On suppose alors que la loi $\mathbb{P}$ de $T$ est quelconque dans un ensemble de lois noté $\mathcal{P}$, sans spécification particulière de la liaison entre la variable $T$ et les variables exogènes. Comme la loi de $T$ est caractérisée par certaines fonctions : fonction de répartition, fonction densité, survie, hasard, etc... ou autres paramètres dépendant de ces fonctions, tels que la moyenne, la médiane, le mode, le mode conditionnel, on se trouve confronté à un problème *d'estimation fonctionnelle*. L'estimateur de Kaplan-Meier de la fonction de survie relève typiquement de cette problématique. Cette approche non paramétrique présente l'avantage de ne présupposer aucune forme particulière de la fonction estimée. La contrepartie sera la nécessité de disposer d'un nombre important d'observations (grande taille de l'échantillon), ceci est dû en fait à l'appartenance de ces paramètres fonctionnels à un espace de dimension infinie. L'approche semi-paramétrique est un compromis entre les deux approches précédents. Le lecteur intéressé pourra se référer à l'article de Begun *et al.* (1983), ou à celui de Welner (1985), ou encore à l'ouvrage de Bickel *et al.*(1987).

La principale source de difficulté dans l'analyse des durées de survie, et pour diverses raisons, est la présence de données incomplètes. Pour de telles observations pratiques, les procédures statistiques classiques ne sont plus valables et on a recours à des techniques statistiques plus élaborées afin de modéliser de telles observations pratiques pour valider les résultats expérimentaux.
Les données tronquées à gauche, modélisent des cadres d'études expérimentales pour des durées de vie qui doivent être "assez grandes" pour être observées. En effet, $T$ doit en fait être plus grande qu'une variable de troncature $Y$ pour pouvoir être observée. Ainsi, les observations ne sont possibles que si $T \geq Y$. Il s'agit d'un modèle qui est tout d'abord apparu en astronomie, où des échantillons sont composés d'objets astraux d'une certaine zone. Les luminosités absolue et apparente d'un objet astral sont respectivement définies comme étant sa brillance observée à une distance fixe et depuis la Terre et l'on n'observe que les objets qui sont suffisamment brillants, c'est-à-dire ceux pour lequels la luminosité $M \geq m$, $m$ étant la variable de troncature : ce sont des données dites *tronquées à gauche*. Dans ce cas, nous disposons de $N$ objets dans l'échantillon, mais nous ne sommes capables d'observer que les $n$ objets suffisamment brillants.
L'autre cas classique de données incomplètes est celui des données dites *censurées à droite*. Ce phénomène de censure modélise des cadres d'études expérimentales, pour certaines maladies, où les patients peuvent êtres perdus de vue suite à un déménagement ou encore suite à un décès non-inhérent à la maladie comme un accident de la route.

A titre d'exemple, analysons aussi la fonction de fiabilité d'une machine M. Dans cette perspective, l'observation portera sur l'étude du fonctionnement de $n$ machines identiques à M et nous noterons $T_1, ..., T_n$ les durées de vie de ces machines. Ces variables (aléatoires) seront supposées positives et de même densité $f$. La fonction de répartition associée sera notée $F$ et la fonction de fiabilité sera $S(t) = \mathbb{P}(T > t) = 1 - F(t)$.

C'est un problème classique qui intéresse par exemple l'industrie automobile quand elle cherche à prévoir la durée de vie d'un modèle. Dans un cadre simplifié où toutes les durées de vie sont observées, ce problème admet des solutions élémentaires et naturelles. Des estimateurs empiriques naturels existent pour la fonction de répartition, la fonction de survie notée $S$ et pour le taux de hasard cumulé $H = -\log(S)$ qui sont, respectivement, l'estimateur de Kaplan-Meier pour $S$ et l'estimateur de Nelson-Aalen pour $H$.

Nous nous intéressons, plus particulièrement, au cadre plus réaliste où l'observation de $T_1, ..., T_n$ est incomplète. Illustrons cette problématique avec l'exemple suivant. Dans un cadre idéal, un fabricant d'automobiles, s'il assure la maintenance de ses véhicules, peut observer l'instant de première panne de chaque véhicule d'un modèle qu'il vend. Il peut alors déterminer aisément si ce modèle est fiable. De manière plus réaliste, il peut arriver que la première panne de certains véhicules ne soit pas observée pour différentes raisons (vente par certains propriétaires, accidents indépendants du fonctionnement du véhicule, ...). Dans ce cadre-là, on parle de modèle censuré : c'est -à-dire à chaque véhicule $i$, $i$=1,2,...,$n$ est associé un couple de v.a. durées $(T_i, C_i)$ dont seule la plus petite est observée, $T_i$ est la durée de survie et $C_i$ est l'instant de censure.
L'estimation de certaines fonctions de ce modèle est un problème beaucoup plus délicat. Pour le résoudre, on note $(Y_1, \delta_1), (Y_2, \delta_2), ..., (Y_n, \delta_n)$ la suite observée où, pour tout $1 \leq i \leq n$,

$$Y_i = T_i \wedge C_i \quad et \quad \delta_i = 1\!\!1_{\{T_i \leq C_i\}}. \tag{1.5}$$

Ici $Y_i$ est la durée observée et $\delta_i$ est une variable binaire représentant la nature de cette durée qui prend la valeur 1 s'il s'agit d'une vraie durée de vie et 0 si c'est une censure.

Ce n'est qu'à la suite de l'article de Kaplan et Meier (1958) que les données censurées ont trouvé le poids qui est le leur dans la réalité.

Les articles et les manuels qui traitent des durées censurées utilisent généralement l'une ou l'autre de deux approches très différentes : soit les méthodes de la statistique classique (Bailey (1979), Kalbfleich et Prentice (1980), Cox et Oakes (1984), Moreau (1984), Bretagnolle et Huber (1988) usant souvent d'arguments combinatoires fins comme chez Guilbaut (1987), soit les processus ponctuels comme Gill (1980), Andersen et Gill (1982), Harrington et Fleming (1982), Cross et Huber (1987).

L'estimateur de Kaplan et Meier (E.K.M.) de la fonction de survie $S$ est défini, dans le cas de non ex-aequo, par

$$1 - \hat{F}_n(t) = \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{1}{n-i+1}\right)^{\mathbf{1}\{Y_{(i)} \leq t\}} & \text{si} \quad t < Y_{(n)}, \\ 0 & \text{si} \quad t \geq Y_{(n)} \end{cases}$$

ou $Y_{(1)}, Y_{(2)}, ..., Y_{(n)}$ les valeurs ordonnées des $Y_i$ et pour chacune des valeurs $Y_{(i)}$ la valeur $\delta_i$ correspondante.

Cet estimateur a des propriétés analogues à celles de la fonction de répartition empirique, en particulier il vérifie un théorème de normalité asymptotique globale ( Breslow et Crowley, 1974). Mais il a aussi d'autres propriétés qui, elles, sont typiques de la présence de censure et ont l'intérêt de donner des idées lorsqu'on cherche à construire d'autres procédures d'estimation en présence de censure, comme cela sera montré un peu plus loin en chapitres 3 et 4 pour différentes situations de dépendance. Le comportement asymptotique de cet estimateur a suscité l'intérêt d'un grand nombre d'auteurs, Breslow et Crowley (1974) furent les premiers à traiter de sa convergence et de sa normalité asymptotique. Pour plus de détails, nous renvoyons au livre de Shorack et Wellner (1986, p. 304).

## 1.2 Description et Contribution de la thèse

La statistique non paramétrique connaît un grand essor chez de nombreux auteurs et dans différents domaines. En effet, celle-ci possède un champ d'application très large permettant, ainsi, l'explication de certains phénomènes mal modélisés jusqu'à présent, tels que les séries chronologiques, et prédire les réalisations futures.

Il faut mentionner, par ailleur que la modélisation statistique par le biais de données incomplètes est largement employée lors d'études pratiques sur les durées de vie. Les données incomplètes les plus couramment utilisées sont les données censurées. Nous avons choisi, au travers de cette thèse, de considérer une durée de vie $T$ censurée à droite par une variable $C$.
L'estimation par échantillonnage pour ces types de variables demande beaucoup de prudence, contrairement aux autres méthodes classiques d'échantillonnage. Parmi les pistes explorées pour confronter cette difficulté de robustesse, on a fait appel à l'estimateur de Kaplan-Meier (1958) qui est d'un emploi très répandu ; son importance nous a donc conduits à lui accorder une grande place dans notre travail.
Dans cette thèse, nous nous intéressons à l'estimation non paramétrique de la densité conditionnelle en utilisant le noyeau récursif , pour des variables censurés
Après avoir construit notre estimateur de la densité conditionnelle dans le cas indépendant, nous avons cherché à déterminer les vitesses exactes de convergence de notre estimateur.
Dans le but de présenter les travaux que nous avons réalisé durant la réalisation de cette thèse, celle-ci est organisé comme suit :
Le chapitre suivant est un chapitre introductif qui présente une étude bibliographique des problèmes liés à l'estimation non paramétrique des paramètres conditionnels que ce soit dans le cadre de dimension finie ou infnie, nous donnons des contexte bibliographiques sur l'estimation non paramètre de la densitè conditionnelle et sur l'estimateur récursif de la densité. Enfin nous allons présenter des résultats obtenu durant cette thése .

Dans le chapitre 2, nous donnons la définition des donnés incomplètes précisament les données censurées. Nous y présentons la définition de l'estimateur de Kaplan-Meir pour la fonction de survie
Dans le troixième chapitre nous considérons une suite d'observation indépendantes et iden-

tiquement distribuées, ce chapitre est rédigés sous forme d'article, nous construisons un estimateur à noyau récursif de la densité conditionnelle où les données sont censurées à droite, nous établissons la convergence uniforme presque sûre sur un compacct avec vitesse et la normalité asymptotique sous des conditions de régularité sur les noyaux et nous étudions les propriétés asymptotiques de cet etimateur. Des simulation permettent de juger la qualité de notre estimateur.

Le quatrième chapitre traite le problème de l'estimation non paramétrique de la même fonction sous des conditions de dépendance faible (mélange fort) conditionnellement à une covariable $X$ de dimension $d$, sous une hypothèse qui combine les deux paramètres de dépendance et de dimension, nous établissons la convergence uniforme presque sûre sur un compact de $\mathbb{R}^d$ de la densité conditionnelle.

Nous finissons ce manuscrit par quelques perspectives de recherche.
Enfin, une annexe génralise un ensemble d'outils probabilistes ainsi que des définitions de certaines caractéristiques pour la variable d'intérêt, utilisées dans cette thèse.

## 1.3 Brève présentation des résultats obtenus

Dans cette section, nous donnons une brève présentation des différents résultats obtenus durant cette thése .

### 1.3.1 Notations :

Soit $(X_n, T_n)_{n\geq 1}$ une suite strictement stationnaire de vecteurs aléatoires définis dans le même espace probabilisé $(\Omega, F, P)$, à valeurs dans $\mathbb{R}^d \times \mathbb{R}$, ayant la même loi que $(X, T)$, considérons pour tout $x \in \mathbb{R}^d$, la densité conditionnele $\varphi(.|x)$ de $T_1$ sachant $X_1 = x$. On note par $(C_i)_{i=1,\cdots,n}$ variables aléatoire de censure indépendantes et identiquement distribués avec fonction de distribution continue $G$. Ainsi On construit nos estimateurs par des variables observés $(X_i, Y_i, \delta_i)_{i=1,\ldots n}$, où $Y_i = T_i \wedge C_i$ et $\delta_i = \mathbb{1}_{\{T_i\leq C_i\}}$ , où $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble $A$.
Notons par $g(.,.)$ la densité de probabilité du couple $(X, T)$ et $\phi(.,.)$ la densité conditionnelle de $T$ sachant $X = x$.
l'estimation du noyau de la densité conditionnelle $\phi(t|x)$ noté $\bar{\phi}_n(t|x)$ définit par :

$$\forall x \in \mathbb{R}^d \text{ et } \forall y \in \mathbb{R} \qquad \bar{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_n^{-1}\delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) L\left(\frac{t - Y_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)},$$

où $K$, $L$ sont des noyaux et $h_n$ est une séquence de nombres réels positives.

La version récursive de notre précédent estimateur de noyau est

$$\widehat{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_i^{-(d+1)}\delta_i\bar{G}_n^{-1}(Y_i)K\left(\frac{x-X_i}{h_i}\right)L\left(\frac{t-Y_i}{h_i}\right)}{\sum_{i=1}^{n} h_i^{-d}K\left(\frac{x-X_i}{h_i}\right)} =: \frac{\hat{g}_n(x,t)}{\ell_n(x)}$$

### 1.3.2 Résultats : Cas iid (unidimensionnel)

**Théoreme 1.3.1** .
*Sous certaines conditions, nous aurons pour la convergence presque sûre :*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x)-\phi(t|x)\right| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right), h_n^2\right)\right\} \quad a.s. \quad as \quad n\to\infty \tag{1.6}$$

**Théoreme 1.3.2** .
*Sous certaines conditions, nous aurons :*

$$\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x)-\phi(t|x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\sigma^2(x,t)\right)$$

*where* $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution,*

$$\sigma^2(x,t) = \theta\frac{\phi(t|x)}{\ell(x)\bar{G}(t)}\int_{\mathbb{R}^d}\int_{\mathbb{R}} K^2(z)L^2(y)dzdy$$

*and* $\mathcal{A} = \{(x,t) \quad \sigma^2(x,t)\neq 0\}$.

Les hypothèses imposées et les preuves des résultats ci-dessus seront données dans le chapitre 2.

### 1.3.3 Résultats : Cas $\alpha$ -mélangeant

**Théoreme 1.3.3** *Sous l'hypothéses détaillées dans le chapitre 3, on aura :*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x)-\phi(t|x)\right| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right), h_n^{+2}\right)\right\} \quad a.s. \quad as \quad n\to\infty \tag{1.7}$$

**Théoreme 1.3.4** .
*Nous aurons Sous les hypothéses détaillées dans le chapitre 3*

$$\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x)-\phi(t|x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\sigma^2(x,t)\right)$$

*où* $\xrightarrow{\mathcal{D}}$ *:la convegence en distribution.*

$$\sigma^2(x,t) = \theta\frac{\phi(t|x)}{\ell(x)\bar{G}(t)}\int_{\mathbb{R}^d}\int_{\mathbb{R}} K^2(z)L^2(y)dzdy$$

*and* $\mathcal{A} = \{(x,t) \quad \sigma^2(x,t)\neq 0\}$.

Les démonstration de ces théorèmes sont détaillées dans le chapitre 3.

# Bibliographie


[1] Ahmad, I., Lin, P.E. (1976). Nonparametric sequential estimation of a multiple regression Function, *Bull. Math. Statist.*, 17, 63–75.

[2] Amiri, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortemement mélangeant, *C. R. Acad. Sci. Paris, Ser*, I 347, 309–314.

[3] Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction, *J. Nonparam. Statist,* 24 (1), 169–186.

[4] Amiri, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, *arXiv :1211.5767v2.*

[5] Andersen, P. et Gill, R. (1982). Cox's regression model for counting processes : a large sample study. Ann. Statist., 10 : 1100-1120.

[6] Beran, R. (1981). Nonparametric regression with randomly censored survival data, *Technical university of Clifornia, Berkeley.*

[7] Bosq, D. (1989). Proprietes des operateurs de covariance empiriques d'un processus stationnaire hilbertien. C. R. Acad. Sci. Paris Ser. I Math. 309, No.14, 873-875.

[8] Bosq, D. (1990). Modele autoregressif hilbertien. Application a la prevision du comportement d'un processus a temps continu sur un intervalle de temps donne. C. R. Acad. Sci. Paris Ser. I Math. 310, No.11, 787-790.

[9] Bosq, D. (2000). Linear processs in function spaces. Theory and Application. Lectures Notes in Statistics. Vol 149, Springer Verlag.

[10] Bosq, D., Delecroix, M. (1985). Nonparametric prediction of a Hilbertspace valued random variable. Stochastic Process. Appl. 19, 271-280.

[11] Breslow, N. et Crowley, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. Ann. Statist., 2 : 437-453.

[12] Carbon, M. and Tran, L. T. (1996). On histograms for linear processes, J. Statist. Plann. Inference , 53 (3), 403-419.

[13] Carbonez, A., Györfi, L., Vander Meulin EC . (1995). Partitioning estimates of a regression function under random censoring, *Statist. & Decisions* , 13, 21–37.

[14] Carroll, J. (1976). On sequential density estimation, *Z. Wahrscheinlichkeits- theorie und Verw. Gebiete* , 36, 137–151.

[15] Collomb, G. (1984). Proprietes de convergence presque complete du predicteur a noyau. Z. W. Gebiete. 66, 441-460.

[16] Collomb, G. (1985). Nonparametric regression : an up to date bibliography Statistics. 16, 309-324.

[17] Dabo-Niang, S., Rhomari, N. (2003). Estimation non parametrique de la regression avec variable explicative dans un espace metrique. C. R., Math., Acad. Sci. Paris. 336, 75-80.

[18] Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data, *Scandi. J. Statist,* 14, 181–197.

[19] Dabrowska, D.M. (1989). Uniform consistency of the kernel conditional Kaplan Meier estimate, *Ann. Statist,* 17, 1157–1167.

[20] Davies, I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist,* 15, 49–54.

[21] Deheuvels, P. (1973). Sur l'Estimation séquentielle de la densité,*C. R. Acad. Sci., Paris, Ser. A,* 276, 1119-1121.

[22] Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C. R. Acad. Sci., Paris,* 278, 1217-1220.

[23] Deheuvels, P., Einmahl, JHJ. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications, *Ann Proba*, 28, 1301–1335.

[24] Devroye, L., Wagner, T.J. (1980). On the $L^1$ convergence of kernel estimators of regression functions with application in discrimination. *Z. Wahrschein. Verw. Get,* 51, 15-25.

[25] Doob, J. (1953). Stochastic process, Wiley New York.
2735-2759.

[26] Fan, J. et Yim, T. H., (2004). A crossvalidation method for estimating conditional densities. Biometrika, 91(4) :819-834.

[27] Ferraty, F. (2010). Special issue on statistical methods and problems in infinite dimensional spaces. J. Multivariate Analysis. 101(2), 305-490. 317-344.

[28] Gannoun, A., Saracco, J., Yu, K. (2003), Nonparametric prediction by conditional median and quantiles. J. Stat. Plann. Inference. 117, No.2, 207-223.

[29] Giné, E., Guillou, A., (2001). On consistency of kernel density estimators for randomly censored data : rates holding uniformly over adaptive intervals. Annales de l'Institut Henri Poincaré, 37, 503-522.

[30] Giné, E., Guillou, A., (2002). Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincaré, 38, 907-921.

[31] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, 58, 13-30.

[32] Isogai, E. (1984). Joint asymptotic normality of nonparametric recursive density estimators at a finite number of distinct points,*J. Japan Statist. Soc.*, 14 (2), 125-

[33] Khardani, S., Lemdani, M., Ould Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship, *J. of the Korean Statistical Society*, 39, 455–469.

[34] Khardani, S., Lemdani, M., Ould Saïd, E. (2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series, *J. Stat. Plann. Inference*, 141, 3426–3436.

[35] Kohler, M., Máthé, K., Pinter, M. (2002). Prediction from randomly Right Censored Data, *J. Multivariate Anal*, (80), 73–100.

[36] Krzÿzak, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation , *IEEE Trans. Inform. Theory*, 38, 1323–1338.

[37] Laksaci, A. (2005). Contribution aux modèles non paramétriques conditionnels pour variables explicatives fonctionnels. Thèse de doctorat, université de Toulouse.

