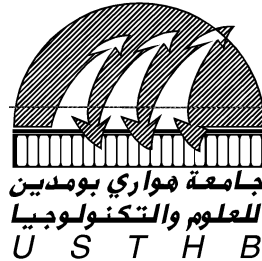


REPUBLIQUE ALGÉRIENNE DEMOCRATIQUE ET POPULAIRE
MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE
UNIVERSITÉ DES SCIENCES ET DE LA TECHNOLOGIE
HOUARI BOUMEDIENNE-ALGER

FACULTÉ DE MATHÉMATIQUES



THÈSE

Présentée pour l'obtention du diplôme de DOCTORAT D'ÉTAT

EN : MATHÉMATIQUES

Spécialité : Probabilités et Statistiques.

par S

Sujet :

ESTIMATION DE LA FONCTION DES QUANTILES
DANS LE MODÈLE DE CENSURE

Soutenue publiquement le 26 octobre 2008, devant le jury composé de :

M. Abdelkader	KHELLADI	Professeur,	U.S.T.H.B.	Président
M. Elias	OULD SAÏD	MCF-HDR,	Univ. du Littoral, France	Directeur de thèse
M. Amar	AISSANI	Professeur,	U.S.T.H.B.	Examineur
Me. Catherine	HUBER	Professeur,	Univ. Paris V, France	Examineur
M. Mounir	MESBAH	Professeur,	Univ. Paris VI, France	Examineur
M. Zaher	MOHDEB	Professeur,	Univ. Constantine	Examineur
Me. Hafida	GUERBYENNE	MCF,	U.S.T.H.B.	Examineur

Résumé

Dans cette thèse, nous nous proposons d'étudier quelques aspects de l'estimation fonctionnelle pour des données incomplètes. Plus précisément, nous nous intéressons à la fonction quantile et à la fonction quantile conditionnelle pour lesquelles nous construisons des estimateurs et nous étudions le comportement asymptotique dans le modèle censuré. Dans un premier temps, nous considérons une suite de v.a. $\{T_i, i \geq 1\}$ stationnaires de fonction de répartition F et fortement mélangeantes, censurées à droite par $\{C_i, i \geq 1\}$ supposées stationnaires et fortement mélangeantes et indépendantes de $\{T_i, i \geq 1\}$, de telle sorte que l'on observe $\{(Y_i, \delta_i), i \geq 1\}$ où $Y_i = T_i \wedge C_i$ et $\delta_i = \mathbb{I}_{\{T_i \leq C_i\}}$ et nous considérons la fonction quantile $\xi_p = \inf \{y; F(y) \geq p\}$ dont nous donnons une représentation forte de son estimateur. Dans un second temps, nous considérons le même modèle que précédemment mais en présence de covariables $\{X_i, i \geq 1\}$. On suppose que $\{(T_i, X_i), i \geq 1\}$ et $\{C_i, i \geq 1\}$ sont indépendants et on s'intéresse à la fonction quantile conditionnelle $\xi_p(x) = \inf \{y; F(y|x) \geq p\}$ pour laquelle nous proposons un nouvel estimateur à noyau convergent tout en précisant la vitesse de convergence. Finalement, nous considérons le quantile conditionnel dans le modèle de censure dans le cas i.i.d., nous définissons un nouvel estimateur lissé. Nous étudions les propriétés asymptotiques de cet estimateur, plus précisément la convergence et la normalité asymptotique avec quelques applications. Ces résultats sont illustrés par des simulations.

Mots clefs : α -mélange, convergence forte, données censurées, estimateur de Kaplan-Meier, fonction des quantiles, quantile conditionnel, normalité asymptotique, représentation forte.

Abstract

In this thesis, we study some aspects of the functional estimation from incomplete data. To be more specific, we study the estimation of the quantile function and the conditional quantile function under the censored model. Firstly, we consider a sequence $\{T_i, i \geq 1\}$ of stationary random variables with a common distribution function F and strongly mixing; censored by $\{C_i, i \geq 1\}$ stationary sequence of mixing random variables and supposed to be independent from $\{T_i, i \geq 1\}$. We observe only the n pairs $\{(Y_i, \delta_i), i \geq 1\}$ where $Y_i = T_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$. We want to estimate the quantile function $\xi_p = \inf \{y; F(y) \geq p\}$. We derive the strong consistency and a Bahadur-type representation of the quantile of the Kaplan-Meier estimator of F . Secondly, we consider the same model as above, but we suppose that we are in presence of an explanatory variable X . We assume that $\{(T_i, X_i), i \geq 1\}$ and $\{C_i, i \geq 1\}$ are independent and we give a kernel estimator of the conditional quantile function $\xi_p(x) = \inf \{y; F(y|x) \geq p\}$ and establish its strong uniform consistency with rate. Finally we introduce a new smooth estimator of the conditional quantile function in the censored model in the i.i.d. case. We show that this estimator converges uniformly almost surely and it is asymptotically normal and we give some applications. Some simulations are given to illustrate our results.

Key words : α -mixing, asymptotic normality, censored data conditional quantile, Kaplan-Meier estimator, quantile function, strong consistency, strong representation.

A ma mère et mon père,
à mes chers frères et
à mon petit neveu Mohamed El Hachemi

Remerciements

Au terme de ce travail, je tiens à remercier vivement Monsieur le Professeur E. Ould-Saïd, qui a bien voulu accepter de diriger ce travail et m'a proposé un sujet très intéressant. J'ai constamment bénéficié de ses suggestions, de ses conseils, de ses encouragements. J'ai trouvé en lui un directeur toujours disponible, ouvert et très soucieux de la valorisation des travaux accomplis. Il m'a guidé et encouragé avec beaucoup de compétence. Ma dette envers lui est immense. Qu'il veuille bien trouver ici, le témoignage de ma profonde gratitude.

Je tiens à remercier vivement le Professeur M. Khelladi, qui m'a fait l'honneur de présider le jury de cette thèse.

Je suis très honorée que Madame et Messieurs les Professeurs C. Huber, H. Guerbyenne, A. Aissani, M. Mesbah et M. Mohdeb aient accepté d'examiner ce travail et faire partie de mon jury.

Il me tient à cœur de remercier le Professeur Mohamed Lemdani, de l'université de Lille II, qui a travers ces commentaires et suggestions a amélioré la présentation du manuscrit.

Je tiens à exprimer ma reconnaissance au Professeur M. Prévost, Directeur du Laboratoire de Mathématiques Pures et Appliquées, Université du Littoral, pour son accueil au laboratoire durant une année.

Mes remerciements vont également au personnel de la Bibliothèque Mathématique Recherche BIBMATH à l'Université des Sciences et de la Technologie de Lille.

Il me tient à cœur d'exprimer ma gratitude à mes frères Smail, Halim, Hacene et principalement Adel, ainsi qu'à mes amis, en particulier Dalila, Abd El Kader, Nassima, Ratiba et Zohra.

A Saida et Farid, j'adresse mes plus vifs remerciements pour m'avoir accueillie au cours des mes séjours à Lille.

Un grand merci à Maghnia et Sihem pour leurs présence et soutien au cours de la dernière année de thèse passée à Lille.

J'exprime ma sympathie à tous les collègues du département de Probabilité et Statistique et de la Faculté de Mathématiques de l'U.S.T.H.B.

Bien entendu 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, pour qui tous les remerciements du monde ne suffiraient pas.

Je remercie aussi les étudiants à qui j'ai dispensé des cours, et surtout ceux dont j'ai eu l'occasion d'encadrer les projets de fin d'études, qui je l'espère ont été aussi enrichissants pour eux qu'ils l'ont été pour moi.

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

Table des matières

Notations	1
1 Introduction	3
2 Données censurées	12
2.1 Durée de vie	12
2.1.1 Les données tronquées	12
2.1.2 Les données censurées	13
2.2 Estimation de la f.d.r. dans le modèle censuré, cas i.i.d.	13
2.3 Propriétés asymptotiques de l'estimateur de Kaplan-Meier	16
3 Données censurées et α-mélangeantes	21
3.1 Données α -mélangeantes	21
3.1.1 Définitions	21
3.1.2 Quelques théorèmes de convergence	22
3.2 Données censurées et α mélangeantes.	24
3.2.1 Propriétés de l'E.K.M.	25
4 La fonction quantile.	29
4.1 Estimation de la fonction quantile pour des données complètes.	29
4.1.1 Cas iid	29
4.1.2 Cas d' α -mélange.	31
4.2 Estimation de la fonction quantile dans le modèle censuré et i.i.d. . .	32
4.3 Représentation forte de la fonction quantile pour des données censurées et α -mélangeantes	34
4.3.1 Introduction	34
4.3.2 Assumptions and Results	36
4.3.3 Auxiliary results and proofs	38
5 La fonction quantile conditionnelle	43
5.1 Introduction	43
5.1.1 Cas de données complètes	44
5.1.2 Cas de données censurées	45
5.2 Convergence presque sûre de l'estimateur à noyau du quantile condi- tionnel pour des variables censurées et dépendantes	47

5.2.1	Introduction	48
5.2.2	Assumptions and main results	48
5.2.3	Simulation study	50
5.2.4	Auxiliary results and proofs	50
5.3	Prevision via le quantile conditionnel dans le modèle censuré	57
5.3.1	Introduction	57
5.3.2	Definition of the new estimator	58
5.3.3	Assumptions and main results	61
5.3.4	Applications	65
	5.3.4.1 <i>Applications to prediction.</i>	65
	5.3.4.2 <i>Confidence intervals.</i>	65
5.3.5	Simulations Studies	66
	5.3.5.1 <i>Consistency</i>	66
	5.3.5.2 <i>Asymptotic normality</i>	66
	5.3.5.3 <i>Confidence curves</i>	68
5.3.6	Auxiliary results and proofs	69
6	Conclusion	82
6.1	Commentaires	82
6.2	Perspectives de recherche	83
	Bibliographie	85

Notations

Nous regroupons ici, les différentes notations et abréviations utilisées tout au long du document.

- $(\Omega, \mathcal{A}, \mathbb{P})$: espace probabilisé.
- i.i.d. : indépendantes et identiquement distribuées.
- f.d.r. : fonction de répartition.
- L.L.I. : Loi du Logarithme Itéré.
- T.C.L. : Théorème Central Limite.
- $\mathbb{I}_A(\cdot)$: fonction indicatrice sur l'ensemble A .
- $\mathcal{N}$: loi normale.
- $\xrightarrow{\mathcal{L}}$: convergence en loi.
- $\xrightarrow{\mathbb{P}}$: convergence en probabilité.
- p.s. : presque sûrement.
- $u \wedge v := \min(u; v)$.
- $(a, b] = \{x; x > a \text{ et } x \leq b\}$.
- $[a, b) = \{x; x \geq a \text{ et } x < b\}$.
- $(a, b) = \{x; x > a \text{ et } x < b\}$.

Les travaux présentés dans ce document ont pour certains fait l'objet de publication. D'autres sont soumis ou en cours de révision.

- OULD-SAÏD, E. AND SADKI, O. (2005). Strong approximation of quantile function for strong mixing and censored processes. *Comm. Statist. Theory and Methods*, **34**, 1449-1459.
- OULD-SAÏD, E. AND SADKI, O. (2008). Prediction via the conditional quantile for right censorship model. A paraître dans Far East Journal of Theoretical Statistics

Chapitre 1

Introduction

La recherche dans les domaines tels que la médecine, l'assurance, la fiabilité, la démographie,... n'ont à première vue rien en commun. Toutefois, dans chacun de ces domaines, on s'intéresse à des variables aléatoires (v.a.) représentant la durée de temps jusqu'à l'apparition d'un certain événement, appelées durées de survie. En médecine, elle est la durée de survie ou le temps de rémission d'un patient, alors qu'en industrie, cette variable peut être la durée séparant deux pannes successives d'une machine. Dans le domaine économique, elle est la durée de chômage ou le temps passé dans un emploi, la durée de vie d'une entreprise, etc...

Soit T , une v.a. représentant la durée de survie d'un individu issu d'une population donnée. T peut être caractérisée de plusieurs manières, la plus importante étant sa fonction de répartition $F(t) = \mathbb{P}[T \leq t]$. Pour des données complètes, c'est-à-dire lorsque T est complètement observée et qu'on dispose d'un échantillon de taille n , $\{T_i; 1 \leq i \leq n\}$ de la v.a. T , un estimateur naturel de F est alors F_n , la fonction de répartition empirique $F_n(t) = \frac{\#\{T_i : T_i \leq t\}}{n}$; les propriétés de $F_n(\cdot)$ sont largement connues dans la littérature statistique (voir par exemple chapitre 19 dans Van der Vaart (1998)).

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 et 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 être plus grande qu'une variable de troncature Z ; ainsi les observations ne sont possibles que pour $T \geq Z$. Il s'agit d'un modèle que l'on peut trouver, par exemple en astronomie où des échantillons sont composés d'objets astraux d'une certaine zone. On s'intéresse à la luminosité M absolue ou apparente d'un objet astral 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 lesquels la luminosité $M \geq m$; m étant la variable de troncature, ce sont les 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 cas se présente par exemple, quand le chercheur ne dispose pas d'assez de temps pour attendre que toutes les observations atteignent l'événement d'intérêt. En industrie, il n'est pas toujours possible d'attendre que toutes les machines tombent en panne et en médecine, il n'est pas d'éthique pour le médecin d'attendre que tous les patients sur lesquels porte l'étude meurent. Ce genre de données est composé de deux parties : d'une part nous avons les observations montrant l'événement d'intérêt à un certain instant (observations non censurées) et d'autre part, les observations qui n'ont pas encore montré l'événement d'intérêt (observations censurées).

Pour estimer la fiabilité d'un élément, on observe la vie de n éléments identiques ; pratiquement, on part donc de la donnée de n variables aléatoires $T_1, T_2, \dots, T_n$ correspondant aux n durées de vie observées et recueillies après la mise en test de n machines à la sortie d'une chaîne de production. Supposons qu'il soit décidé d'arrêter le test des n machines à une date C fixée *a priori*. Les observations dont nous disposons ne sont plus les durées de vie $T_1, T_2, \dots, T_n$, mais seulement les durées de vie inférieures à C pour les machines déjà tombées en panne. Pour les autres machines, la durée de vie observée est supérieure à C . En réalité nous disposons des observations $Y_1, Y_2, \dots, Y_n$ définies par $Y_i = T_i \wedge C$; il s'agit alors d'un modèle censuré de type I. Une généralisation de ce modèle est le modèle censuré de type II, dans lequel les dates auxquelles les censures sont effectuées sont aléatoires. L'intérêt pratique d'un tel modèle est évident, car lorsque nous observons le fonctionnement de plusieurs grands systèmes, les dates à partir desquelles nous perdons l'information ne peuvent être déterminées à l'avance.

Les données censurées à droite sont fréquemment utilisées lors d'études sur les durées de vie. Dans une étude médicale où on s'intéresse à l'évolution d'un certain type de cancer, certains patients peuvent être censurés parce que l'étude fut arrêtée avant que leur cancer ne progresse, ou décédés d'une autre cause que la maladie en question, ou encore perdus de vue pour cause de déménagement.

Un exemple, très souvent cité dans la littérature statistique médicale est celui des données de Freireich de 1963 (voir Dreesbeke *et al.* 1989) obtenues lors d'un essai thérapeutique qui avait pour but de comparer les durées de rémission, exprimées en semaines, de sujets atteints de leucémie selon qu'ils ont reçu de la 6-mercaptopurine ou un placebo. Douze sujets parmi les 42 considérés sont perdus de vue et on ne dispose pas de leurs durées de rémission mais seulement des dates de leurs sorties de l'étude. On dit que les perdus de vue ont été censurés, et ce problème de censure demande un traitement particulier. En effet si l'on se contentait d'éliminer les observations incomplètes, on perdrait beaucoup d'information et l'échantillon qui en résulterait ne serait plus représentatif de la même population.

Dans le modèle censuré, chacune des observations de l'échantillon consiste en un

couple de valeurs : (Y, δ) avec Y la durée observée et δ une variable binaire représentant la nature de la durée observée qui prend la valeur 1 s'il s'agit d'une vraie durée de vie et 0 si c'est une censure.

Autrement dit, à chaque individu i ; $i = 1, 2, \dots, n$ est associé un couple de durée (T_i, C_i) , de v.a., dont seule la plus petite est observée; T_i est la durée de survie et C_i est l'instant de censure. On sait de plus quelle est la nature de cette durée : si $\delta_i = 1$ c'est une survie, si $\delta_i = 0$ c'est une censure.

En juin 1958, E. Kaplan et P. Meier publient dans le journal de l'American Statistical Association un article intitulé "Non parametric estimation from incomplete observations". Dans ce travail, Kaplan et Meier définissent l'estimateur "PL" de la fonction de survie $S(t) = \mathbb{P}(T < t)$ [PL pour produit limite, ainsi nommé car il s'agit de la limite de la décomposition de la probabilité conditionnelle en produit de probabilités], le comparent à l'estimateur sur l'échantillon réduit (c'est à dire calculé sur la population des individus observables tout au long de l'expérience), étudient son espérance et sa variance et montrent ainsi l'intérêt de l'estimateur PL.

Notons $Y_{(1)}, Y_{(2)}, \dots, Y_{(n)}$ les valeurs ordonnées des Y_i et pour chacune des valeurs $Y_{(i)}$ la valeur δ_i correspondante. D'où le schéma suivant :

durée de vie $Y_{(i)}$	$\leftarrow$ — $\rightarrow$ $\leftarrow$ — — — $\rightarrow$ $\leftarrow$ — — — — — $\rightarrow$ $\leftarrow$ — —
	$Y_{(1)}$ $Y_{(2)}$ $Y_{(3)}$ $Y_{(4)}$
nature de la survie δ_i	1 0 0 0
valeur observée T_i	$T_1 = Y_{(1)}$ $T_2 > Y_{(2)}$ $T_3 > Y_{(3)}$

A l'instant t , l'intervalle correspondant est celui de $Y_{(i)}$, parmi les n individus il reste R_i individus à risque ou présent (ni décédés, ni censurés). La probabilité de survie à

l'instant t , est estimée par $\hat{S}_n(t) = \prod_{i=1, Y_i < t, T_i < C_i}^n \left(\frac{R_i - 1}{R_i} \right)$

La formule se simplifie agréablement si l'on peut supposer qu'il n'y a pas d'ex-aequo, et nous retrouvons l'expression classique de l'estimateur de Kaplan Meier (1958) :

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

Depuis, de nombreux auteurs se sont intéressés au domaine, mais reste cependant un sujet d'actualité comme le prouvent les recents papiers de Cao *et al* (2005) et Ould Saïd (2006). Les premiers travaux ont porté sur le cas où les durées de survie et les instants de censure sont deux suites indépendantes de v.a. indépendantes et identiquement distribuées. L'estimateur de Kaplan-Meier (E.K.M) $\hat{F}_n$, de la fonction de répartition F possède beaucoup de propriétés analogues à celles de la fonction de répartition empirique.

Le comportement asymptotique a suscité l'intérêt d'un grand nombre d'auteurs ; Bres-

low et Crowley (1974) furent les premiers à traiter de la convergence de l'E.K.M. et de sa normalité asymptotique ; leurs résultats furent repris par Aalen (1978), Gill (1980, 1983), Shorack-Wellner (1986).

Un premier type d'approximation du processus $\left(\widehat{F}_n - F\right)$ par une moyenne de v.a. i.i.d. fut obtenu par Burke, Csörgő et Horváth (1981, 1986) avec un reste de l'ordre de $O\left(\frac{\log^2 n}{n^{1/2}}\right)$; un second type d'approximation fut établi par Lo et Singh (1986) avec un taux de $O\left(\left(\frac{\log n}{n}\right)^{3/4}\right)$ presque sûrement ; Ce taux a été ensuite amélioré à $O\left(\frac{\log n}{n}\right)$ par Lo, Mack et Wang (1989).

Breslow et Crowley (1974), Gill (1980), Burke *et al*(1981, 1986), Major et Rejtő (1988) et Csörgő (1996) se sont intéressés à la normalité asymptotique du processus $\sqrt{n}\left(\widehat{F}_n - F\right)$. le livre de Shorack et Wellner (1986) donne le détail de certains résultats.

Nous rappelons, dans le premier chapitre de cette thèse, les principaux résultats de la littérature concernant l'estimateur de Kaplan-Meier.

Dans la réalité, il arrive souvent que les durées de vie étudiées présentent une certaine forme de dépendance. Les données dépendantes sont naturellement présentes dans beaucoup de disciplines comme la biologie, la médecine, les sciences sociales, le terme générique couvre actuellement une multitude de structures de données.

Dans le cas d'un long suivi médical, il n'est pas rare d'observer des données dépendantes et censurées, comme c'est le cas de patients diabétiques suivis pour voir l'évolution de la maladie et de ses complications. Dans " the Diabetes Control and Complications Trial", research group 1986, les patients sont aléatoirement répartis en deux groupes, le premier recevant le traitement standard et second le traitement expérimental. Le but de l'étude est d'établir une relation entre le contrôle de glycémie et des complications vasculaire d'un diabétique. Dans cet essai thérapeutique, un des principaux résultat est le suivi de l'examen du fond d'œil des yeux des patients. Pour les patients de chaque groupe, chaque œil est considéré comme un élément de l'échantillon et l'E.K.M. est calculé sur la base de l'échantillon de tous les yeux. Mais les deux yeux d'un même malade sont naturellement corrélés et deux yeux de deux malades sont indépendants.(voir Hansen K.W. 1996) Les données ne sont alors pas indépendantes mais on peut considérer que, asymptotiquement il y a indépendance.

Mathématiquement cette dépendance est modélisée par les données mélangeantes. Il existe plusieurs types de mélanges qui sont définis à partir de coefficients, notés, selon le cas, α , β , Φ , Ψ , ou ρ . Ces coefficients sont nuls pour des données indépendantes mais sont non nuls lorsqu'il y a dépendance. Parmi toutes ces formes de

mélanges, l' α -mélange est la plus faible et donc la moins restrictive. Toute suite de v.a. β , Φ , Ψ , ou ρ -mélangeante sera donc forcément α -mélangeante. En particulier les processus AR et ARMA, qui sont largement utilisés dans l'analyse des séries temporelles, sont α -mélangeantes avec des coefficients géométrique et exponentiel respectivement. Auestad et Tjøstheim (1990) fournissent des discussions intéressantes sur le rôle de l' α -mélange pour l'identification du modèle dans le cadre de l'analyse des séries temporelles non linéaires.

Une importance de plus en plus croissante est portée au cas où les durées de survie et (ou) les instants de censure présentent une forme de dépendance, en raison des diverses applications dans différents domaines. Voekel et Crowley (1984) utilisent l'approche des données dites semi-markoviennes pour établir un modèle pour des données de survie sur le cancer. On peut trouver plus de détails sur les applications de l'E.K.M. pour des données dépendantes en médecine et en agriculture dans Wei *et al.* (1989), Koehler (1995), Koehler et Symanawski (1995) et Kang et Koehler (1997).