[38] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur a noyau de la densite conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P. 51, 69-80.

[39] Laksaci, M. and Maref, F. (2009). Conditional cumulative distribution estimation and its applications

[40] Laksaci, A., Madani, F., Rachdi, M. (2010). Kernel conditional density estimation when the regressor is valued in a semi metric space. International Statistical Review. (In press).

*Journal of probability and statistical sciences*, **13**, Pages 47-56.

[41] Lecoutre, J. P., Ould-Said, E. (1993). Estimation de la fonction de hasard pour un processus fortement melangeant avec censure. Publ. Inst. Statist. Univ. Paris. 37, No.1-2, 59-69.

[42] Liang, H.Y. and Baek, J. (2004). Asymptotic normality of recursive density estimates under some dependence assumptions, Metrika 60, 155-166.

[43] Loannides, D., Matzner-Lober, E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. J. Nonparametr. Stat. 16, 515-524.

[44] Loeve, M. (1963). Probability Theory. Third Edition. Van Nostranr Princeton.

[45] Louani, D., Ould-Said, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. J. Nonparametric Statist. 11, No.4, 413-442.

[46] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. Monthly Notices Roy. Astronom Soc., 155, 95118.

[47] Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, IEEE Trans. Inform. Theory, 32, 254-267,

[48] Masry, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes, *Statist. Probab. Lett.*, 5, 249-254.

[49] Masry, E. and Györfi, L. (1987). Strong consistency and rates for recursive probability density estimators of stationary processes, *J. Multivariate Anal.*, 22, 79- 93.

[50] Masry, E. and Fan,J., (1997). Local polynomial estimation of recursive function for mixing processes, *Scandinave Journal of Statistics*, 24, 165–179.

[51] Mokkadem, A., Pelletier, M., Thiam, B. (2006) Large and moderate deviations principles for recursive kernel estimator of a multivariate density and its partial derivatives. *Serdica Math. J.* **32** , Pages 323–354.

[52] Nadaraya, E. (1964). On estimating regression. Theory Prob. Appl. 10, 186-196.

[53] Nguyen, T., Saracco, J. (2010). Estimation récursive en régression inverse par tranches, *Journal de la société française de statistique*, 151(2), 19–46.

[54] Ould-Said, E., Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. Nonparametric Statistics, 17, 797-806.

[55] Ould-Said, E., Tatachak, A. (2009a). On the nonparametric estimation of the simple mode under random left-truncation model. Romanian Journal of Pure and Applied Mathematics, 54, 243-266.

[56] Ould-Said, E., Tatachak, A. (2009b). Strong consistency rate for the kernel mode under strong mixing hypothesis and left truncation. J. Comm. Statist. Theory Me- thods, 38, 1154-1169.

[57] Quintela-del-Rio, A., Vieu, P. (1997). A nonparametric conditional mode estimate. J. Nonparametr. Statist. 8, No.3, 253-266.
2801.

[58] Rhomari, N. (2002). Approximation et inegalites exponentielles pour les sommes de vecteurs aleatoires dependants. C. R. Acad. Sci. Paris. 334, 149-

[59] Rio, E. (2000). Théorie asymptotique des processus aléaltoires faiblement dépendants, vol 31 Mathematics and applications Springer-Verlag, Berlin.

[60] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In Multivariate Analysis II, Ed. P.R. Krishnaiah. Academic Press, New York and London.

[61] Roussas, G. G. (1968). On some properties of nonparametric estimates of probability density functions. Bull. Soc. Math. Greece (N.S.) 9 29-43.

[62] Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl*, 36 (1), 107–116.

[63] Roussas, G. G. (1991). Kernel estimates under association : strong uniform consistency. Statist. Probab. Lett. 12, No.5, 393-403.

[64] Roussas, G.G. and Tran, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions, Annals of Statist. 20 (1), 98-120.

[65] Roussas, G.G. (1992). Ecxact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation, *J. Nonparametr. Stat.* 1, 171-195.

[66] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. Stat. Probab . Lett. 7, No.5, 407-412.

[67] Sarda, P., Vieu, P. (2000). Kernel regression. In : M. Schimek (ed.) Smoothing and regression ; Approaches, Computation, and Application. Wiley Series in Pprobability and Statistics, *Wiley, New York.*

[68] Schwartz, S.C. (1967). Estimation of a probability density by an orthogonal series, Ann. Math. Statist. 38, 1261-1265.

[69] Shorack, G.R. and Wellner, J.A. (1986). Empirical processes with applications to statistics. *Wiley, New-York.*

[70] Takahata, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, Bull. Tokyo Gakugei Univ. Nat. Sci. Ser. IV, 11-32.

[71] Tran, L.T (1989). Recursive density estimation under dependence. IEEE Trans. Inform. Theory 35 (5), 1103-1008.

[72] Tran, L. T. (1994). Density estimation for time series by histograms, J. Statist. Plann. Inference , 40 (1), 61-79.

[73] Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. J. Multivariate Anal. 39 (2), 324-347. Watson, G. S. (1964). Smooth regression analysis. Sankhya Ser. A. 26, 359-372.

[74] Walk, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Ann. Inst. Statist. Math.*, 53 (4), 691–707.

[75] Watson, G.S. (1964). Smooth regression analysis. Shakhya Ser. A 26, 359-372.

[76] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, Pages 316-327.

[77] Wertz, W. (1985), Sequential and recursive estimators of the probability density. *Statistics* **16**, Pages 277-295.

[78] Wolverton, C.T. and Wagner, T.J. (1969), Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory* **15**, Pages 258-265.

[79] Yamato, H. (1971), Sequential estimation of a continuous probability density function and mode. *Bull. Math. Satist.* **14**, Pages 1-12.

[80] Youndjé, E., (1993), Estimation non paramétrique de la densité conditionnelle par la méthode du noyau. Thèse de Doctarat, Université de Rouen.

# Chapitre 2

# Données incomplètes

L'un des problèmes important dans l'analyse de survie est la modélisation de l'influence de des certains nombre de facteurs sur la fonction de survie. En effet, dans la plus part des situations il y a des individus qu'on n'arrive pas à observer sont événement d'interêt. Un exemple souvent considéré est lorsque on s'intéresse de savoir comment le temps de survie chez les fumeurs. Plus précisément, on étude la relation entre la variable d'intérêt et l'âge auquel la personne a commencé à fumer. Parfois on se trouve avec un fumeur qui décède avant le début de l'étude est systématiquement exclu de l'échantillon . Par contre, on peut tomber avec des cas alternatifs où un fumeur non reste vivant même après la fin de l'étude. Dans les deux situations nous avons un manque d'information et un certain nombre d'observations incomplètes. On trouvera dans la littérature plusieurs de modèles adaptés permettant de modéliser ces phénomènes .
La durée de vie est une variable aléatoire $Y$ positive. Ces types des variables sont observés dans plusieurs domaines tels que l'économie, médecine, biologie, épidémiologie, astronomie , . . . . En pratique, la notion de durée de vie est quantifiée par le temps écoulé jusqu'a la survenue d'un évènement ( par exemple, première panne pour une machine, décès ou rechute pour un malade, embauche pour un chômeur,. . . ). Lorsque les durées de vie sont observées dans leur totalité, on parle de données complètes, dans le cas contraire les données sont incomplètes et nécessitent un traitement statistique particulier. Dans cette thèse nous nous intéressons où cas des données incomplètes.

Les premiers résultats sur l'analyse des données de survie à (1693) par Halley. Ce dernier, a étudié des relevés d'état civil de Londres. Typiquement, le développement de l'analyse de survie a connu un grand sucées après la seconde guerre mondiale. La plus part des études sont basées sur des modèles paramétriques avec des lois exponentielles ou de Weibull. Dans ce dernier temps, les applications sur les données médicales ont eu le privilège en analyse de survie. A ce sujet l'aspect non paramétrique est le plus utilisés. Dans ce contexte, Kaplan-Meier (1958) est parmi les premiers qui ont développés des estimateurs non-paramétriques pour la fonction de survie. Nous renvoyons à Cox (1972) pour le cas semi-paramétriques. Le modèle de régression est l'un des principaux thèmes de l'analyse de survie. En particulier si, on observe une variable $T$ censurée à droite c'est à dire que l'observation prend fin avant que l'évènement ne survienne, alors il est difficile de supposer que le temps d'apparition d'un évènement est une réalisation d'un processus aléatoire d'un associée à une distribution particulière ce que justifier l'importance de la modélisation non paramétrique . Nous referons aux (Bailey (1979), Kalbfleich et Prentice (1980), Cox et Oakes (1984), Moreau (1984), Bretagnolle et Huber (1988), Gill (1980), Andersen et Gill (1982), Harrington et Fleming (1982), Cross et Huber (1987) pour une étude bibliographique de base le traitement statistique et/ou stochastique de données incomplète. Il est à noter que la principale source de difficulté dans l'analyse des durées de vie, est le fait que la variable d'intérêt n'est pas complètement observée.

## 2.1 Données censurées

Dans cette partie nous rappelons en détaille les définitions de base des données censurées ainsi que des exemples qu'on trouve souvent dans la littérature de base. En particulier en Bio-statistique, lorsque, on s'intéresse au cas à la durée de vie d'un groupe d'individus. Par exemple, on considère des patients atteints d'un cancer de la peau qui ont tous dû subir une opération chirurgicale . Dans ce cas on s'intéresse à leur durée de vie à partir de cette opération. Cette étude contient également une variable explicative donnant des informations sur le patient comme par exemple : le sexe ou l'âge du patient ainsi que le moment de l'opération. Comme ces individus sont arrivés à divers instants certains étaient encore en vie à la fin de l'étude et on ne peut donc pas connaître leur durée de vie exacte. De même, un certain nombre de patients sont décédés par des raisons indépendante de la maladie ce qui nous empêche de connaître leur durée de vie réelle. Dans chacun de ces cas, on ne connaît donc que la date minimum entre leur durée de vie effective et, dans le premier cas, la date de fin d'étude, dans le deuxième cas, la date de décès causée par un évènement indépendant de la tumeur.
Dans cet exemple on n'observe que le minimum entre la variable d'intérêt $Y$ et une autre variable $C$ dite variable de censure. Ceci signifie que l'on observe seulement la variable $T = Y \wedge C$. On utilise également l'indicateur de censure, $\delta = \mathbb{I}_{Y \leq C}$, afin de s'informer si notre variable observée correspond bien à la variable d'intérêt ou à la variable de censure. Ces observations sont liées à une variable explicative $X$ qui fournit des informations sur $Y$ .

Il est clair que ces modèles des données incomplètes requièrent l'utilisation de techniques adaptées pour prendre en considération les observations censurées sans perdre trop d'information sur $Y$ . En pratique, on distingue trois types de censures **Censure à droite**
Une durée de survie est dite censurée à droite si l'individu n'a pas connu l'évènement d'intérêt à sa dernière visite. La censure à droite est l'exemple le plus fréquent d'observation incomplète en analyse de survie, et a largement été décrit dans la littérature (Anderson, Borgan et Keiding (1993)). Formellement, la durée de survie d'un évènement est définie par le couple $(T; C)$ où

$$Y = \inf(T, , C)$$

et

$$\delta = \begin{cases} 1 & \text{si} \quad T \leq C \\ 0 & \text{si} \quad T > C \end{cases},$$

Avec la durée de vie $T$ et le temps de censure supposés indépendants. C'est-à-dire, on observe le véritable temps de survie que s'il est inférieur à $C$. Dans ce cas la donnée n'est pas censurée et $\delta = 1$. Si $\delta = 0$, la donnée est dite censurée àà droite : au lieu d'observer $T$, on observe une valeur $C$ avec pour seule information le fait que $T$ soit supérieur à $C$. C'est la censure de type 1.

**Censure de type 2 : attente.**
On décide d'observer les durées de survie des $n$ patients jusqu'à ce que $r$ d'entre eux soient décédés et d'arrêter l'étude à ce moment-là. La date de fin d'expérience devient alors aléatoire, le nombre d'évènement étant, quand à lui, non aléatoire. Ce modèle est souvent utilisé dans les études de fiabilité.

**Censure de type 3 : aléatoire.**
C'est typiquement ce modèle qui est utilisé pour les essais thérapeutiques. Dans ce type d'expérience, la date d'inclusion du patient dans l'étude est fixé, mais la date de fin d'observation est inconnue (celle-ci correspond, par exemple, à la durée d'hospitalisation du patient). Ici, le nombre d'événement observés et la durée totale de l'expérience sont aléatoires.

**Exemple 2.1.1** *On s'intéresse au temps de survie de personnes atteintes d'une maladie. On fixe le temps d'étude et à la fin de ce temps certaines personnes sont encore vivantes. Pour ces personnes, tout ce que l'on sait est que leur temps de survie dépasse le temps observé, ce sont des données à droite de type 1.*

**Censure à gauche**
Une durée de survie est dite censurée à gauche si l'individu a déjà connu l'évènement d'intérêt avant l'entrée dans l'étude. Formellement, la durée de survie pour un individu est définie par le couple $(Y; \delta)$ où

$$Y = \max(T, C)$$

et

$$\delta = \begin{cases} 1 & \text{si} \quad T > C \\ 0 & \text{si} \quad T \leq C \end{cases},$$

Avec la durée de vie et le temps de censure $C$ supposés indépendants. Si $\delta = 1$, le sujet subit l'évènement et est observé. Si $\delta = 0$, le sujet est dit censuré à gauche : au lieu d'observer $T$, on observe une valeur $C$ avec pour seule information le fait que $T$ soit inférieur à $C$.

**Exemple 2.1.2** *Sur le même (exemple 2.1.1) que précédemment, on ne peut pas toujours savoir le moment exact du déclenchement de la maladie, pour certaines personnes, on sait seulement que leur âge est inférieur à leur âge au moment de l'étude. Ces données sont censurées à gauche.*

## 2.2 Estimation de la fonction de survie

Dans cette section nous rappelons les estimations de base de la fonction de survie. En effet, soit $(\Omega, A, \mathbb{P})$ un espace probabilisé. Soit $T_1, \cdots, T_n, \cdots$ une suite de variable aléatoire (v.a) positives indépendantes et identiquement distribuées (i.i.d) de fonction de répartition (f.d.r.) commune $F$. Soit $C_1, \cdots, C_n, \cdots$ 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 $T_i$. Soit $(Y_i, \delta_i)_{i=1,\cdots,n}$ l'échantillon réellement observé.
Posons $F(t) = \mathbb{P}(T > t)$, $\bar{G}(t) = \mathbb{P}(C > t)$, alors les $Y_i$ sont de f.d.r. $L$ donnée par $\bar{L} = \mathbb{P}(Y > t) = \bar{F}(t)\bar{G}(t)$. Posons aussi $\tau_H = \sup\{t/\bar{H}(t) > 0\}$ où $H$ est une f.d.r.
Définissons un processus ponctuel "d'événement" par

$$N_i(t) = \mathbb{P}(Y_i \leq t, \delta_i = 1).$$

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

$$\bar{Y}(t) = \sum_{i=1}^{n} Y_i(t),$$

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

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

### 2.2.1 Estimateur de Kaplan-Meier de la fonction de survie

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 $t_1 < t$

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

pour $t_2 < t_1$ on a $\bar{F}(t_1) = \mathbb{P}(Y > t_1/Y > t_2)\bar{F}(t_1)$ 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

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

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

$$\hat{q}_i = \frac{M_i}{R_i} = \frac{\textit{nombre de morts } Y_i}{\textit{nombre de sujets risque}}$$

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 $Y_i$ et donc $M_i = 1$, si $\delta_i = 0$ c'est qu'il ya une censure en $Y_i$

$$\hat{p_i} = \begin{cases} 1 - \frac{1}{R_i} & \textit{en cas de mort en } Y_{(i)}, \\ 1 & \textit{en cas de censure.} \end{cases}$$

Il est clair que $R_i = n - i + 1$. On On obtient finalement l'EKM pour la fonction de survie de la variable durée de vie $T$

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

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

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

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

**Remarque 2.2.1**

$$\bar{F}_n(t) =: \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{\delta_{(i)}}{n-i+1}\right)^{\mathbb{I}_{\{Y_{(i)} \leq t\}}} & si \quad t < Y_{(n)} \\ 0 & si \quad t \geq Y_{(n)} \end{cases},$$

*et*

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

### Propriétés asymptotiques de l'estimateur de Kaplan-Meier

L'E.K.M. possède beaucoup de propriétés analogues à celles de la fonction de répartition empirique. Les propriétés asymptotiques de l'E.K.M. ont été étudiées par plusieurs auteurs (voir e.g. Peterson (1977), Andersen et al. (1993), Shorack et Wellner (1986)). Nous commençons par ses propriétés de convergence.

**Théoreme 2.2.1** *Si la survie $T$, de fonction de répartition $F$ et la censure $C$, de fonction de répartition $G$ sont indépendantes, alors*

$$\sup_{0 \leq t \leq \tau_H} |F_n(t) - F(t)| \longrightarrow 0 \; p.s. \; quand \; n \longrightarrow \infty.$$

**Preuve.** (voir Shorack et Wellner 1986, p.304 ). ■
Définissons les processus suivants, pour tout $0 \leq t \leq \tau_H$

$$Z_{(t)} = \sqrt{n}\,(F_n(t) - F(t))$$

Nous avons les résultats de normalité asymptotique suivants :

**Théoreme 2.2.2** (Breslow et Crowley, 1974)
*Si la survie $T$, de fonction de répartition $F$ et la censure $C$, de fonction de répartition $G$ sont indépendantes, alors,*

$$Z_n \xrightarrow{\mathcal{L}} Z \text{ quand } n \longrightarrow \infty$$

*où $Z$ est un processus gaussien centré de fonction de covariance*

$$\sigma(s,t) = Cov(Z(s), Z(t)) = (1 - F(s))(1 - F(t)) \int_0^{s\wedge t} \frac{dH(x)}{(1 - F(x))^2}$$

La première représentation forte de $F_n - F$ par une moyenne de v.a. iid a été obtenue par Burke et al. (1981, 1986) avec un reste d'ordre $O(n^{-1/2}\log^2 n)$ . 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 terme négligeable d'ordre $O\left((n^{-1}\log n)^{3/4}\right)$. Ce taux a été ensuite amélioré à $O(n^{-1}\log n)$ par Lo et al. (1989).

# Chapitre 3

# Nonparametric conditional density estimation for censored data based on a recursive kernel

Ce chapitre à fait l'objet d'une publication à la revue Electronic Journal of Statistics.

## On the strong iniform consistency of the conditional density for censored data based on a recursive kernel[1]

**Sihem Semmar and Salah Khardani**

*École Nationale d'Ingénieurs de Monastir (ENIM),Laboratoire de Physique-Mathématique, Fonctions Spéciales et Applications (P.M.F.S.A) Hammam-Sousse*
*Laboratoire de Statistique et Processus Stochastique , Université de Sidi Bel Abbès, BP 89 Sidi Bel Abbès 22000. Algeria*

**abstract** Let $(T_n)_{n\geq 1}$ be a sequence random variables (rvs) of interest distributed as $T$. In censorship models the rv $T$ is subject to random censoring by another rv $C$. We consider the problem of estimating its conditional density function ,given a vector of covariates X. and establish its almost sure convergence with rate under an $\alpha$-mixing condition.

**AMS 2000 subject classificcations :** Primary 62G05, 62G07, 62G08, 62G20, 62H12.
**Keywords and phrases :** Asymptotic normality, censored data, condi- tional density, kernel estimator, recursive estimation, Kaplan-Meier esti- mator, uniform almost sure convergence.

1. Article écrit en collaboration avec Khardani Salah paru en EJS (2014)

## 3.1 Introduction

Studying the relationship between a response variable and a explanatory variable is one of the most important statistical analysis. Usually this relationship is modeled with the regression function. However, it is well known, this nonparametric model is not efficient in some pathological situations. For instance, the multi-modal densities case, the case where the expected value might be nowhere near a mode or for situations in which confidence intervals are preferred to point estimates. In all these case the conditional density is a pertinent model to explore this relationship. The main purpose of this paper is to study this nonparametric model when the response variable is subject to censoring, by using a kernel recursive estimation method.

Noting that the nonparametric modeling of censored data is intensively discussed in the recent statistical literature. It dates back to Beran (1981), who introduced a class of nonparametric regression estimators for the conditional survival function in the presence of right-censoring. Dabrowska (1987,1989) studied the asymptotic properties of the distribution and quantiles functions estimators. Kohler *et al.* (2002) gave a simpler proof in the randomly right-censoring case for kernel, nearest neighbor, least squares and penalized least squares estimates. Further results was obtained by Khardani *et al.* (2010, 2012). Concerning the nonparametric conditional model, we cite for the conditional model in both (iid and mixing case) and for conditional quantiles function. In this vast variety of papers, the authors use the Nadaraya-Watson techniques as estimation method which is a particular case of the recursive kernel estimator considered in this paper. Moreover, this last has various advantages over the kernel method. We deal with recursive kernel estimators where, by recursive we mean that the estimator calculated from the first $n$ observations, say $f_{n+1}$, is only a function of $f_n$ and the $(n+1)^{-th}$ observation. As is well known, the recursive property is particularly interesting when the sample data are obtained by mean of some observational mechanism that allows an increase in the sample size over time. This situation is usual in many control and supervision problems and, above all, in time series analysis. In the above cases, the recursive estimates allow us to update the estimations as additional observations are obtained, unlike non-recursive methods where estimates must be completely recalculated when each additional item of data received. From a practical point of view, this iterative procedure provides an important saving in computational time and memory, since the updating of the estimates is independent of the previous sample size. It is not the case for the basic kernel estimator which has to be computed again on the whole sample. Recursive estimators show good theoretical properties, from the point of view of mean square error (small variance) and almost sure convergence.

The first recursive modifications of the Nadaraya-Watson estimate have been introduced by Ahmad and Lin (1976) and Devroye and Wagner (1980) say (AL) and (DW). In complete data, Kernel recursive estimators have been introduced by Wolvertone and Wagner (1969) and Yamato (1972). Next Davies (1973), Deheuvels (1974), Wertz (1985) and Roussas (1992) have independently studied the rates of the almost sure convergence of particular recursive density estimates.

The law of the iterated logarithm of the recursive density estimator was established by

Wegman and Davies (1979) and Roussas and Tran (1992). For other works on recursive density estimation, the reader is referred to the papers of (A L) , and Carroll (1976). Recently, in a context of $\alpha-$mixing processes, Amiri (2012) gave the exact asymptotic quadratic error of a general family of kernel estimator, whose (AL) and (DW) are particular cases. The asymptotic normality of the same family is obtained by Amiri (2013).
The recursive regression estimator for identically independent distributed (i.i.d.) random variables has been studied by many authors among whom we quote (A L) , Krzyzak(1992) and Walk (2001) for a nonparametric approach and Nguyen and Saracco(2010) for semi-parametric models. In the strong mixing case, Roussas(1990) derived the uniform almost sure convergence for (DW), while Roussas and Tran (1992) showed its asymptotic normality. Masry and Fan(1997) studied some properties of local polynomial regression for dependent data.
Despite this great importance the recursive kernel estimation of censored nonparametric has not yet been fully explored. The present work is the first contribution that consider a recursive estimate in censored data. The main aim of this contribution is to study the asymptotic properties of the recursive kernel estimator of the conditional density and its derivatives, under random right censoring. Specifically, the asymptotic properties stated are the strong convergence and the asymptotic normality of these estimators. The paper is organized as follows. We present our model in Section 2. In Section 3 we introduce notations, assumptions and we state the main results. Finally, the proofs of the main results are relegated to Section 4 with some auxiliary results with their proofs.

## 3.2 Presentation of estimates

Consider $n$ pairs of independent random variables $(X_i, T_i)$ for $i = 1, \ldots, n$ that we assume drawn from the pair $(X, T)$ which is valued in $\mathrm{R}^d \times \mathrm{R}$. In this paper we consider the problem of nonparametric estimation of the conditional density of $Y$ given $X = x$ when the response variable $Y_i$ are rightly censored. Furthermore, we denote by $(C_i)_{i=1,\ldots n}$ the censoring random variables which are supposed independent and identically distributed with a common unknown continuous distribution function $G$.Thus, we construct our estimators by the observed variables $(X_i, Y_i, \delta_i)_{i=1,\ldots,n}$, where $Y_i = T_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$ , where $\mathbb{1}_A$ denotes the indicator function of the set $A$.
To follow the convention in biomedical studies, we assume that $(C_i)_{1 \leq i \leq n}$ and $(T_i, X_i)_{1 \leq i \leq n}$ are independent; this condition is plausible whenever the censoring is independent of the modality of the patients.
The cumulative distribution function $G$, of the censoring random variables, is estimated by Kaplan Meier (1958) estimator defined as follows

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^{n} \left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{\mathbb{1}_{\{Y_{(i)} \leq t\}}} & \text{if} \quad t < Y_{(n)} \\ 0 & \text{otherwise} \end{cases},$$

which is known to be uniformly convergent to $\bar{G}$.

Given i.i.d. observations $(X_1, Y_1, \delta_1), \dots (X_n, Y_n, \delta_n)$ of $(X, Y, \delta)$, the kernel estimate of the conditional density $\phi(t|x)$ denoted $\bar{\phi}_n(t|x)$, is defined by

$$\forall x \in \mathbb{R}^d \text{ and } \forall y \in \mathbb{R} \qquad \bar{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_n^{-1} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) L\left(\frac{t - Y_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right)},$$

where $K$, $L$ are a kernels and $h_n$ is a sequence of positive real numbers. Note that this last estimator has been recently used by Khardani (2010).
A recursive version of the previous kernel estimator is defined by

$$\widehat{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_i^{-(d+1)} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_i}\right) L\left(\frac{t - Y_i}{h_i}\right)}{\sum_{i=1}^{n} h_i^{-d} K\left(\frac{x - X_i}{h_i}\right)} =: \frac{\hat{g}_n(x,t)}{\ell_n(x)}$$

where

$$\hat{g}_n(x,t) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+1}} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_i}\right) L\left(\frac{t - Y_i}{h_i}\right), \tag{3.1}$$

and

$$\ell_n(x) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right), \ \forall x \in \mathrm{IR}^d.$$

**Remark 3.2.1** *the Kaplan-Meir estimator is not recursive and the use of such estimator can slightly penalizes the efficiency of our estimator in term of computational time.*

## 3.3 Assumptions and main results

Throughout the paper, we put $h_n^- = \inf_{i=1\dots n} h_i$, $h_n^+ = \sup_{i=1\dots n} h_i$ and all along the paper, when no confusion is possible, we denote by $M$ and/or $M'$ any generic positive constant. Further, we will denote by $F(\cdot)$ (resp. $G(\cdot)$) the distribution function of $T$ (resp. of $C$) and by $\tau_F$ (resp. $\tau_G$) the upper endpoints of the survival function $\bar{F}$ (resp. of $\bar{G}$). In the following we assume that $\tau_F < \infty$, $\bar{G}(\tau_F) > 0$ and $C$ is independent to $(X, T)$. We also assume that there exist a compact set $\mathcal{C} \subset \mathcal{C}_0 = \{x \in \mathbb{R}^d \ \ \ell(x) > 0\}$ where $\ell$ is the density of the explicative variable $X$, and $\Omega$ be a compact set such that $\Omega \subset (-\infty, \tau]$ where $\tau < \tau_F \wedge \tau_G$.
We introduce the following assumptions :
***Assumption A1.***
The kernels $K$ and $L$ are Lipschitz continuous functions and compactly supported satisfy .

$$\int_{\mathbf{R^d}} u_l K(u) du = 0 \ \text{ for } l = 1, \dots, d \text{ with } u = (u_1, \dots, u_d)^T \quad \text{and} \quad \int_{\mathbf{R}} v L(v) dv = 0$$

***Assumption A2.***

**(i)** The marginal density $\ell(\cdot)$ is twice differentiable and satisfies a Lipschitz condition. Furthermore $\ell(x) > \Gamma$ for all $x \in \mathcal{C}$ and $\Gamma > 0$. Where $\mathcal{C}$ is a compact set of $\mathbb{R}$.

**(ii)** The joint density $g(\cdot,\cdot)$ of $(X,T)$ is bounded function twice differentiable.

**Remark 3.3.1** *Assumptions **A1** and **A2** are usually used in non censoring kernel estimation method. The independence assumption between $(C_i)_i$ and $(X_i,T_i)_i$, may seem to be strong and one can think of replacing it by a classical conditional independence assumption between $(C_i)_i$ and $(T_i)_i$ given $(X_i)_i$. However considering the latter demands an a priori work of deriving the rate of convergence of the censoring variable's conditional law (see Deheuvels and Einmahl (2000)). Moreover our framework is classical and was considered by Carbonez et al. (1995) and kohler et al.(2002), among others.*

### 3.3.1 Uniform strong consistency results with rate of convergence

In order to give the rate of the uniform almost sure convergence of our estimate we need the following additional assumptions :
***Assumption C.***

**(i)** The sequences $h_n^+$ and $h_n^-$ satisfy $\lim_{n\to\infty} h_n^+ + \frac{\log n}{nh_n^{-d+1}} = 0$ as $n \to \infty$.

**(ii)** $\lim_{n\to\infty} n^\beta h_n^- = \infty$ for some $\beta > 0$.

**Théoreme 3.3.2** .
*Under Assumptions* **A1**, **A2** *and* **C** *we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x) - \phi(t|x)\right| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right), h_n^{+^2}\right)\right\} \quad a.s. \quad as \quad n \to \infty \tag{3.2}$$

**Remark 3.3.3** *Observe that, although the expression of the convergence rate is not classic in nonparametric statistic data analysis, this convergence rate is identifiable to the usual rate in the kernel method case where, for all i, we have $h_i = h_n = h_n^- = h_n^+$.*

***Proof of Theorem 3.3.2***

Set

$$\widetilde{g}_n(x,t) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^{(d+1)}}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)$$

with $K_i(x) = K\left(\frac{x-X_i}{h_i}\right)$, $L_i(t) = L\left(\frac{t-Y_i}{h_i}\right)$.
Now, the proof of this Theorem is based on the following decomposition

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x)-\phi(t|x)\right| &\leq \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\frac{\widehat{g}_n(x,t)}{\ell_n(x)}-\frac{\tilde{g}_n(x,t)}{\ell_n(x)}\right|+\left|\frac{\tilde{g}_n(x,t)}{\ell_n(x)}-\frac{\mathbb{E}\tilde{g}_n(x,t)}{\ell_n(x)}\right| \\
&+ \left|\frac{\mathbb{E}\tilde{g}_n(x,t)}{\ell_n(x)}-\frac{g(x,t)}{\ell_n(x)}\right|+\left|\frac{g(x,t)}{\ell_n(x)}-\frac{g(x,t)}{\ell(x)}\right| \\
&\leq \frac{1}{\inf_{x\in\mathcal{C}}\ell_n(x)}\left\{\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\widehat{g}_n(x,t)-\tilde{g}_n(x,t)|+\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-g(x,t)|\right. \\
&+ \left.+\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\phi(t|x)|\sup_{x\in\mathcal{C}}|\ell(x)-\ell_n(x)|\right\} \qquad (3.3)
\end{aligned}$$

So, the proof of this Theorem is a direct consequence of Lemmas 4.4.2– 4.4.4.

**Lemma 3.3.4** *Under Assumption* **C**, **A1** *and* **A2** *(ii), we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-g(x,t)|=O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right),h_n^{+^2}\right)\right\} \quad a.s. \quad as \quad n\to\infty.$$

**Lemma 3.3.5** . *Under Assumption* **C**, **A1** *and* **A2** *(i), we have*

$$\sup_{x\in\mathcal{C}}|\ell(x)-\ell_n(x)|=O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{-d}}}\right),h_n^{+^2}\right)\right\} \quad a.s. \quad as \quad n\to\infty.$$

**Lemma 3.3.6** . *Under Assumption* **C**, **A1** *and* **A2** *(ii), we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\widehat{g}_n(x,t)-\widetilde{g}_n(x,t)|=O\left\{\left(\sqrt{\frac{\log\log n}{n}}\right)\right\} \quad a.s. \quad as \quad n\to\infty.$$

### 3.3.2 Asymptotic normality

Now, we study the asymptotic normality of our estimate. To do that, we replace condition **C** by the following assumption :
***Assumption N.***

- **(i)** $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n\left(\frac{h_n}{h_i}\right)^k=\theta_k$.
- **(ii)** $h_n^{(d+1)}\log\log n=o(1)$, $\lim_{n\to\infty}nh_n^{(d+1)}h_n^{+^4}=0$ and $\lim_{n\to\infty}nh_n^{(d+1)}=\infty$

**Théoreme 3.3.7** .
*Under Assumptions **A1**,**A2** and* **N**, *we have, for any* $(x,t)\in\mathcal{A}$,

$$\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x)-\phi(t|x)\right)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\sigma^2(x,t)\right)$$

*where* $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution,*

$$\sigma^2(x,t) = \theta_{d+1}\frac{\phi(t|x)}{\ell(x)\bar{G}(t)}\int_{\mathbb{R}^d}\int_{\mathbb{R}} K^2(z)L^2(y)dzdy$$

*and* $\mathcal{A} = \{(x,t) \quad \sigma^2(x,t) \neq 0\}$.

**Corollary 3.3.8** *Based on* $\overline{G}_n(\cdot), \hat{\phi}_n(\cdot|x)$ *and* $\ell_n(x)$ *we easily get a plug-in estimator* $\hat{\sigma}_n^2(x,t)$ *for* $\sigma^2(x,t)$ *which, under the assumptions of Theorem 4.4.6, gives a confidence interval of asymptotic level* $1-\alpha$ *for* $\phi(t|x)$

$$\left[\hat{\phi}_n(t|x) - \frac{u_{1-\alpha/2}\hat{\sigma}_n(x,t)}{\sqrt{nh^{d+1}}}, \hat{\phi}_n(t|x) + \frac{u_{1-\alpha/2}\hat{\sigma}_n(x,t)}{\sqrt{nh^{d+1}}}\right]$$

*where* $u_{1-\alpha/2}$ *denotes the* $(1-\alpha/2)$*-quantile of the standard normal distribution.*

***Proof of Theorem 4.4.6***

It is clear that

$$\begin{aligned}
\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x) - \phi(t|x)\right) &= \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right] \\
&+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] \\
&+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)\right] \\
&+ \sqrt{nh_n^{(d+1)}}\frac{g(x,t)}{\ell_n(x)\ell(x)}\left[\ell(x) - \ell_n(x)\right].
\end{aligned} \tag{3.4}$$

Thus, The proof of Theorem 4.4.6 can be deduced directly from the following Lemmas

**Lemma 3.3.9** *Under the Hypotheses of Theorem 4.4.6, we have*

$$\sqrt{nh_n^{(d+1)}}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right] \to 0 \quad a.s. \quad as \quad n\to\infty, \tag{3.5}$$

$$\sqrt{nh_n^{(d+1)}}\left[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)\right] \to 0 \qquad as \quad n\to\infty \tag{3.6}$$

*and*

$$\sqrt{nh_n^{(d+1)}}\left(\ell_n(x) - \ell(x)\right) \to 0 \quad in\ probability \quad as \quad n\to\infty. \tag{3.7}$$

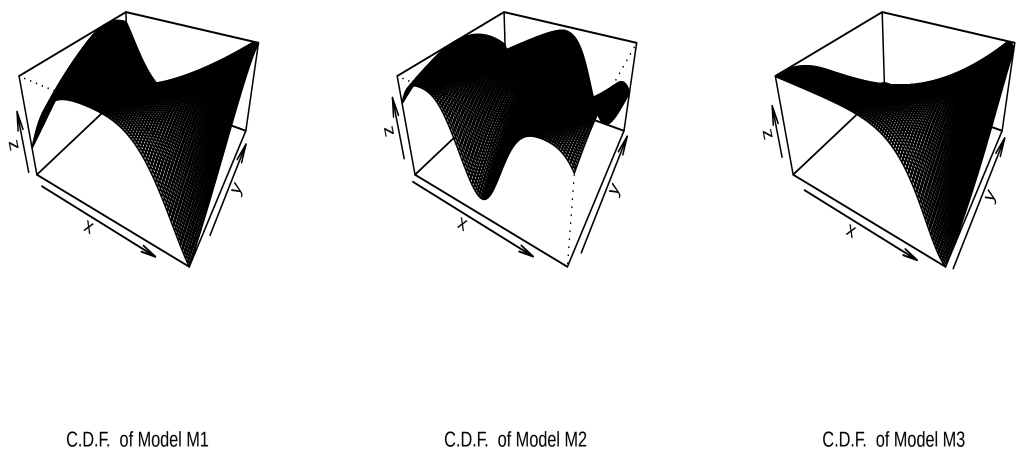


FIGURE 3.1 – The C.D.f.

**Lemma 3.3.10** *Under Assumption* **A1**, **A2** *and* **N**(*i*), *we have*

$$\left(nh_n^{(d+1)}\right)^{\frac{1}{2}} \left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma'^2\right)$$

*where* $\sigma'^2(x,t) = \theta \dfrac{g(x,t)}{\bar{G}(t)} \displaystyle\int_{\mathbb{R}^d} \int_{\mathbb{R}} K_i^2(z) L_i^2(y) dz dy$

## 3.4 Numerical study

In this short section we compare the finite-sample performance of the recursive kernel method and the classical kernel via a Monte Carlo study. For this comparison study, we consider the same models used in Lemdani et al. (2013) that is

$$\begin{array}{ccc} M1 & Y = X^2 + 1 + \epsilon & \text{parabolic case,} \\ M2 & Y = sin(1.5X) + \epsilon & \text{sinus case,} \\ M3 & Y = exp(X - 0.2) + \epsilon & \text{exponential case} \end{array}$$

where the random variables $X$ and $\epsilon$ are i.i.d. and follow respectively the the normal distribution $N(0,1)$. and $N(0,\sigma)$.
It is clear that the conditional density expression is closely related to the distribution of $\epsilon$. Thus, the conditional densities are respectively

In order to control the effect of the censoring in the efficiency of both estimators we variate the the percentage of censoring for each models by considering a various censoring distributions. Precisely, we generate the the censoring variables $C$ by an exponential distribution $\mathcal{C}(\lambda_1)$ shifted by $\lambda_2$ (for the exponential model), by a normal distribution $N(0, \sigma_1)$ (for sinus case) and by $N(0, \sigma_2)$ (for parabolic case). Thus, the behavior of both estimators is evaluated over a several parameters, such as the sample size $n$, the percentage of censoring $\tau$ controlled by $(\lambda_1, \lambda_2, \sigma_1, \sigma_2)$, the dimension of the regressors $d$ and the type of model $M$.. For sake of shortness, we consider the unidimensional case, we fixe the sample size $n = 200$, we took $\sigma = 0.2$, we consider three censoring type ( $\tau = 10$, $\tau = 40$ and $\tau = 70$ ). The test of the performance of both estimators is described by the following averaged squared errors

$$MSE(KERNEL) = \frac{1}{n}\sum_{i=1}^{n}(\bar{\phi}_n(Y_i|X_i) - \phi(Y_i|X_i))^2$$

and recursive

$$MSE(RECURSIVE) = \frac{1}{n}\sum_{i=1}^{n}(\widehat{\phi}_n(Y_i|X_i) - \phi(Y_i|X_i))^2$$

Now, for our practical study, we use the Gaussian kernel and we consider the well-known smoothing parameter defined by $h_{n,i} = \sigma_n^2 i^{-1/5}$ where

$$\sigma_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \quad \text{and} \quad \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

The obtained results are given in Table 1. It is clear from Table 1 that the recursive method is slightly better than the classical kernel method. However, the main advantage of the recursive method is that considerably faster than the classical one for the three models. In particular, it reduces sensibly the computation time in function of the sample size and the kind of models. Overall, both methods give a satisfactory level of accuracy and the latter is strongly dependent to the censoring rate.

| Model | $\tau$ | MSE(KERNEL) | MSE(RECURSIVE) |
|---|---|---|---|
| M1 | 10 | 0.41 | 0.22 |
| | 40 | 0.64 | 0.55 |
| | 70 | 1.74 | 1.97 |
| M2 | 10 | 0.59 | 0.36 |
| | 40 | 0.33 | 0.30 |
| | 70 | 1.80 | 1.84 |
| M3 | 10 | 0.79 | 0.29 |
| | 40 | 1.32 | 1.18 |
| | 70 | 2.17 | 2.65 |

Table 1 $MSE$-Results.