Il est important de savoir si l'E.K.M. garde ses propriétés de consistance et de normalité asymptotique entre autres pour des données dépendantes. Ying et Wei (1994) étudient la consistance et la normalité asymptotique de l'E.K.M. dans le cas du Φ -mélange; Cai et Printice (1995), Kang et Koehler(1997) et Cai et Roussas (1998) traitent de la normalité asymptotique de l'E.K.M. sous différentes formes de dépendance.

Dans cette thèse, nous nous sommes intéressé au modèle censuré dans le cas d' α -mélange ou mélange fort. C'est un thème très privilégié ces dernières années. Cai (1998, 2001) a établi la représentation forte et la normalité asymptotique de l'E.K.M., il montre la consistance forte et donne le taux de convergence et étudie le processus stochastique associé à l'E.K.M. qui converge faiblement vers un processus gaussien de structure de covariance spécifique. Après avoir rappelé les définitions des différents types de mélange, nous présentons au chapitre 2, quelques résultats récents relatifs à l'étude du E.K.M. et ses propriétés de convergence dans le cas d' α -mélange.

L'estimation des paramètres fonctionnels présents dans le cas du modèle de censure a donné lieu à une très riche littérature, nous citons par exemple les travaux de Lecoutre et Ould Saïd (1995b) portant sur l'estimation de la fonction de hasard, Lemdani et Ould Saïd (2001) se sont intéressé à l'estimation de la fonction de risque pour des données tronquées et censurés, Ould Saïd et Cai (2005) ont traité le mode conditionnel.

La fonction des quantiles simples est une caractéristique très importante en statistique et qui a suscité l'intérêt d'un grand nombre d'auteurs. La fonction des quantiles n'est autre que l'inverse de la f.d.r. $F : F^{-1}(p) = \inf\{t; f(t) \geq p\}$ pour tout $p \in (0, 1)$; la médiane et les quartiles n'en sont que des cas particuliers.

Pour ne citer qu'un exemple, des diverses applications de la fonction quantile, nous citons le VaR (the Value at Risk) utilisé dans le domaine des finances, la VaR est

généralement utilisé pour mesurer le risque du marché d'un portefeuille d'instruments financiers et correspond au montant des pertes qui ne devrait pas être dépassé avec une probabilité donnée. Pour tout α dans $(0, 1)$, $\text{VaR}(\alpha)$ n'est rien d'autre que le quantile d'ordre $(1 - \alpha)$ de la distribution du retour de l'actif.

Pour des données complètes, la fonction quantile est estimée par l'inverse généralisé de la fonction de répartition $F_n^{-1}(\cdot)$; ses propriétés asymptotiques ont été rappelées en particulier par R.J. Serfling (1980), M. Csörgö (1983), Shorack et Wellner (1986) et Van der Vaart (1998).

Plusieurs auteurs se sont intéressés à l'estimation de la fonction quantile pour des données complètes mais dépendantes, nous citons en particulier Sen (1972) qui donna une représentation de type Bahadur pour une suite de v.a. Φ -mélangeante. Babu et Sing (1978) établirent la représentation de type Bahadur; au sens que la différence entre l'estimateur et la fonction quantile peut s'écrire sous la forme d'une moyenne de v.a. plus un reste de l'ordre de $O\left(\left(\frac{\log n}{n}\right)^{\frac{3}{4}}\right)$ pour F_n^{-1} sous l'hypothèse de mé-

lange fort lorsque le coefficient de mélange est exponentiel. Yoshihara (1995) montra le résultat lorsque les v.a. sont uniformément bornées et pour un taux de mélange polynomial; Sun (2006) améliore le résultat de Yoshihara.

La représentation forte du quantile convenablement normalisé fut obtenue par Fotopoulos *et al* (1994) et amélioré par Yu (1996). Dans le modèle censuré, le cas i.i.d. Sander (1975) fut le premier à suggérer d'estimer $F^{-1}(p)$ par $\hat{\xi}_{p,n} = \inf\{t; \hat{F}_n(t) \geq p\}$ et étudia la convergence faible. Reid (1981) s'intéressa à l'estimation de la médiane. Horvath (1982) étudia l'approximation forte du quantile, K.F. Cheng (1984) en donna une représentation presque sûre. Aly *et al* (1985) donnèrent une approximation du processus $\sqrt{n}(\hat{\xi}_{p,n} - F^{-1}(p))$ par un processus de Kiefer. Lio *et al* (1986), Padgett et Thombs (1989) et Lio et Padgett (1997) proposèrent des estimateurs à noyau de la fonction quantile. Xiang (1995) traita l'inefficacité de l'estimateur à noyau du quantile.

Notre premier travail porte sur la fonction quantile dans le cas où les durées de survie et les instants de censure sont deux suites fortement mélangeantes.

L'estimateur de ξ_p est défini par $\xi_{p,n} = \hat{F}_n^{-1}(p) = \inf\{y; \hat{F}_n(y) \geq p\}$; où $\hat{F}_n$ est l'E.K.M. de F . Notre hypothèse principale porte sur le coefficient de mélange qui doit être de l'ordre de $O\left(e^{-\log^{1+\zeta} n}\right)$ pour un certain $\zeta > 0$. Sous cette condition, nous éta-

blissons la convergente presque sûre $\hat{F}_n^{-1}$. Puis nous donnons une représentation forte de la différence $\hat{F}_n^{-1}(p) - F^{-1}(p)$ comme une moyenne de v.a. avec un reste d'ordre de $o(\sqrt{\log \log n/n})$. Si de plus F est deux fois continûment dérivable en $F^{-1}(p)$, alors

le reste devient $O\left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}}\right)$ p.s. où $\beta > 0$ et $\lambda > 0$. Nous remarquons que ce taux est différent de $O\left(\frac{\log n}{n}\right)$ obtenu par Lo *et al* (1989) pour le cas i.i.d. car notre

approche utilise l'approximation forte de l'E.K.M. par une moyenne de v.a. avec un reste de l'ordre de $\frac{\log^{-\lambda} n}{\sqrt{n}}$.

L'application du Théorème 1, Théorème 2 de Cai (2001) et du Théorème 18.5.4 d'Ibragimov et Linnick (1981), nous permet de d'établir la normalité asymptotique de la suite $\sqrt{n} \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right)$, en donnant de manière explicite le terme de variance.

Lors d'études pratiques sur les durées de vie, par exemple en médecine, les chercheurs ne se contentent pas de relever uniquement la durée de vie observée; mais ils procèdent au relevé de nombreuses données telles que les antécédants familiaux, le taux de glycémie dans la sang, l'âge, la densité de globules rouges, etc... Ces données ne sont généralement pas indépendantes de la durée de vie, certaines d'entre elles pouvant être des circonstances aggravantes pour la maladie considérée. La connaissance de ces données pour un patient, peut permettre une estimation plus fiable de sa durée de vie. Dans un tel cas, en regroupant toutes les mesures annexes dans un vecteur de dimension d , noté X , il sera plus intéressant d'estimer le quantile conditionnel de T sachant X que le quantile de T sans d'autre information.

Dans le problème de la prévision, la médiane conditionnelle, qui est un cas particulier du quantile conditionnel, par sa robustesse offre une bonne alternative aux méthodes standards basées sur la moyenne. En absence de censure, nous citons en particulier les travaux de Samanta (1989) qui étudia la consistance forte et la normalité asymptotique d'un estimateur à noyau du quantile conditionnel dans le cas i.i.d., Honda (2000) traita l'estimation de la fonction quantile conditionnelle pour de processus α -mélangeants et Berlinet *et al.* (2001) qui donnèrent des résultats de convergence et la normalité asymptotique de la fonction quantile conditionnelle dans le cas dépendant avec une application au cas du mélange fort.

Le premier travail concernant l'E.K.M. généralisé, en présence de covariables revient à Beran (1981) et Miller-Halpern (1982). Dabrowska (1987, 1989) établit la consistance uniforme de l'estimateur à noyau de la fonction de hasard cumulative conditionnelle puis la convergence faible de $\sqrt{nh_n} \left(\hat{F}_n(\cdot) - F(\cdot) \right)$; où h_n est une suite spécifique, elle obtient ainsi des résultats analogues au théorème 4 de Breslow et Crowley (1974) et Dabrowska (1995) traita l'estimation de la fonction de survie conditionnelle lorsque la variable d'intérêt et la covariable sont toutes les deux sujettes à des censures. Ganzález Manteiga et Cadarso Suárez C (1994) étudièrent aussi les propriétés asymptotiques de l'E.K.M. généralisé.

Dabrowska (1992) donna une représentation de type Bahadur d'un estimateur à noyau du quantile conditionnel dans le modèle censuré et i.i.d., Lecoutre et Ould Saïd (1995a) traitèrent la convergence d'un estimateur lissé du quantile conditionnel dans le modèle censuré dans le cas de l' α -mélange. Kohler *et al.* (2002) s'intéressèrent à la prévision pour des données censurées par la moyenne et récemment Ould Saïd (2006) établit la convergence presque sûre uniforme avec vitesse de convergence d'un estimateur à noyau de la fonction quantile.

Notre deuxième travail a été consacré à l'estimation des quantiles conditionnels ; on est en présence d'une covariable X et pour tout $x \in \mathbb{R}$, on note par $F(.|x)$ la f.d.r. conditionnelle de T_1 sachant $X = x$ et on observe la suite $\{(Y_i, \delta_i, X_i); i \geq 1\}$. Dans le deuxième paragraphe du chapitre 4, nous supposons que la suite $\{(Y_i, \delta_i, X_i); i \geq 1\}$ est α -mélangeante et que $\{C_i; i \geq 1\}$ et $\{(X_i, T_i); i \geq 1\}$ sont indépendantes. Pour tout $p \in (0, 1)$, le quantile conditionnel d'ordre p de F sachant que $X = x$ est donné par $\xi_p(x) = \inf \{y; F(y|x) \geq p\}$. Nous considérons un estimateur à noyau $F_n(.|x)$ de la f.d.r. conditionnelle $F(.|x)$. Remarquons que cet estimateur est le même estimateur que celui de Ould-Saïd (2006) et est analogue à celui de Kohler *et al* (2002) en remplaçant Y_i par $\mathbb{I}_{\{Y_i \leq t\}}$. Un estimateur naturel de $\xi_p(x)$ est donné par

$$\xi_{p,n}(x) = \inf \{t \in \mathbb{R} : F_n(t|x) \geq p\}.$$

Sous des hypothèses de régularité, nous obtenons la vitesse de convergence presque sûre sur un compact, de l'estimateur de la f.d.r. conditionnelle ainsi que celui du quantile conditionnel. Nous obtenons en particulier le taux de convergence de l'ordre

$$\text{de } O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n^2}}, h_n \right\} \right).$$

Nous faisons remarquer que notre vitesse de convergence est assez proche du taux optimal dans le cas i.i.d.

A échantillon fini, nous étudions, à l'aide de simulations les performances de cet estimateur pour un modèle particulier.

Finalement, au cours de notre troisième travail, nous considérons le cas où $\{X_i, i \geq 1\}$, $\{T_i, i \geq 1\}$ et $\{C_i, i \geq 1\}$ sont des suites de v.a. i.i.d. et telles que $\{C_i, i \geq 1\}$ et $\{(X_i, T_i), i \geq 1\}$ sont indépendantes. Alors que d'autres auteurs, comme Dabrowska (1992), font l'hypothèse que $\{T_i, i \geq 1\}$ et $\{C_i, i \geq 1\}$ sont conditionnellement indépendants sachant X . Notre condition est dans la pratique très plausible chaque fois que la censure ne relève pas de l'état du patient. De plus celle-ci offre l'avantage de nous permettre de construire un estimateur sans biais de $F(.|.)$.

Nous construisons un nouvel estimateur lissé $F_n(.|.)$ pour la f.d.r. conditionnelle $F(.|.)$ et par conséquent un estimateur de la fonction quantile conditionnelle. Nous établissons que $F_n(.|.)$ est consistant, puis nous montrons la normalité asymptotique du processus $(F_n(.|.) - F(.|.))$ convenablement standardisé et nous déduisons que l'estimateur de la fonction quantile conditionnelle converge presque sûrement et que convenablement standardisé, il est asymptotiquement gaussien tout en explicitant le terme de la variance ; ce qui nous permet d'appliquer nos résultats à la prévision et de construire des intervalles de confiance. Des simulations viennent illustrer ces résultats et prouvent leur performance à échantillon fini.

Cette thèse est organisée en 6 chapitres. Le deuxième chapitre propose une présentation des données incomplètes puis la définition et quelques propriétés de l'E.K.M. dans le cas i.i.d. nous y énonçons les principaux résultats de Breslow et Crowley (1974) sur

la convergence de l'estimateur de Kaplan-Meier, puis nous donnons l'approximation forte établie par Lo et Singh (1985).

Après un rappel sur les différents types de mélange, nous donnons dans le troisième chapitre les versions de certains théorèmes classiques en statistiques pour des données α -mélangeantes, en particulier les inégalités de type exponentiel de Bosq (1998) et la loi du log itéré établie par Cai et Roussas (1992); nous abordons l'étude de l'E.K.M. et ses propriétés de convergence dans le cas de l' α -mélange. Nous présentons essentiellement les travaux de Cai. Z (1998, 2001) sur l'E.K.M. pour des données censurées et α -mélangeantes

Dans le chapitre 4 nous nous intéressons à la fonction quantile simple, nous présentons d'abord une représentation forte de l'estimateur empirique du quantile et la représentation de type Bahadur pour des données complètes et i.i.d.; puis nous présentons les travaux de Yoshihara (1995) sur la représentation de type Bahadur lorsque les v.a. sont uniformément bornées et α -mélangeantes avec un coefficient de mélange polynomial. Nous donnons par la suite une représentation forte de la fonction quantile dans le modèle de censure dans le cas où les durées de vie et les instants de censure sont α -mélangeants; ces résultats sont repris dans leurs versions parues.

Dans le chapitre 5, nous supposons qu'une autre source d'information est disponible au travers d'une v.a. X de densité ℓ , corrélée avec la variable d'intérêt T la loi qui nous intéresse est celle de T sachant $X = x$; nous présentons brièvement les travaux de Dabrowska (1992) sur l'estimation par le noyau de la fonction quantile conditionnelle puis ceux d'Ould-Saïd (2006) dans le cas i.i.d.

Nous donnons, par la suite dans un premier temps, un résultat sur la convergence forte d'un estimateur à noyau pour le modèle censuré dans le cas de l' α -mélange.

Dans un second temps nous établissons la consistance et la normalité asymptotique d'un estimateur à noyau du quantile conditionnelle dans le modèle de censure avec des données i.i.d. avec des applications à la prévision; ces résultats sont illustrés par des simulations qui confirment la performance de nos estimateurs à taille finie.

les deux derniers résultats que nous présentons ici sont soumis à parution et nous avons pris la liberté de les présenter en anglais.

Le dernier chapitre présente les conclusions de ce travail de thèse, ainsi que quelques extensions possibles.

Chapitre 2

Données censurées

2.1 Durée de vie

Nous appelons durée de vie une v.a. généralement positive T . Au niveau des applications, cette variable représente la durée passée dans un état donné (la durée de chômage, la durée de l'activité, la durée d'attente à un guichet, la durée de vie d'un malade, le temps de remission d'un malade,...) ou la durée séparant deux événements (la durée séparant deux pannes successives, la durée passée entre deux emplois, ...). Les v.a. durées de vie trouvent des applications dans pratiquement tous les champs d'applications des statistiques ; la médecine, l'économie, la fiabilité, la démographie, l'assurance, etc...

Lorsqu'on dispose d'un échantillon $\{T_i; 1 \leq i \leq n\}$ de la v.a. T tel que chaque donnée T_i de l'échantillon, durée de vie du i^{eme} individu, est observée dans sa totalité nous dirons que les données sont complètement observées ou il s'agit de données complètes.

Dans ce cas un estimateur naturel de la f.d.r. F de T est alors l'estimateur empirique F_n défini par $F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{T_i \leq y\}}$. Les propriétés de $F_n(\cdot)$ sont largement diffusées dans la littérature statistique.

La principale source de difficulté dans l'analyse des durées de vie et pour diverses raisons, est la présence de données incomplètes ; pour lesquelles la variable d'intérêt n'est pas complètement observée pour toutes les données de l'échantillon. Nous présentons brièvement deux cas de données incomplètes ; les données tronquées et les données censurées.

2.1.1 Les données tronquées

Lors d'une étude pratique sur les durées de vie, il n'est pas rare que la variable d'intérêt ne soit pas observable quand elle est inférieure à un seuil C ou une v.a. C , nous dirons qu'on est en présence d'une troncature gauche. La troncature élimine donc de l'étude une partie de l'échantillon, ce qui aura pour conséquence que l'analyse pourra porter seulement sur la loi conditionnelle de T sachant que $T \geq C$. La troncature gauche apparaît, par exemple dans la recherche d'objets cachés qui devront être assez

grands pour être détectés comme les réserves de pétrole, les astres lointains,... Dans ce cas, nous disposons de N objets dans l'échantillon, mais nous ne sommes capables d'observer que les n objets suffisamment grands (supérieurs à C).

Lyndell-Bell (1971) proposa une estimation non paramétrique de la f.d.r. de T et Woodroffe (1985) posa le cadre de l'estimation dans le modèle de troncature.

2.1.2 Les données censurées

Un autre cas où les données incomplètes apparaissent est celui des données censurées à droite ; il se présente quand le chercheur ne dispose pas d'assez de temps pour attendre que toutes les observations atteignent l'événement d'intérêt. Il s'agit du cas où la variable d'intérêt n'est pas observée sur toute la durée de l'étude.

En effet pour tester la fiabilité d'un système, on met en état de fonctionnement n systèmes du même type, et note la durée de vie T jusqu'à la première panne. La durée de l'étude est C .

T étant la variable d'intérêt, si au lieu d'observer T on observe une constante C et si on sait que $T > C$, on dit qu'il y a une censure droite de type I.

Il arrive souvent que la censure soit propre à l'élément de l'échantillon qui est censuré. C'est le cas lors d'une étude médicale où des patients sont perdus de vue pour des raisons différentes ; pour cause d'un déménagement ou d'un décès dû à une cause autre que la maladie sur laquelle porte l'étude ; et à des moments différents. Notons qu'il s'agit du type de censure le plus fréquemment utilisé.

Étant donné une v.a. T , on dit qu'il y a censure aléatoire droite (ou censure de type II) s'il existe une v.a. C , telle que au lieu d'observer T on observe : $Y = T \wedge C$ et $\delta = \mathbb{1}_{\{T \leq C\}}$.

δ sert en fait à connaître la nature de l'observation, il indique si l'on est face à une observation réelle ($\delta = 1$) ou à une censure ($\delta = 0$).

Autrement dit, à chaque individu i ; $i = 1, 2, \dots, n$ est associé un couple de durées (T_i, C_i) , de v.a. positives, dont seule la plus petite est observée ; T_i est la durée de survie et C_i est la censure, mais seule $Y_i = T_i \wedge C_i$ est observée. On sait de plus quelle est la nature de cette durée : si $\delta_i = 1$ c'est une survie, si $\delta_i = 0$ c'est une censure.

2.2 Estimation de la f.d.r. dans le modèle censuré, cas i.i.d.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Soit $T_1, \dots, T_n, \dots$ une suite de v.a. positives indépendantes et identiquement distribuées (i.i.d.) de f.d.r. commune F et de densité continue f .

Soit $C_1, \dots, C_n, \dots$ une suite de v.a. de censure, positives, i.i.d. de f.d.r. G .

Généralement et conformément aux études biomédicales, les v.a. C_i sont supposées être indépendantes des T_i .

Soit $\{(Y_1, \delta_1), \dots, (Y_n, \delta_n)\}$ l'échantillon réellement observé où

$Y_i = T_i \wedge C_i$ et $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$.

Les Y_i sont de f.d.r. H donnée par $\overline{H}(t) = 1 - H(t) = \overline{F}(t)\overline{G}(t), \forall t \geq 0$.

Pour toute f.d.r. L , nous noterons par $\tau_L = \sup\{t : L(t) < 1\}$.

Nous notons par H_1 et H_0 les fonctions de répartition des v.a. non censurées et censurées

$$H_1(t) = \mathbb{P}(Y_i \leq t, \delta_i = 1) = \mathbb{P}(T_i \leq t \wedge C_i) = \int_0^t (1 - G(s)) dF(s)$$

et

$$H_0(t) = \mathbb{P}(Y_i \leq t, \delta_i = 0).$$

Clairement $\overline{H}(t) = \overline{H}_0(t) + \overline{H}_1(t)$.

Avant d'aborder l'estimation de la f.d.r. pour des données censurées, nous avons besoin donner quelques définitions :

$Y_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \geq t\}}$, le nombre de sujets présents jusqu'à t ,

$N_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \geq t, \delta_i = 1\}}$, le nombre d'observations non censurées inférieures ou égales à t

et $M_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \geq t, \delta_i = 0\}}$, le nombre d'observations censurées inférieures ou égales à t .

Rappelons que la fonction de hasard cumulative d'une f.r.d. L est définie, pour tout $t \leq \tau_L$ par

$$\Lambda(t) = \int_0^t \frac{dL(s)}{1 - L(s^-)}.$$

La fonction de hasard cumulative détermine de manière unique la f.d.r. L par la relation suivante

$$1 - L(t) := \mathcal{E}(-\Lambda)(t) = \exp(-\Lambda^c(t)) \prod_{s \leq t} (1 - \Delta\Lambda(s))$$

où $\Delta\Lambda(s) = \Lambda(s) - \Lambda(s-)$ et $\Lambda^c(s) = \Lambda(s) - \sum_{s \leq t} \Delta\Lambda(s)$.

Remarquons que pour L continue $\Lambda(t) = -\log(1 - L(t))$.

Notons que la fonction de hasard cumulative de F s'écrit sous la forme

$$\Lambda(t) = \int_0^t \frac{(1 - G(s))dF(s)}{(1 - G(s))(1 - F(s-))} = \int_0^t \frac{dH_1(s)}{H(s-)}$$

ceci suggère d'estimer la fonction de hasard cumulative par

$$\hat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)} = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{Y_{(i)} \leq t\}}}{1 - N_n(Y_{(i)}^-)}$$

où

$$dN_n(s) = N_n(s) - N_n(s^-).$$

Cet estimateur est connu sous le nom de l'estimateur de Nelson-Aalen de la fonction de hasard cumulative.