## 3.5 Proofs of the intermediates results

**Proof of Lemma 4.4.2 :** We write

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-g(x,t)| \le \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-\mathbb{E}\tilde{g}_n(x,t)| + \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)|.$$

For the quantity $\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)|$, we use the fact that, for all measurable function $\varphi$ and for all $i=1,\dots n$.

$$\mathbb{1}_{\{T_1\le C_1\}}\varphi(Y_1) = \mathbb{1}_{\{T_1\le C_1\}}\varphi(T_1).$$

Then,

$$\begin{aligned}
\mathbb{E}\tilde{g}_n(x,t) &= n^{-1}\sum_{i=1}^n \frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_i}\right)\delta_i\bar{G}^{-1}(T_i)L\left(\frac{t-T_1}{h_i}\right)\right\}\\
&= n^{-1}\sum_{i=1}^n \frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_i}\right)\bar{G}^{-1}(T_i)L\left(\frac{t-T_1}{h_i}\right)\mathbb{E}\left[\mathbb{1}_{\{T_i\le C_i\}}|X_i,T_i\right]\right\}\\
&= n^{-1}\sum_{i=1}^n \frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_1}\right)L\left(\frac{t-T_1}{h_1}\right)\right\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)| &\le n^{-1}\sum_{i=1}^n\left|\int_{\mathbf{R^d}}\int_{\mathbf{R}}K(u)L(v)[g(x-h_iu,t-h_iv)-g(x,t)]dudv\right|\\
&\le Mn^{-1}\sum_{i=1}^n h_i^2 \le Mh_n^{+^2}.
\end{aligned}$$

Therefore,

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)| = O(h_n^{+^2}).$$

Now, concerning the quantity $\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-\mathbb{E}\tilde{g}_n(x,t)|$ we use the compactness property of the sets $\mathcal{C}$ and $\Omega$ to write that, for some $(x_k)_{1\le k\le\lambda_n}$ and $(t_j)_{1\le j\le\kappa_n}$,

$$\mathcal{C}\subset\bigcup_{k=1}^{\lambda_n}\mathcal{B}(x_k,a_n) \qquad \text{and} \qquad \Omega\subset\bigcup_{j=1}^{\kappa_n}B(t_j,b_n)$$

where $\lambda_n\sim a_n^{-d}$ and $\kappa_n\sim b_n^{-1}$ with $a_n=b_n=n^{-(d+1)\beta-1/2}$.
Now, for any $x\in\mathcal{C}$ and $t\in\Omega$, we set by $\tilde{k}(x)=\arg\min_k\|x_k-x\|$ and $\tilde{j}(t)=\arg\min_j|t-t_j|$
Then, for any $(x,t)\in\mathcal{C}\times\Omega$, we can write

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-\mathbb{E}\tilde{g}_n(x,t)| &\le \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\tilde{g}_n(x,t)-\tilde{g}_n(x,t_{\tilde{j}})\right| + \sup_{x\in\mathcal{S}}\sup_{t\in\Omega}\left|\mathbb{E}\tilde{g}_n(x,t_{\tilde{j}})-\mathbb{E}\tilde{g}_n(x,t)\right|\\
&+ \max_j\sup_{x\in\mathcal{C}}\left|\tilde{g}_n(x,t_j)-\tilde{g}_n(x_{\tilde{k}},t_j)\right| + \max_j\sup_{x\in\mathcal{C}}\left|\mathbb{E}\tilde{g}_n(x_{\tilde{k}},t_j)-\mathbb{E}\tilde{g}_n(x,t_j)\right|\\
&+ \max_k\max_j|\tilde{g}_n(x_k,t_j)-\mathbb{E}\tilde{g}_n(x_k,t_j)|\\
&=: \mathcal{T}_{1,n}+\mathcal{T}_{2,n}+\mathcal{T}_{3,n}+\mathcal{T}_{4,n}+\mathcal{T}_{5,n}. \qquad (3.8)
\end{aligned}$$

Concerning $(\mathcal{T}_{1,n})$ : We use the Lipschitizian condition on $L$ to get

$$
\begin{aligned}
\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\tilde{g}_n(x,t)-\tilde{g}_n(x,t_{\tilde{j}})\right| &\leq \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\frac{1}{n}\sum_{i=1}^{n}h_i^{-(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)\left|L_i(t)-L_i(t_{\tilde{j}})\right| \\
&\leq \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}C\left|t-t_{\tilde{j}}\right|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i}h_i^{-(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x) \\
&\leq Mb_n\frac{1}{n}\sum_{i=1}^{n}h_i^{-(d+2)} \\
&\leq M\frac{b_n}{h_n^{-d+2}}. \qquad (3.9)
\end{aligned}
$$

So, under Assumption **C(ii)**, we have

$$
(\mathcal{T}_{1,n})=O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right).
$$

By using the same arguments as those used $\mathcal{T}_{1,n}$ we obtain

$$
(\mathcal{T}_{2,n})=O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right),\quad (\mathcal{T}_{3,n})=O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right)\ \text{and}\quad (\mathcal{T}_{4,n})=O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right).
$$

Finally, in order to study $\mathcal{T}_{5,n}$ we use Bernstein's inequality. To do that, we put, for $1\leq i\leq n$, $1\leq k\leq\lambda_n$, and $1\leq j\leq\kappa_n$

$$
U_i=U_i(x_k,t_j):=\left\{h_i^{-(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)-\mathrm{E}\left[h_i^{-(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right]\right\}.
$$

Using the fact that the kernels $K$ and $L$ are bounded, we get

$$
|U_i|\leq Ch_i^{-(d+1)}\leq Ch_n^{-(d+1)}=M.
$$

Moreover, by a similar ideas to those used in the first part of this Lemma, we show that

$$
\begin{aligned}
Var(U_i) &\leq C\int_{\mathbf{R^d}}\int_{\mathbf{R}}\frac{1}{h_i^{-(d+1)}}K^2(u)L^2(v)g(x_k-rh,t_j-sh)dudv. \\
&\leq Ch_i^{-(d+1)}\leq Ch_n^{-(d+1)}:=\sigma.
\end{aligned}
$$

Hence, by Bernstein's inequality (see Hoeffding,1963), it follows that for all $\epsilon>0$ :

$$
\mathrm{P}\left\{\left|n^{-1}\sum_{i=1}^{n}U_i\right|>\epsilon\right\} \leq 2\exp\left\{-\left(\frac{n\epsilon}{M}\right)h\left(\frac{M\epsilon}{\sigma^2}\right)\right\} \qquad (3.10)
$$

where $h(u) = 3u/(6+2u)$ for all $u > 0$.
Now, taking $\epsilon = \epsilon_0 \left(\frac{\log n}{nh_n^{-(d+1)}}\right)^{1/2}$, we have for any $(k, j)$, we obtain

$$\begin{aligned} \mathrm{P}\left\{\left|n^{-1}\sum_{i=1}^{n} U_i\right| > \epsilon\right\} &\leq 2\exp\left\{-\frac{3\epsilon_0^2 \log n}{3c + c_0}\right\} \\ &\leq 2n^{-C\epsilon_0^2} \end{aligned}$$

Thus,

$$\mathbb{P}\left\{\max_{k=1,\ldots,\lambda_n} \max_{j=1,\ldots,\kappa_n} \left|\sum_{i=1}^{n} U_i(x_k, t_j)\right| > \epsilon\right\} \leq \mathcal{C}\lambda_n \kappa_n n^{-C\epsilon_0}. \tag{3.11}$$

Consequently, Borel-Cantelli's lemma and an appropriate choice of $\epsilon_0$ allows us to write that :

$$\mathcal{T}_{5,n} = O\left(\sqrt{\frac{\log n}{nh_n^{-(d+1)}}}\right). \tag{3.12}$$

**Proof of Lemma 4.4.5** Firstly, we write

$$\begin{aligned} \sup_{x\in\mathcal{C}} |\ell_n(x) - \ell(x)| &\leq \sup_{x\in\mathcal{C}} |\ell_n(x) - \mathbb{E}[\ell_n(x)]| + \sup_{x\in\mathcal{C}} |\mathbb{E}[\ell_n(x)] - \ell(x)| \\ &=: \mathcal{L}_{1n} + \mathcal{L}_{2n}. \end{aligned}$$

The first term $\mathcal{L}_{1n}$ is very close to the last part of Lemma 4.4.2. So, by a standard analytical argument we get,

$$\mathcal{L}_{2n} = O(h_n^{+^2}). \tag{3.13}$$

While the proof of the second term For $\mathcal{L}_{2n}$ follows the same lines as in Lemma 4.4.2. Therefore, we get

$$\mathcal{L}_{1n} = O_{a.s.}\left(\frac{\log\ n}{nh_n^{-d}}\right)^{1/2}$$

which completes the proof of this Lemma.

**Proof of Lemma 4.4.4** It is clear that

$$\begin{aligned} |\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| &\leq \sum_{i=1}^{n} \left|\frac{1}{nh_i^{(d+1)}} \delta_i K_i(x) L_i(t) \left(\frac{1}{\bar{G}(Y_i)} - \frac{1}{\bar{G}_n(Y_i)}\right)\right| \\ &\leq \frac{\sup_{t\leq\tau_F} |\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau)} \widetilde{g}_n(x,t) \end{aligned} \tag{3.14}$$

Since $\bar{G}_n(\tau) > 0$, in conjunction with the SLLN and the LIL on the censoring law (see formula (4.28) in Deheuvels and Einmahl, 2000)), the result is an immediate consequence of

Lemma 4.4.2.

**Proof of Lemma 4.4.8**

— **Proof of 3.5**
Similarly to the previous Lemma, we have

$$\sqrt{nh_n^{(d+1)}}\,|\widehat{g}_n(x,t)-\widetilde{g}_n(x,t)| \leq \sqrt{nh_n^{(d+1)}}\frac{\sup_{t\leq\tau_F}|\bar{G}_n(t)-\bar{G}(t)|}{\bar{G}_n(\tau)}\widetilde{g}_n(x,t).$$

Further, as

$$\sup_{t\leq\tau_F}|\bar{G}_n(t)-\bar{G}(t)| = O_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right)$$

then

$$\sqrt{nh_n^{(d+1)}}\left(\frac{\sup_{t\leq\tau_F}\bar{G}_n(t)-\bar{G}(t)}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\right) = O_{a.s.}\left(\sqrt{\log\log nh_n^{(d+1)}}\right).$$

From **N(ii)** we obtain that

$$\sqrt{nh_n^{(d+1)}}\frac{\sup_{t\leq\tau_F}\bar{G}_n(t)-\bar{G}(t)}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)} = o_{a.s.}(1)$$

The latter combined with the results of Lemma4.4.2 allows us to complete the proof of 3.5.

— **Proof of 3.6** It is shown in the first part of Lemma 4.4.2, that

$$[\mathbb{E}(\tilde{g}_n(x,t))-g(x,t)] = O(h_n^{+^2}).$$

Thus,

$$\sqrt{nh_n^{(d+1)}}\,[\mathbb{E}(\tilde{g}_n(x,t))-g(x,t)] = O(\sqrt{nh_n^{d+1}(h_n^{+^4}})$$

which goes to zero under the second part of Assumption **N(ii**

— **Proof of 3.7** By a simple analytical arguments we show that

$$Var\left(\ell_n(x)-\ell(x)\right) = O\left(n^{-1}\sum_{i=1}^{n}h_i^d\right) \text{ and } \mathbb{E}\left[\ell_n(x)-\ell(x)\right] = O(h_n^{+^2}).$$

Now, Assumption **N**(ii) gives $nh_n^{d+1}(h_n^{+^4} \to 0$ and Assumption **N**(i) implies that

$$nh_n^{d+1}Var\left(\ell_n(x)-\ell(x)\right) = o(1).$$

So,

$$\sqrt{nh_n^{d+1}}\,(\ell_n(x)-\ell(x)) \to 0 \quad \text{in probability} \quad as \quad n\to\infty.$$

**Proof of Lemma 4.4.7** : The proof of this Lemma is based on the version of the central limit Theorem given in (Loève (1963), p. 275) where the main point is to calculate the following limit

$$nh_n^{d+1}Var\left[\tilde{g}_n(x,t)\right] \to \sigma'^2(x). \tag{3.15}$$

Indeed, we have

$$\begin{aligned}
nh_n^{d+1}Var\left[\tilde{g}_n(x,t)\right] &= \frac{nh_n^{(d+1)}}{n^2}Var\left[\sum_{i=1}^{n} h_i^{-(d+1)}\bar{G}^{-1}(Y)K_i(x)L_i(t)\mathbb{1}_{\{T_1\leq C_1\}}\right] \\
&= \frac{h_n^{(d+1)}}{n}\sum_{i=1}^{n} h_i^{-2(d+1)}\mathbb{E}\left[\bar{G}^{-2}(T)K_i^2(x)L_i^2(t)\mathbb{E}\left[\mathbb{1}_{T_1\leq C_1}|X_1,T_1\right]\right] \\
&- \frac{h_n^{(d+1)}}{n}\sum_{i=1}^{n} h_i^{-2(d+1)}\left[\mathbb{E}\left\{\bar{G}^{-1}(T)K_i(x)L_i(t)\mathbb{E}\left[\mathbb{1}_{T_1\leq C_1}|X_1,T_1\right]\right\}\right]^2 \\
&:= \nabla_n^1 + \nabla_n^2.
\end{aligned}$$

Observe that

$$\nabla_n^2 = h_n^{(d+1)}\mathbb{E}^2\left[\tilde{g}_n(x,t)\right].$$

Once again, we use the result of Lemma 4.4.2 to show that $\nabla_n^2 = o(1)$.
Now, concerning the first term $\nabla_n^1$, we have

$$\nabla_n^1 = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right)^{(d+1)}\int_{\mathbb{R}^d}\int_{\mathbb{R}}\frac{K^2(z)L^2(y)}{\bar{G}(t-yh_i)}g(x-zh_i,t-yh_i)dzdy.$$

The continuity of the functions $\bar{G}$ and $g$ permit to write

$$\nabla_n^1 = \frac{g(x,t)}{\bar{G}(t)}\int_{\mathbb{R}^d}\int_{\mathbb{R}}K^2(z)L^2(y)\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right)^{(d+1)}\right) + o\left(\frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right)^{(d+1)}\right)$$

From Assumption **A.1**(ii), we obtain the claimed result (3.15).
Let's now prove our asymptotic result. To do that we put

$$\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] = \sum_{i=1}^{n} w_{i,n}(x)$$

where

$$w_{i,n}(x) = \frac{\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}}{nh_i^{d+1}}\left\{\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t) - \mathbb{E}\left(\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right)\right\}$$

and we prove that for some $\beta > 2$

$$\frac{\sum_{i=1}^{n}\mathbb{E}\left[|w_{i,n}(x)|^{\beta}\right]}{\left(Var\left(\sum_{i=1}^{n} w_{i,n}(x)\right)\right)^{(\beta)/2}} \to 0.$$

Indeed, set $\psi_{i,n}^{\beta}(x) = \mathbb{E}\left|w_{i,n}(x)\right|^{\beta}$. Applying the Cr-inequality (see Loève, 1963, p. 155)

$$\begin{aligned}
\psi_{i,n}^{\beta}(x) &= \mathbb{E}\left|\frac{\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}}{nh_i^{d+1}}\left[\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t) - \mathbb{E}\left(\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right)\right]\right|^{\beta} \\
&\leq 2^{\beta-1}\frac{\left(nh_n^{(d+1)}\right)^{\frac{\beta}{2}}}{n^{\beta}(h_i^{d+1})^{\beta}}\mathbb{E}\left(\left|\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right|^{\beta}\right) \\
&+ 2^{\beta-1}\frac{\left(nh_n^{(d+1)}\right)^{\frac{\beta}{2}}}{n^{\beta}(h_i^{d+1})^{\beta}}\left|\mathbb{E}\left(\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right)\right|^{\beta}.
\end{aligned} \tag{3.16}$$

Furthermore, by a standard arguments, we show that

$$\mathbb{E}\left(\left|\delta_i\bar{G}^{-1}(Y)K_i(x)L_i(t)\right|^{\beta}\right) = O(h_i^{(d+1)})$$

and

$$\left|\mathbb{E}\left(\delta_i\bar{G}^{-1}(Y)K_i(x)L_i(t)\right)\right|^{\beta} = O(h_i^{\beta(d+1)}).$$

Therefore,

$$\psi_n^{\beta}(x) = \sum_{i=1}\psi_{i,n}^{\beta}(x) = O\left(n^{1-\beta/2}h_n^{(d+1)(1-\beta/2)}\right).$$

Because of $1-\beta/2<0$ we have $\psi_n^{\beta}(x)\to 0$ which implies that

$$\lim_{n\to\infty}\frac{\sum_{i=1}^{n}\mathbb{E}\left|\psi_{i,n}(x)\right|^{\beta}}{\left(Var\left(\sum_{i=1}^{n}w_{i,n}(x)\right)\right)^{(\beta)/2}} \to 0.$$

The proof of this Lemma is now complete.

## Acknowledgements

The authors are grateful to two anonymous referees whose careful reading and appropriate remarks gave them the opportunity to improve the quality of the paper.

# Bibliographie


[1] Ahmad, I., Lin, P.E. (1976). Nonparametric sequential estimation of a multiple regression Function, *Bull. Math. Statist.*, 17, 63–75.

[2] Amiri, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortemement mélangeant, *C. R. Acad. Sci. Paris, Ser*, I 347, 309–314.

[3] Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction, *J. Nonparam. Statist,* 24 (1), 169–186.

[4] Amiri, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, *arXiv :1211.5767v2.*

[5] Beran, R. (1981). Nonparametric regression with randomly censored survival data, *Technical university of Clifornia, Berkeley.*

[6] Carroll, J. (1976). On sequential density estimation, *Z. Wahrscheinlichkeits- theorie und Verw. Gebiete* , 36, 137–151.

[7] Carbonez, A., Györfi, L., Vander Meulin EC . (1995). Partitioning estimates of a regression function under random censoring, *Statist. & Decisions* , 13, 21–37.

[8] Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data, *Scandi. J. Statist,* 14, 181–197.

[9] Dabrowska, D.M. (1989). Uniform consistency of the kernel conditional Kaplan Meier estimate, *Ann. Statist,* 17, 1157–1167.

[10] Davies, I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist,* 15, 49–54.

[11] Deheuvels, P., Einmahl, JHJ. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications, *Ann Proba*, 28, 1301–1335.

[12] Devroye, L., Wagner, T.J. (1980). On the $L^1$ convergence of kernel estimators of regression functions with application in discrimination. *Z. Wahrschein. Verw. Get,* 51, 15-25.

[13] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, 58, 13-30.

[14] Khardani, S., Lemdani, M., Ould Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship, *J. of the Korean Statistical Society*, 39, 455–469.

[15] Khardani, S., Lemdani, M., Ould Saïd, E. (2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series, *J. Stat. Plann. Inference*, 141, 3426–3436.

[16] Kohler, M., Máthé, K., Pinter, M. (2002). Prediction from randomly Right Censored Data, *J. Multivariate Anal*, (80), 73–100.

[17] Krzÿzak, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation , *IEEE Trans. Inform. Theory*, 38, 1323–1338.

[18] Loève, M. (1963). Probability theory, *New York : Springer-Verlag*.

[19] Masry, E. and Fan,J., (1997). Local polynomial estimation of recursive function for mixing processes, *Scandinave Journal of Statistics*, 24, 165–179.

[20] Nguyen, T., Saracco, J. (2010). Estimation récursive en régression inverse par tranches, *Journal de la société française de statistique*, 151(2), 19–46.

[21] Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl*, 36 (1), 107–116.

[22] Roussas, G.G., Tran, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions , *Annals of Statist*, 20 (1), 98–120.

[23] Walk, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Ann. Inst. Statist. Math.*, 53 (4), 691–707.

[24] Wegman, J. Davies, I. (1979). Remarks on some recursive estimators of a probability density, *Ann. Statist*, 7, 316–327.

[25] Wertz,W. (1985). Sequential and recursive estimators of the probability density, *Statistics*, 16, 277–295.

[26] Wolverton, C. and Wagner, T.J. (1969). Asymptotically optimal discriminant functions for pattern classifcation, *IEEE Trans. Inform. Theory*, 15, 258–265.

# Chapitre 4

# On the strong iniform consistency of the conditional density for censored data based on a recursive kernel

## 4.1 Introduction

Conditional density plays an important role in exploring relationship between responses and covariates. In particular it is used to pricing financial derivatives and estimating parameters in financial models . Estimating conditional density has been extensively studied in the literature. We cite, for instance, Tjøstheim (1994), Hyndman et al.(1996) Fan et al.(1996),...).In the last two decades, the kernel estimation of the conditional density estimation has received lot of attention. For example, Bashtannyk and Hyndman (2001) and Hyndman and Yao (2001) proposed several simple and useful rules for selecting bandwidths for the conditional density estimation. Hall et al. (2004) applied the cross-validation technique to estimate the conditional density. Fan and Yim (2004) proposed a consistent data-driven bandwidth selection procedure in estimating the conditional density functions. De Gooijer and Zerom (2003) introduced a so-called re-weighted Nadaraya-Watson estimator for the conditional density function. However, all these papers assume that the observations are complete.
The nonparametric modeling of censored data is intensively discussed in the recent statistical literature. The first contribution dates back to Beran (1981), who introduced a class of nonparametric regression estimators for the conditional survival function in the presence of right-censoring. Dabrowska (1987,1989) establish the asymptotic properties of the distribution and quantiles functions estimators. Kohler *et al.* (2002) gave a simpler proof in the randomly right-censoring case for kernel, nearest neighbor, least squares and penalized least squares estimates.Further results was studied by Khardani *et al.* (2010, 2012).
Concerning the practical aspects of statistical analysis of censored dependent data we give the following usual examples. In the clinical trials domain, it is frequently happens that the patients from the same hospital have correlated survival times due to unmeasured variables like the quality of the hospital equipment. An example of such data can be found in Lipsitz and Ibrahim (2000) For more examples in a real data, the reader can refer to Wei and Lin

(1989), Cai and Prentice (1995).
The main purpose of this paper is to study a certain class of the recursive estimation. We recall that an estimator is said to be 'recursive' if its value calculated from the first $n$ observations, say $f_n$, is only a function of $f_{n-1}$ and the $n$th observation. In this estimation method the estimator can be updated with each new observation added to the database. This recursive property is clearly useful in sequential investigations and also for a fairly large sample size, since addition of a new observation means the non-recursive estimators must be entirely recomputed. Despite this great importance the recursive kernel estimation of censored dependent data has not yet been fully explored. The main aim of the present work is to study the asymptotic properties of the recursive kernel estimator of the conditional density under random right censoring. Specifically, the asymptotic properties stated are the strong convergence and the asymptotic normality of these estimators when the observation satisfy the strong mixing condition .
The paper is organized as follows. In Section2 we recall our kernel conditional density estimator in the censorship model with some notations.In Section3 the assumptions and main results are given. Finally,the proofs of the main results are relegated to Section4 with some auxiliary results with their proofs.

## 4.2 Preamble

Let$(X_n, T_n)_{n\geq 1}$ be a $\mathrm{R}^d \times \mathrm{R}$ valued stationary strongly mixing process defined on probability space $(\Omega, A, \mathbb{P})$. In this paper we consider the model of random right censorship. Let $(T_i)_{1\leq i\leq n}$ be the survival times and suppose that they form a stationary $\alpha-$mixing sequence with common unknown continuous distribution function (df) $F$ with density $g$. In many situations we observe only censored lifetimes of the items under study. Thait is, assuming that $(C_i)_{1\leq i\leq n}$ is a sample independent and identically distributed (iid) censoring random variables (rvs) with common continuous df $G$, we observe only the $n$ pairs $(Y_i, \delta_i)$ with

$$Y_i = T_i \wedge C_i \quad and \quad \delta_i = \mathbb{1}_{\{T_i\leq C_i\}}. \quad 1 \leq i \leq n \tag{4.1}$$

where $\mathbb{1}_A$ denotes the indicator function of the set $A$. To follow the convention in biomedical studies, we assume that $(C_i)_{1\leq i\leq n}$ and $(T_i, X_i)_{1\leq i\leq n}$ are independent ; this condition is plausible whenever the censoring is independent of the modality of the patients.
Throughout the paper, $x = (x_1, \cdot, x_d) \in \mathbb{R}^d$. For any $x$, denote by $\phi(.|x)$ the conditional density function of $Y$ given $X = x$.
Given a sequence of observations $(X_1, Y_1, \delta_1), \ldots (X_n, Y_n, \delta_n)$ of $(X, Y, \delta)$, the kernel estimate of the conditional density $\phi(t|x)$ denoted $\bar{\phi}_n(t|x)$, is defined by

$$\forall x \in \mathbb{R}^d \text{ and } \forall y \in \mathbb{R} \qquad \bar{\phi}_n(t|x) = \frac{\displaystyle\sum_{i=1}^{n} h_n^{-1}\delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x - X_i}{h_n}\right) L\left(\frac{t - Y_i}{h_n}\right)}{\sum_{i=1}^{n} K\left(\frac{x-X_i}{h_n}\right)}, \tag{4.2}$$

where $K$, $L$ are a kernels and $h_n$ is a sequence of positive real numbers. Note that this last estimator has been recently used by Khardani (2010).In practice $\overline{G}(.)$ is unknown, hence it is not possible to use (4.2) as an estimator. One way to overcome this difficulty is to replace $\overline{G}(.)$ by Kaplan and Meier(1958) estimate $\overline{G}_n(.)$ given by

$$\bar{G}_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1-\delta_{(i)}}{n-i+1}\right)^{1\!\!1_{\{Y_{(i)}\leq t\}}} & \text{if} \quad t < Y_{(n)} \\ 0 & \text{otherwise} \end{cases},$$

which is known to be uniformly convergent to $\bar{G}$ and $Y_{(1)} < Y_{(2)} < \cdot < Y_{(n)}$ are the order statistics of $(Y_{(i)})_{1\leq i\leq n}$ and $\delta_{(i)}$ is the concomitant of $Y_{(i)}$.
A recursive version of the previous kernel estimator is defined by

$$\widehat{\phi}_n(t|x) = \frac{\sum_{i=1}^n h_i^{-(d+1)} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x-X_i}{h_i}\right) L\left(\frac{t-Y_i}{h_i}\right)}{\sum_{i=1}^n h_i^{-d} K\left(\frac{x-X_i}{h_i}\right)} =: \frac{\hat{g}_n(x,t)}{\ell_n(x)}$$

where

$$\hat{g}_n(x,t) := \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{d+1}} \delta_i \bar{G}_n^{-1}(Y_i) K\left(\frac{x-X_i}{h_i}\right) L\left(\frac{t-Y_i}{h_i}\right), \tag{4.3}$$

and

$$\ell_n(x) := \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{x-X_i}{h_i}\right), \ \forall x \in \mathbb{R}^d.$$

**Remark 4.2.1** *the Kaplan-Meir estimator is not recursive and the use of such estimator can slightly penalizes the efficiency of our estimator in term of computational time.*

**Remark 4.2.2** *The joint pdf $\ell(.)$ is not affected by censoring and is therefore consistently estimated by $\ell_n(.)$*

## 4.3 Assumptions and main results

Throughout the paper, when no confusion is possible, we denote by $C$ and/or $C'$ any generic positive constant and we put $C'h_n \leq h_i \leq Ch_n; \quad \forall 1 \leq i \leq n$. Further, we will denote by $F(\cdot)$ (resp. $G(\cdot)$) the distribution function of $T$ (resp. of $C$) and by $\tau_F$ (resp. $\tau_G$) the upper endpoints of the survival function $\bar{F}$ (resp. of $\bar{G}$). In the following we assume that $\tau_F < \infty$, $\bar{G}(\tau_F) > 0$ and $C$ is independent to $(X,T)$. We also assume that there exist a compact set $\mathcal{C} \subset \mathcal{C}_0 = \{x \in \mathbb{R}^d \ \ \ell(x) > 0\}$ where $\ell$ is the density of the explicative variable $X$, and $\Omega$ be a compact set such that $\Omega \subset (-\infty, \tau]$ where $\tau < \tau_F \wedge \tau_G$. From now, for any function $\varphi$ and $j \in \mathbb{N}$, let $\varphi^j$ denote the jth-order derivative of $\varphi$(with $\varphi^{(0)} = \varphi$). As we deal with the $\alpha-$mixig case, we recall its definition

**Definition 4.3.1** *Let $(Z_i)$ be a sequence of rvs. Given a positive integer $n$, set*

$$\alpha(n) = \sup_k \sup |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)|, A \in F_1^k(Z) \text{ and } B \in F_{k+n}^\infty(Z),$$

*where $F_1^k(Z)$ is the $\sigma$-field of events generated by $Z_j, i \leq j \leq k$.*
*The sequence is said to be $\alpha$-mixing if the mixing coefficient $\alpha(n) \to 0 \to$ as $n \to\to \infty$.*

There exist many processes fulfilling the strong mixing property. We quote, here, the usual ARMA processes wich are geometrically strongly mixing, *i.e.*, there exist $\rho \in (0,1)$ and $a > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq a\rho^n$ (see, e.g, Jones(1978)). The threshold models, the EXPRAR models(see, Ozaki(1979)), the simple ARCH models (see Engle(1982)), their GARCH extension( see Bollerslev(1986)) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions.
In order to give the rate of the uniform almost sure convergence of our estimate we need the following assumptions :
**A1** : $(T_n, X_n)_{n\geq 1}$ is stationary $\alpha$-mixing sequence of rvs, with coefficient $\alpha(n)$.
**A2** :The mixing coefficient satisfies $\alpha(n) = O(n^{-\nu}), \quad \nu > 0$
**B1** :The joint pdf $g(\cdot,\cdot)$ of $(X,T)$ is bounded function and continuously differentiable up to order 3.
**B2** :The joint pdf $g_{1.j}$ of $((X_1,T_1),(X_j,T_j))$ exists and satisfies

$$\sup_{R^{d+1}\times R^{d+1}} |g_{1.j}(\cdot,\cdot) - g(\cdot)g(\cdot)| \leq C \leq \infty \quad for\ any \quad j \geq 2$$

.

**B3** :The marginal density $\ell(\cdot)$ is twice differentiable and satisfies a Lipschitz condition. Furthermore $\ell(x) > \Gamma$ for all $x \in \mathcal{C}$ and $\Gamma > 0$. Where $\mathcal{C}$ is a compact set of $\mathbb{R}$.
**B4** :The joint pdf $\ell_{1.j}$ of $((X_1,X_j))$ exists and satisfies

$$\sup_{R^d\times R^d} |\ell^*_{1.j}(\cdot,\cdot) - \ell(\cdot)\ell(\cdot)| \leq C \leq \infty \quad for\ any \quad j \geq 2$$

**K1** :The kernels $K$ and $L$ are Lipschitz continuous functions and compactly supported.
**K2** : $\int_{\mathbf{R^d}} u_l K(u)du = 0$, for $l = 1,\dots,d$ with $u = (u_1,\dots,u_d)^T$ and $\int_{\mathbf{R}} vL(v)dv = 0$
**H1** : There exists $\eta > 0$ such that, $Cn^{\frac{4+(d+1)^2\beta+d-\nu+\eta}{\nu+1}} \leq h_n^{d+1}$ with $\nu > 4 + (d+1)^2\beta + d$

**Remark 4.3.2** *It is clear that the considered assumptions are standard in this context of censoring times series analysis. In particular, ours conditions cover all the axes of our study. Indeed, Conditions **A1** and **A2** control the correlation condition by defining the mixing type as well as its coefficient rate. Assumptions **B1** − **B4** define our nonparametric model. Such assumptions are usually used in asymptotic theory of kernel estimation method. Conditions **K1** − **K2**are mils hypotheses to characterize the kernel function of our estimate. It is well documented in nonparametric analysis of censored data that the independence assumption between $(C_i)_i$ and $(X_i,T_i)_i$, may seem to be strong and one can think of replacing it by a classical conditional independence assumption between $(C_i)_i$ and $(T_i)_i$ given $(X_i)_i$ (see, Khardani et al. (2014) for more discussion ).*

### 4.3.1 Uniform strong consistency results with rate of convergence

**Théoreme 4.3.3** *Under Assumptions* ***A1,A2,K1,K2, B1 - B4*** *and* ***H***

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x)-\phi(t|x)\right| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right),h_n^2\right)\right\} \quad a.s. \quad as \quad n\to\infty \quad (4.4)$$

## 4.4 Auxilary results and proofs

**Lemma 4.4.1** *(Rio,2000,pp.87) Let $U_i, i\geq 1$ be a stationary sequenceof centered real rvs, with strong mixing coefficient $\alpha(n)=O(n^{-v}), v>1$, such that $\|U_1\|_\infty<+\infty$. Then for each $r>1$ and $\epsilon>0$*

$$\mathbb{P}\left\{\left|\sum_{i=1}^n U_i\right|>\epsilon\right\}\leq C\left(1+\frac{\epsilon^2}{rS_n^2}\right)^{-r/2}+nCr^{-1}\left(\frac{2r}{\epsilon}\right)^{v+1},$$

*Where $S_n^2=\sum_{1\leq i\leq n}\sum_{1\leq j\leq n}|cov(U_i,U_j)|$ and where the constant $C$ does not depend on $n$.*

**Lemma 4.4.2** *Under Assumptions* **B1**, **B3** *and* **K2**, *we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)| = O(h_n^2) \qquad as \quad n\longrightarrow\infty.$$

**Proof.** The proof is standard, in the sense that is not affected by the dependence structure. Using the fact that, for all measurable function $\varphi$ and for all $i=1,\dots n$.

$$1\!\!1_{\{T_1\leq C_1\}}\varphi(Y_1)=1\!\!1_{\{T_1\leq C_1\}}\varphi(T_1).$$

Then,

$$\begin{aligned}
\mathbb{E}\tilde{g}_n(x,t) &= n^{-1}\sum_{i=1}^n\frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_i}\right)\delta_i\bar{G}^{-1}(T_i)L\left(\frac{t-T_1}{h_i}\right)\right\}\\
&= n^{-1}\sum_{i=1}^n\frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_i}\right)\bar{G}^{-1}(T_i)L\left(\frac{t-T_1}{h_i}\right)\mathbb{E}\left[1\!\!1_{\{T_i\leq C_i\}}|X_i,T_i\right]\right\}\\
&= n^{-1}\sum_{i=1}^n\frac{1}{h_i^{d+1}}\mathbb{E}\left\{K\left(\frac{x-X_1}{h_1}\right)L\left(\frac{t-T_1}{h_1}\right)\right\}
\end{aligned}$$

Therefore,

$$\begin{aligned}
|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)| &\leq n^{-1}\sum_{i=1}^n\left|\int_{\mathbf{R^d}}\int_{\mathbf{R}}K(u)L(v)[g(x-h_iu,t-h_iv)-g(x,t)]dudv\right|\\
&\leq Mn^{-1}\sum_{i=1}^n h_i^2\leq Mh_n^2.
\end{aligned}$$

Therefore,

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\mathbb{E}\tilde{g}_n(x,t)-g(x,t)| = O(h_n^2).$$

**Lemma 4.4.3** *Under Assumptions of Theorem 4.3.3 we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-\mathbb{E}\tilde{g}_n(x,t)| = O\left\{\left(\frac{\log n}{nh_n^{(d+1)}}\right)^{1/2}\right\} \quad a.s. \qquad as \quad n\longrightarrow\infty.$$

**Proof.** The compactness property of the sets $\mathcal{C}$ and $\Omega$ allows us to write that, for some $(x_k)_{1\le k\le\lambda_n}$ and $(t_j)_{1\le j\le\kappa_n}$,

$$\mathcal{C}\subset\bigcup_{k=1}^{\lambda_n}\mathcal{B}(x_k,a_n) \qquad \text{and} \qquad \Omega\subset\bigcup_{j=1}^{\kappa_n}B(t_j,b_n)$$

where $\lambda_n\sim a_n^{-d}$ and $\kappa_n\sim b_n^{-1}$ with $a_n=b_n=n^{-(d+1)\beta-1/2}$.
Now, for any $x\in\mathcal{C}$ and $t\in\Omega$, we set by $\tilde{k}(x)=\arg\min_k\|x_k-x\|$ and $\tilde{j}(t)=\arg\min_j|t-t_j|$
Now setting

$$Z_i(x_k,t_j):=n^{-1}\left\{h_i^{(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)-\mathrm{E}\left[h_i^{(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right]\right\} \tag{4.5}$$

Then, for any $(x,t)\in\mathcal{C}\times\Omega$, we can write

$$\begin{aligned}\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-\mathbb{E}\tilde{g}_n(x,t)| &\le \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\tilde{g}_n(x,t)-\tilde{g}_n(x,t_{\tilde{j}})\right|+\sup_{x\in\mathcal{S}}\sup_{t\in\Omega}\left|\mathbb{E}\tilde{g}_n(x,t_{\tilde{j}})-\mathbb{E}\tilde{g}_n(x,t)\right|\\ &+ \max_j\sup_{x\in\mathcal{C}}\left|\tilde{g}_n(x,t_j)-\tilde{g}_n(x_{\tilde{k}},t_j)\right|+\max_j\sup_{x\in\mathcal{C}}\left|\mathbb{E}\tilde{g}_n(x_{\tilde{k}},t_j)-\mathbb{E}\tilde{g}_n(x,t_j)\right|\\ &+ \max_k\max_j|\tilde{g}_n(x_k,t_j)-\mathbb{E}\tilde{g}_n(x_k,t_j)|\\ &=: \mathcal{T}_{1,n}+\mathcal{T}_{2,n}+\mathcal{T}_{3,n}+\mathcal{T}_{4,n}+\mathcal{T}_{5,n}.\end{aligned} \tag{4.6}$$