L'estimateur de Kaplan-Meier de la fonction de survie $1 - F(t)$ est obtenu par

$$\begin{aligned} 1 - \hat{F}_n(t) &= \mathcal{E} \left(-\hat{\Lambda}_n \right) (t) \\ &= \exp \left(-\Lambda_n^c(t) \right) \prod_{s \leq t} (1 - \Delta \Lambda_n(s)) \\ &= \prod_{s \leq t} \left(1 - \frac{dN_n(s)}{Y_n(s)} \right) \end{aligned}$$

et peut se mettre sous la forme

$$1 - \hat{F}_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{\delta_i}{n-i+1} \right)^{\mathbb{I}_{\{Y_{(i)} \leq t\}}} & \text{si } t < Y_{(n)}, \\ 0 & \text{si } t \geq Y_{(n)}. \end{cases} \quad (2.1)$$

L'estimateur de Kaplan-Meier $\hat{G}_n$ de G s'obtient de la même manière par

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

Remarque 2.1. La variance asymptotique de $\hat{F}_n(t)$ est

$$C(t) = \int_0^t \frac{dH_1(s)}{(1 - H(s^-))^2} = \int_0^t \frac{dF(s)}{(1 - F(s^-))^2 (1 - G(s^-))}$$

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

$$\begin{aligned} C_n(t) &= \int_0^t \frac{dN_n(s)}{(1 - Y_n(s^-))^2} \\ &= \frac{1}{n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{Y_{(i)} \leq t\}}}{\left(1 - N_n(Y_{(i)}^-) \right)^2}. \end{aligned} \quad (2.2)$$

$C_n(t)$ est un estimateur consistant de $C(t)$; Klein (1991) montra que C_n est plus performant que d'autres estimateurs de la variance de $\hat{F}_n$.

2.3 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. Nous commençons par ses propriétés de convergence.

Théorème 2.1. (*Shorack et Wellner 1986, p. 304*).

Si F et G sont deux fonctions de répartition sur $[0, +\infty)$ alors

(i) $\sup_{0 \leq t \leq \tau_H} |\hat{F}_n(t) - F(t)| \rightarrow 0$ p. s. quand $n \rightarrow +\infty$

(ii) Pour tout τ fixé, $\tau < \tau_H$ $\sup_{0 \leq t \leq \tau} |\hat{\Lambda}_n(t) - \Lambda(t)| \rightarrow 0$ p.s. quand $n \rightarrow +\infty$

Preuve Le théorème de Glivenko Cantelli, nous permet d'établir que

$$\sup_{0 \leq t \leq \tau_H} \left| \frac{1}{n} N_n(s) - H_1(s) \right| \rightarrow 0 \text{ p.s.}, \quad \sup_{0 \leq t \leq \tau_H} \left| \frac{1}{n} Y_n(s) - \bar{H}(t) \right| \rightarrow 0 \text{ p.s.}$$

$$\text{et } \sup_{0 \leq t \leq \tau_H} \left| \frac{1}{n} Y_n(s^-) - H_1(s^-) \right| \rightarrow 0 \text{ p.s. quand } n \rightarrow \infty.$$

Par conséquent

$$\begin{aligned} |\hat{\Lambda}_n(t) - \Lambda(t)| &= \left| \int_0^t \frac{dN_n(s)}{Y_n(s^-)} - \int_0^t \frac{dH_1(s)}{1 - H(s^-)} \right| \\ &\leq \int_0^t \left| \frac{1}{1 - Y_n(s^-)} - \frac{1}{1 - H(s^-)} \right| dN_n(s) + \left| \int_0^t \frac{1}{1 - H(s^-)} d(N_n(s) - H_1(s)) \right| \\ &\leq o(1) + \left| \int_0^t \frac{1}{1 - H(s^-)} d(N_n(s) - H_1(s)) \right| \\ &\leq o(1). \end{aligned}$$

La dernière inégalité s'obtient grâce à la loi forte des grands nombres.

Pour tout $0 \leq t \leq \tau_H$ on a

$$\frac{1 - \hat{F}_n(t)}{1 - F(t)} = 1 - \int_0^t \frac{1 - \hat{F}_n(s^-)}{1 - F(s)} d(\hat{\Lambda}_n(s) - \Lambda(s)).$$

Nous avons alors

$$\begin{aligned} |\hat{F}_n(t) - F(t)| &\leq |1 - F(t)| \left| \int_0^t \frac{1 - \hat{F}_n(s^-)}{1 - F(s)} d(\hat{\Lambda}_n(s) - \Lambda(s)) \right| \\ &\leq \left| \int_0^t \frac{1 - \hat{F}_n(s^-)}{1 - F(s^-)} \frac{1}{1 - \Delta\Lambda(s)} d(\hat{\Lambda}_n(s) - \Lambda(s)) \right| \\ &\leq \left| \int_0^t \frac{1 - \hat{F}_n(s^-)}{1 - F(s^-)} dK_n \right| \end{aligned}$$

où

$$K_n(t) = \int_0^t \frac{d\left(\widehat{\Lambda}_n(s) - \Lambda(s)\right)}{1 - \Delta\Lambda(s)}.$$

Donc

$$\begin{aligned} \left| \widehat{F}_n(t) - F(t) \right| &\leq \left| \frac{1 - \widehat{F}_n(t)}{1 - F(t)} K_n(t) - \int_0^t K_n(s) d\left(\frac{1 - \widehat{F}_n(s-)}{1 - F(s-)} \right) \right| \\ &\leq \frac{1}{1 - F(\tau)} |K_n(t)| + \sup_{0 \leq t \leq \tau} |K_n(t)| \int_0^t d\left(\frac{1 - \widehat{F}_n(s-)}{1 - F(s-)} \right) \\ &\leq \sup_{0 \leq t \leq \tau} |K_n(t)| O(1) \quad p.s. \end{aligned}$$

D'autre part

$$\begin{aligned} |K_n(t)| &= \left| \int_0^t d\left(\widehat{\Lambda}_n(s) - \Lambda(s)\right) + \int_0^t \frac{\Delta F(s)}{1 - F(s)} d\left(\widehat{\Lambda}_n(s) - \Lambda(s)\right) \right| \\ &\leq \left| \widehat{\Lambda}_n(t) - \Lambda(t) \right| + \sum_{s \leq t; \Delta F > 0} \frac{\Delta F(s)}{1 - F(s)} \Delta\left(\widehat{\Lambda}_n(s) - \Lambda(s)\right). \end{aligned}$$

Par conséquent

$$\sup_{0 \leq t \leq \tau} |K_n(t)| \leq \sup_{0 \leq t \leq \tau} \left| \widehat{\Lambda}_n(t) - \Lambda(t) \right| + 2 \sup_{0 \leq t \leq \tau} \left| \widehat{\Lambda}_n(t) - \Lambda(t) \right| \frac{F(\tau)}{1 - F(\tau)} \rightarrow 0 \quad p.s.$$

D'où $\widehat{F}_n(t) \rightarrow F(t)$ p.s. quand $n \rightarrow \infty$. ■

Définissons les processus suivants, pour tout $0 \leq t \leq \tau_H$

$$Z_n(t) = \sqrt{n} \left(\widehat{\Lambda}_n(t) - \Lambda(t) \right)$$

et

$$X_n(t) = \sqrt{n} \left(\widehat{F}_n(t) - F(t) \right).$$

Nous avons les résultats de normalité asymptotique suivants :

Théorème 2.2. (Breslow et Crowley, (1974))

Si $\tau < +\infty$ et F et G sont continues alors $Z_n \xrightarrow{\mathcal{L}} Z$ quand $n \rightarrow +\infty$
où Z est un processus Gaussien centré tel que

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

Théorème 2.3. (*Breslow et Crowley, 1974*).

Sous les mêmes hypothèses que le théorème précédent, $X_n \xrightarrow{\mathcal{L}} X$ quand $n \rightarrow +\infty$ où X est un processus gaussien centré tel que

$$\text{cov}(X(t), X(s)) = (1 - F(s))(1 - F(t)) \int_0^s \frac{dH(y)}{(1 - F(y))^2}.$$

La preuve du théorème 2.3 est basée sur une approximation du processus $X_n(t)$ par une expression linéaire de $\bar{Y}_n$ et $\bar{N}_n$ et des applications des résultats standards sur la convergence faible des fonctions de répartition empiriques. Shorack et Wellner (1986) utilisent le théorème central limite pour les martingales :

ils écrivent la martingale $M_n(t) = \sqrt{n} \left(N_n(t) - \int_0^t (1_H(s)) d\Lambda(s) \right)$ sous la forme $\frac{1}{\sqrt{n}} \sum_{i=1}^n M_{ni}(t)$.

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

Pour cela, pour tous réels positifs z et t et $\delta \in \{0, 1\}$, posons

$$\xi(z, \delta, t) = \bar{F}(t) \left[g(s \wedge t) + \frac{1}{\bar{H}(z)} \mathbb{1}_{\{z \leq t, \delta=1\}} \right]$$

où

$$g(x) = \int_0^x \frac{d\bar{H}_1(s)}{(\bar{H}(s))^2}.$$

les v.a. ξ_i sont uniformément bornées sur $[0, \tau_H]$, elles sont centrées et

$$\text{cov}(\xi_i(s), \xi_i(t)) = \bar{F}(s)\bar{F}(t) \int_0^{s \wedge t} \frac{dH_1(x)}{(\bar{H}(x))^2}$$

.

Théorème 2.4. (*Lo et Singh, 1985*).

Supposons que F est continue, alors

$$\hat{F}_n(t) - F(t) = \frac{1}{n} \sum_{i=1}^n \xi(z_i, \delta_i, t) + r_n(t)$$

où

$$\sup_{0 \leq t < \tau} |r_n(t)| = O \left(\left(\frac{\log n}{n} \right)^{3/4} \right)$$

p.s., pour tout $\tau < \tau_H$.

Afin de démontrer ce théorème, nous avons besoin de quelques résultats intermédiaires

Lemme 2.1. *Si F est continue, alors presque sûrement*

$$\sup_{0 \leq t \leq \tau_H} \left| \int_0^t \left(Y_n^{-1} - \overline{H}^{-1} \right) d(N_n - \overline{H}_1) \right| = O \left(\left(\frac{\log n}{n} \right)^{3/4} \right).$$

preuve Pour simplifier, supposons que F et G sont continues.

L'intervalle $[0, \tau]$ peut être subdivisé en k_n intervalles $[x_i, x_{i+1}]$; $i = 1, \dots, k_n$ où

$k_n = O \left(\sqrt{\frac{n}{\log n}} \right)$ et $0 = x_1 < x_2 < \dots < x_{k_n} < x_{k_n+1} = \tau$ tels que

$$|H(x_{i+1}) - H(x_i)| \leq C \left(\sqrt{\frac{\log n}{n}} \right)$$

où C est une constante.

On montre alors que

$$\begin{aligned} & \left| \int_0^t \left(Y_n^{-1} - \widehat{H}^{-1} \right) d(N_n - \overline{H}_1) \right| \\ & \leq k_n \sup_{0 \leq x \leq \tau} |Y_n^{-1}(x) - \overline{H}_1(x)| \max_{1 \leq i \leq k_n} |N_n(x_{i+1}) - N_n(x_i) - \overline{H}_1(x_{i+1}) + \overline{H}_1(x_i)| \\ & + 2 \max_{1 \leq i \leq k_n} \sup_{y \in [x_i, x_{i+1}]} \left| Y_n^{-1}(y) - Y_n^{-1}(x_i) - \overline{H}^{-1}(y) + \overline{H}^{-1}(x_i) \right| := A + B \end{aligned}$$

Pour estimer le terme B , on subdivise chaque intervalle $[x_i, x_{i+1}]$ en sous intervalles $[x_{ij}, x_{i(j+1)}]$, $j = 1, \dots, \delta_n$ tels que

$$|\overline{H}(x_{i(j+1)}) - \overline{H}(x_{ij})| = O \left(\left(\frac{\log n}{n} \right)^{3/4} \right)$$

uniformément en i et j et $\delta_n = O \left(\left(\frac{n}{\log n} \right)^{3/4} \right)$.

Comme

$$\sup |Y_n - \overline{H}|^2 = O \left(\frac{\log n}{n} \right),$$

on a

$$\begin{aligned}
& \sup_{y \in [x_i, x_{i+1}]} \left| Y_n^{-1}(y) - Y_n^{-1}(x_i) - \bar{H}^{-1}(y) + \bar{H}^{-1}(x_i) \right| \\
& \leq \sup_{y \in [x_i, x_{i+1}]} \left| (Y_n(y) - \bar{H}(y)) \bar{H}^{-2}(y) - (Y_n(x_i) - \bar{H}(x_i)) \bar{H}^{-2}(x_i) \right| + O\left(\frac{\log n}{n}\right) p.s. \\
& \leq \sup_{y \in [x_i, x_{i+1}]} \bar{H}^{-2}(x_{i+1}) |Y_n(y) - Y_n(x_i) - \bar{H}(y) + \bar{H}(x_i)| + O\left(\frac{\log n}{n}\right) p.s. \\
& \leq C \max_{1 \leq j \leq \delta_n} |Y_n(x_{ij}) - Y_n(x_i) - \bar{H}(x_{ij}) + \bar{H}(x_i)| + O\left(\frac{\log n}{n}\right) p.s.
\end{aligned}$$

Appliquons l'inégalité de type exponentiel suivante

si $\eta_1, \dots, \eta_n$ sont des v.a. i.i.d. centrées telles que $|\eta_i| \leq C$ et $\text{var}(\eta_i) = \sigma^2$ alors $\forall z > 0$ et d ; tel que $Cz \leq d$ et $nz\sigma^2 \leq d^2$

$$\mathbb{P} \left(\sum_1^n |\eta_i| \geq 3d \right) \leq 2e^{-z}$$

en prenant $C = 1$, $\sigma^2 = O\left(\sqrt{\frac{\log n}{n}}\right)$, $z = O(\log n)$, nous obtenons

$$\max_{i \leq k_n, j \leq \delta_n} \mathbb{P}(|Y_n(x_{ij}) - Y_n(x_i) - \bar{H}(x_{ij}) + \bar{H}(x_i)|) = O(n^{-3}) \blacksquare$$

Remarque 2.2. Si l'on suppose que F est continue, il est alors naturel de considérer un estimateur lissé $\tilde{F}_n$ de F . De tels estimateurs apparaissent comme des intégrales d'estimateurs non paramétriques de la densité $\hat{f}_n$ définis par

$$\hat{f}_n(x) = \frac{1}{h_n} \int K\left(\frac{x-u}{h_n}\right) d\hat{F}_n(u)$$

où K est une densité de probabilité, appelée noyau et h_n une suite de réels telle que $h_n \rightarrow 0$ quand $n \rightarrow \infty$. L'estimateur de Kaplan-Meier peut être représenté par

$$\tilde{F}_n(t) = \int_{-\infty}^t \hat{f}_n(x) dx.$$

Dans le cas non censuré et i.i.d. les performances asymptotiques de $\tilde{F}_n$ ont été montrées par Reis(1981) et Falk(1983) qui trouvent que $\tilde{F}_n$ est meilleur que F_n pour un choix approprié du noyau K , de la fenêtre h_n et des conditions de régularité sur F .

Chapitre 3

Données censurées et α –mélangeantes

3.1 Données α –mélangeantes

3.1.1 Définitions

Souvent, les données traitées en statistique présentent une certaine forme de dépendance. On peut citer les processus ARMA, les processus GARCH, Les chaînes de Markov qui sont dépendants mais sont asymptotiquement indépendants. Afin de mesurer cette dépendance entre les σ –algèbres, différents coefficients sont utilisés. Plus précisément, soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé et $\mathcal{B}$ et $\mathcal{C}$ deux sous σ –algèbres de $\mathcal{A}$.

1. Le coefficient de mélange fort (ou d' α –mélange, introduit par Rosenblatt 1956)

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup \left\{ \left| \mathbb{P}(B \cap C) - \mathbb{P}(B) \mathbb{P}(C) \right| ; B \in \mathcal{B} \text{ et } C \in \mathcal{C} \right\}.$$

Ce coefficient est tel que $0 \leq \alpha(\mathcal{B}, \mathcal{C}) \leq 1/4$.

2. Le coefficient de régularité absolue (ou de β –mélange, introduit par Kolmogorov 1931)

$$\beta(\mathcal{B}, \mathcal{C}) = \sup \left\{ \frac{1}{2} \sum_{i=1}^I \sum_{j=1}^J \left| \mathbb{P}(B_i) \mathbb{P}(C_j) - \mathbb{P}(B_i \cap C_j) \right| \right\}$$

le sup est pris sur toutes les partitions $(B_i)_{i \in I}, (C_j)_{j \in J}$ de Ω avec $B_i \in \mathcal{B}, C_j \in \mathcal{C}$.
Le coefficient $\beta(\mathcal{B}, \mathcal{C})$ est tel que $0 \leq \beta(\mathcal{B}, \mathcal{C}) \leq 1$.

3. Le coefficient de mélange uniforme (ou de Φ –mélange, introduit par Ibragimov 1962)

$$\Phi(\mathcal{B}, \mathcal{C}) = \sup \left\{ \left| \mathbb{P}(C) - \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)} \right| ; B \in \mathcal{B}, \mathbb{P}(B) \neq 0 \text{ et } C \in \mathcal{C} \right\}.$$

$\Phi(\mathcal{B}, \mathcal{C})$ vérifie les inégalités $0 \leq \Phi(\mathcal{B}, \mathcal{C}) \leq 1$.

4. Le coefficient de Ψ - mélange (introduit par Blum *et al* 1963)

$$\Psi(\mathcal{B}, \mathcal{C}) = \sup\{1 - \frac{\mathbb{P}(B \cap C)}{\mathbb{P}(B)\mathbb{P}(C)}; B \in \mathcal{B}, \mathbb{P}(B) \neq 0, C \in \mathcal{C}, \mathbb{P}(C) \neq 0\}.$$

5. Le coefficient de ρ - mélange (introduit par Hirschfeld 1935, Gebelein 1941)

$$\rho(\mathcal{B}, \mathcal{C}) = \sup\{|corr(X, Y)|; X \in L^2(\mathcal{B}) \text{ et } Y \in L^2(\mathcal{C})\}.$$

Supposons que $c(\cdot, \cdot)$ désigne un coefficient de mélange arbitrairement pris parmi ceux précédemment définis (α , β , Φ , Ψ ou ρ).

Soit $\{Y_i, i \geq 1\}$ une suite de v.a. et soit $\mathcal{F}_i^k$ la σ algèbre engendrée par $\{Y_j; i \leq j \leq k\}$. Etant donné un entier positif l , posons pour tout k fixé, $k \geq 1$;

$$c(l) = c(\mathcal{F}_1^k, \mathcal{F}_{k+l}^{+\infty}).$$

La suite de v.a. $\{Y_i, i \geq 1\}$ est dite c -mélangeante si

$$\lim_{l \rightarrow \infty} c(l) = 0.$$

La dépendance faible est alors "mesurée" en termes de coefficients de mélange entre les tribus engendrées par les variables avant l'instant k et après l'instant $(k + l)$; ces coefficients étant nuls quand les tribus sont indépendantes.

Ces notions de mélange ne sont pas, bien évidemment équivalentes, mais il existe différentes implications entre elles :

$$\Psi - \text{mélange} \Rightarrow \Phi - \text{mélange} \Rightarrow \begin{cases} \beta - \text{mélange} \Rightarrow \alpha - \text{mélange} \\ \rho - \text{mélange} \Rightarrow \alpha - \text{mélange}. \end{cases}$$

Ainsi l' α -mélange est la plus faible et la moins restrictive de toutes ces formes de mélange. Toute suite de v.a. mélangeante sera alors forcément α -mélangeante. De plus les processus α -mélangeants sont très présents dans la pratique statistique, en particulier les processus AR et ARMA, qui sont largement utilisés dans l'analyse des séries temporelles, sont α -mélangeantes avec des coefficients géométrique et exponentiel respectivement. Cela explique donc que l'estimation pour des données mélangeantes est souvent restreinte dans la littérature au cas de données α -mélangeantes.

3.1.2 Quelques théorèmes de convergence

Des versions analogues aux théorèmes classiques en statistique mathématique ont été établis pour les processus α -mélangeants. Nous citons en particulier les inégalités de type exponentiels et L.I.L. qui nous serviront de manière fondamentale.

Théorème 3.1. *Bosq (1998). Soit $(X_t, t \geq \mathbb{Z})$ un processus réel centré α -mélangeant tel que*

$\sup_{1 \leq t \leq n} |X_t| \leq b$, et $S_n = \sum_{i=1}^n X_i$ alors

(i) pour tout entier $q \in [1, \frac{n}{2}]$ et tout $\epsilon > 0$

$$\mathbb{P}(|S_n| > n\epsilon) \leq 4\exp\left(-\frac{\epsilon^2}{8b^2}q\right) + 22\left(1 + \frac{4b}{\epsilon}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) \quad (3.1)$$

(ii) pour tout entier $q \in [1, \frac{n}{2}]$ et tout $\epsilon > 0$

$$\mathbb{P}(|S_n| > n\epsilon) \leq 4\exp\left(-\frac{\epsilon^2}{8v^2(q)}q\right) + 22\left(1 + \frac{4b}{\epsilon}\right)^{1/2} q\alpha\left(\left[\frac{n}{2q}\right]\right) \quad (3.2)$$

où $v^2(q) = \frac{2}{p^2}\sigma^2(q) + \frac{b\epsilon}{2}$ avec $p = \frac{n}{2q}$ et

$$\begin{aligned} \sigma^2(q) &= \max_{0 \leq j \leq 2q-1} \mathbb{E}[(jp + 1 - ip)X_{[jp]+1} + X_{[jp]+2} + \dots \\ &\quad + X_{[(j+1)p]} + ((j+1)p - [(j+1)p])X_{[(j+1)p]+1}]^2. \end{aligned}$$

La preuve de ce théorème, comme beaucoup d'autres résultats sur les processus α -mélangeants, est basée sur la technique de couplage qui consiste à approximer les v.a. α -mélangeantes par des v.a. indépendantes. la preuve de (i) est basée sur l'inégalité de Hoeffding et celle de (ii) sur l'inégalité de Bernstein.

L'estimation dans le cadre des données α -mélangeantes nécessite généralement des hypothèses quant à l'ordre du coefficients de mélange. Cai et Roussas(1992) considèrent une suite de v.a. $\{Y_n, n \geq 1\}$ stationnaire α -mélangeante de f.d.r. inconnue F et de coefficient de mélange $\alpha(n) = O(n^{-\nu})$ pour $\nu > 3$. C'est sous ces hypothèses qu'ils obtiennent une L.L.I. pour la fonction de répartition empirique.