Concerning $(\mathcal{T}_{1,n})$ : We use the Lipschitizian condition on $L$ to get

$$\begin{aligned}\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\tilde{g}_n(x,t)-\tilde{g}_n(x,t_{\tilde{j}})\right| &\le \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\frac{1}{n}\sum_{i=1}^n h_i^{(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)\left|L_i(t)-L_i(t_{\tilde{j}})\right|\\ &\le \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}C\left|t-t_{\tilde{j}}\right|\frac{1}{n}\sum_{i=1}^n\frac{1}{h_i}h_i^{-(d+1)}\delta_i\bar{G}^{-1}(Y_i)K_i(x)\\ &\le Mb_n\frac{1}{n}\sum_{i=1}^n h_i^{-(d+2)}\\ &\le M\frac{b_n}{h_n^{d+2}}.\end{aligned} \tag{4.7}$$

So, we have

$$(\mathcal{T}_{1,n})=O\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right).$$

By using the same arguments as those used $\mathcal{T}_{1,n}$ we obtain

$$(\mathcal{T}_{2,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right), \quad (\mathcal{T}_{3,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right) \text{ and } \quad (\mathcal{T}_{4,n}) = O\left(\sqrt{\frac{\log n}{nh_n^{(d+1)}}}\right).$$

Now, in order to study $\mathcal{T}_{5,n}$ we put, for $1 \leq i \leq n$, $1 \leq k \leq \lambda_n$, and $1 \leq j \leq \kappa_n$

$$U_i = U_i(x_k, t_j) := nh_i^{(d+1)} Z_i(x_k, t_j).$$

Firstly, we calculate

$$S_n^2 = \sum_{i=1}^{n} Var(U_i) + \sum_{l \neq i} |cov(U_i, U_l)| =: \sum_{i=1}^{n} Var(U_i) + S_n^{2*} \tag{4.8}$$

Where $S_n^{2*}$ is the sum of covariances. Using

$$\mathbb{1}_{\{T_1 \leq C_1\}} \varphi(Y_1) = \mathbb{1}_{\{T_1 \leq C_1\}} \varphi(T_1), \tag{4.9}$$

the conditional expectation properties and a change of variables, we have

$$\begin{aligned}
Var(U_i) &= Var\left(\delta_i \bar{G}^{-1}(Y_i) K\left(\frac{x_k - X_i}{h_i}\right) L\left(\frac{t_j - T_i}{h_i}\right)\right) \\
&= \left[\mathbb{E}\left(\delta_1^2 \bar{G}^{-2}(Y_1) K^2\left(\frac{x_k - X_1}{h_i}\right) L^2\left(\frac{t_j - T_1}{h_i}\right)\right) - \mathbb{E}^2\left(\delta_i \bar{G}^{-1}(Y_i) K\left(\frac{x_k - X_i}{h_i}\right) L\left(\frac{t_j - T_i}{h_i}\right)\right)\right] \\
&\leq \mathbb{E}\left[\bar{G}^{-2}(Y_1) K^2\left(\frac{x_k - X_1}{h_i}\right) L^2\left(\frac{t_j - T_1}{h_i}\right) \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq C_1\}} / X_1 T_1\right]\right] \\
&\leq h_i^{(d+1)} \int_{R^d} \int_R \frac{K^2(r) L^2(s)}{\bar{G}(\tau)} g(x_k - rh_i, t_j - sh_i) dr ds
\end{aligned}$$

Under Assumptions **B4** we obtain

$$\sum_{i=1}^{n} Var(U_i) = \sum_{i=1}^{n} h_i^{(d+1)} = O(nh_n^{(d+1)}) \tag{4.10}$$

Then from 4.5,using again the conditional expectation, we get under assumption, for $l \neq i$

$$\begin{aligned}
cov(U_i, U_l) &= \mathbb{E}\left[\frac{\delta_i \delta_l}{\bar{G}(Y_i)\bar{G}(Y_l)} K_i(x_k) L_i(t_j) K_l(x_k) L_l(t_j)\right] - \mathbb{E}^2\left[\frac{\delta_1}{\bar{G}(Y_1)} K(x_k) L(t_j)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\frac{\delta_i \delta_l / T_i T_l}{\bar{G}(Y_i)\bar{G}(Y_l)}\right) K_i(x_k) L_i(t_j) K_l(x_k) L_l(t_j)\right] - \mathbb{E}^2\left[\frac{\delta_1}{\bar{G}(Y_1)} K(x_k) L(t_j)\right] \\
&= h_i^{2(d+1)} \int_{R^d} \int_R \int_{R^d} \int_R \left[g\left(x_k - s_1 h_i, t_j - r_1 h_i, x_k - s_2 h_i, t_j - r_2 h_i\right)\right. \\
&- \left. g\left(x_k - s_1 h_i, t_j - r_1 h_i\right) g\left(x_k - s_2 h_i, t_j - r_2 h_i\right)\right] K(s_1) L(r_1) K(s_2) L(r_2) ds_1 dr_1 ds_2 dr_2
\end{aligned} \tag{4.11}$$

which yields, under Assumptions **B2**

$$cov(U_i, U_l) = O(h_n^{2(d+1)}) \tag{4.12}$$

uniformly on $i$ and $l$.
Now, following Masry (1986), we define the sets
$E_1 = \{(i,l) \textit{ such that } 1 \leq \|i-l\| \leq \xi_n\}$ and $E_2 = \{(i,l) \textit{ such that } \xi_n+1 \leq \|i-l\| \leq n-1\}$.
where $\xi_n$ as $n \to \infty$ at a slow rate, that is $\xi_n = o(n)$. Let $S_n^{2*} = \jmath_{1,n} + \jmath_{2,n}$ where $\jmath_{1,n}$ and $\jmath_{2,n}$ be the sums of covariances over $E_1$ and $E_2$, respectively. We then get from 4.12

$$\jmath_{1,n} = \sum_{(i,l)\in E_1} cov(U_i, U_l) = O(nh_n^{2(d+1)}\xi_n) \tag{4.13}$$

For $\jmath_{2,n}$ we use the modified Davydov inequality for mixing processes(see Rio,p.10, Formula 1.12a). This leads, for all $i \neq l$, to

$$|cov(U_i, U_l)| \leq C\alpha(|i-l|)$$

we then get, from Assumption **A2**

$$\jmath_{2,n} \leq C \sum_{i=1}^{n} \sum_{\xi_n+1<|i-l|\leq n-1} |i-l|^{-v} = O(n\xi_n^{-v+1}) \tag{4.14}$$

Therefore equalizing the rates in 4.13 and 4.14, we get the optimal choice (minimizing the covariance) $\xi_n = Ch_n^{(-2(d+1)/v)}$ yielding

$$S_n^{2*} = \jmath_{1,n} + \jmath_{2,n} = O(nh_n^{2(d+1)((v-1)/v)})$$

wiche combined to (4.8) and (4.10) yields, then

$$S_n^2 \sim nh_n^{(d+1)} \tag{4.15}$$

Now we are in position to apply Lemma 4.1.

Then, for $\epsilon > 0$, applying Lemma 4.4.1, we have, for any $(k, j)$

$$\begin{aligned} \mathbb{P}\left\{\left|\sum_{i=1}^{n} Z_i(x_k, t_j)\right| > \epsilon\right\} \leq \mathbb{P}\left\{\left|\sum_{i=1}^{n} U_i\right| > nh^{(d+1)}\epsilon\right\} &\leq c\left(1 + \frac{n^2h^{2(d+1)}\epsilon^2}{16rS_n^2}\right)^{-r/2} + ncr^{-1}\left(\frac{2r}{nh^{(d+1)}\epsilon}\right)^{\nu+1} \\ &=: c\left(\Psi_{1,n} + \Psi_{2,n}\right) \end{aligned} \tag{4.16}$$

Now, we apply (4.16) with

$$\epsilon = \epsilon_0 \left(\frac{\log n}{nh^{(d+1)}}\right)^{1/2} \qquad \text{and} \qquad r = \mathcal{C}\log n^2$$

and use the Taylor series expansion of $\log(1+x)$ and (4.16) to get

$$\Psi_{1,n} \leq \mathcal{C} \exp\left(-\mathcal{C}\epsilon_0^2 \log n\right) = \mathcal{C}n^{-\mathcal{C}\epsilon_0^2} \quad \text{and} \quad \Psi_{2,n} \leq \mathcal{C}\epsilon_0^{-(\nu+1)} n^{\frac{1-\nu}{2}} (\log n)^{\frac{3\nu+1}{2}} h^{\frac{(d+1)(\nu+1)}{2}}. \tag{4.17}$$

On the other hand, using $\mathbb{P}\left(\cup_i A_i\right) \leq \sum_i \mathbb{P}(A_i)$ we can write

$$\mathbb{P}\left\{\max_{k=1,\ldots,\lambda_n} \max_{j=1,\ldots,\kappa_n} \left|\sum_{i=1}^n Z_i(x_k, t_j)\right| > \epsilon\right\} \leq \mathcal{C}\lambda_n\kappa_n\left(\Psi_{1,n} + \Psi_{2,n}\right). \tag{4.18}$$

Now (4.17), under **H1**, yields $\sum_n \lambda_n\kappa_n\Psi_{1,n} < \infty$. Finally, applying Borel-Cantelli's lemma to (4.18) yields

$$\mathcal{I}_{5,n} = O_{a.co.}\left(\left(\frac{\log n}{nh_n^{d+1}}\right)^{1/2}\right) \tag{4.19}$$

**Lemma 4.4.4** . *Under Assumptions of Theorem 4.3.3 we have*

$$\sup_{x\in\mathcal{C}}\sup_{t\in\Omega} |\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| = O\left\{\left(\sqrt{\frac{\log\log n}{n}}\right)\right\} \quad a.s. \quad as \quad n \to \infty. \tag{4.20}$$

**Proof.**It is clear that

$$\begin{aligned} |\widehat{g}_n(x,t) - \widetilde{g}_n(x,t)| &\leq \sum_{i=1}^n \left|\frac{1}{nh_i^{(d+1)}}\delta_i K_i(x)L_i(t)\left(\frac{1}{\bar{G}(Y_i)} - \frac{1}{\bar{G}_n(Y_i)}\right)\right| \\ &\leq \frac{\sup_{t\leq\tau_F}|\bar{G}_n(t) - \bar{G}(t)|}{\bar{G}_n(\tau)}\widetilde{g}_n(x,t) \end{aligned} \tag{4.21}$$

Since $\bar{G}_n(\tau) > 0$, in conjunction with the SLLN and the LIL on the censoring law (see formula (4.28) in Deheuvels and Einmahl, 2000)), the result is an immediate consequence of the previous Lemmas.

**Lemma 4.4.5** . *Under Assumptions of Theorem 4.3.3 we have*

$$\sup_{x\in\mathcal{C}} |\ell(x) - \ell_n(x)| = O\left\{\max\left(\left(\sqrt{\frac{\log n}{nh_n^d}}\right), h_n^2\right)\right\} \quad a.s. \quad as \quad n \to \infty.$$

**Proof.** Firstly, we write

$$\begin{aligned} \sup_{x\in\mathcal{C}} |\ell_n(x) - \ell(x)| &\leq \sup_{x\in\mathcal{C}} |\ell_n(x) - \mathbb{E}[\ell_n(x)]| + \sup_{x\in\mathcal{C}} |\mathbb{E}[\ell_n(x)] - \ell(x)| \\ &=: \mathcal{L}_{1n} + \mathcal{L}_{2n}. \end{aligned}$$

The term $\mathcal{L}_{2n}$ is very close to the last part of Lemma 4.4.2. So, by a standard analytical argument we get,

$$\mathcal{L}_{2n} = O(h_n^2). \tag{4.22}$$

While the proof of the second term For $\mathcal{L}_{1n}$ follows the same lines as in Lemma 4.4.3. Therefore, we get

$$\mathcal{L}_{1n} = O_{a.s.}\left(\frac{\log n}{nh_n^d}\right)^{1/2}$$

which completes the proof of this Lemma.
**Proof of Theorem 4.3.3.**
Set

$$\widetilde{g}_n(x,t) := \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^{(d+1)}}\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)$$

with $K_i(x) = K\left(\frac{x-X_i}{h_i}\right)$, $L_i(t) = L\left(\frac{t-Y_i}{h_i}\right)$.
Now, the proof of this Theorem is based on the following decomposition

$$\begin{aligned}
\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\widehat{\phi}_n(t|x)-\phi(t|x)\right| &\leq \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}\left|\frac{\widehat{g}_n(x,t)}{\ell_n(x)}-\frac{\tilde{g}_n(x,t)}{\ell_n(x)}\right| + \left|\frac{\tilde{g}_n(x,t)}{\ell_n(x)}-\frac{\mathbb{E}\tilde{g}_n(x,t)}{\ell_n(x)}\right| \\
&+ \left|\frac{\mathbb{E}\tilde{g}_n(x,t)}{\ell_n(x)}-\frac{g(x,t)}{\ell_n(x)}\right| + \left|\frac{g(x,t)}{\ell_n(x)}-\frac{g(x,t)}{\ell(x)}\right| \\
&\leq \frac{1}{\inf_{x\in\mathcal{C}}\ell_n(x)}\left\{\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\widehat{g}_n(x,t)-\tilde{g}_n(x,t)| + \sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\tilde{g}_n(x,t)-g(x,t)|\right. \\
&+ \left. +\sup_{x\in\mathcal{C}}\sup_{t\in\Omega}|\phi(t|x)|\sup_{x\in\mathcal{C}}|\ell(x)-\ell_n(x)|\right\} \qquad (4.23)
\end{aligned}$$

So, the proof of this Theorem is a direct consequence of Lemmas (4.4.2-4.4.5).

### 4.4.1 Asymptotic normality

Now, we study the asymptotic normality of our estimate which is given in the following Theorem. For this purpose we replace **H1**
**Assumptions N1 :**

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{h_n}{h_i}\right) = \theta \text{ and } \quad n^{(1-\nu)/(1+\nu)} \leq h_n^{d+1} \leq n^{1/(3-2\nu)}$$

and we added
**Assumptions N2 :**

$$\lim nh_n^{d+5} \to 0$$

**Théoreme 4.4.6** .
*Under Assumptions* ***K1-K2,A1-A2,B1-B4, N1-N2*** *we have, for any* $(x,t)\in\mathcal{A}$*,*

$$\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x)-\phi(t|x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\sigma^2(x,t)\right)$$

*where* $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution,*

$$\sigma^2(x,t) = \theta \frac{\phi(t|x)}{\ell(x)\bar{G}(t)} \int_{\mathbb{R}^d}\int_{\mathbb{R}} K^2(z)L^2(y)dzdy$$

*and* $\mathcal{A} = \{(x,t) \quad \sigma^2(x,t) \neq 0\}$.

***Proof of Theorem 4.4.6***

It is clear that

$$\begin{aligned}
\sqrt{nh_n^{(d+1)}}\left(\hat{\phi}_n(t|x) - \phi(t|x)\right) &= \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right] \\
&+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] \\
&+ \frac{\sqrt{nh_n^{(d+1)}}}{\ell_n(x)}\left[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)\right] \\
&+ \sqrt{nh_n^{(d+1)}}\frac{g(x,t)}{\ell_n(x)\ell(x)}\left[\ell(x) - \ell_n(x)\right]. \qquad (4.24) \\
&= \frac{\mathcal{J}_1(x,t) + \mathcal{J}_2(x,t) + \mathcal{J}_3(x,t)}{\ell_n(x)} + \sqrt{nh_n^{(d+1)}}\frac{g(x,t)}{\ell_n(x)\ell(x)}\left[\ell_n(x) - \ell(x)\right]
\end{aligned}$$

Thus, The proof of Theorem 4.4.6 can be deduced directly from the following Lemmas

**Lemma 4.4.7** *Under Assumptions* ***K1-K2,A1-A2,B1-B2, N1*** *we have*

$$\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma'^2\right)$$

*where* $\sigma'^2(x,t) = \theta \dfrac{g(x,t)}{\bar{G}(t)} \displaystyle\int_{\mathbb{R}^d}\int_{\mathbb{R}} K_i^2(z)L_i^2(y)dzdy$

**Proof.** Let, for any $i \geq 1$ ad $(x,t) \in \mathbb{R}^d \times \mathbb{R}$,

$$\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}\left[\tilde{g}_n(x,t) - \mathbb{E}(\tilde{g}_n(x,t))\right] = \sum_{i=1}^{n} W_{i,n}(x)$$

where

$$W_{i,n}(x) = \frac{\left(nh_n^{(d+1)}\right)^{\frac{1}{2}}}{nh_i^{d+1}}\left\{\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t) - \mathbb{E}\left(\delta_i\bar{G}^{-1}(Y_i)K_i(x)L_i(t)\right)\right\}$$

The proof of this lemma is based on the central limit theorem of Liebscher (2001)(Corollary 2.2, p. 196) which restes on the asymptotic behavior of the following quantity

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}[W_{i,n}(x)^2] \tag{4.25}$$

and the additional assumptions

$$\begin{cases} \textit{There exists a sequence } \tau_n = o(\sqrt{n}) \textit{ such that } \tau_n \leq (\max_{i=1,\dots,n} C_i)^{-1} \\ \textit{where } C_i = ess\sup_{\omega\in\Omega}|W_{i,n}(x)|, \\ \textit{and } \frac{n}{\tau_n}\alpha(\epsilon\tau_n) \to 0 \textit{ for all} \epsilon > 0, \end{cases}. \tag{4.26}$$

$$\begin{cases} \textit{There exist a sequence } (m_n) \textit{ of a positive integers tending to } \infty \textit{ such that,} \\ nm_n\gamma_n = o(1) where \gamma_n := \max_{1\leq i\neq j\leq n}\left(\mathbb{E}\left[W_{i,n}(x)W_{j,n}(x)\right]\right), \\ \left(\sum_{j=m_n+1}^{\infty}\alpha(j)\right)\Sigma_{i=1}^{n}C_i = o(1) \end{cases}. \tag{4.27}$$

We start by evaluating the limit of (4.25). To do that, let us remark that

$$\begin{aligned}
\mathbb{E}\left[W_{i,n}(x)^2\right] &= \frac{h_n^{d+1}}{n}\sum_{i=1}^{n}\frac{1}{h_i^{2(d+1)}}\mathbb{E}\left[\frac{1}{\bar{G}^2(T_1)}K^2\left(\frac{x-X_i}{h_i}\right)L^2\left(\frac{t-Y_i}{h_i}\right)\mathbb{E}\left(\mathbb{1}\{T_1\leq C_1\}/X_1,T_1\right)\right] \\
&- \frac{h_n^{d+1}}{n}\sum_{i=1}^{n}\frac{1}{h_i^{2(d+1)}}\mathbb{E}^2\left[\frac{1}{\bar{G}^2(T_1)}K^2\left(\frac{x-X_i}{h_i}\right)L^2\left(\frac{t-Y_i}{h_i}\right)\mathbb{E}\left(\mathbb{1}\{T_1\leq C_1\}/X_1,T_1\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^{n}\frac{h_n^{(d+1)}}{h_i^{2(d+1)}}\int_{\mathbb{R}^d}\int_R\frac{1}{\bar{G}(v)}\left[K\left(\frac{x-u}{h_i}\right)L\left(\frac{t-v}{h_i}\right)\right]^2 g(u,v)dudv \\
&- \frac{1}{n}\sum_{i=1}^{n}\frac{h_n^{(d+1)}}{h_i^{2(d+1)}}\int_{\mathbb{R}^d}\int_R\left[K\left(\frac{x-u}{h_i}\right)L\left(\frac{t-v}{h_i}\right)g(u,v)dudv\right]^2 \\
&\to^{\infty} \theta\frac{g(x,t)}{\bar{G}(t)}\int_{\mathbb{R}^d}\int_R\left[K(s)L(r)\right]^2 dsdr
\end{aligned} \tag{4.28}$$