Théorème 3.2. *Cai et Roussas (1992).*

Soit F_n la f.d.r. empirique basée sur $Y_1, \dots, Y_n$, une suite de v.a. stationnaire α -mélangeante de f.d.r. inconnue F alors

$$\limsup_{n \rightarrow +\infty} \left[\left(\frac{n}{2 \log \log n} \right)^{1/2} \sup_{y \in \mathbb{R}} |F_n(y) - F(y)| \right] = 1 \quad p.s.$$

Pour la preuve de ce théorème, nous avons du théorème 3 de Dhompongsa(1984), que nous donnons tel qu'il est donné dans Cai et Roussas (1992). Pour cela définissons le processus empirique

$$R(s, t) = \sum_{k \leq t} g_k(s), \quad s \in \mathbb{R}, \quad t \geq 0 \quad (3.3)$$

où $g_k(s) = \mathbb{1}_{\{Y_k \leq s\}} - F(s)$

La structure de covariance est donnée par

$$\Gamma(s, t) = \mathbb{E}[g_1(s)g_1(t)] + \sum_{k=2}^{+\infty} \mathbb{E}[g_1(s)g_k(t)] + \sum_{k=2}^{+\infty} \mathbb{E}[g_k(s)g_1(t)] \quad (3.4)$$

Théorème 3.3. *Dhompongsa (1984).*

Soit $\{Y_n, n \geq 1\}$ une suite de v.a. stationnaire α -mélangeante de f.d.r. inconnue F et de coefficient de mélange $\alpha(n) = O(n^{-\nu})$ pour un certain $\nu > 3$. Alors, sans changer la répartition du processus empirique défini par (3.3), nous pouvons le redéfinir sur un espace de probabilité plus riche sur lequel il existe un processus de Kiefer $\{K(s, t), s \in \mathbb{R}, t \geq 0\}$ de fonction de covariance

$$\mathbb{E}(K(s, t)K(s', t')) = \Gamma(s, s')\Gamma(t, t');$$

où $\Gamma(., .)$ est donnée par (3.4), tel que , pour in certain $\lambda > 0$ ne dépendant que de ν

$$\sup_{0 \leq t \leq T} \sup_{s \in \mathbb{R}} |R(s, t) - K(s, t)| = O(T^{1/2}(\log T)^{-\lambda}) \quad p.s. \text{ pour tout } T > 0.$$

Sous des conditions suffisantes en particulier sur le coefficient d' α -mélange, des extensions du théorème central limite peuvent être trouvées dans Ibragimov et Linnick (1981), ainsi qu'une extension du TCL fonctionnel de Donsker aux processus stationnaires et α -mélangeants peut être trouvée dans Doukhan (1994) et Rio (2000).

3.2 Données censurées et α mélangeantes.

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé.

Soit $\{T_i; i \geq 1\}$ une suite de v.a. stationnaires, de f.d.r. commune F et de densité continue f . Soit $\{C_i; i \geq 1\}$ la suite des v.a. de censure, stationnaires et de f.d.r. G . Conformément aux études sur les durées de vie dans le domaine biomédical, nous supposons que les suites $\{T_i; i \geq 1\}$ et $\{C_i; i \geq 1\}$ sont indépendantes.

L'échantillon réellement observé est $\{(Y_1, \delta_1), \dots, (Y_n, \delta_n)\}$ où

$Y_i = T_i \wedge C_i$ et $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$.

Dans le cadre de données censurées et α mélangeantes, la fonction de hasard cummulative Λ , la f.d.r. H des v.a. Y_i , les fonctions H_1 et H_0 et les processus N_n, Y_n et M_n sont définis comme dans le cas i.i.d. (voir le paragraphe 1.2). Nous rappelons ici ces définitions afin de faciliter le référencement :

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s-)}$$

$$H(t) = \mathbb{P}(Y_i \leq t), \quad \bar{H}(t) = \bar{F}(t)\bar{G}(t),$$

$$H_1(t) = \mathbb{P}(Y_i \leq t, \delta_i = 1), \quad H_0(t) = \mathbb{P}(Y_i \leq t, \delta_i = 0),$$

$$N_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq t, \delta_i = 1\}}, \quad M_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \leq t, \delta_i = 0\}} \quad \text{et}$$

$$Y_n(t) = \sum_{i=1}^n \mathbb{1}_{\{Y_i \geq t\}}.$$

Nous rappelons aussi les définitions de l'estimateur de Nelson-Aalen (voir Nelson (1972) et Aalen (1978)) de $\Lambda(t)$ et de l'estimateur de Kaplan-Meier de F donnés par

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)}$$

et

$$1 - \widehat{F}_n(t) = \prod_{s \leq t} \left(1 - \frac{dN_n(s)}{Y_n(s)}\right) \quad (3.5)$$

où $dN_n(t) = N_n(t) - N_n(t^-)$ et $N_n(t^-) = \lim_{\varepsilon \rightarrow 0^+} N_n(t - \varepsilon)$.

3.2.1 Propriétés de l'E.K.M.

Cai (2001) donne des résultats de convergence faible et forte de l'E.K.M. sous des hypothèses sur le coefficient de mélange et dans le cas où les durées de survie $\{T_i; i \geq 1\}$ et les instants de censure $\{C_i; i \geq 1\}$ sont deux suites indépendantes de v.a. α -mélangeantes stationnaires.

Théorème 3.4. *Cai (2001).*

Supposons que $\{T_i; i \geq 1\}$ est une suite de v.a. stationnaire et α -mélangeante de coefficient de mélange α_1 ,

$\{C_i; i \geq 1\}$ est une suite de v.a. stationnaire et α -mélangeante de coefficient de mélange α_2 et indépendante de $\{T_i; i \geq 1\}$. Alors

$$\sup_{0 \leq t \leq \tau_H} \left| \widehat{F}_n(t) - F(t) \right| \xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty \quad (3.6)$$

et

$$\sup_{0 \leq t \leq Y_{(n)}} \left| \widehat{F}_n(t) - F(t) \right| \xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty \quad (3.7)$$

où $Y_{(n)} = \max_{1 \leq i \leq n} Y_i$

Si de plus $\sum_{n=3}^{+\infty} \log n (\log \log n)^{1+\delta} \alpha(n)/n < +\infty$ où $\alpha(n) = \max(\alpha_1(n), \alpha_2(n))$

alors (3.6) et (3.7) ont lieu presque sûrement.

Afin de prouver ce théorème, nous avons besoin d'un résultat qui concerne le caractère α -mélangeant d'un couple de v.a.

Lemme 3.1. (*Cai 2001 lemme 2.*)

Soit $\{X_j; j \geq 1\}$ et $\{Y_j; j \geq 1\}$ deux suites de v.a. α -mélangeantes de coefficients de mélange $\alpha_1(\cdot)$ et $\alpha_2(\cdot)$ respectivement, alors $\{(X_j, Y_j); j \geq 1\}$ est une suite α -mélangeante de coefficient de mélange $4\alpha(n)$, où $\alpha(n) = \max\{\alpha_1(n), \alpha_2(n)\}$ et il en est de même, en particulier pour la suite $\{X_j \wedge Y_j; j \geq 1\}$.

Preuve du Théorème 3.4

Soit $\alpha_0 = \mathbb{P}(T_i \leq C_i) = \int_0^{+\infty} F(z) dG(z)$.

Clairement $\alpha_0^{-1} H_1(t)$ est la f.d.r. conditionnelle de Y_i sachant $\{\delta_i = 1\}$.

Cai (2001) montre que

$$\begin{aligned} \sup_{0 \leq t \leq \tau_H} |\bar{Y}_n(t) - \bar{H}(t)| &\xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty \\ \text{et } \sup_{0 \leq t \leq \tau_H} |\bar{N}_n(t) - \bar{H}_1(t)| &\xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty. \end{aligned} \quad (3.8)$$

Si de plus $\alpha(n) = O(n^{-\nu})$; $\nu > 3$ alors (3.8) ont lieu presque sûrement.

Le Lemme 2 de Gill (1980), nous permet de déduire que

$$\begin{aligned} \sup_{0 \leq t \leq \tau} |\hat{F}_n(t) - F(t)| &\xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty \\ \text{et } \sup_{0 \leq t \leq \tau} |\hat{F}_n(t) - F(t)| &\xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty, \end{aligned} \quad (3.9)$$

pour $\tau < \tau_H$.

Si on considère le cas $F(\tau_H) = 1$, alors pour tout $0 < \epsilon < 1$ prenons $\tau < \tau_H$ tel que $1 - \epsilon < F(\tau) < 1$;

donc pour tout $t \in [\tau, \tau_H]$

$$\hat{F}_n(\tau-) \leq \hat{F}_n(t) \leq 1 \quad \text{et} \quad 1 - \epsilon < F(\tau-) = F(\tau) \leq F(t) < 1$$

Donc $\sup_{\tau \leq t \leq \tau_H} |\hat{F}_n(t) - F(t)| < \max\{\epsilon, 1 - \hat{F}_n(\tau-)\}$; d'où de la première relation de (3.9)

$$\sup_{0 \leq t \leq \tau_H} |\hat{F}_n(t) - F(t)| \xrightarrow{\mathbb{P}} 0 \quad \text{quand } n \rightarrow +\infty$$

et

$$\sup_{0 \leq t \leq \tau_H} |\hat{F}_n(t) - F(t)| \xrightarrow{p.s.} 0 \quad \text{quand } n \rightarrow +\infty$$

est obtenue de la même manière en considérant la deuxième relation de (3.9) ■

Cai(2001) établit une représentation forte du K.M.E. que nous donnons ci dessous sous les hypothèses du théorème 3.4 et pour $\alpha(n) = O(n^{-\nu})$; $\nu > 3$

$\sup_{0 \leq t \leq \tau} |\hat{F}_n(t) - F(t)| = O(\log \log n/n)^{1/2}$ pour tout $0 \leq t \leq \tau$;

et étudie aussi sa normalité asymptotique et montre que

$\sqrt{n} [\hat{F}_n(\cdot) - F(\cdot)]$ converge faiblement vers $[1 - F(\cdot)] W(\cdot)$ où $W(\cdot)$ est un processus gaussien centré sur $D[0, \tau]$ (pour tout $\tau < \tau_H$ et $F(\tau) < 1$) tel que $W(0) = 0$ et la structure de covariance de $W(\cdot)$ est donnée par

$$\sigma^2(s, t) = \int_0^s \int_0^t \frac{dM(u, v)}{\bar{H}(u)\bar{H}(v)}, \quad (3.10)$$

avec $M(s, t)$ est définie par

$$M(s, t) = M_{11}(s, t) - \int_0^s M_{12}(t, v) d\Delta(v) - \int_0^s M_{12}(s, u) d\Delta(u) + \int_0^s \int_0^t M_{22}(u, v) d\Delta(u) d\Delta(v)$$

avec

$$M_{11}(s, t) = cov(\mathbb{I}_{\{T_1 \leq s \wedge C_1\}}, \mathbb{I}_{\{T_1 \leq t \wedge C_1\}}) + \sum_{j=2}^{\infty} [cov(\mathbb{I}_{\{T_1 \leq s \wedge C_1\}}, \mathbb{I}_{\{T_j \leq t \wedge C_j\}}) + cov(\mathbb{I}_{\{T_1 \leq t \wedge C_1\}}, \mathbb{I}_{\{T_j \leq s \wedge C_j\}})]$$

$$M_{12}(s, t) = cov(\mathbb{I}_{\{T_1 \wedge C_1 \leq s\}}, \mathbb{I}_{\{T_1 \leq t \wedge C_1\}}) + \sum_{j=2}^{\infty} [cov(\mathbb{I}_{\{T_1 \leq s \wedge C_1\}}, \mathbb{I}_{\{T_j \wedge C_j \leq t\}}) + cov(\mathbb{I}_{\{T_1 \leq t \wedge C_1\}}, \mathbb{I}_{\{T_j \wedge C_j \leq s\}})]$$

et

$$M_{22}(s, t) = cov(\mathbb{I}_{\{T_1 \wedge C_1 \leq s\}}, \mathbb{I}_{\{T_1 \wedge C_1 \leq t\}}) + \sum_{j=2}^{\infty} [cov(\mathbb{I}_{\{T_1 \wedge C_1 \leq s\}}, \mathbb{I}_{\{T_j \wedge C_j \leq t\}}) + cov(\mathbb{I}_{\{T_1 \wedge C_1 \leq t\}}, \mathbb{I}_{\{T_j \wedge C_j \leq s\}})]$$

Remarque 3.1. *L'estimateur (2.2), de la variance de $\hat{F}_n$, obtenu par la formule de Greenwood n'est pas valable dans le cas d'observations dépendantes, des versions modifiées ont été proposées par Ying et Wei (1994) et Kang et Koehler(1997). Cai (2001) a construit un estimateur de $\sigma^2(t, t)$ donnée par (3.10) en commençant par estimer $M(u, v)$ par*

$$\widehat{M}_n(u, v) = \frac{1}{n - l + 1} \sum_{j=0}^{n-l} \tilde{S}_j(l, u) \tilde{S}_j(l, v),$$

où, pour tout $0 \leq j \leq n - l$,

$$\tilde{S}_j(l, u) = \frac{1}{\sqrt{l}} \sum_{k=j+1}^{j+l} [\delta_k \mathbb{I}_{\{Z_k \leq u\}} - \int_0^u \mathbb{I}_{\{Z_k \geq z\}} d\hat{\Lambda}_n(z)]$$

et l est choisi tel que $l = l_n \rightarrow \infty$ et $l_n = o\left((\log \log n/n)^{-1/2}\right)$. Puis il considéra l'estimateur $n\widehat{V}_n(t)$ pour la variance de $\sqrt{n}\widehat{F}_n(t)$, où

$$\widehat{V}_n(t) = n[1 - \widehat{F}_n(t)]^2 \int_0^t \int_0^t \frac{d\widehat{M}_n(u, v)}{Y_n(u)Y_n(v)}$$

et il montra que c'est un estimateur consistant.

Chapitre 4

La fonction quantile.

Considérons une v.a. Y de f.d.r. F .

Pour tout $0 \leq p \leq 1$, le p^{ieme} quantile ξ_p de F est défini par

$$\xi_p = F^{-1}(p) = \inf\{y : F(y) \geq p\}. \quad (4.1)$$

La médiane $F^{-1}(1/2)$ et les quartiles $F^{-1}(1/4)$, $F^{-1}(1/2)$ et $F^{-1}(3/4)$ très utilisés en statistique, n'en sont que des cas particuliers. L'importance capitale de cette caractéristique en statistique mathématique n'est plus à prouver de nos jours.

ξ_p possède la propriété fondamentale : pour tout $x \in \mathbb{R}$ et $0 \leq p \leq 1$

$$F(x) \geq p \Leftrightarrow \xi_p \leq x$$

Ce qui implique : pour tous x et $y \in \mathbb{R}$

$$F(x) \leq p \leq F(y) \Leftrightarrow x \leq \xi_p \leq y$$

4.1 Estimation de la fonction quantile pour des données complètes.

4.1.1 Cas iid

Un estimateur naturel de ξ_p est l'estimateur empirique du quantile basé sur un échantillon de taille n , noté $F_n^{-1}(p)$ et qui est le p^{ieme} quantile de la f.d.r. empirique

$$F_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq y\}};$$

$$F_n^{-1}(p) = \inf\{y; F_n(y) \geq p\}.$$

Nous rappelons dans ce paragraphe quelques résultats importants. Le théorème suivant traite de la consistance forte de $F_n^{-1}(p)$

Théorème 4.1. (*Serfling 1980, p. 75*).

Soit $0 < p < 1$. Si $F^{-1}(p)$ est l'unique solution de l'équation $F(x) = p$, alors

$$F_n^{-1}(p) \xrightarrow{p.s.} F^{-1}(p).$$

La preuve de ce théorème est fondée sur deux points essentiels :
 $F^{-1}(p)$ est l'unique solution de l'équation $F(x) = p$, alors

$$F(F^{-1}(p) - \epsilon) \leq p \leq F(F^{-1}(p) + \epsilon) .$$

et F_n est fortement consistant pour F , alors

$$F_n(F^{-1}(p) - \epsilon) \xrightarrow{p.s.} F(F^{-1}(p) - \epsilon)$$

et

$$F_n(F^{-1}(p) + \epsilon) \xrightarrow{p.s.} F(F^{-1}(p) + \epsilon) ;$$

pour tout $\epsilon > 0$.

Dans ce qui suit, nous donnons une inégalité de type exponentiel pour la différence

$F_n^{-1}(p) - F^{-1}(p)$ et qui s'obtient grâce à l'inégalité de Hoeffding (1963).

Théorème 4.2. (*Serfling 1980, p.75*).

Soit $0 < p < 1$. Si $F^{-1}(p)$ est l'unique solution de l'équation $F(x) = p$, alors pour tout $\epsilon > 0$

$$\mathbb{P}(|F_n^{-1}(p) - F^{-1}(p)| > \epsilon) \leq 2e^{-2n\delta_\epsilon^2} \quad \forall n \geq 1$$

où $\delta_\epsilon = \min\{F(F^{-1}(p) + \epsilon), p - F(F^{-1}(p) - \epsilon)\}$.

Bahadur (1966) introduit une élégante représentation de F_n^{-1} comme somme de v.a. i.i.d. avec un reste de l'ordre de $O(n^{-1/4}(\log n)^{1/2}(\log \log n)^{1/4})$. De cette représentation découlent certaines propriétés importantes du quantile empirique comme la loi du logarithme itéré et la normalité asymptotique.

Nous donnons une version légèrement modifiée de la représentation de Bahadur.

Théorème 4.3. (*Kiefer 1967*).

Soit $0 < p < 1$. Supposons que F est deux fois différentiable en $\xi_p = F^{-1}(p)$ avec $f(\xi_p) = F'(\xi_p) > 0$, alors avec la probabilité 1,

$$F_n^{-1}(p) = \xi_p + \frac{p - F_n(\xi_p)}{f(\xi_p)} + R_n$$

où $R_n = O(n^{-3/4}(\log \log n)^{3/4})$ quand $n \rightarrow \infty$.

Alors que la vitesse de Bahadur (1966) était avec la probabilité 1 :

$$R_n(p) = O(n^{-3/4}(\log n)^{1/2}(\log \log n)^{1/4}), \quad n \rightarrow \infty.$$

Afin d'avoir une représentation sous la forme d'une moyenne de v.a., posons :

$$\mu_i = \xi_p + \frac{p - \mathbb{I}_{\{Y_i \leq F^{-1}(p)\}}}{f(F^{-1}(p))}, \quad i = 1, 2, \dots$$

ainsi la conclusion du théorème précédent peut être exprimée comme ceci : avec probabilité 1

$$F_n^{-1}(p) = \frac{1}{n} \sum_{i=1}^n \mu_i + O\left(n^{-3/4} (\log n)^{3/4}\right), \quad n \rightarrow \infty. \quad (4.2)$$

Ceci nous permet donc, avec la L.L.I. pour la f.d.r., d'établir celle pour le quantile (voir remarque (viii) de Serfling (1980) p. 93).

Sous les hypothèses du Théorème 4.3, nous avons, avec la probabilité 1 :

$$\limsup_{n \rightarrow \infty} \pm \frac{n^{1/2} (F_n^{-1}(p) - F^{-1}(p))}{(2 \log \log n)^{1/2}} = \frac{[p(1-p)]^{1/2}}{f(F^{-1}(p))}.$$

De la représentation (4.2) découle aussi la normalité asymptotique de $F_n^{-1}(p)$; sous les mêmes hypothèses que le Théorème 4.3, nous avons

$$\sqrt{n} (F_n^{-1} - \xi_p) \xrightarrow{D} \mathcal{N}\left(0, \frac{p(1-p)}{f^2(\xi_p)}\right).$$

4.1.2 Cas d' α -mélange.

Plusieurs auteurs se sont intéressés à l'estimation de la fonction quantile pour des données complètes mais dépendantes, nous citons en particulier Sen (1972) qui donna une représentation du type de Bahadur pour une suite de v.a. Φ -mélangeante. Babu et Sing(1978) ont établi la représentation de Bahadur pour $F_n^{-1}(p)$ sous l'hypothèse de mélange fort lorsque le coefficient de mélange est exponentiel.

Nous donnons dans le théorème suivant la représentation de type Bahadur pour la fonction quantile simple, établie par Yoshihara (1995) lorsque les v.a. sont uniformément bornées et pour un taux de mélange polynomial.

Théorème 4.4. (*Yoshihara, 1995*).

Soit $\{Y_n; n \geq 0\}$ une suite de v.a. stationnaires, à valeurs dans $[0, 1]$, de f.d.r. commune F de densité continue et bornée en un voisinage de ξ_p et telle que

$$0 < f(\xi_p) < +\infty \quad \text{et} \quad f'(\xi_p) > 0.$$

Si $\{Y_n; n \geq 0\}$ est α -mélangeante telle qu'il existe $\epsilon > 0$ pour lequel $\alpha(n) = O(n^{-5/2-\epsilon})$, alors

$$|\sqrt{n} (F_n^{-1}(p) - \xi_p) f(\xi_p) + \sqrt{n} (F_n(\xi_p) - p)| = O(n^{-1/4} \log n) \quad p.s.$$

et

$$\frac{\sqrt{n} f(\xi_p)}{\sigma_p} (F_n^{-1}(p) - \xi_p) \xrightarrow{D} \mathcal{N}(0, 1)$$

$$\text{où } \sigma_p^2 = \mathbb{E}(g_p^2(Y_0)) + 2 \sum_{k=1}^{\infty} \mathbb{E}(g_p(Y_0) g_p(Y_k)),$$

avec $g_t(u) = \mathbb{1}_{[0,1]}(u) - F(t)$ pour tout $u \in [0, 1]$ et $t \in [0, 1]$.

La preuve du théorème est basée sur des inégalités sur les moments de suites de v.a. stationnaires bornées, en particulier les résultats de Billingsley (1968, p. 199) et ceux de Roussas (1988) sur les suites de v.a. Φ -mélangeantes.

Sun (2006) étend ce résultat au cas de v.a. fortement mélangeante non nécessairement bornées et de coefficient de mélange d'ordre $O(n^{-\beta})$, avec $\beta > 0$. Pour cela, il utilise les inégalités de type exponentiel (3.1) de Bosq (1998) et il montre que

$$F_n^{-1}(p) = \xi_p - (F_n(\xi_p) - p) / f(\xi_p) + o(n^{-3/4} \log n) \quad p.s.$$

4.2 Estimation de la fonction quantile dans le modèle censuré et i.i.d.

De manière analogue aux données complètes, la fonction des quantiles d'une f.d.r. F est donnée par la relation (4.1) que nous rappelons

$$\xi_p = F^{-1}(p) = \inf\{y : F(y) \geq p\}, \text{ pour tout } p \in (0, 1).$$

De même $F^{-1}(p)$ est estimé par

$$\xi_{p,n} = \hat{F}_n^{-1}(p) = \inf\{y \in \mathbb{R}; \hat{F}_n(y) \geq p\} \quad (4.3)$$

où $\hat{F}_n$ est l'E.K.M. donné par (2.1).

Nous rappelons dans ce qui suit quelques résultats sur le comportement asymptotique de $\xi_{p,n}$ et on se contentera de donner quelques idées de base des preuves.

Théorème 4.5. (*K.F. Cheng, 1984*).