It follows that

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}[W_{i,n}(x)^2] = \theta\frac{g(x,t)}{\bar{G}(t)}\int_{\mathbb{R}^d}\int_R\left[K(s)L(r)\right]^2 dsdr \tag{4.29}$$

Concerning (4.26), using the boundness $K$ and $L$ to show that $C_i = O\left(\frac{1}{\sqrt{nh_n^{(d+1)}}}\right)$. Therefore, we can take $\tau_n = \sqrt{\frac{nh_n^{(d+1)}}{\log n}}$.
Furthermore, this choice gives, for all $\epsilon > 0$

$$\begin{aligned}
\frac{n}{\tau_n}\alpha(\epsilon\tau_n) &\leq C\left(n^{1-(\nu+1)/2}h_n^{-(d+1)(\nu+1)/2}(\log n)^{(\nu+1)/2}\right) \\
&\leq C\left(n^{d(\nu+1)(3\nu-\nu^2)/2(\nu-1)}(\log n)^{(\nu+1)/2}\right) \to 0 \quad \textit{since } \nu > 3
\end{aligned}$$

Let us derive (4.27). For $1 \leq i < j \leq n$ we have

$$\begin{aligned}
\mathbb{E}\left[|W_{i,n}(x)W_{j,n}(x)|\right] &= \left|\left(\frac{(nh_n^{(d+1)})^{1/2}}{nh_i^{(d+1)}}\right)^2 \left[\int_{\mathbb{R}^d}\int_R\int_{\mathbb{R}^d}\int_R K\left(\frac{x-u}{h_i}\right)L\left(\frac{t-v}{h_i}\right)K\left(\frac{x-s}{h_i}\right)L\left(\frac{t-r}{h_i}\right)\right.\right. \\
&\times \left. g_{1,j-i+1}(u,v,s,r)dudvdsdr - \left[\int_{\mathbb{R}^d}\int_R K\left(\frac{x-u}{h_i}\right)L\left(\frac{t-v}{h_i}\right)g(u,v)dudv\right]^2\right]\Bigg| \\
&= (n^{-1}h_n^{(d+1)}) \qquad (4.30)
\end{aligned}$$

we have $\forall i \neq j$

$$\gamma_n = \max_{1\leq i\neq j\leq n}\left(\mathbb{E}\left[|W_{i,n}(x)W_{j,n}(x)|\right]\right) = O(n^{-1}h_n^{(d+1)})$$

On the other hand, using the fact that

$$\sum_{j\geq x+1} j^{-\nu} = \left[(\nu-1)x^{\nu-1}\right]^{-1}$$

to write

$$\sum_{j=m_n+1}^{\infty}\alpha(j) \leq \sum_{j=m_n}^{\infty}\alpha(j) \leq \int_{t\geq m_n} t^{-\nu}dt = \frac{m_n^{1-\nu}}{\nu-1}$$

thus,

$$\left(\sum_{j=m_n+1}^{\infty}\alpha(j)\right)\Sigma_{i=1}^n C_i = O\left(\frac{m_n^{1-\nu}}{\nu-1}\sqrt{\frac{n}{h_n^{(d+1)}}}\right)$$

choosing $m_n = \left[\left(\frac{h_n^{(d+1)}}{n\log n}\right)^{1/2(1-\nu)}\right]$ where [.] denote the function integer part. It is clear that, under (**N1**). In addition, if we replace $m_n$ by its expression, we obtain

$$\left(\sum_{j=m_n+1}^{\infty}\alpha(j)\right)\Sigma_{i=1}^n c_i = O\left((\log n)^{-1/2}\right) = o(1)$$

and under hypothesis (**N1**) we have

$$\begin{aligned}
m_n\gamma_n &\leq Cn^{-1-1/2(1-\nu)}h_n^{(d+1)(1+1/2(1-\nu))}(\log n)^{-1/2(1-\nu)} \\
&\leq Cn^{(2\nu-3)/2(1-\nu)}h_n^{(d+1)(3-2\nu)/2(1-\nu)}(\log n)^{-1/2(1-\nu)} \\
&= o(1)
\end{aligned}$$

Now, the lemma can be easily deduced from (4.25)-(4.27) and corollary 2.2 of Liebscher(2001).

**Lemma 4.4.8** *Under assumptions of Theorem 4.4.6, we have*

$$\sqrt{nh_n^{(d+1)}}\left[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)\right] \to 0 \quad a.s. \quad as \quad n\to\infty \qquad (4.31)$$

$$\sqrt{nh_n^{(d+1)}}\,[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)] \to 0 \qquad as \quad n \to \infty \tag{4.32}$$

$$\sqrt{nh_n^{(d+1)}}\,(\ell_n(x) - \ell(x)) \to 0 \quad in \quad probability \quad as \quad n \to \infty \tag{4.33}$$

**Proof of 4.31** Similarly to the previous Lemma, we have

$$\sqrt{nh_n^{(d+1)}}\,[\widehat{g}_n(x,t) - \tilde{g}_n(x,t)] \leq \sqrt{nh_n^{(d+1)}}\left(\frac{\sup_{t\leq\tau_F}\bar{G}_n(t) - \bar{G}(t)}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\right)\sum_{i=1}^{n}\left|\frac{1}{nh_i^{(d+1)}}K_i(x)L_i(t)\right|$$

Now, as

$$\sup_{t\leq\tau_F}|\bar{G}_n(t) - \bar{G}(t)| = O_{a.s.}\left(\sqrt{\frac{\log\log n}{n}}\right)$$

then

$$\sqrt{nh_n^{(d+1)}}\left(\frac{\sup_{t\leq\tau_F}\bar{G}_n(t) - \bar{G}(t)}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)}\right) = O_{a.s.}\left(\sqrt{\log\log nh_n^{(d+1)}}\right)$$

From **N1** we obtain that

$$\sqrt{nh_n^{(d+1)}}\frac{\sup_{t\leq\tau_F}\bar{G}_n(t) - \bar{G}(t)}{\bar{G}_n(\tau_F)\bar{G}(\tau_F)} = o_{a.s.}(1)$$

The latter combined with the results of Lemma4.4.2 and 4.4.3 allows us to complete the proof of this assertion.

**Proof of 4.32** : It is shown in Lemma 4.4.2, that

$$[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)] = O(h_n^2).$$

Thus,

$$\sqrt{nh_n^{(d+1)}}\,[\mathbb{E}(\tilde{g}_n(x,t)) - g(x,t)] = O(\sqrt{nh_n^{d+5}})$$

which goes to zero under Assumption **N2**
**Proof of 4.33** : By Lemma 4.4.4, we have

$$\sqrt{nh_n^{(d+1)}}\,(\ell_n(x) - \ell(x)) = O_{a.s.}\left(O\left\{\max\left(\left(\sqrt{h_n\log n}\right), \sqrt{nh_n^{d+5}}\right)\right\}\right). \tag{4.34}$$

Now, Assumption **N2** gives $\sqrt{nh_n^{d+5}} \to 0$ and moreover it implies that $h_n = o\left(n^{-1/(d+5)}\right)$. So,

$$\sqrt{h_n\log n} = o\left(\sqrt{n^{-1/(d+5)}\log n}\right) = o_{a.s.}(1)$$

# Bibliographie


[1] Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction, *J. Nonparam. Statist,* 24 (1), 169-186.

[2] Amiri, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, *arXiv :1211.5767v2.*

[3] Bradley, R., (2007). Introduction to Strong Mixing Conditions. Volume 1-3 : Kendrick Press.

[4] Bollerslev, T., (1986). General autoregressive conditional heteroskedasticity, J. Economt. 31 , 307-327.

[5] Cai, Z. (1998). Asymptotic properties of Kaplan-Meier estimator for censored dependent data. Statist. Probab. Lett., 37 : 381-389.

[6] Cai, J. and Prentice, R. L., (1995). Estimating equations for hazard ratio parameters based on correlated failure time data, Biometrika 82, 151-164.

[7] Carbon, M. and Tran, L. T. (1996). On histograms for linear processes, J. Statist. Plann. Inference , 53 (3), 403-419.

[8] Carbonez, A., Györfi, L., Van der Meulen, E.C., (1995). Partitioning estimates of a regression function under random censoring. Statistics Decisions, 13, 21-37.

[9] Chernoff, H., (1964). Estimation of the mode. Annals of the Institute of Statistical Mathematics, 16, 31-41.

[10] Chow, Y.S., Teicher, H., (1997) Probability Theory. Independence, Interchangeability, Martingales. New York : Springer.

[11] Dedecker, J. et Prieur, C. (2005). New dependence coefficient. Examples and Applications to statistics. Probab. Theory Related Fields. 132 : 203-236. 42.

[12] Deheuvels, P., Einmahl, J.H.J., (2000). Functional limit laws for the increments of Kaplan- Meier product-limit processes and applications. The Annals of Probability, 28, 1301-1335.

[13] Dudley, R.M., (1999). Uniform Central Limit Theorems. Cambridge, UK : Cambridge University Press.

[14] Engle, R.F., (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflation. Econometrics, 50, 4, 987- 1007.

[15] Fan, J. et Yim, T. H., (2004). A crossvalidation method for estimating conditional densities. Biometrika, 91(4) :819-834.

882.

[16] Giné, E., Guillou, A., (2001). On consistency of kernel density estimators for randomly censored data : rates holding uniformly over adaptive intervals. Annales de l'Institut Henri Poincaré, 37, 503-522.

[17] Giné, E., Guillou, A., (2002). Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincaré, 38, 907-921.

[18] Gooijer, J.G. and Zerom, D., (2003). On conditional density dstimation, Stat. Neerl. 57 159-176.

[19] Hall,P. Racine,J. Li,Q. (2004). Cross-validation and the estimation of conditional probability densities, J. Amer. Statist. Assoc. 99 (468) 1015-1026.

[20] Hyndman, R. and Yao, Q. (2002). Nonparametric estimation and symmetry tests for conditional density functions. J. Nonparametr. Stat., 14, Pages 259-278.

[21] Jones, D.A. (1978). Nonlinear autoregressive processes. Proc. Roy. Soc. London A., 360, 7-95.

[22] Kaplan, E.M., Meier, P., (1958). Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 53, 457-481.

[23] Khardani, S. ; Lemdani, M. and Ould Saïd, E. (2010) Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship. J. Korean Statist. Soc. 39 , 455-469.

[24] Khardani, S. ; Lemdani, M. and Ould Saïd, E. (2011) Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series. J. Statist. Plann. Inference 141 , 3426-3436

[25] Khardani, S. ; Lemdani, M. and Ould Saïd, E. (2012) On the strong uniform consistency of the mode estimator for censored time series. Metrika 75, 229-241.

[26] Kohler, M., Máthé, K., Pinter, M. (2002). Prediction from randomly Right Censored Data, *J. Multivariate Anal*, (80), 73-100.

[27] Konakov, V.D., (1974). On the asymptotic normality of the mode of multidimensional distributions. Theory of Probability and its Applications, 19, 794-799.

[28] Liebscher, E., (2001). Estimation of the density and the regression function under mixing condition. Statist. Decisions, 19 : 9-26.

[29] Lipsitz, SR, Ibrahim, JG, 2000. Estimation with Correlated Censored Survival Data with Missing Covariates. Biostatistics, 1, 315-327.

[30] Loève, M., (1963). Probability Theory, New York : Springer-Verlag.

[31] Louani, D., (1998). On the asymptotic normality of the kernel estimators of the function and its derivatives under censoring. Communications in Statistics. Theory and Methods, 27, 2909-2924.

[32] Louani, D., Ould Said, E., (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. Journal of Nonparametric Statistics, 11, 413-442.

[33] Mehra, K.L., Ramakrishnaiah, Y.S., Sashikala, P., (2000). Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode. Annals of the Institute of Statistical Mathematics, 52, 630-645.

[34] Nadaraya, E.N., (1965). On nonparametric estimates of density functions and regression curves. Theory of Probability and its Applications, 10, 186-190.

[35] Ould Said, E., (1993). Estimation non paramétrique du mode conditionnel. Application à la prévision. Comptes Rendus de l'Académie des Sciences, Paris, Série I, tome 316, 943-947.

[36] Ould Said, E., Cai, Z., (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. Journal of Nonparametric Statistics, 17, 797-806.

[37] Ozaki,T. 1979.Nonlinear time series models for nonlinear random vibrations, Technical report, Univ. of Manchester.

[38] Polonik, W. Yao, Q. (2000). Conditional minimum volume predictive regions for stochastic processes, J. Amer. Statist. Assoc. 95 509-519.

[39] Rio, E., 2000. Théorie asymptotique des processus aléatoires faiblement dépendants. Mathématiques et applications, 31 Springer-Verlag, New York.

[40] Samanta, M., (1973). Nonparametric estimation of the mode of a multivariate density. South African Statistical Journal, 7, 109-117.

[41] Samanta, M., Thavaneswaran, A., (1990). Nonparametric estimation of the conditional mode. Communications in Statistics. Theory and Methods, 16, 4515-4524.

[42] Silverman, B.W., (1986). Estimation for statistics and data analysis, Monographs on Statistics and Applied Probability, London : Chapman and Hall.

[43] Talagrand, M., (1996). New concentration inequalities in product spaces. Invent. Math., 126, 505-563.

[44] Tjøstheim, D., (1994). Non-linear time series : a selective review, Scand. J. Statist. 21 97-130.

[45] Van der Vaart, A.W., Wellner, J.A., (1996). Weak Convergence and Empirical Processes, Berlin : Springer.

[46] Van Ryzin, J., (1969). On strong consistency of density estimates. The Annals of Mathematical Statistics, 40, 1765-1772.

[47] Vieu, P., (1996). A note on density mode function. Statistics and Probability Letters, 26, 297- 307.

[48] Wei, L. J., Lin, D. Y., Weissfeld, L., (1989). Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. Journal of the American Statistical Association, 84, 1065-1073.

# Chapitre 5

# Conclusion et perspective

## 5.1 Conclusion

Dans cette contribution, nous avons abordé une approche qui est fondée sur l'estimation récursive de la fonction de densitée conditionnelle, nous avons travailé sur des données censurées. Deux cas ont été étudié, le premier s'agit des variables indépendantes et identiquement distribuées tandis que le deuxième cas s'agit des variables $\alpha$ mélangentes.

Dans la première partie, nous avons construit pour ce modèle de censure un estimateur récursif de la densté conditionnelle. Ensuite nous avons prouvé sa convergence uniforme presue sûre, sur un compact avec taux de convegence et nous avons établit la normalité asymptotique.

Dans la deuxième partie, nous avons considéré les même estimateurs récursives définit dans le premier cas l'orsque les données présentent une forme de dépendance nous avons traité le cas de l'$\alpha$- mélange. De plus notre estimateur qu'on a donné est utilisable pour d'autre modèle nom paramétrique. Nous illustrons les résultats obtenu par des simulations et applications réelle.

Nos résultats obtenus ont permis de confirmer que les estimateurs récursifs sont statistiquement meilleurs que leurs versions non récursives, au regard du gain de temps calcul réalisé, les noyaux récursifs présentent un avantage décisif.

En conclusion, la récursivité améliore l'estimateur à noyau habituel au sens de la variance. Ainsi, l'augmentation du biais apporté par la récursivité n'est pas compensée par la diminution de la variance et se traduit par une légère détérioration de l'EQM. Toutefois les estimateurs récursifs atteignent bien les vitesses optimales et conservent un avantage décisif en terme de rapidité de calcul.

Nous pensons par le travail présenté dans cette thèse avoir donné un outil statistique pouvant s'appliquer à un grand nombre d'étude pratique.

## 5.2 Perspective

Pour conclure les travaux de cette thèse, nous exposons dans ce qui suit, quelques développements futurs possibles en vue d'améliorer et d'étendre nos résultats.

**Cas des données fonctionnelles :**

La littérature sur les données censurées est trés riche mais reste encore d'actualité notamment en ce qui concerne l'estimation des paramétres fonctionnels présents dans ce modèle, Pour cela nous pensons qu'il est possible d'adapter nos résultats sur la densité conditionnel dans le cas ou la variable explicative est fonctionnelle.

**L'estimation récursive de la densitée conditionnelle au cas spatialé :**

Dans notre travail on a confirmé l'éfficacité des méthodes récursives. La généralisaton de nos résutats au cadre spatial est une suite logique à suivre,nos données peuvent être étendus à des données spatiales. Nous pensons que cette génération peut être facilement atteinte par l'adaptation des résultats de Rachdi et Niang et Yao (2011), Laksaci et Mechab (2014).

**Choix du paramétre de lissage :**

Pour évaluer l'optimalité de la vitesse de convergence nous nous sommes basés sur le critère de la convergence uniforme presqque sûre, établit dans le cas i.i.d par Stute (1982). Il serait intéréssant de regarder si de tels résultat demeurent vrais pour des données tronqué.

# Chapitre 6

# Annexe : Quelques outils de probabilités

## 6.1 Notations et définitions

Dans ce chapitre, nous allons ouvrir une parenthèse assez rapide pour rappeler quelques définitions et théorèmes utilisés le long de cette thèse.

### 6.1.1 Convergence en probabilité

**Definition 6.1.1** *Soit* $(X, X_n), n \geq 1$ *une suite de variables aléatoires réelles, définies sur le même espace de probabilité* $(\Omega, \mathcal{F}, \mathbb{P})$*. La suite* $(X_n)$ *converge en probabilitÃ© vers* $X$ *si :*

$$\forall \epsilon > 0, \qquad \lim_{n \to \infty} \mathbb{P}(|X_n - X| \geq \epsilon) = 0$$

### 6.1.2 Limite d'un produit

**Proposition 6.1.1** *Soient*$X_n, Y_n$ *deux suites de variables aléatoires réelle. Si* $X_n \longrightarrow X$*,* $Y_n \longrightarrow Y$ *en probabilité, alors* $X_n Y_n \longrightarrow XY$ *en probabilité.*

Soit $X, (X_n, n \geq 1)$ une suite de variables aléatoires à valeurs dans espace métrique $(E, d)$.

**Definition 6.1.2** *La suite* $X_n \longrightarrow X$ *en probabilité si*

$$\forall \epsilon > 0, \quad \lim_{n \to \infty} \mathbb{P}(d(X_n, X) \geq \epsilon) = 0$$

## 6.2 Convergence presque sûre

**Definition 6.2.1** *La suite* $(X_n)$ *converge presque sûrement (p.s) vers* $X$*, si*

$$\left\{ w : lim_{n \to \infty} X_{n(w)=(w)} \right\}$$

*a une probabilité 1.*

## 6.3 Théorème de la limite centrale

**Théoreme 6.3.1** *Soit* $(X_n)$ *un échantillon iid d'une loi de moyenne* $m$ *et de variance* $\sigma^2$. *La convergence suivante a lieu en loi, lorsque* $n \to \infty$

$$R_n := \sqrt{n}\frac{\bar{X} - m}{\sigma} = \frac{S_n - nm}{\sigma\sqrt{n}} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

**Théorème de limite centrale de Loève :**

**Théoreme 6.3.2** *Soit* $(X_n)_{n\in\mathbb{Z}}$ *une suite de variables aléatoires indépendantes de même loi centrée. Notons* $S_n = \sum_{k=1}^{n} X_k$ *la somme partielle et* $s_n = \sigma(S_n)$ *son écart-type*

1. *Si* $\frac{1}{s_n^{2+\delta}} \sum_{k=1}^{n} \mathbb{E}\left[|X_k|^{2+\delta}\right] \longrightarrow 0$ *pour* $\delta > O$, *alors*

$$\frac{S_n}{s_n} \xrightarrow{\mathcal{L}} \mathcal{N}(0,1)$$

*où* $\mathcal{L}$ *désigne la convergence en loi, et* $\mathcal{N}(0,1))$ *est la loi gaussienne centrée réduite*

**Théorème central limite de Doukhan et al. (1994) :**

Soit $(X_n)_{n\in\mathbb{Z}}$ une suite stationnaire et fortement mélangeante de variables aléatoires réelles centrées, de suite de coefficients de mélange $(\alpha_n)_{n\geq 0}$. Notons $Q_X$ la fonction quantile de la variable $X$ définie par $Q_x(u) = \inf\{t : \mathbb{P}(|X| > t) \leq u\}$, et $S_n = \sum_{i=1}^{n} X_i$ la somme partielle

**Théoreme 6.3.3** *Soit* $(X_n)_{n\in\mathbb{Z}}$ *une suite stationnaire de variables aléatoires centrées vérifiant :*

$$\int_0^1 \alpha^{-1}(u)(Q_{X_0}(u))^2 < +\infty$$

*alors, la série* $\sum_{n\in\mathbb{Z}} \mathbb{E}\left[X_0 X_n\right]$ *converge vers* $\sigma^2 \geq 0$ *et* $n^{-1}var(S_n)$ *converge aussi vers* $\sigma^2$. *De plus, si* $\sigma^2 > 0$, *alors* $\frac{S_n}{\sqrt{n}}$ *converge vers une loi normale centrée de variance* $\sigma^2$.

**Théorème central limite de Liebscher (2001) :**

Soit $(Y_{n,i}; i = 1, \cdots, n)$ une suite de variables aléatoires centrées, $\alpha$-mélangeantes, et $T_n = \sum_{i=1}^{n} Y_{n,i}$
Notons

$$\Gamma_n := \max_{i=1,\cdots,n} \mathbb{E}\left[Y_{n,i}^2\right]$$

$$\gamma_n := \max_{j,k1\leq j,k\leq n, j\neq k} \left(\mathbb{E}\left[|Y_{n,j}Y_{n,k}|\right] + \mathbb{E}\left[|Y_{n,j}|\right]\mathbb{E}\left[|Y_{n,k}|\right]\right)$$

**Condition** $C(p)$ : Supposons, $2 < p < \infty$ : $\Gamma_n = O(n^{-1}), \sum_{k=1}^{\infty} k^{2/(p-2)}\alpha(k) < \infty$, et $\mathbb{E}\left[|Y_{n,i}|^p\right]^{1/p} < \infty$ $(i = 1, \cdots, n, n \in \mathbb{N})$ . Il existe une suite de nombres réels positifs $m_n$ qui tend vers $\infty$ tel que :

$$nm_n\gamma_n = o(1),$$

et

$$\left(\sum_{m_n+1}^{\infty} j^{2/(p-2)}\alpha(j)\right)^{1-2/p}\sum_{i=1}^{n}(\mathbb{E}\left[|Y_{n,i}|^p\right])^{2/p} = o(1)$$

**Théoreme 6.3.4** *Supposons que la condition $C(p)$ est vérifiée pour quelques $p > 2$, supposons aussi qu'il existe une suite de nombres réels positifs $\tau_n$, tel que $\tau_n = o(\sqrt{n})$, tel que*

$$\sum_{i=1}^{n}\mathbb{E}\left[|Y_{n,i}|^2\right]\mathbb{1}(|Y_{n,i}| > \tau_n^{-1}) = o(1)$$

*et*

$$\frac{n}{\tau_n}\alpha(\bar{\epsilon}\tau_n) \longrightarrow 0, \forall\bar{\epsilon}$$

*De plus, supposons que*

$$\lim_{n\to\infty}\sum_{i=1}^{n}\mathbb{E}\left[|Y_{n,i}|^2\right] = \sigma^2 > 0$$

*alors*

$$T_n \xrightarrow{\mathcal{D}} \mathcal{N}(0,\sigma^2), \qquad lorsque \quad n \longrightarrow \infty$$

## 6.4 Inégalités exponentielles

**Lemme 6.4.1** ***(Rio 2000, p. 87)***
*Soit $\{U_i, i \in \mathbb{N}\}$ une suite de variables aléatoires réelles centrées et de coefficient de mélange fort $\alpha(n) = O(n^{-\nu}), \nu > 1$, et $\forall n \in \mathbb{N}, 1 \le i \le n \qquad |U_i| < +\infty$. Alors pour tout $r > 1$ :*

$$\mathbb{P}\left\{\left|\sum_{i=1}^{n}U_i\right| > \epsilon\right\} \le C\left(1+\frac{\epsilon^2}{rS_n^2}\right)^{-r/2} + nCr^{-1}\left(\frac{2r}{\epsilon}\right)^{\nu+1}$$

*avec $S_n^2 = \sum_{1\le i,j\le n}|cov(U_i,U_j)|$.*

**Lemme 6.4.2** ***(Borel-Cantelli)***

1. *Soit $A_n$ une suite d'événements*
   *Si*
   $$\sum_{n=1}^{\infty}\mathbb{P}(A_n) < +\infty,$$
   *alors* $\mathbb{P}(\limsup A_n) = 0$.
2. *On suppose maintenant que les événements $(A_n)$ sont indépendants. Si*
   $$\sum_{n=1}^{\infty}\mathbb{P}(A_n) = +\infty,$$
   *alors* $\mathbb{P}(\limsup A_n) = 1$.

**Inégalité** $C_r$

$$\mathbb{E}\left[|X+Y|^r\right] \leq c_r \mathbb{E}\left[|X|^r\right] + c_r \mathbb{E}\left[|Y|^r\right]$$

où $C_r = 1(resp.2^{r-1})$ selon que $r \leq 1(resp.r \geq 1)$

**Remarque 6.4.1** *La convergence presque sûre entraîne la convergence en probabilité.*

## 6.5 Loi forte des grands nombres

**Théoreme 6.5.1** ***(loi forte des grands nombres)***
*Soit $X, X_n$ une suite de variables aléatoires indépendantes et de même loi. Alors $\bar{X}$ converge p.s. si et seulement si* $\mathbb{E}\left[|X|\right] < +\infty$

# Bibliographie


[1] Ahmad, I., Lin, P.E. (1976). Nonparametric sequential estimation of a multiple regression Function, *Bull. Math. Statist.*, 17, 63–75.

[2] Amiri, A. (2009). Sur une famille paramétrique d'estimateurs séquentiels de la densité pour un processus fortemement mélangeant, *C. R. Acad. Sci. Paris, Ser*, I 347, 309–314.

[3] Amiri, A. (2012). Recursive regression estimators with application to nonparametric prediction, *J. Nonparam. Statist,* 24 (1), 169–186.

[4] Amiri, A. (2013). Asymptotic normality of recursive estimators under strong mixing conditions, *arXiv :1211.5767v2.*

[5] Andersen, P. et Gill, R. (1982). Cox's regression model for counting processes : a large sample study. Ann. Statist., 10 : 1100-1120.

[6] Beran, R. (1981). Nonparametric regression with randomly censored survival data, *Technical university of Clifornia, Berkeley.*

[7] Bosq, D. (1989). Proprietes des operateurs de covariance empiriques d'un processus stationnaire hilbertien. C. R. Acad. Sci. Paris Ser. I Math. 309, No.14, 873-875.

[8] Bosq, D. (1990). Modele autoregressif hilbertien. Application a la prevision du comportement d'un processus a temps continu sur un intervalle de temps donne. C. R. Acad. Sci. Paris Ser. I Math. 310, No.11, 787-790.

[9] Bosq, D. (2000). Linear processs in function spaces. Theory and Application. Lectures Notes in Statistics. Vol 149, Springer Verlag.

[10] Bosq, D., Delecroix, M. (1985). Nonparametric prediction of a Hilbertspace valued random variable. Stochastic Process. Appl. 19, 271-280.

[11] Breslow, N. et Crowley, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. Ann. Statist., 2 : 437-453.

[12] Carroll, J. (1976). On sequential density estimation, *Z. Wahrscheinlichkeits- theorie und Verw. Gebiete* , 36, 137–151.

[13] Carbonez, A., Györfi, L., Vander Meulin EC . (1995). Partitioning estimates of a regression function under random censoring, *Statist. & Decisions* , 13, 21–37.

[14] Collomb, G. (1984). Proprietes de convergence presque complete du predicteur a noyau. Z. W. Gebiete. 66, 441-460.

[15] Collomb, G. (1985). Nonparametric regression : an up to date bibliography Statistics. 16, 309-324.

[16] Dabo-Niang, S., Rhomari, N. (2003). Estimation non parametrique de la regression avec variable explicative dans un espace metrique. C. R., Math., Acad. Sci. Paris. 336, 75-80.

[17] Dabrowska, D.M. (1987). Nonparametric regression with censored survival time data, *Scandi. J. Statist,* 14, 181–197.

[18] Dabrowska, D.M. (1989). Uniform consistency of the kernel conditional Kaplan Meier estimate, *Ann. Statist,* 17, 1157–1167.

[19] Davies, I. (1973). Strong consistency of a sequential estimator of a probability density function, *Bull. Math. Statist,* 15, 49–54.

[20] Deheuvels, P. (1973). Sur l'Estimation séquentielle de la densité,*C. R. Acad. Sci., Paris, Ser. A,* 276, 1119-1121.

[21] Deheuvels, P. (1974). Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité, *C. R. Acad. Sci., Paris,* 278, 1217-1220.

[22] Deheuvels, P., Einmahl, JHJ. (2000). Functional limit laws for the increments of Kaplan-Meier product-limit processes and applications, *Ann Proba*, 28, 1301–1335.

[23] Devroye, L., Wagner, T.J. (1980). On the $L^1$ convergence of kernel estimators of regression functions with application in discrimination. *Z. Wahrschein. Verw. Get,* 51, 15-25.

[24] Doob, J. (1953). Stochastic process, Wiley New York.

[25] M. Ezzahrioui, E. Ould-SaÃ¯d. (2006). On the asymptotic properties of a nonparametric estimator of the conditional mode for functional dependent data. Under revision in Statistica Neerlandika.

[26] M. Ezzahrioui, E. Ould-SaÃ¯d. (2008a). Asymptotic normality of the kernel estimators of the conditional mode for functional data. J. Nonparametric Statist., 20, 3-18.

[27] M. Ezzahrioui, E. Ould-SaÃ¯d. (2008b). Asymptotic normality of the kernel estimators of the conditional quantile in the normed space. Far East J. Theoretical Statist., 25, 15-38.

[28] M. Ezzahrioui, E. Ould-SaÃ¯d. (2008c). Asymptotic results of the kernel estimator of the conditional quantile in the normed space under -mixing hypothesis. Comm. Statist. Theory and Methods 37, 2735-2759.

[29] Ferraty, F. (2010). Special issue on statistical methods and problems in infinite dimensional spaces. J. Multivariate Analysis. 101(2), 305-490. 317-344.

[30] Ferraty, F. Laksaci, A., Vieu, P. (2005). Functional time series pre- diction via conditional mode estimation. C. R. Math. Acad. Sci. Paris. 340, 389-392.

[31] Ferraty, F. Laksaci, A., Vieu, P. (2006). Estimation of some characteristics of the conditional distribution in nonparametric functional models. Statistical Inference for Stochastic Processes. 9, 47-76.

[32] Ferraty, F., Van Keilegom, I., Vieu, P. (2009). On the validity of the bootstrap in nonparametric functional regression. Scand. J. Stat. (In press).

[33] Ferraty, F., Vieu, P. (2000). Dimension fractale et estimation de la regression dans des espaces vectoriels semi-normes. C. R. Acad. Sci., Paris. 330, No.2, 139-142.

[34] Ferraty, F., Vieu, P. (2011). Kernel regression estimation for functional data. In the Oxford Handbook of Functional Data Analysis (Ed. F. Ferraty and Y. Romain). Oxford University Press.

[35] Gannoun, A., Saracco, J., Yu, K. (2003), Nonparametric prediction by conditional median and quantiles. J. Stat. Plann. Inference. 117, No.2, 207-223.

[36] Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables, *J. Amer. Statist. Assoc.*, 58, 13-30.

[37] Isogai, E. (1984). Joint asymptotic normality of nonparametric recursive density estimators at a finite number of distinct points,*J. Japan Statist. Soc.*, 14 (2), 125-

[38] Khardani, S., Lemdani, M., Ould Saïd, E. (2010). Some asymptotic properties for a smooth kernel estimator of the conditional mode under random censorship, *J. of the Korean Statistical Society*, 39, 455–469.

[39] Khardani, S., Lemdani, M., Ould Saïd, E. (2011). Uniform rate of strong consistency for a smooth kernel estimator of the conditional mode for censored time series, *J. Stat. Plann. Inference*, 141, 3426–3436.

[40] Kohler, M., Máthé, K., Pinter, M. (2002). Prediction from randomly Right Censored Data, *J. Multivariate Anal*, (80), 73–100.

[41] Krzÿzak, A. (1992). Global convergence of the recursive kernel regression estimates with applications in classification and nonlinear system estimation , *IEEE Trans. Inform. Theory*, 38, 1323–1338.

[42] Laksaci, A. (2005). Contribution aux modèles non paramétriques conditionnels pour variables explicatives fonctionnels. Thèse de doctorat, université de Toulouse.

[43] Laksaci, A. (2007). Convergence en moyenne quadratique de l'estimateur a noyau de la densite conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P. 51, 69-80.

[44] Laksaci, M. and Maref, F. (2009). Conditional cumulative distribution estimation and its applications

[45] Laksaci, A., Madani, F., Rachdi, M. (2010). Kernel conditional density estimation when the regressor is valued in a semi metric space. International Statistical Review. (In press).

*Journal of probability and statistical sciences*, **13**, Pages 47-56.

[46] Lecoutre, J. P., Ould-Said, E. (1993). Estimation de la fonction de hasard pour un processus fortement melangeant avec censure. Publ. Inst. Statist. Univ. Paris. 37, No.1-2, 59-69.

[47] Liang, H.Y. and Baek, J. (2004). Asymptotic normality of recursive density estimates under some dependence assumptions, Metrika 60, 155-166.

[48] Loannides, D., Matzner-Lober, E. (2004). A note on asymptotic normality of convergent estimates of the conditional mode with errors-in-variables. J. Nonparametr. Stat. 16, 515-524.

[49] Loeve, M. (1963). Probability Theory. Third Edition. Van Nostranr Princeton.

[50] Louani, D., Ould-Said, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. J. Nonparametric Statist. 11, No.4, 413-442.

[51] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. Monthly Notices Roy. Astronom Soc., 155, 95118.

[52] Masry, E. (1986). Recursive probability density estimation for weakly dependent stationary processes, IEEE Trans. Inform. Theory, 32, 254-267,

[53] Masry, E. (1987). Almost sure convergence of recursive density estimators for stationary mixing processes, *Statist. Probab. Lett.*, 5, 249-254.

[54] Masry, E. and Györfi, L. (1987). Strong consistency and rates for recursive probability density estimators of stationary processes, *J. Multivariate Anal.*, 22, 79- 93.

[55] Masry, E. and Fan,J., (1997). Local polynomial estimation of recursive function for mixing processes, *Scandinave Journal of S ta tistics*, 24, 165–179.

[56] Mokkadem, A., Pelletier, M., Thiam, B. (2006) Large and moderate deviations principles for recursive kernel estimator of a multivariate density and its partial derivatives. *Serdica Math. J.* **32** , Pages 323–354.

[57] Nadaraya, E. (1964). On estimating regression. Theory Prob. Appl. 10, 186-196.

[58] Nguyen, T., Saracco, J. (2010). Estimation récursive en régression inverse par tranches, *Journal de la société française de statistique*, 151(2), 19–46.

[59] Ould-Said, E. (1997). A note on ergodic processes prediction via estimation of the conditional mode function. Scand. J. Statist. 24, 231-239.

[60] Ould-Said, E., Cai Z. (2005). Strong uniform consistency of nonparametric estimation of the censored conditional mode function. Nonparametric Statistics, 17, 797-806.

[61] Ould-Said, E., Tatachak, A. (2009a). On the nonparametric estimation of the simple mode under random left-truncation model. Romanian Journal of Pure and Applied Mathematics, 54, 243-266.

[62] Ould-Said, E., Tatachak, A. (2009b). Strong consistency rate for the kernel mode under strong mixing hypothesis and left truncation. J. Comm. Statist. Theory Me- thods, 38, 1154-1169.

[63] Quintela-del-Rio, A., Vieu, P. (1997). A nonparametric conditional mode estimate. J. Nonparametr. Statist. 8, No.3, 253-266.
2801.

[64] Rhomari, N. (2002). Approximation et inegalites exponentielles pour les sommes de vecteurs aleatoires dependants. C. R. Acad. Sci. Paris. 334, 149-

[65] Rio, E. (2000). Théorie asymptotique des processus aléaltoires faiblement dépendants, vol 31 Mathematics and applications Springer-Verlag, Berlin.

[66] Rosenblatt, M. (1969). Conditional probability density and regression estimators. In Multivariate Analysis II, Ed. P.R. Krishnaiah. Academic Press, New York and London.

[67] Roussas, G. G. (1968). On some properties of nonparametric estimates of probability density functions. Bull. Soc. Math. Greece (N.S.) 9 29-43.

[68] Roussas, G.G. (1990). Nonparametric regression estimation under mixing conditions, *Stochastic Process. Appl*, 36 (1), 107–116.

[69] Roussas, G. G. (1991). Kernel estimates under association : strong uniform consistency. Statist. Probab. Lett. 12, No.5, 393-403.

[70] Roussas, G.G. and Tran, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions, Annals of Statist. 20 (1), 98-120.

[71] Roussas, G.G. (1992). Ecxact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation, *J. Nonparametr. Stat.* 1, 171-195.

[72] Samanta, M. (1989). Non-parametric estimation of conditional quantiles. Stat. Probab . Lett. 7, No.5, 407-412.

[73] Sarda, P., Vieu, P. (2000). Kernel regression. In : M. Schimek (ed.) Smoothing and regression ; Approaches, Computation, and Application. Wiley Series in Pprobability and Statistics, *Wiley, New York.*

[74] Schwartz, S.C. (1967). Estimation of a probability density by an orthogonal series, Ann. Math. Statist. 38, 1261-1265.

[75] Shorack, G.R. and Wellner, J.A. (1986). Empirical processes with applications to statistics. *Wiley, New-York.*

[76] Takahata, H. (1980). Almost sure convergence of density estimators for weakly dependent stationary processes, Bull. Tokyo Gakugei Univ. Nat. Sci. Ser. IV, 11-32.

[77] Tran, L.T (1989). Recursive density estimation under dependence. IEEE Trans. Inform. Theory 35 (5), 1103-1008.

[78] Vieu, P. (1991). Quadratic errors for nonparametric estimates under dependence. J. Multivariate Anal. 39 (2), 324-347. Watson, G. S. (1964). Smooth regression analysis. Sankhya Ser. A. 26, 359-372.

[79] Walk, H. (2001). Strong universal pointwise consistency of recursive regression estimates, *Ann. Inst. Statist. Math.*, 53 (4), 691–707.

[80] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density. *Ann. Statist.* **7**, Pages 316-327.

[81] Wertz, W. (1985), Sequential and recursive estimators of the probability density. *Statistics* **16**, Pages 277-295.

[82] Wolverton, C.T. and Wagner, T.J. (1969), Asymptotically optimal discriminant functions for pattern classification. *IEEE Trans. Inform. Theory* **15**, Pages 258-265.

[83] Yamato, H. (1971), Sequential estimation of a continuous probability density function and mode. *Bull. Math. Satist.* **14**, Pages 1-12.