Soit $0 \leq p_0 < 1$ tel que $\xi_{p_0} < \tau_H$. Supposons que F admet une dérivée seconde F'' bornée sur $[0, \xi_{p_0} + \delta)$, pour un certain $\delta > 0$ et tel que $\inf_{0 \leq p \leq p_0} f(\xi_p) > 1$, alors

$$\sup_{0 \leq p \leq p_0} |\xi_{p,n} - \xi_p - \frac{p - \hat{F}_n(\xi_p)}{f(\xi_p)}| = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) \quad p.s. \quad \text{quand } n \rightarrow +\infty.$$

Pour établir ce théorème, K.F. Cheng montre en premier que pour tout $0 \leq \tau \leq \tau_H$ et si f est bornée sur $[0, \tau]$, alors

$$\sup_{0 \leq x, y \leq \tau; |x-y| \leq a_n} |(\hat{F}_n(y) - \hat{F}_n(x)) - (F(y) - F(x))| = O\left(\left(\frac{\log n}{n}\right)^{3/4}\right) \quad p.s. \quad \text{quand } n \rightarrow +\infty;$$

$$\text{avec } a_n = O\left(\left(\frac{\log n}{n}\right)^{1/2}\right).$$

Puis il montre que

$$\sup_{0 \leq p \leq p_0} |\xi_{p,n} - \xi_p| = O\left(\left(\frac{\log n}{n}\right)^{1/2}\right) \quad p.s. \quad \text{quand } n \rightarrow +\infty.$$

Lorsque l'espace de probabilité $(\Omega, A, \mathbb{P})$ est suffisamment riche et sous l'hypothèse que F soit dérivable de dérivée f continue, Aly *et al.* (1986) établissent l'approximation forte du processus quantile

$$\rho_n(p) = n^{1/2} f(\xi_p) (\xi_p - \xi_{p,n})$$

par un processus de Kiefer généralisé. Pour cela il montre d'abord le résultat suivant : On peut définir un processus de Kiefer $K(t, s)$ tel que pour tout $\tau < \tau_F$

$$\sup_{t \leq \tau} |\beta_n(t) - n^{1/2} K(t, n)| = O(n^{-1/2} (\log n)^{5/2}) \quad p.s.,$$

où $\beta_n(t) = n^{1/2} (\hat{F}_n(t) - F(t))$, avec $\mathbb{E}[K(t, s)] = 0$

$$\text{et } \mathbb{E}[K(t, s)K(t', s')] = (s \wedge s')(1 - F(t))(1 - F(t')) \int_{-\infty}^{t \wedge t'} (1 - F(s))^{-2} dF * (s)$$

où $F * (t) = \mathbb{P}((T_i \wedge C_i) \leq t, \delta_i = 1)$.

Duquel il déduisent que pour tout p_0 tel que $\inf_{0 \leq p \leq p^*} f(\xi_p) > 0$, on a alors l'approximation

$$\sup_{0 \leq p \leq p_0} |\rho_n(p) - n^{1/2} K^*(p, n)| \xrightarrow{P} 0$$

où $K^*(p, s) = K(\xi_p, s)$.

Remarque 4.1. *Beaucoup d'auteurs ont considéré des estimateurs lissés du quantile ; nous citons en particulier :*

Padgett (1986), Lio, Padgett and Yu (1986), Lio and Padgett(1987) proposèrent un estimateurs à noyau de la fonction quantile , ils définissent cet estimateur par

$$Q_n(p) = \frac{1}{h_n} \sum_{i=1}^n Z_{(i)} \int_{S_{i-1}}^{S_i} K\left(\frac{t-p}{h_n}\right) dt$$

où h_n est une suite de réels positifs tendant vers zéro, K une densité de probabilité bornée à support compact et

$$S_0 = 0, \quad S_i = \sum_{j=1}^i \hat{F}_n(Z_{(j+1)}); \quad i = 1, \dots, n-1 \text{ et } S_n = 1.$$

Ils étudièrent les propriétés asymptotiques de $Q_n(p)$ sous des conditions sur la fenêtre h_n , la fonction noyau K et la f.d.r. F .

Padgett et Thombs (1989) proposèrent un autre estimateur lissé, il commencèrent par estimer la f.d.r. F par

$$F_n(t) = \int_{-\infty}^t W\left(\frac{t-x}{h_n}\right) d\hat{F}_n(x),$$

$$\text{avec } W(t) = \int_{-\infty}^t K(u) du;$$

puis ils prirent l'estimateur $Q_n(p) = \inf\{t \in \mathbb{R}; F_n(t) \geq p\}$. Ils établissent sa consistance et sa normalité asymptotique.

4.3 Représentation forte de la fonction quantile pour des données censurées et α -mélangeantes

Nous considérons une suite de durées de survie $\{T_1, T_2, \dots\}$ stationnaire et fortement mélangeante de coefficient de mélange α_1 censurées par $\{C_1, C_2, \dots\}$ supposée aussi stationnaire et fortement mélangeantes de coefficient de mélange α_2 et telles que les suites $\{T_i; i \geq 1\}$ et $\{C_i; i \geq 1\}$ soient indépendantes.

Il est clair du lemme 3.1 que les v.a. $Z_i = T_i \wedge C_i; i \geq 1$ sont α -mélangeantes et de coefficient de mélange $4\alpha(n)$, où $\alpha(n) = \max\{\alpha_1(n), \alpha_2(n)\}$.

Nous montrons la consistance et nous donnons une représentation forte de l'estimateur (4.3) sous des conditions sur le coefficient de mélange.

Les résultats présents dans cette partie ont été publiés dans *Communications in Statistics Theory and Methods*. Nous présentons ici, dans la version parue, les parties de cet article où sont donnés les résultats principaux et leurs preuves.

4.3.1 Introduction

Consider a sequence of strictly stationary random variables (rv) $T_1, T_2, \dots$ with common unknown absolutely continuous distribution function (df) F . These rv are regarded as the lifetimes of the items under study. In many situations, due to possible withdrawals of the items from the life testing experimentation, the lifetimes may not be directly observable. Instead, we observe only censored lifetimes. That is, assuming that $C_1, C_2, \dots, C_n$ are n censoring rv with common unknown continuous df G , we observe only the n pairs $\{(Z_i, \delta_i), i = 1, 2, \dots, n\}$, where $Z_i = T_i \wedge C_i$ and $\delta_i = \mathbb{I}_{\{T_i \leq C_i\}}$ (where $\mathbb{I}_A$ denotes the indicator function of the set A).

Throughout this paper, to ensure the identifiability of the model, we suppose that the sequences $(T_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ are mutually independent, but each of the sequences satisfies some kind of dependence such as α -mixing. Now, it is known (see Lemma 2 in Cai (2001)) that the Z_i are α -mixing (with an appropriate coefficient α) with common df $H(t) = 1 - (1 - F(t))(1 - G(t))$ (by independence). It is clear that the uncensored model is the special case of the censored model with $G = 0$. Consider the following two stochastic processes

$$N_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \leq t, \delta_i=1\}} = \sum_{i=1}^n \mathbb{I}_{\{T_i \leq t \wedge C_i\}}$$

(the number of uncensored observations less than or equal to t), and

$$Y_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \geq t\}}.$$

Then, the well known Kaplan-Meier (1958) estimator (KME) based on the n pairs $\{(Z_i, \delta_i), 1 \leq i \leq n\}$ is given by

$$1 - \widehat{F}_n(x) = \prod_{s \leq x} \left(1 - \frac{dN_n(s)}{Y_n(s)} \right) \quad (4.4)$$

where $dN_n(t) = N_n(t) - N_n(t^-)$ and $N_n(t^-) = \lim_{\varepsilon \rightarrow 0^+} N_n(t - \varepsilon)$. The cumulative hazard function $\Lambda(t)$ is defined by

$$\Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s^-)}. \quad (4.5)$$

Clearly, $\Lambda(t) = -\log(1 - F(t))$ in our case.

Let $F_*(t) = \mathbb{P}(Z_i \leq t, \delta_i = 1) = \int_0^{+\infty} F(t \wedge s) dG(s) = \int_0^t (1 - G(s)) dF(s)$.

From (4.5) it can be shown that

$$\Lambda(t) = \int_0^t \frac{dF_*(s)}{\overline{H}(s)},$$

where $\overline{H} = 1 - H$.

Since $N_n(t)$ and $Y_n(t)$ are clearly the empirical estimates of $F_*(t)$ and $\overline{H}(t)$ respectively, the empirical cumulative hazard function is given by

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)},$$

which is referred to as the Nelson-Aalen estimator of $\Lambda(t)$ in the literature.

In the case where $(T_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ are two independent α -sequences, Cai (1998, 2001) established the strong representation and asymptotic normality for (4.4), with uniform consistency and gave the rates. He also studied a stochastic process generated by (4.4), which converges weakly to a Gaussian process with a specified covariance structure. For more details on the applications of the KME for dependent data in medicine and agriculture, we refer the reader to the papers by Wei *et al.* (1989), Kohler (1994), Kohler and Symanovski (1995) and Kang and Kohler (1997).

For $0 < p < 1$, the p th quantile of $F(t)$ is defined by

$$F^{-1}(p) = \inf\{x \in \mathbb{R}; F(x) \geq p\} \quad (4.6)$$

and the sample estimator of $F^{-1}(p)$ is defined by

$$\widehat{F}_n^{-1}(p) = \inf\{x \in \mathbb{R}; \widehat{F}_n(x) \geq p\}, \quad (4.7)$$

where $\widehat{F}_n(x)$ is the KME defined in (4.4).

In the independent framework with no censoring, the properties of estimator (4.7) (where $\widehat{F}_n$ is replaced by the empirical df F_n) have been extensively studied (see e.g.

Scörgő (1983) or Shorack and Wellner (1986)). Under ϕ -mixing condition (for the definition see Doukhan (1996)), the Bahadur representation was obtained by Sen (1972) and the extension to the α -mixing case was obtained by Yoshihara (1995). The strong approximation of the quantile process defined by $\rho_n(s) = n^{1/2}f(F^{-1}(s))(F^{-1}(s) - F_n^{-1}(s))$, where f is density function and F_n^{-1} is sample quantile function and F_n , was obtained by Fotopoulos *et al.* (1994) and was later improved by Yu (1996).

For a censored model with $(T_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ independent and identically distributed (iid) sequences and mutually independent, Padgett and Thombs (1989) stated the strong consistency and asymptotic normality for a smooth estimator of (4.6). Sander (1975) obtained some asymptotic properties, and Csörgő (1983) and Cheng (1984) discussed strong approximation results with some applications for (4.7).

In this paper, we establish the pointwise consistency and give a strong representation of Kaplan-Meier quantile function $\hat{F}_n^{-1}(\cdot)$ under a strong-mixing hypothesis.

4.3.2 Assumptions and Results

Now we introduce some assumptions that are used to state our results gathered below for easy reference.

A1. $(T_i)_{i \geq 1}$ is a sequence of stationary α -mixing random variables with continuous df F and mixing coefficient $\alpha_1(n)$.

A2. $(C_i)_{i \geq 1}$ is a sequence of stationary α -mixing random variables with continuous df G and mixing coefficient $\alpha_2(n)$. Moreover the censoring times are independent of $(T_i)_{i \geq 1}$.

A3. $\alpha(n) = O\left(e^{-\log^{1+\zeta} n}\right)$ for some $\zeta > 0$, with $\alpha(n) = \max(\alpha_1(n), \alpha_2(n))$.

Remark 4.2. In the work of Cai (1998), the assumption on the mixing coefficient is $\alpha(n) = O(n^{-\nu})$ for some $\nu > 3$ which is a little less quick than our assumption **A3**. Therefore the results of Cai (1998) remains valid under our assumptions. Furthermore, the polynomial rate can not be used. Indeed, this choice made the bounded term $\mathcal{J}_2$ in (4.16) as a general term of divergent series for all choice of q_n and ε_n , where ε_n is a bandwidth of the grid of $[0, \tau]$ and its choice is imposed by the strong representation of the df.

This assumption is needed in order to use the strong representation of the Kaplan-Meier estimator obtained by Cai (1998) which comes from Dhompongsa's (1984) result. We point out that we do not require the censoring times $(C_i)_{i \geq 1}$ to be mutually independent, which is done by Lecoutre and Ould-Saïd (1995a, b), Cai (1998) and Cai and Roussas (1998).

For any df $W(\cdot)$, define the (possibly infinite) time $\tau_W = \sup\{x; W(x) < 1\}$, and let $\tau < \tau_H$. It is clear by independence that $\tau_H = \tau_F \wedge \tau_G$.

Let $F_*(t) = \mathbb{P}(Z_i \leq t, \delta_i = 1) = \int_0^{+\infty} F(t \wedge s) dG(s) = \int_0^t (1 - G(s)) dF(s)$.

Let $g(x) = \int_0^x (\overline{H}(u))^{-2} dF_*(u)$ and for positive reals z and x , and $\delta \in \{0, 1\}$ let

$$\xi(z, \delta, x) = g(z \wedge x) - \mathbb{I}_{\{z \leq x, \delta=1\}} / \overline{H}(z). \quad (4.8)$$

It can be shown that $\mathbb{E}[\xi(z_i, \delta_i, x)] = 0$ and $\text{cov}(\xi(z_i, \delta_i, s), \xi(z_i, \delta_i, t)) = g(s \wedge t)$.

Now, we are in position to give our main results. The first result gives the strong consistency of $\hat{F}_n^{-1}$.

Lemma 4.1. *Let $0 < p < 1$. Under **A1-A3**, assuming that F is continuous and that $F^{-1}(p)$ is the unique solution of $F(x) = p$, then*

$$\hat{F}_n^{-1}(p) \longrightarrow F^{-1}(p) \text{ a.s.}$$

The strong representation of the sample quantile is given in the following

Theorem 4.6. *Let $0 < p < 1$. Under **A1-A3**, assuming that F is Lipschitz continuous and continuously differentiable at $F^{-1}(p)$ with $F'(F^{-1}(p)) = f(F^{-1}(p)) > 0$, then*

$$\begin{aligned} \hat{F}_n^{-1}(p) - F^{-1}(p) &= \frac{p - \hat{F}_n(F^{-1}(p))}{f(F^{-1}(p))} + o(a_n) \text{ a.s.} \\ &= o_{\mathbb{P}}(n^{-1/2}) \end{aligned}$$

where $a_n = \sqrt{\log \log n / n}$. In addition, if F is twice continuously differentiable at $F^{-1}(p)$, then

$$\hat{F}_n^{-1}(p) - F^{-1}(p) = \frac{p - \hat{F}_n(F^{-1}(p))}{f(F^{-1}(p))} + O\left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}}\right) \text{ a.s. where } \beta > 0 \text{ and } \lambda > 0.$$

Now we present an implication for our strong representation result. It is clear to see that $\{\xi(Z_i, \delta_i, x)\}$ defined in (4.8) is a sequence of stationary α -mixing bounded random variables. Applying Theorem 1, Cai's Theorem 2 and Theorem 18.5.4 in Ibragimov and Linnick (1981) we have

Theorem 4.7. *Under the assumptions of Theorem 4.6, we have*

$$\sqrt{n} \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2)$$

$$\text{where } \Sigma^2 = \frac{(1-p)^2}{(f(F^{-1}(p)))^2} \left\{ \text{Var}(\xi(Z_1, \delta_1, F^{-1}(p))) + 2 \sum_{j=2}^{\infty} \text{cov}(\xi(Z_1, \delta_1, F^{-1}(p)), \xi(Z_j, \delta_j, F^{-1}(p))) \right\},$$

$$\text{Var}(\xi(Z_1, \delta_1, F^{-1}(p))) = \int_0^{F^{-1}(p)} \frac{F_*(du)}{\overline{H}^2(u)}$$

and $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution.

Remark 4.3. *In the α -mixing case, we can not achieve the same rate as in the iid case. The main reason is that our approach utilizes the strong approximation of Kaplan-Meier estimator which is expressed as a mean of random variables with negligible remainder term of order $\log^{-\lambda} n/n^{1/2}$. This is not as sharp as in iid case, for which the order is $\log n/n$ (see, Lo, Mack and Wang (1989)).*

4.3.3 Auxiliary results and proofs

Proof of Lemma 4.1

The proof is mainly based on Theorem 1 ii) in Cai (2001) and the properties of the quantile function. Therefore, it is omitted. $\blacksquare$

To prove Theorem 4.6, we will need some lemmas. The first lemma deals with the behavior of the difference between $\hat{F}_n(\hat{F}_n^{-1}(p))$ and p .

Lemma 4.2. *Let $0 < p_0 \leq p_1 < 1$. Under **A1-A3**, assuming that F is continuous, then*

$$\sup_{p_0 \leq p \leq p_1} \left| \hat{F}_n(\hat{F}_n^{-1}(p)) - p \right| = O(b_n) \quad a.s. \quad (4.9)$$

where $b_n = \log^{-\lambda} n/n^{1/2}$ with $\lambda > 0$.

Proof :

The strong representation of Kaplan-Meier estimator in the strong mixing case obtained by Cai (Theorem 2, 1998) gives

$$\hat{F}_n(x) - F(x) = (1 - F(x))S_n(x) + R_n(x)$$

with $S_n(x) = -\frac{1}{n} \sum_{i=1}^n \xi(Z_i, \delta_i, x)$ and $\sup_{x \geq 0} R_n(x) = O(b_n) \quad a.s.$

Let $x = \hat{F}_n^{-1}(p)$, we have

$$\begin{aligned} \sup_{p_0 \leq p \leq p_1} \left| \hat{F}_n(\hat{F}_n^{-1}(p)) - p \right| &\leq \sup_{\hat{F}_n^{-1}(p_0) \leq x \leq \hat{F}_n^{-1}(p_1)} \left| \hat{F}_n(x) - \hat{F}_n(x^-) \right| \\ &\leq \sup_{\hat{F}_n^{-1}(p_0) \leq x \leq \hat{F}_n^{-1}(p_1)} \left[|S_n(x) - S_n(x^-)| + R'_n(x) \right] \end{aligned}$$

where

$$\sup_{\hat{F}_n^{-1}(p_0) \leq x \leq \hat{F}_n^{-1}(p_1)} |R'_n(x)| = \sup_{\hat{F}_n^{-1}(p_0) \leq x \leq \hat{F}_n^{-1}(p_1)} |R_n(x) - R_n(x^-)| = O(b_n). \quad (4.10)$$

Now, we have

$$\begin{aligned}
S_n(x) - S_n(x^-) &= \frac{1}{n} \left| \sum_{i=1}^n \xi(Z_i, \delta_i, x) - \sum_{i=1}^n \xi(Z_i, \delta_i, x^-) \right| \\
&\leq \frac{1}{n} \left| \sum_{i=1}^n \frac{\mathbb{I}_{\{X_i \leq x\}} - \mathbb{I}_{\{X_i < x\}}}{\overline{H}(X_i)} \right| \\
&= \frac{1}{n \overline{H}(x)} \sum_{i=1}^n \mathbb{I}_{\{X_i = x\}}.
\end{aligned}$$

Or $\frac{1}{\overline{H}(x)}$ is bounded by $\frac{1}{\overline{H}(\tau_H)}$ then we have

$$\sup_{\hat{F}_n^{-1}(p_0) \leq x \leq \hat{F}_n^{-1}(p_1)} [|S_n(x) - S_n(x^-)|] = o(n^{-1+\nu})$$

for all $0 < \nu < 1/2$ (see Lemma 1 in Lemdani *et al.* (2005), for the ex-aequo terms in the α -mixing case), which combined with (4.10) give (4.9). $\blacksquare$

Lemma 4.3. *Under A1-A3, assuming that $F' = f$ is continuous and strictly positive on $[F^{-1}(p_0) - \delta, F^{-1}(p_1) + \delta]$, for some $\delta > 0$. Then*

$$\sup_{p_0 \leq p \leq p_1} |\hat{F}_n^{-1}(p) - F^{-1}(p)| = O(a_n) \quad a.s. \quad (4.11)$$

and

$$\sup_{p_0 \leq p \leq p_1} |\hat{F}_n^{-1}(p) - F^{-1}(p)| = O_{\mathbb{P}}(n^{-1/2}). \quad (4.12)$$

Proof :

From Theorem 2 in Cai (2001) and for $\tau < \tau_F$, we have

$$\sup_{0 \leq x \leq \tau} |\hat{F}_n(x) - F(x)| = O(a_n) \quad a.s.$$

We may write

$$\hat{F}_n(\hat{F}_n^{-1}(p)) = F(\hat{F}_n^{-1}(p)) + O(a_n) \quad a.s.$$

A Taylor expansion of F in the neighborhood of $F^{-1}(p)$ yields

$$\hat{F}_n(\hat{F}_n^{-1}(p)) = F(F^{-1}(p)) + [\hat{F}_n^{-1}(p) - F^{-1}(p)] f(\xi^*) + O(a_n)$$

where ξ^* is between $F^{-1}(p)$ and $\hat{F}_n^{-1}(p)$.

Hence

$$\hat{F}_n^{-1}(p) - F^{-1}(p) \leq \frac{[\hat{F}_n(\hat{F}_n^{-1}(p)) - F(F^{-1}(p))]}{f(\xi^*)} + O(a_n).$$

By Lemma 4.1, we have

$$\begin{aligned} \sup_{p_0 \leq p \leq p_1} \left| \hat{F}_n^{-1}(p) - F^{-1}(p) \right| &\leq \frac{\sup_{p_0 \leq p \leq p_1} \left| \left[\hat{F}_n \left(\hat{F}_n^{-1}(p) \right) - F \left(F^{-1}(p) \right) \right] \right|}{\inf_{\xi \in [F^{-1}(p); \hat{F}_n^{-1}(p)]} f(\xi)} + O(a_n) \\ &= O(b_n) + O(a_n) = O(a_n) \quad a.s. \end{aligned}$$

Now, we introduce the following lemma due to Bosq (1998) given in the slightly different version which is needed in the forthcoming lemma.

Lemma 4.4. *Let $(U_i)_{i \in \mathbb{N}}$ be a sequence of stationary strongly mixing centered rv, such that*

$$\forall n \in \mathbb{N}, \forall i \in \mathbb{N}, \sup_{1 \leq i \leq n} |U_i| \leq d \quad \text{and for each integer } q \in [1; n/2] \quad \text{and each } \varepsilon > 0$$

$$\mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n U_i \right| > \varepsilon \right) \leq 4 \exp \left(-\frac{\varepsilon^2 q}{8} \left(\frac{8q^2 \sigma^2(q)}{n^2} + \frac{d\varepsilon}{2} \right)^{-1} \right) + 22 \left(1 + \frac{4d}{\varepsilon} \right)^{1/2} q \alpha \left(\left\lceil \frac{n}{2q} \right\rceil \right)$$

where

$$\sigma^2(q) = \frac{K_1 n}{2q} \mathbb{E} [U_1^2]$$

The next lemma provides an order of magnitude of the oscillation behaviour of $\hat{F}_n - F$.

Lemma 4.5. *Let $\lambda_n = \text{const.} b_n$. Under **A1-A3** and assuming that F is Lipschitz continuous on $[0, \tau]$. Then*

$$\sup_{|s-t| \leq \lambda_n} \left| \left(\hat{F}_n(s) - F(s) \right) - \left(\hat{F}_n(t) - F(t) \right) \right| = O \left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}} \right) \quad a.s.$$

where $\beta > 0$ and $\lambda > 0$.

Proof :

We first show the statement for $\hat{\Lambda}_n - \Lambda$ rather than $\hat{F}_n - F$. The strong representation of the cumulative hazard function gives (see : Cai 1998, p. 386, 387)

$$\begin{aligned} \hat{\Lambda}_n(s) - \hat{\Lambda}_n(t) - \Lambda(s) + \Lambda(t) &= \int_s^t \frac{d\bar{N}_n(x)}{\bar{H}(x)} - \int_s^t \frac{\bar{Y}_n(x)}{(\bar{H}(x))^2} dF_*(x) + O(b_n) \quad a.s. \\ &= \int_s^t \frac{d(\bar{N}_n(x) - F_*(x))}{\bar{H}(x)} - \int_s^t \frac{\bar{Y}_n(x) - \bar{H}(x)}{(\bar{H}(x))^2} dF_*(x) + O(b_n) \quad a.s. \\ &=: A - B + O(b_n) \quad a.s. \end{aligned} \tag{4.13}$$

To deal with B , we treat $\bar{Y}_n(t)$ as an empirical df associated to (Z_i, δ_i) and from Theorem 1 in Cai (1998), we have

$$\sup_{t \geq 0} |\bar{Y}_n(t) - \bar{H}(t)| = O(a_n) \quad a.s.$$

Making use of Lipschitz continuity of F , B is bounded by

$$O\left(\frac{\sqrt{\log \log n}}{n \log^\lambda n}\right) = o\left(\frac{\log^{1-\lambda} n}{n}\right) \quad a.s. \quad (4.14)$$

Now in order to deal with A , put $\varepsilon_n = K_0 \log^{-\beta-\lambda} n / \sqrt{n}$ with $\beta > 0$ and $\lambda > 0$, and consider the grid

$$x_{n,j} = j\varepsilon_n; \quad j = 0, 1, \dots, k_n = \left\lceil \frac{\tau}{\varepsilon_n} \right\rceil, \quad x_{n,k+1} = \tau.$$

For each $x \in [0, \tau]$ choose $x_{n,j}$ such that $x_{n,j} \leq x \leq x_{n,j+1}$. Using easy notations we put $\underline{y} := x_{n,j}$ and $\bar{y} := x_{n,j+1}$.

Then

$$\begin{aligned} \int_s^{\bar{y}} \frac{d(\bar{N}_n(x) - F_*(x))}{\bar{H}(x)} &\leq \int_{\underline{s}}^{\bar{y}} \frac{d(\bar{N}_n(x) - F_*(x))}{\bar{H}(x)} + O(\varepsilon_n) \\ &\equiv \eta_n + O(\varepsilon_n). \end{aligned} \quad (4.15)$$

The η_n can be considered as a mean of α -mixing rv that is $\eta_n = \frac{1}{n} \sum_{i=1}^n \varsigma_i$ with

$$\varsigma_i = \int_{\underline{s}}^{\bar{y}} \frac{d(N^i(x) - F_*(x))}{\bar{H}(x)} \quad \text{and} \quad N^i(x) = \mathbb{I}_{\{Z_i \leq x\}}.$$

We have

$$\begin{aligned} |\varsigma_i| &\leq \int_{\underline{s}}^{\bar{y}} \left| \frac{d(N^i(x) - F_*(x))}{\bar{H}(x)} \right| \\ &\leq \frac{1}{\bar{H}(\tau_H)} \left(1 + \lambda_n \sup_{x \in [0, \tau]} f_* \right) =: d. \end{aligned}$$

A simple algebraic calculation gives

$$\mathbb{E}[\varsigma_i] = 0,$$

$$\text{Var}[\varsigma_i] \leq \int_{\underline{s}}^{\bar{y}} \frac{dF_*(x)}{\bar{H}^2(x)} \leq K_1 \cdot \lambda_n = O\left(\frac{\log^{-\lambda} n}{n^{1/2}}\right).$$

Now we apply Lemma 4.4 for $U_i = \varsigma_i$ with $\varepsilon = \varepsilon_n$ and $q = q_n$.

$$\begin{aligned} \mathbb{P}\left(\frac{1}{n} \left| \sum_{i=1}^n \varsigma_i \right| > \varepsilon_n\right) &\leq 4 \exp \left\{ -\frac{\varepsilon_n^2 q_n}{8} \left[\frac{8q_n^2 \sigma^2(q_n)}{n^2} + \frac{d\varepsilon_n}{2} \right]^{-1} \right\} + 22q_n \left(1 + \frac{4d}{\varepsilon_n} \right)^{1/2} \alpha \left(\left\lceil \frac{n}{2q_n} \right\rceil \right) \\ &:= \mathcal{J}_1 + \mathcal{J}_2. \end{aligned} \quad (4.16)$$

Now taking $q_n = K_2 n^\kappa$ for some $1/2 < \kappa < 1$ we have, for n large enough, $\frac{q_n^2 \sigma^2(q_n)}{n^2}$ is of order $n^{\kappa-3/2}(\log n)^{-\lambda}$ and is therefore asymptotically negligible than $\frac{d\varepsilon_n}{2}$ (since $\kappa \in (1/2, 1)$). Then, for n large enough

$$\mathcal{J}_1 \leq 4 \exp \left\{ K_3 \frac{n^{\kappa-1/2}}{\log^{\beta+\lambda} n} \right\}. \quad (4.17)$$

The second term of (4.16) can be bounded by

$$\mathcal{J}_2 \leq n^{\kappa-1/4} \log^{\frac{\beta+\lambda}{2}} n \exp \left\{ -(1-\kappa)^{1+\zeta} (\log K_4 n)^{1+\zeta} \right\}. \quad (4.18)$$

Each bounded term in (4.17) and (4.18) can be made $o(n^{-5/2})$ for an appropriate choose of constants.

Now we come back to (4.15). Since the number of the η_n 's is at most $O(n)$, (4.15), (4.16), (4.17) and (4.18) in conjunction with Borel-Cantelli's lemma implies that almost surely

$$\max_n |\eta_n| \leq \varepsilon_n \text{ a.s.}$$

Now this fact, (4.13), (5.42) and (4.15) allow us to conclude for the upper bound of the oscillation modulus of $\hat{\Lambda}_n - \Lambda$. The lower bound can be derived similarly by considering in (4.15) $\bar{s}$ and $\bar{t}$ instead of $\underline{s}$ and $\bar{t}$ respectively. The corresponding result for $\hat{F}_n - F$ is similar to the proof of Theorem 2 in Cai (1998). Therefore, it is omitted. ■

Proof of Theorem 4.6 Using lemmas 4.3 and 4.5, a Taylor expansion of F gives

$$\begin{aligned} \hat{F}_n \left(\hat{F}_n^{-1}(p) \right) - \hat{F}_n \left(F^{-1}(p) \right) &= F \left(\hat{F}_n^{-1}(p) \right) - F \left(F^{-1}(p) \right) O \left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}} \right) \\ &= \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right) f \left(F^{-1}(p) \right) \\ &+ o \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right) + O \left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}} \right). \end{aligned} \quad (4.19)$$

Then the result follows from Lemma 2. ■

Under the hyphotesis made on F , one can expand (4.19) further to get

$$O \left[\left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right)^2 \right] \text{ instead of } o \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right).$$

The remainder term is therefore $O \left(\frac{\log^{-\beta-\lambda} n}{\sqrt{n}} \right)$. Since the bounds leading to (4.19) are uniform in $p_0 \leq p \leq p_1$, then also the error bounds hold uniformly in the Taylor expansion under the stated regularity assumptions on F . ■

Chapitre 5

La fonction quantile conditionnelle

5.1 Introduction

Considérons deux v.a. continues T et X , T est la variable d'intérêt et X est la covariable ou la variable explicative, pouvant prendre des valeurs dans $\mathbb{R}^d, d \geq 1$.

Soit $F(.,.)$ la f.d.r. conjointe de T et X ; c'est à dire pour tout $t \in \mathbb{R}$ et $x \in \mathbb{R}^d$,

$$F(t, x) = \mathbb{P}[T \leq t, X \leq x];$$

avec $X \leq x \Leftrightarrow X_1 \leq x_1, X_2 \leq x_2, \dots, X_d \leq x_d$.

Pour tout $x \in \mathbb{R}^d$, la f.d.r. conditionnelle de T sachant $X = x$ est donnée par

$$\forall t \in \mathbb{R}, F(t|x) = \mathbb{P}(T \leq t | X = x) = \mathbb{E}[\mathbb{1}_{\{T \leq t\}} | X = x] \quad (5.1)$$

et peut se mettre sous la forme $F(t|x) =: \frac{F_1(t, x)}{\ell(x)}$ où ℓ est la densité marginale de X .

Pour tout $p \in (0, 1)$, la fonction quantile conditionnelle de T sachant $X = x$ est définie par

$$\xi_p(x) = F^{-1}(p|x) = \inf\{t \in \mathbb{R}; F(t|x) \geq p\} \quad (5.2)$$

et satisfait lorsque $F(.,.)$ est continue

$$F(\xi_p(x)|x) = p. \quad (5.3)$$

une caractérisation alternative du quantile conditionnel $\xi_p(x)$ est obtenue sous forme d'un problème d'optimisation

$$\xi_p(x) = \arg \min_{\theta \in \mathbb{R}} \mathbb{E} \left[(2p - 1)(T_{n+m} - \theta) + |T_{n+m} - \theta| \mid X_n = x \right].$$

Il est clair qu'un estimateur de (5.2) s'obtient en estimant (5.1).

Plusieurs approches ont été développées pour l'estimation des quantiles conditionnels. L'approche paramétrique, généralement fondée sur des considérations de normalité de la distribution conditionnelle, est souvent mal adaptée à la réalité des données pratiques dont on ne connaît pas la loi sous-jacente. De nombreux travaux récents ont été menés pour l'estimation non paramétrique des quantiles conditionnels aussi bien dans le cadre théorique que sur le plan des applications.

5.1.1 Cas de données complètes

En absence de censure, un estimateur linéaire de $F(t|x)$ est donné par

$$F_n(t|x) = \sum_{i=1}^n W_{in}(x) \mathbb{I}_{\{T_i \leq t\}}, \quad (5.4)$$

où les poids $W_{in}(\cdot)$ sont des fonctions de x mesurables dépendant uniquement de $X_1, X_2, \dots, X_n$.

L'estimateur à noyau de $F(t|x)$ s'obtient avec les poids de Nadaraya-Watson, définis par

$$W_{in}(x) = \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{h_n}\right)}. \quad (5.5)$$

La fonction K est une densité de probabilité, appelée noyau et h_n est une suite de réels positifs tendant vers 0 et appelée fenêtre.

Il en découle naturellement un estimateur de $\xi_p(x)$ de la forme suivante

$$\xi_{p,n}(x) = F_n^{-1}(p|x) = \inf\{t \in \mathbb{R}; F_n(t|x) \geq p\}.$$

Une version plus "lisse" de l'estimateur de la f.d.r. conditionnelle $F(\cdot|.)$, peut être obtenue en remplaçant dans (5.4) la fonction indicatrice par une densité de probabilité φ , l'estimateur correspondant, est appelée estimateur par noyau produit et est défini, pour tout $t \in \mathbb{R}$ comme suit

$$F_n(t|x) = \frac{\sum_{i=1}^n K\left(\frac{x-X_i}{a_n}\right) \Psi\left(\frac{t-T_i}{b_n}\right)}{\sum_{j=1}^n K\left(\frac{x-X_j}{a_n}\right)}, \quad (5.6)$$

où Ψ est la f.d.r. associée à la densité de probabilité φ .

La qualité de cet estimateur est étroitement liée au choix des deux paramètres de lissage a_n et b_n ; beaucoup d'auteurs se sont intéressés à ce problème, nous citons en particulier Yu et Jones (1998) qui proposent une méthode empirique.

Samanta(1989) établit, pour des données i.i.d., la convergence forte et la normalité asymptotique de l'estimateur du quantile obtenu en prenant dans (5.6) le noyau $\varphi = K$ et les fenêtres a_n et b_n égales.

Berlinet *et al.* (2001) considèrent l'estimateur $\xi_{p,n}(x) = F_n^{-1}(p|x)$ en prenant

$$F_n(t|x) = \frac{\sum_{i=1}^n K_0\left(\frac{x-X_i}{h_n}\right) \left(\int_{-\infty}^t K_1\left(\frac{z-T_i}{h_n}\right) dz\right)}{\sum_{j=1}^n K_0\left(\frac{x-X_j}{h_n}\right)},$$

pour des données dépendantes. Sous des hypothèses, en particulier sur les noyaux K_0 et K_1 , la fenêtre h_n , les densités de probabilités $f(.,.)$ et ℓ et sur la structure de dépendance, ils établissent la convergence et la normalité asymptotique de $\xi_{p,n}(x)$, en particulier dans le cas de l' α -mélange.

D'autres méthodes d'estimation furent considérées par d'autres auteurs, nous citons en particulier Honda (2000) qui utilise la méthode des polynômes locaux pour construire un estimateur uniformément convergent et asymptotiquement normal pour le quantile conditionnel dans le cas de l' α -mélange.

5.1.2 Cas de données censurées

Soit la v.a. T , représentant la durée de vie d'un individu censurée à droite par la v.a. C . Lors d'une étude expérimentale, on relève de plus d'autres données relatives à l'état de l'individu, comme par exemple la pression artérielle, le taux de cholestérol, etc..., pour une étude médicale; dans un tel cas, on regroupe toutes les mesures annexes dans la covariable, notée X .

La f.d.r. de T sachant $X = x$ est donnée par (5.1) qu'on suppose dérivable par rapport à t et de dérivée $f(t|x)$.

Sous l'hypothèse que T et C sont conditionnellement indépendants sachant X et on observe les v.a. $Y_i = T_i \wedge C_i$, $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$ et X_i pour $i \geq 1$ tels que les (Y_i, δ_i, X_i) sont i.i.d., Beran (1981) proposa d'estimer les fonctions de survie suivantes

$H_1(t|x) = \mathbb{P}(Y > t, \delta = 1|X = x)$ et $H_2(t|x) = \mathbb{P}(Y > t|X = x)$

par des estimateurs à noyau de type Nadaraya-Watson (Nadaraya 1964; Watson 1964)

$$\hat{H}_{in}(t|x) = (\ell_n^{-1}(x)) H_{in}(t, x),$$

où

$$H_{1n}(t, x) = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{1}_{\{Y_i > t, \delta_i = 1\}} K\left(\frac{x - X_i}{h_n}\right)$$

$$H_{2n}(t, x) = \frac{1}{nh_n} \sum_{i=1}^n \mathbb{1}_{\{Y_i > t\}} K\left(\frac{x - X_i}{h_n}\right)$$

et $\ell_n(x)$ est l'estimateur de Parzen-Rozenblatt (Parzen (1962); Rozenblatt (1971)) de

la densité $\ell(x)$ de la covariable; $\ell_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)$.

Les estimateurs de la fonction de hasard cumulative $\Lambda(t|x)$ et la fonction de survie $1 - F(t|x)$ sont respectivement obtenus par

$$\Lambda_n(t|x) = - \int_0^t \frac{H_{1n}(ds|x)}{H_{2n}(s^-|x)},$$

$$1 - F_n(t|x) = \prod_{s < t} \{1 - \Lambda_n(ds|x)\}. \quad (5.7)$$

Les propriétés de $F_n(t|x)$ ont été étudiées par beaucoup d'auteurs en particulier Beran (1981). Dabrowska (1987, 1989) montre que $F_n(t|x)$ est uniformément convergent et que convenablement normalisé, il converge vers un mouvement brownien. Ces résultats sont redémontrés par Van Keilegom et Veraverbeke (1997) avec un choix optimal de la fenêtre h_n .

Lecoutre et Ould-Saïd (1995) établissent la convergence uniforme presque complète de l'estimateur (5.7) pour des données censurées fortement mélangeantes.

Dabrowska (1992) considère l'estimateur du quantile conditionnel défini par

$$\xi_{p,n}(x) = \inf\{t; F_n(t|x) \geq p\}. \quad (5.8)$$

En utilisant le résultat de Stute (1984) sur la majoration de type exponentiel du module de variation des processus empiriques, Dabrowska (1992) donne une représentation de type Bahadur pour le quantile conditionnel. Sous des hypothèses de régularité sur la densité $\ell(x)$, les fonctions $H_i(t|x)$ et le noyau K , elle montre en particulier que

$$\xi_{p,n}(x) - \xi_p(x) = \frac{p - F_n(\xi_p(x)|x)}{f(\xi_p(x)|x)} + R_n(p, x),$$

avec $R_n(p, x)$ de l'ordre de $O\left(\left(\frac{\log n}{nh_n}\right)^{3/4}\right)$.

Van Keilegom et Veraverbeke (1997, 1998) introduisent la méthode du bootstrap sur les estimateurs (5.7) et (5.8) et étudient la convergence et la normalité asymptotique des estimateurs bootstrapés de la f.d.r. conditionnelle et du quantile conditionnel. Gannoun *et al.* (2005) utilisent la méthode dite de la constante locale "local constant kernel estimation" pour estimer le quantile conditionnel et ils établissent sa convergence presque sûre et sa normalité asymptotique.

Sous l'hypothèse que (T, X) et C sont indépendants, Ould-Saïd (2006) construit un estimateur de la f.d.r. conditionnelle (5.1) en suivant l'idée de Carbonez *et al.* (1995) : soit $\theta : \mathbb{R}^2 \rightarrow \mathbb{R}$ une application mesurable et on considère le problème de l'estimation de $\mathbb{E}[\theta(X, T)]$ à partir de l'échantillon $\{(Y_i, \delta_i, X_i); i = 1, \dots, n\}$.

Un estimateur sans biais de $\mathbb{E}[\theta(X, T)]$ est donné par $\frac{1}{n} \sum_{i=1}^n \frac{\delta_i \theta(X_i, Y_i)}{\bar{G}(Y_i)}$.

Dans le cas particulier où $\theta(X, T) = (\mathbb{1}_{\{T \leq t\}}|X)$ on obtient un estimateur sans biais de $F(\cdot|x)$.

Alors un estimateur à noyau de $F(t|x)$ est donnée par :

$$\begin{aligned} \tilde{F}_n(t|x) &= \sum_{i=1}^n W_{i,n}(x) \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}(t)} \\ &= \frac{\frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}(t)} K\left(\frac{x - X_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} \end{aligned} \quad (5.9)$$

Mais en pratique la f.d.r. G des censures est souvent inconnue, donc il est impossible d'utiliser l'estimateur (5.9). Par conséquent, on remplace G par son estimateur de Kaplan-Meier G_n donné par

$$1 - G_n(t) = \begin{cases} \prod_{i=1}^n \left(1 - \frac{1-\delta_i}{n-i+1}\right)^{\mathbb{I}_{\{Y_i \leq t\}}} & \text{if } t < Y_{(n)}, \\ 0 & \text{if } t \geq Y_{(n)}. \end{cases}$$

Un estimateur calculable de $F(\cdot|\cdot)$ est obtenu par

$$\begin{aligned} F_n(t|x) &= \sum_{i=1}^n W_{in}(x) \frac{\delta_i \mathbb{I}_{\{Y_i \leq t\}}}{G_n(Y_i)} \\ &= \frac{\frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{Y_i \leq t\}}}{G_n(Y_i)} K\left(\frac{x-X_i}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \end{aligned} \quad (5.10)$$

et qui peut se mettre sous la forme

$$F_n(t|x) =: \frac{F_{1,n}(x, t)}{\ell_n(x)},$$

où $\ell_n(\cdot)$ est l'estimateur à noyau de $\ell(\cdot)$ la densité des covariables, K est une densité de probabilité et h_n est une suite de réels positifs qui décroît vers zéro quand n tend vers l'infini.

Le quantile conditionnel est alors, estimé par

$$\xi_{p,n}(x) = \inf\{t \in \mathbb{R} : F_n(t|x) \geq p\}.$$

Ould-Saïd (2006) montre, sous des hypothèses de régularité sur $F(\cdot, \cdot)$, $F(\cdot|\cdot)$, $f(\cdot|\cdot)$, $\ell(\cdot)$, h_n et le noyau K , que (5.10) converge uniformément presque sûrement tout en précisant la vitesse de convergence et il déduit que

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \max\left\{\sqrt{\frac{\log \log n}{nh_n}}, h_n^2\right\} \quad p.s.$$

où Ω est un compact de $\mathbb{R}$ inclu dans le support de $\ell(\cdot)$.

5.2 Convergence presque sûre de l'estimateur à noyau du quantile conditionnel pour des variables censurées et dépendantes

Soit $\{(T_i, C_i, X_i), i \geq 1\}$ une suite de variables aléatoires stationnaires, de sorte que seules les v.a. $Y_i = T_i \wedge C_i$, $\delta_i = \mathbb{I}_{\{T_i \leq C_i\}}$ et X_i sont observées. Nous supposons que $\{C_i, i \geq 1\}$ et $\{(X_i, T_i), i \geq 1, \}$ sont indépendantes et que la suite $\{(Y_i, \delta_i, X_i), i \geq 1\}$ est fortement mélangeante de coefficient de mélange α .

Nous supposons que pour tout x fixé dans $\mathbb{R}$, $F(\cdot|x)$ est une fonction réelle continûment dérivable.

Soit $p \in (0, 1)$, on considère le problème de l'estimation du paramètre $\xi_p(x)$ défini par

$$F(\xi_p(x)|x) = p.$$

Les résultats présentés dans cette partie sont une extension des résultats d'Ould-Saïd (2006) au cas dépendant. Nous présentons ici, dans la version soumise, les parties de cet article où sont donnés les résultats principaux et leurs preuves.

5.2.1 Introduction

In mathematical statistics, there is a voluminous research on the quantile estimation. It is well known that the unconditional/conditional quantile function can give a good description of the data (see, e.g. Chaudhuri, Doksum and Samarov (1997)) especially the median which is highly robust against outliers. Many distinguished papers deal with this topic, to quote a few of them; in the independent and identically distributed (iid) framework without censoring, several authors considered this problem; for example, Mehra, Rao and Upadrashta (1991), Chaudhuri (1991a, 1991b), Fan, Hu and Truong (1994) and Xiang (1996). Honda (2000) dealt with the α -mixing process and proved the uniform convergence and asymptotic normality of an estimate of the conditional quantile using the local polynomial fitting method.

In the censored case, Dabrowska (1992) established a Bahadur type representation of kernel conditional quantile estimator, and Xiang (1995) treated the deficiency of the sample quantile estimator with respect to a kernel estimate using coverage probability. Qin and Wu (2001) gave a new estimator for a conditional quantile by using local linear fitting method when some auxiliary information is available. The asymptotic normality has been established. Ould-Saïd and Sadki (2005) established the strong representation of the sample quantile when the data are strongly mixing. Finally, Ould-Saïd (2006) established the strong consistency with rates for the same estimator of this paper in the iid case.

Our interest is to extend the obtained result by Ould-Saïd (2006) in the iid censorship model to the strong mixing case. Our results are stated under some suitable conditions upon the kernel function, the smoothing parameter and the distribution function (df) that appear in this model. The Kaplan-Meier estimator (KME) (see Kaplan and Meier (1958)), for the df of the censoring random variables (rv's), is used to soften the effect of the first censoring values.

5.2.2 Assumptions and main results

Let Ω be a compact set of $\mathbb{R}$ included in $\Omega_0 = \{x \in \mathbb{R} | \ell(x) > 0\}$, $\tau_F = \sup\{t; F(t) < 1\}$ and τ be a positive real number such as $\tau < \tau_F$.

Now introduce some assumptions, gathered below for easy reference, needed to state

our results.

A1. The mixing coefficient α is such that $\alpha(n) = O(n^{-\nu})$ for some $\nu > 3$.

A2. The bandwidth h_n satisfies $h_n \rightarrow 0$, and $nh_n \rightarrow +\infty$, and $\sqrt{\frac{\log \log n}{n}} = O(h_n^2)$.

A3. The kernel K is bounded and symmetric and has a compact support. It is also Hölderian of order γ for some $\gamma > 0$ and satisfies $\int_{\mathbb{R}} |uK(u)| du < \infty$ and $\int_{\mathbb{R}} u^2 |K(u)| du < \infty$

and its first derivative $K'(u)$ exists for all $u \in \mathbb{R}$ and is such as $\int_{\mathbb{R}} |K'(u)| du < \infty$.

A4. $F_1(.,.)$ is differentiable to order 2 with respect to the first component and

$$\sup_{(x,t) \in \Omega \times (0,\tau]} \left| \frac{\partial^2}{\partial x^2} F_1(x,t) \right| < \infty.$$

A5. The marginal density $\ell(.)$ has bounded second derivative and $\ell(x) < \gamma_0$ for all $x \in \Omega$ for some $\gamma_0 > 0$.

A6. K and F_1 satisfies $\int_{\mathbb{R}} |K(u)| F_1(u,t) du < \infty$ and $\int_{\mathbb{R}} K^2(u) F_1(u,t) du < \infty$.

A7. The conditional df $F(.,.)$ has bounded first derivative $f(.,.)$ and there exists $\gamma_1 > 0$ such that $f(t|x) > \gamma_1$ for all $(x,t) \in \Omega \times (0,\tau)$.

Our first result stated in Proposition 5.1, is the uniform almost sure convergence with a rate of the conditional df $F(.,.)$ estimator. The second result deals with the strong uniform convergence with a rate of the kernel conditional quantile estimator and is given in Theorem 5.1.

Proposition 5.1. *Under assumptions A1-A6, we have, for n large enough*

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)| = O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n^2}}, h_n \right\} \right) \quad a.s.$$

Theorem 5.1. *Under assumptions A1-A6 and if for each fixed $p \in (0,1)$, the function ξ_p satisfies*

$$\forall \varepsilon > 0, \exists \beta > 0, \forall \eta_p : \Omega \rightarrow \mathbb{R}, \sup_{x \in \Omega} |\xi_p(x) - \eta_p(x)| \geq \varepsilon \implies \sup_{x \in \Omega} |F(\xi_p(x)|x) - F(\eta_p(x)|x)| \geq \beta,$$

then, we have

$$\lim_{n \rightarrow \infty} \sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = 0 \quad a.s.$$

In addition, if A7 holds we have, for n sufficiently large

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n^2}}, h_n \right\} \right) \quad a.s.$$

Remark 5.1. Many authors assume that $\{T_i, i \geq 1, \}$ and $\{C_i, i \geq 1\}$ are conditionally independent given X , see for example Dabrowska (1992), but here we assume that $\{C_i, i \geq 1\}$ and $\{(X_i, T_i), i \geq 1, \}$ are independent; this condition is plausible whenever the censoring is independent of the modality of the patients under study. Furthermore, it is very usefull to give an unbiased estimator of $F(t|x)$.

Remark 5.2. If we choose $h_n = \left(\frac{\log \log n}{n}\right)^{\frac{1}{4}}$ then, we get the following rate

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O\left(\left(\frac{\log \log n}{n}\right)^{\frac{1}{4}}\right).$$

Under restrictive hypotheses upon $F_1(.,.)$ and $K(.)$ we can get the following rate

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O\left(\left(\frac{\log \log n}{n}\right)^{\frac{1}{3}}\right)$$

which is very close to the i.i.d. case.

5.2.3 Simulation study

We consider the strong mixing bidimensional process generated by :

$$\begin{cases} X_i = \rho X_{i-1} + \sqrt{1 - \rho^2} \epsilon_i, \\ T_i = X_{i+1} \quad i = 1, 2, \dots, n, \end{cases}$$

where $0 < \rho < 1$, (ϵ_i) is a white noise sequence and X_0 is a standard normal rv independent of (ϵ_i) . Clearly, the process (X_n) is an AR(1) and given $X_1 = x$, we have $T_1 = \rho x + \sqrt{1 - \rho^2} \epsilon_2$, hence it is distributed as a Gaussian rv $\mathcal{N}(\rho x, 1 - \rho^2)$. We also simulate n i.i.d. rv C_i exponentially distributed with parameter $\lambda = 0.125$. Clearly, the process (T_n, X_n, C_n) is stationary and strong mixing. Then we compute our estimator with the observed data (X_i, Y_i, δ_i) , where $Y_i = T_i \wedge C_i$ and $\delta_i = \mathbb{1}_{\{T_i \leq C_i\}}$, and we use the standard Gaussian kernel. We choose h_n , the bandwidth according to assumption **A2**. We take several values of n and in each case, we draw two curves corresponding to the conditional quantile estimator $\xi_{p,n}(x)$ for $x \in [-2, 2]$ and the theoretical conditional quantile $\xi_p(x)$ for $p = 1/2$ and $\rho = 0.8$

From the curves, it is clear that our estimator has good behavior in particular when n increase.

5.2.4 Auxiliary results and proofs

We need some auxiliary results to prove the main results.

The first Lemma deals with the behavior of the difference between $G_n(t)$ and $G(t)$.

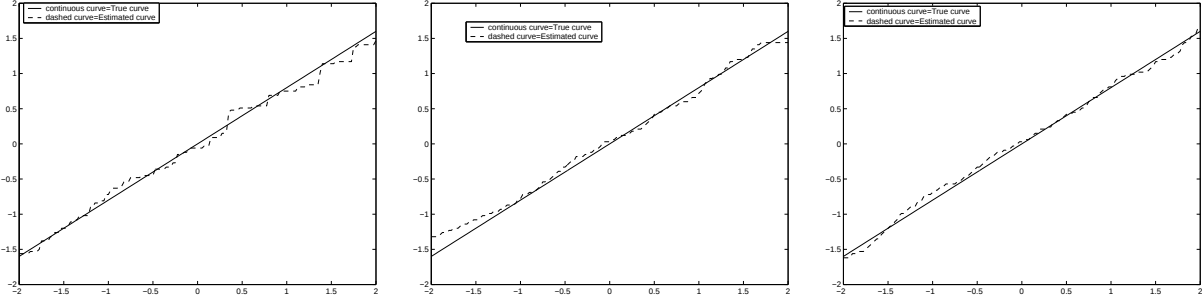


FIG. 5.1 – $n = 100, n = 500$ and $n = 1000$ respectively.

Lemma 5.1. *Under assumption A1 we have for n large enough*

$$\sup_{0 < t \leq \tau} |G_n(t) - G(t)| = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ a.s.}$$

proof : Let

$$N_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq t, \delta_i = 1\}} = \sum_{i=1}^n \mathbb{I}_{\{T_i \leq t \wedge C_i\}}$$

be the number of uncensored observations less than or equal to t ,

$$M_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Y_i \leq t, \delta_i = 0\}} = \sum_{i=1}^n \mathbb{I}_{\{C_i \leq t \wedge T_i\}}$$

be the number of censored observations less than or equal to t and

$$Y_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Y_i \geq t\}}.$$

be the number of observations great than or equal to t . Let

$$\Lambda(t) = \int_0^t \frac{dG(s)}{1 - G(s)}.$$

Clearly $\Lambda(t) = -\log(1 - G(t))$.

The empirical estimator of $\Lambda(t)$ is given by

$$\Lambda_n(t) = \int_0^t \frac{dM_n(s)}{Y_n(s)}.$$

Theorem 3.2 in Cai and Roussas (1992) permits to have

$$\sup_{t \geq 0} |\bar{Y}_n(t) - \bar{H}(t)| = O\left(\sqrt{\frac{\log \log n}{n}}\right) \text{ a.s.}$$

and

$$\sup_{t \geq 0} |\overline{M}_n(t) - \overline{L}_0(t)| = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad a.s.$$

where $\overline{Y}_n(t) = \frac{Y_n(t)}{n}$, $\overline{M}_n(t) = \frac{M_n(t)}{n}$, H is the common df of Y_i and $L_0(t) = \mathbb{P}(Y_i > t, \delta_i = 0)$.
Then we obtain that

$$\sup_{0 < t \leq \tau} |\Lambda_n(t) - \Lambda(t)| = O\left(\sqrt{\frac{\log \log n}{n}}\right) \quad a.s.$$

Therefore, Lemma 1 of Breslow and Crowley (1974) yields

$$\sup_{0 < t \leq \tau} |-\log(1 - G_n(t)) - \Lambda_n(t)| = O\left(\frac{1}{n}\right) \quad a.s.$$

Using the Taylor expansion, we obtain

$$G_n(t) - G(t) = e^{-\Lambda_n^*(t)} [\Lambda_n(t) - \Lambda(t)] + e^{-\Lambda_n^{**}(t)} [-\log(1 - G_n(t)) - \Lambda_n(t)]$$

where $\Lambda_n^*(t)$ is between $\Lambda_n(t)$ and $\Lambda(t)$ and $\Lambda_n^{**}(t)$ is between $-\log(1 - G_n(t))$ and $\Lambda_n(t)$.

This allows us to conclude. ■

Note that $\tilde{F}_n(t|x)$ can be written as

$$\tilde{F}_n(t|x) = \frac{\tilde{F}_{1,n}(x, t)}{\ell_n(x)} \quad (5.11)$$

where

$$\tilde{F}_{1,n}(x, t) := \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i \mathbb{I}_{\{Y_i \leq t\}}}{\overline{G}(Y_i)} K\left(\frac{x - X_i}{h_n}\right).$$

The second Lemma deals with the behavior of the difference between $F_{1,n}$ and $\tilde{F}_{1,n}$.

Lemma 5.2. *Under assumptions **A1**, **A2** and **A3** we have for n large enough*

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_{1,n}(x, t) - \tilde{F}_{1,n}(x, t)| = O(h_n) \quad a.s.$$

proof : we have

$$\begin{aligned} |F_{1,n}(x, t) - \tilde{F}_{1,n}(x, t)| &\leq \frac{1}{nh_n} \sum_{i=1}^n \left| \delta_i \mathbb{I}_{\{Y_i \leq t\}} K\left(\frac{x - X_i}{h_n}\right) \right| \left| \frac{1}{G_n(Y_i)} - \frac{1}{\overline{G}(Y_i)} \right| \\ &\leq \frac{1}{h_n} \frac{\|K\|_\infty}{\overline{G}_n(\tau) \overline{G}(\tau)} \sup_{0 < t \leq \tau} |G_n(t) - G(t)|. \end{aligned}$$

Since $\overline{G}(\tau) > 0$, in conjunction with Lemma 4.1, we have

$$\left| F_{1,n}(x, t) - \tilde{F}_{1,n}(x, t) \right| \leq \frac{c_1 \|K\|_\infty}{\overline{G}^2(\tau)} \frac{1}{h_n} \left(\frac{\log \log n}{n} \right)^{1/2} \quad a.s.$$

where c_1 is a positive constant. Assumption **A2** concludes the proof. $\blacksquare$

The following Lemma is Lemma 2 in Ould-Saïd (2006) and therefore its proof is omitted.

Lemma 5.3. *Under assumptions **A2**, **A3**, **A4** and **A6** we have for n large enough*

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] - F_1(x, t) \right| = O(h_n^2) \quad a.s.$$

Now, we introduce the following lemma due to Bosq (1998) needed in the forthcoming Lemma.

Lemma 5.4. *(Bosq 1998).*

Let $(U_i)_{i \in \mathbb{N}}$ be a sequence of strongly mixing centered rv, such that $\forall n \in \mathbb{N}, \forall i \in \mathbb{N}, 1 \leq i \leq n, |U_i| \leq d$ then for each $q \in [1, \frac{n}{2}]$ and for each $\varepsilon > 0$

$$\mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n U_i \right| > \varepsilon \right) \leq 4 \exp \left(-\frac{\varepsilon^2 q}{8} \left[\frac{8q^2 \sigma^2(q)}{n^2} + \frac{d\varepsilon}{2} \right]^{-1} \right) + 22 \left(1 + \frac{4d}{\varepsilon} \right)^{\frac{1}{2}} q \alpha \left(\left\lceil \frac{n}{2q} \right\rceil \right)$$

where $\sigma^2(q) = \frac{K_1 n}{2q} \mathbb{E}[U_1^2]$.

Lemma 5.5. *Under assumptions **A1**, **A2**, **A3**, **A5** and **A6** we have for n large enough*

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \tilde{F}_{1,n}(x, t) - \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] \right| \longrightarrow 0, \quad a.s. \text{ as } n \rightarrow \infty.$$

Furthermore

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \tilde{F}_{1,n}(x, t) - \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] \right| = O \left(\sqrt{\frac{\log \log n}{nh_n}} \right) \quad a.s.$$

Proof Ω is a compact set, then it can be covered by a finite number s_n of intervals of length h_n^β , where $\beta > \frac{\gamma+1}{\gamma}$, and centered at some points $x_k, k = 1, \dots, s_n$. Since Ω is bounded, there exists a constant M such that $s_n \leq Mh_n^{-\beta}$. Now, for all fixed $t \in (0, \tau]$ and $x \in \Omega$, put

$$\tilde{\Delta}_i(x) = \Delta_i(x) - \Delta_i(x_k),$$

for all $i = 1, \dots, n$, where

$$\Delta_i(x) := \frac{1}{nh_n} \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\overline{G}(Y_i)} K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left[\frac{1}{nh_n} \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\overline{G}(Y_i)} K \left(\frac{x - X_i}{h_n} \right) \right].$$

Clearly, we have

$$|\Delta_i(x)| \leq |\Delta_i(x_k)| + |\tilde{\Delta}_i(x)|.$$

By **A3**, it follows that there exists a constant C such that

$$\sup_{x \in \Omega} \left| \sum_{i=1}^n \tilde{\Delta}_i(x_k) \right| \leq 2Ch_n^{\beta\gamma - (\gamma+1)}.$$

The assumptions on h_n and β imply that

$$\sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \sum_{i=1}^n \tilde{\Delta}_i(x_k) \right| = o(1) \text{ a.s.}$$

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

$$\mathbb{P} \left\{ \max_{k=1, \dots, s_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} \leq \sum_{k=1}^{s_n} \mathbb{P} \left\{ \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\}. \quad (5.12)$$

The rv's $U_i := \Delta_i(x_k)$ are centered and

$$|U_i| \leq \frac{1}{nh_n} \frac{1}{\bar{G}(\tau)} \left| K \left(\frac{x_k - X_i}{h_n} \right) - \mathbb{E} \left[K \left(\frac{x_k - X_i}{h_n} \right) \right] \right| \leq \frac{2C}{nh_n} \bar{K} := d_n.$$

where $\bar{K} = \sup_{x \in \Omega} |K(x)|$ and C is a positive constant.

Then

$$\text{Var}(U_i) = \mathbb{E} \left[\left(\frac{1}{nh_n} \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}(Y_i)} K \left(\frac{x_k - X_i}{h_n} \right) \right)^2 \right] - \left(\mathbb{E} \left[\frac{1}{nh_n} \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}(Y_i)} K \left(\frac{x_k - X_i}{h_n} \right) \right] \right)^2 =: \mathcal{I}_1 + \mathcal{I}_2.$$

The second term has been calculated. Whereas the first one can be developed as follows

$$\mathcal{I}_1 = \mathbb{E} \left[\frac{1}{n^2 h_n^2} \frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}^2(Y_i)} K^2 \left(\frac{x_k - X_i}{h_n} \right) \right] = \mathbb{E} \left[\frac{1}{n^2 h_n^2} K^2 \left(\frac{x_k - X_i}{h_n} \right) \mathbb{E} \left[\frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}^2(Y_i)} \mid X_i \right] \right].$$

For all $u \in \Omega$, using the properties of conditional expectation and change of variables, we have

$$\mathbb{E} \left[\frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}^2(Y_i)} \mid X_i = u \right] = \mathbb{E} \left[\mathbb{E} \left[\frac{\delta_i \mathbb{1}_{\{Y_i \leq t\}}}{\bar{G}^2(Y_i)} \mid T_i \right] \mid X_i = u \right] = \mathbb{E} \left[\frac{\mathbb{1}_{\{T_i \leq t\}}}{\bar{G}(T_i)} \mid X_i = u \right].$$

Then

$$\begin{aligned}
\mathcal{I}_1 &= \int \frac{1}{n^2 h_n^2} K^2 \left(\frac{x_k - u}{h_n} \right) \mathbb{E} \left[\frac{\mathbb{1}_{\{T_i \leq t\}}}{\overline{G}(T_i)} \mid X_i = u \right] f(u) du \\
&\leq \frac{1}{n^2 h_n^2} \frac{1}{\overline{G}(\tau)} \int K^2 \left(\frac{x_k - u}{h_n} \right) \mathbb{E} [\mathbb{1}_{\{T_i \leq t\}} \mid X_i = u] f(u) du \\
&\leq \frac{1}{n^2 h_n^2} \frac{1}{\overline{G}(\tau)} \int K^2 \left(\frac{x_k - u}{h_n} \right) F(t \mid u) f(u) du \\
&\leq \frac{1}{n^2 h_n^2} \frac{1}{\overline{G}(\tau)} \int K^2 \left(\frac{x_k - u}{h_n} \right) F_1(u, t) du \\
&\leq \frac{1}{n^2 h_n} \frac{1}{\overline{G}(\tau)} \int K^2(v) F_1(x_k - v h_n, t) dv.
\end{aligned}$$

Assumption **A6** yields that

$$\mathcal{I}_1 = O \left(\frac{1}{n^2 h_n} \right).$$

Then

$$Var(U_i) = O \left(\frac{1}{n^2 h_n^2} \right) + O \left(\frac{1}{n^2 h_n} \right) = O \left(\frac{1}{n^2 h_n^2} \right).$$

Now we apply Lemma 4.4, for U_i with $\epsilon = \epsilon_n$ and $q = q_n$.

$$\mathbb{P} \left(\frac{1}{n} \left| \sum_{i=1}^n U_i \right| > \epsilon \right) \leq 4 \exp \left(-\frac{\epsilon^2 q}{8} \left[\frac{8q^2 \sigma^2(q)}{n^2} + \frac{d\epsilon}{2} \right]^{-1} \right) + 22 \left(1 + \frac{4d}{\epsilon} \right)^{\frac{1}{2}} q \alpha \left(\left\lceil \frac{n}{2q} \right\rceil \right) =: A_1 + A_2.$$

Now taking $q_n = K_2 n^\kappa$ for some $\frac{1}{2} < \kappa < \frac{\nu-1}{\nu+1}$; we have for n large enough $\frac{q_n^2 \sigma^2(q_n)}{n^2}$ is of order $n^{\kappa-3} h_n^2$ and therefore asymptotically negligible than $\frac{d\epsilon_n}{2}$ (which is of order $n^{-1} h_n^{-1}$).

Then for n large enough

$$A_1 \leq 4 \exp \left[\frac{-C \epsilon_n^2 n^{1+\kappa} h_n}{8} \right].$$

The second term A_2 can be bounded by

$$A_2 \leq C \left(1 + \frac{2C\overline{K}}{n h_n \epsilon_n} \right)^{\frac{1}{2}} n^{-\nu(1-\kappa)+\kappa}$$

where C is a universal constant.

Then (5.12) becomes

$$\mathbb{P} \left\{ \max_{k=1, \dots, s_n} \left| \sum_{i=1}^n \Delta_i(x_k) \right| > \epsilon \right\} \leq M h_n^{-\beta} (A_1 + A_2) =: \mathcal{J}_1 + \mathcal{J}_2 \quad (5.13)$$

where

$$\mathcal{J}_1 = 4Mh_n^{-\beta} n^{4\exp\left[\frac{-C\epsilon_n^2 n^{1+\kappa} h_n}{8}\right]}$$

and

$$\begin{aligned} \mathcal{J}_2 &= CMh_n^{-\beta} \left(1 + \frac{2C\overline{K}}{nh_n\epsilon_n}\right)^{\frac{1}{2}} n^{-\nu(1-\kappa)+\kappa} \\ &= Ch_n^{-\beta} n^{-\nu(1-\kappa)+\kappa} + Ch_n^{-\beta-1} n^{-\nu(1-\kappa)+\kappa-1} + O\left(h_n^{-\beta-1} n^{-\nu(1-\kappa)+\kappa-1}\right) \\ &= o\left(h_n^{-\beta} n^{-\nu(1-\kappa)+\kappa}\right). \end{aligned}$$

Assumptions **A1**, **A2** and the choice of κ , yield that $\mathcal{J}_1$ and $\mathcal{J}_2$ are the general terms of convergent series; then by Borel-Cantelli's Lemma, the first term of (5.13) goes to zero almost surely. Otherwise, if we replace ϵ_n by $\epsilon_0 \sqrt{\frac{\log \log n}{nh_n^2}}$ for some $\epsilon_0 > 0$ in all steps of the Lemma, we have

$$\sqrt{\frac{nh_n^2}{\log \log n}} \sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \tilde{F}_{1,n}(x, t) - \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] \right| = o(1) \text{ a.s. } n \rightarrow \infty,$$

which completes the proof of Lemma 5.5. ■

Proof of Proposition 5.1 We use the following classical decomposition

$$\begin{aligned} \sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)| &\leq \frac{1}{\inf \ell_n(x)} \left\{ \sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_{1,n}(x, t) - \tilde{F}_{1,n}(x, t)| \right. \\ &\quad + \sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \tilde{F}_{1,n}(x, t) - \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] \right| \\ &\quad + \sup_{x \in \Omega} \sup_{0 < t \leq \tau} \left| \mathbb{E} \left[\tilde{F}_{1,n}(x, t) \right] - F_1(x, t) \right| \\ &\quad \left. + \sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_1(x, t)| \gamma_0^{-1} \sup_{x \in \Omega} |\ell(x) - \ell_n(x)| \right\}. \end{aligned}$$

Assumptions **A1**, **A2**, **A3** and **A5** yield that

$$\sup_{x \in \Omega} |\ell(x) - \ell_n(x)| = O \left(\max \left(\sqrt{\frac{\log \log n}{nh_n^2}}, h_n \right) \right)$$

(see Cai and Roussas (1992), Theorem 4.1).

In conjunction with Lemmas 5.1, 5.2, 5.3 and 5.5 we conclude the proof of the proposition 4.1. ■

Proof of Theorem 5.1 : We have

$$\begin{aligned}
|F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x)| &\leq |F_n(\xi_{p,n}(x)|x) - F(\xi_{p,n}(x)|x)| \\
&+ |F_n(\xi_{p,n}(x)|x) - F(\xi_p(x)|x)| \\
&\leq \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)| + \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)| \\
&\leq 2 \sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)|.
\end{aligned} \tag{5.14}$$

The proposition concludes the first part of the theorem. For the second part, a Taylor expansion of $F(\cdot|x)$ in a neighborhood of ξ_p , implies that

$$F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x) = (\xi_{p,n}(x) - \xi_p(x)) f(\xi_p^*(x)|x),$$

where ξ_p^* is between ξ_p and $\xi_{p,n}$ and $f(\cdot|x)$ is the conditional density of T given $X = x$.

From the behavior of $F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x)$ as n goes to infinity, it is easy to obtain asymptotic results for the sequence $(\xi_{p,n}(x) - \xi_p(x))$ and prove why the rates of convergence are the same for $\xi_{p,n}(x)$ and $F(\xi_{p,n}(x)|x)$. If $f(\xi_p^*(x)|x) \neq 0$ was not satisfied, we should increased the order of Taylor expansion. Then, by (5.44), we have

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| |f(\xi_p^*(x)|x)| \leq 2 \sup_{x \in \Omega} \sup_{0 < t \leq \tau} |F_n(t|x) - F(t|x)|.$$

The result follows from **A7** and the proposition. ■

5.3 Prevision via le quantile conditionnel dans le modèle censuré

L'estimateur du quantile conditionnel obtenu à partir de (5.10) ne pouvant pas être asymptotiquement normale. S'inspirant des travaux de Berlinet *et al* (2001) et Carbonez *et al* (1995), nous construisons un nouvel estimateur de (5.1) et par conséquent de (5.2) dans le modèle de censure dans le cas où les $(T_i, C_i, X_i); i \geq 1$ sont i.i.d.

Les résultats présentés dans cette partie sont soumis à publication. Nous présentons ici, dans la version soumise, les parties de cet article où sont donnés les résultats principaux et leurs preuves.

5.3.1 Introduction

It is well known from the robustness literature that the mean is sensible to outliers (see Hampel *et al.* (1986)); it may be sensible to use the median, which is a particular case of the quantile, rather than the mean to forecast future since the median is highly resistant against outliers. The nonparametric estimation of conditional quantile had received a great interest since 1969, when Roussas (1969) showed the convergence

and asymptotic normality of kernel estimates under Markov assumptions. For independent and identically distributed (iid) random variables (rv), Stone (1977) showed the weak consistency of kernel estimates. The uniform consistency was studied by Schlee (1982) for strong mixing case. Samanta (1989) proved the asymptotic normality in the iid case. Many other authors considered this problem; without pretending to the exhaustiveness, we quote Battacharya and Gangopadhyay (1990), Jones and Hall (1990), Mehra *et al.* (1991), Chandhuri (1991), Fan *et al.* (1994), Welsh (1996) and Xiang (1996). Honda (2000) dealt with the α -mixing case and proved the uniform convergence and asymptotic normality of an estimate of the conditional quantile using polynomial fitting method. Berlinet *et al.* (2001) showed the asymptotic normality of convergent estimates of conditional quantile by considering the particular case of stationary α -mixing process. Gannoun *et al.* (2003) gave a smooth nonparametric conditional median predictor, based on double kernel methods and established its asymptotic normality and proposed an extension to the conditional quantile.

In censoring case, Beran (1981) introduced a nonparametric estimate of the conditional survival function and proved some consistency results which were later exposed and extended by Dabrowska (1987, 1989) in the iid case and Lecoutre and Ould-Saïd (1995) studied the consistency in the strong mixing case. Dabrowska (1992) established a Bahadur representation of kernel quantile estimator and Xiang (1995) obtained the deficiency of sample quantile estimator with respect to a kernel estimator using coverage probability. Leconte *et al.* (2002) built two classes of estimators of the conditional distribution function and the quantile function and showed under some conditions that the two classes are equal. Some simulation have been driven to show that one is better than other in the sens of the mean square error. Further results, including bootstrap approximations have been gotten by Van Keilegom and Veraverbeke (1997, 1998). Recently, Gannoun *et al.* (2005) studied the asymptotic properties of an estimator of the conditional quantile using polynomial. Other large samples properties of the conditional distribution have been studied extensively in the literature (see e.g. González Manteiga and Cadarso Suárez (1994), Stute (1996) and Van Keilegom and Veraverbeke (1996, 2001)). Here we provide consistent and asymptotically normal conditional quantile estimate under the condition C and (T, X) are independent as in the recent paper of Ould-Saïd (2006) (see also Carbonnez *et al.* (1995), Kohler *et al.* (2002)), who established a strong uniform convergence rate of a kernel conditional quantile estimator under censorship model.

5.3.2 Definition of the new estimator

Throughout this paper we assume that

$$\tau_F < \tau_G, \quad (5.15)$$

$$C \text{ and } (X, T) \text{ are independent.} \quad (5.16)$$

Remark 5.1. In view of (5.15) many authors used the milder condition $\tau_F \leq \tau_G$. As we need to prove some uniform results which imply a sufficient rate of convergence of G_n (see Lemma 2) we have to consider a set of values of Y_i which do not include τ_G (a uniform rate for G_n is given in Deheuvels and Einmahl (2000) for $t < \tau_G$).

The condition (5.16) is plausible whenever the censoring is independent of the modality of the patients. This condition is slightly strong that the one usually used that is : T and C are independent conditionally given X (see e.g. Dabrowska (1987, 1989, 1992)). However the condition (5.16) is very useful to give an unbiased estimator of $F_1(t, x)$ (when the weights are uniform) who intervenes in our methodology of construction of the estimate of $F(t|x)$ (see below).

It is clear that an estimator of (5.2) will be obtained by estimating the conditional distribution function (5.1), thus it suffices to estimate $F_1(x, t)$ and $\ell(x)$. The density $\ell(x)$ is not affected by the censoring and therefore can be estimated consistently by the well-known kernel estimator. Furthermore, an unbiased estimator of $F_1(x, t) =$

$\mathbb{E}(\mathbb{1}_{\{T \leq t, X \leq x\}})$ is given by an average mean $\hat{F}_{1,n}(x, t) = \frac{1}{n} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} \mathbb{1}_{\{Y_i \leq t, X_i \leq x\}}$.

Indeed, using the properties of the conditional expectation and (5.16), we get

$$\begin{aligned} \mathbb{E}[\hat{F}_{1,n}(x, t)] &= \mathbb{E}\left[\frac{\delta_1}{\bar{G}(Y_1)} \mathbb{1}_{\{Y_1 \leq t, X_1 \leq x\}}\right] \\ &= \mathbb{E}\left[\mathbb{E}\left[\frac{\delta_1}{\bar{G}(Y_1)} \mathbb{1}_{\{Y_1 \leq t, X_1 \leq x\}} \middle| X_1, T_1\right]\right] \\ &= \mathbb{E}\left[\frac{\mathbb{1}_{\{T_1 \leq t, X_1 \leq x\}}}{\bar{G}(T_1)} \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq C_1\}} \middle| X_1, T_1\right]\right] \\ &= \mathbb{E}\left[\mathbb{1}_{\{T_1 \leq t, X_1 \leq x\}}\right] = F_1(x, t). \end{aligned}$$

Now, instead to use the uniform weight $\frac{1}{n}$ for all $\{Y_i, \delta_i, X_i\}$ we use the Nadaraya-Watson weight

$$\begin{aligned} W_{i,n}(x) &= \frac{K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \\ &= \frac{\frac{1}{nh_n} K\left(\frac{x-X_i}{h_n}\right)}{\ell_n(x)}, \end{aligned} \tag{5.17}$$

where K is a probability density function (so called kernel function), h_n is a sequence of positive real numbers which goes to zero as n goes to infinity (so called bandwidth) and $\ell_n(\cdot)$ is the well known kernel estimator of $\ell(\cdot)$ and instead to the step function $\mathbb{1}_{\{Y_i \leq t\}}$, we use a smooth df H .

Constructing an appropriate estimator is then obtained by adapting the weights

(5.17), in order to put more emphasis on large values of the interest variable T which are more censored than small values.

Ould Saïd (2006) considered the following weights $W_{i,n}(x) = \frac{1}{nh_n} K\left(\frac{x - X_i}{h_n}\right) \frac{\delta_i}{\overline{G}(Y_i)\ell_n(x)}$, and he established a strong uniform convergence of the corresponding kernel conditional quantile estimator; but this estimator is not derivable then we can not establish its asymptotical normality.

Let us define a smooth estimate of (5.1) by

$$\tilde{F}_n(t|x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\overline{G}(Y_i)} K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{t - Y_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)}. \quad (5.18)$$

Recall that (5.18) can be rewritten as :

$$\tilde{F}_n(t|x) = \frac{\tilde{F}_{1,n}(x, t)}{\ell_n(x)} \quad (5.19)$$

where

$$\tilde{F}_{1,n}(x, t) = \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i}{\overline{G}(Y_i)} K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{t - Y_i}{h_n}\right),$$

and

$$\ell_n(x) = \frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right).$$

In practice G is usually unknown, hence it is impossible to use the estimator (5.18). Then, we replace G by the Kaplan-Meier estimate G_n given by

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

Therefore the feasible estimator of the conditional df $F(\cdot|.)$ is given by

$$F_n(t|x) = \frac{\sum_{i=1}^n \frac{\delta_i}{\overline{G}_n(Y_i)} K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{t - Y_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right)} =: \frac{F_{1,n}(x, t)}{\ell_n(x)} \quad (5.20)$$

where

$$F_{1,n}(x, t) = \frac{1}{nh_n} \sum_{i=1}^n \frac{\delta_i}{\overline{G}_n(Y_i)} K\left(\frac{x - X_i}{h_n}\right) H\left(\frac{t - Y_i}{h_n}\right).$$

Then a natural estimator of $\xi_p(\cdot)$, is given by

$$\xi_{p,n}(x) = \inf \{t; F_n(t|x) \geq p\} \quad (5.21)$$

which satisfy

$$F_n(\xi_{p,n}(x)|x) = p. \quad (5.22)$$

We define the first partial derivative with respect to second component of $F_{1,n}(x, t)$ and $\tilde{F}_{1,n}(x, t)$ respectively by

$$\frac{\partial F_{1,n}(x, t)}{\partial t} = F'_{1,n}(x, t) = \frac{1}{nh_n^2} \sum_{i=1}^n \frac{\delta_i}{\bar{G}_n(Y_i)} K\left(\frac{x - X_i}{h}\right) H'\left(\frac{t - Y_i}{h}\right),$$

and

$$\frac{\partial \tilde{F}_{1,n}(x, t)}{\partial t} = \tilde{F}'_{1,n}(x, t) = \frac{1}{nh_n^2} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K\left(\frac{x - X_i}{h}\right) H'\left(\frac{t - Y_i}{h}\right)$$

where H' is the derivative of H .

The conditional densities estimators are given by

$$f_n(t|x) = \frac{F'_{1,n}(x, t)}{\ell_n(x)}$$

and

$$\tilde{f}_n(t|x) = \frac{\tilde{F}'_{1,n}(x, t)}{\ell_n(x)}.$$

In order to study the behavior of the random variable $(\xi_p(x) - \xi_{p,n}(x))$, we make use of the properties in (5.3) and (5.22) and Taylor expansion, one gets

$$F(\xi_p(x)|x) - F(\xi_{p,n}(x)|x) = (\xi_p(x) - \xi_{p,n}(x)) f(\xi_{p,n}^*(x)|x) \quad (5.23)$$

where $\xi_{p,n}^*(x)$ lies between $\xi_p(x)$ and $\xi_{p,n}(x)$. It is clear that equation (5.23) shows that from the behavior of $(F(\xi_p(x)|x) - F(\xi_{p,n}(x)|x))$, it is easy to obtain asymptotic results for the sequence $(\xi_p(x) - \xi_{p,n}(x))$. If $f(\xi_{p,n}^*(x)|x) \neq 0$ was not satisfied, we should have increase the order of Taylor expansion.

5.3.3 Assumptions and main results

Our assumptions are gathered together for easy references.

Let $\Omega_0 = \{x \in \mathbb{R} / \ell(x) > 0\}$, Ω be a compact set such that $\Omega \subset \Omega_0$ and $\mathcal{C}$ be a compact set included in $] -\infty, \tau_F]$.

A1 The kernel K satisfy :

(i) K is strictly positive valued, bounded, with compact support Ω and there exists constant M^* , and m^* such that $0 < M^* < \infty$, $0 < m^* < \infty$, $\sup_{\Omega} K(u) = M^*$, and $\inf_{\Omega} K(u) = m^*$,

(ii) It is Hölderian of order γ for some $\gamma > 0$,

(iii) $\int_{\mathbb{R}} uK(u)du = 0$, $\int_{\mathbb{R}} |u|K(u)du < +\infty$ and $\int_{\mathbb{R}} u^2K(u)du < \infty$,

$$(iv) \quad \int_{\mathbb{R}} K^2(u) du = \kappa < +\infty.$$

A2 The bandwidth h_n satisfy :

$$(i) \quad \sqrt{\frac{\log \log n}{n}} = o(h_n^2),$$

$$(ii) \quad nh_n^5 \longrightarrow 0.$$

A3 The conditional distribution function $F(t|x)$ has positive first derivative with respect to t , for all $x \in \Omega$, denoted $f(t|x)$ and satisfy

(i) The Lipschitz condition of order 1 with respect to t and x , such that

$$\forall(t_1, t_2) \in \mathbb{R}^2, \quad \forall(x_1, x_2) \in \Omega^2, \quad |F(t_1|x_1) - F(t_2|x_2)| \leq C(|x_1 - x_2| + |t_1 - t_2|),$$

(ii) $\int_{\mathbb{R}} |t|f(t|x)dt < +\infty$, for all $x \in \Omega$ (that is the conditional density $f(\cdot|x)$ has a finite first moment),

(iii) There exist a constant $\gamma_1 > 0$ such that $f(t|x) > \gamma_1$ for all $x \in \Omega$ and $t \in \mathbb{R}$.

A4 The distribution function H has a first derivative H' which is positive and bounded such that

(i) There exists two constants $0 < M < \infty$, and $0 < m < \infty$, $\sup_{\mathbb{R}} H'(t) = M$ and $\inf_{\mathbb{R}} H'(t) = m$.

$$(ii) \quad \int_{\mathbb{R}} H'(t)dt = 1 \quad \text{and} \quad \int_{\mathbb{R}} |t|H'(t)dt < +\infty.$$

A5 The marginal density $\ell(\cdot)$ satisfies the Lipschitz condition and there exists $\gamma_0 > 0$ such that $\ell(x) > \gamma_0$ for all $x \in \Omega$.

A6 The distribution function of the censored rv, G has bounded first derivative g .

Comments on the Assumptions.

Assumptions A1 (i), (ii) and (iii) are quite usual in kernel estimation. Condition A1 (iv) intervenes in the variance terms of (5.20) and (5.21). Assumption A2 (i) is needed in the study of the behavior of the differences between $F_{1,n}$ and $\tilde{F}_{1,n}$ and between $f_n(\cdot|x)$ and $\tilde{f}_n(\cdot|x)$, and in the proof of the convergence to zero of the bias term of $\tilde{F}_n(\cdot, \cdot)$ in Lemma 5.8 Assumption A3 (i) is put for technical convenience. The existence of the upper and the lower bounds of H' , A4 (i) will be used in the proof of the asymptotic normality. Assumption A5 intervenes in the convergence of the kernel density estimator ℓ_n . A3 (ii) and A6 are additional assumptions to get the asymptotic variance term.

Our first result deals with the uniform almost sure convergence with rate of the conditional d.f. estimator (5.20) and is stated in Proposition 5.2. The uniform almost sure convergence of the conditional quantile estimator (5.21) and its rate will be derived in Theorem 5.2. Next, in Proposition 5.3, we state the asymptotic normality of the estimator (5.20) suitably normalized, then it suffices to prove the asymptotic normality of $(F_n(\xi_{n,p}(x)|x) - F_n(\xi_p(x)|x))$ and the convergence in probability of the sequence $f_n(\xi_{p,n}^*(x)|x)$ to obtain the asymptotic normality of the estimator of the conditional quantile ξ_p , by Slutsky's theorem.

Proposition 5.2. *Under assumption A1 (i) – (iii), A2 (i), A3 (i), A4 and A5,*

and for n large enough, we have

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_n(t|x) - F(t|x)| = O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n}}, h_n \right\} \right) \quad a.s. \quad (5.24)$$

Theorem 5.2. Under assumptions **A1** (i) – (iii), **A2** (i), **A3** (i), (iii), **A4** and **A5**, if the conditional density satisfies $\inf_{x \in \Omega} f(\xi_p(x)|x) > 0$ then for n large enough, we have

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n}}, h_n \right\} \right) \quad a.s. \quad (5.25)$$

Remark 5.3. The uniform positiveness assumption on the conditional density (in Theorem 1) implies the uniform unicity of the conditional quantile that is

$$\forall \varepsilon > 0, \exists \beta > 0, \forall \eta_p : \Omega \rightarrow \mathbb{R}, \quad \sup_{x \in \Omega} |\xi_p(x) - \eta_p(x)| \geq \varepsilon \Rightarrow \sup_{x \in \Omega} |F(\xi_p(x)|x) - F(\eta_p(x)|x)| \geq \beta. \quad (5.26)$$

On the other hand, assuming the sole (5.26) guarantees the consistency of the conditional quantile but does not give a rate for it.

Remark 5.4. If we choose $h_n = O \left(\frac{\log \log n}{nh_n} \right)^{1/5}$ which is the optimal bandwidth in density estimation, then for each fixed $p \in (0, 1)$ and for n large enough, we have

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \left(\frac{\log \log n}{nh_n} \right)^{1/5}.$$

Under the following mild modifications upon assumptions :

A2 (i) $\sqrt{\frac{\log \log n}{n}} = o(h_n^3),$

A3 (i) the df $F(\cdot|x)$ satisfies the Lipschitz condition of order 2 with respect to t and x ,

A4 (ii) $\int_{\mathbb{R}} |t|^2 H'(t) dt < +\infty,$

we get the following rate $O \left(\max \left\{ \sqrt{\frac{\log \log n}{nh_n}}, h_n^2 \right\} \right) \quad a.s.$ Then if we choose $h_n =$

$O \left(\frac{\log \log n}{nh_n} \right)^{1/5}$, we have $\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| = O \left(\frac{\log \log n}{nh_n} \right)^{2/5}$, which is the rate expected By Mehra et al. (1991) and the same rate obtained by Xiang (1996) and Ould-Saïd (2006). Here we point out that in Gannoun et al. (2005), there is neither uniform result nor speed of convergence.

Proposition 5.3. Under the assumptions **A1-A6** and for any $x \in \Omega_0$ such that $\ell(x) > 0$, we have

$$(nh_n)^{1/2} (F_n(t|x) - F(t|x)) \xrightarrow{\mathcal{D}} \mathcal{N} (0, (\sigma(t, x))^2) \quad \text{as} \quad n \rightarrow \infty \quad (5.27)$$

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

$$(\sigma(t, x))^2 = \frac{\kappa F(t|x) (1 - \overline{G}(t)F(t|x))}{\ell(x)\overline{G}(t)}.$$

Theorem 5.3. *Let $p \in (0, 1)$ such that $p = F(\xi_p(x)|x)$. Under the assumptions **A1-A6**, for any $x \in \Omega_0$ such that $\Sigma(x, \xi_p(x)) \neq 0$, we have*

$$\left(\frac{nh_n}{(\Sigma(x, \xi_p(x)))^2} \right)^{1/2} (\xi_p(x) - \xi_{p,n}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty \quad (5.28)$$

where

$$(\Sigma(x, \xi_p(x)))^2 = \frac{\kappa p (1 - p \overline{G}(\xi_p(x)))}{f^2(\xi_p(x)|x) \ell(x) \overline{G}(\xi_p(x))}.$$

5.3.4 Applications

5.3.4.1 Applications to prediction.

It is well known, from the robustness theory that the median is more robust than the mean, therefore the conditional median, $\mu(x) = \xi_{1/2}(x)$, is a good alternative to the conditional mean as a predictor for a variable Y given $X = x$. Note that the estimation of $\mu(x)$ is given by $\mu_n(x) = \xi_{1/2,n}(x)$. Using this considerations and section 3, we want to predict the non observed r.v. Y_{n+1} (which corresponds to some modality of our problem), from available data $X_1, \dots, X_n$. Given a new value X_{n+1} , we can predict the corresponding response Y_{n+1} by

$$Y_{n+1} = \mu_n(X_{n+1}) = \xi_{1/2,n}(X_{n+1}).$$

Applying the above theorem, we have the following corollary

Corollary 5.1.

$$\left(\frac{nh_n}{(\Sigma(X_{n+1}, \xi_{1/2}(X_{n+1})))^2} \right)^{1/2} (\xi_{1/2,n}(X_{n+1}) - \xi_{1/2}(X_{n+1})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1), \quad \text{as } n \rightarrow \infty.$$

5.3.4.2 Confidence intervals.

Using a plug-in method by replacing $\ell, \overline{G}, f$ and $\xi_{1/2}$ by their estimates $\ell_n, \overline{G}_n, f_n$ and $\xi_{1/2,n}$ respectively, we get from corollary 1,

Corollary 5.2.

$$\left(\frac{nh_n}{(\Sigma_n(x, \xi_{1/2,n}(x)))^2} \right)^{1/2} (\xi_{1/2,n}(x) - \xi_{1/2}(x)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1) \quad \text{as } n \rightarrow \infty. \quad (5.29)$$

From this corollary, we get for each fixed $\eta \in (0, 1)$, the following approximate $(1 - \eta)\%$ confidence interval

$$\xi_{1/2,n}(x) \pm t_{1-\eta/2} \times \frac{\Sigma_n(x, \xi_{1/2,n}(x))}{\sqrt{nh_n}}$$

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

Remark 5.5. *In the proof of Proposition 5.6 (and therefore Theorem 5.6), we use Hoeffding's inequality which is a pointwise exponential inequality and we use a standard idea about covering a compact set by a finite number of intervals to get the uniformity. Another interesting idea is to use Vapnik-Cervonenkis theory by using Pollard's inequality (see Devroye et al (1996) theorem 29.1) together with bounds on the covering number, which gives the uniformity straightforwardly. However, the choice of the kernels in the last case must satisfies K1 condition of Giné and Guillou (2002) which is more restrictive than the Hölderian functions. In this case, one finds a rate of convergence of the order $\left(\frac{\log n}{nh_n^2}\right)^{1/2}$, which is weaker than our result.*

5.3.5 Simulations Studies

We have conducted a numerical study to examine the performance of our estimator. The first subsection deals with the consistency of the conditional quantile estimator $\xi_{p,n}(x)$ given in (5.21). Whereas the second looks at how good is the asymptotic normality when we deal with a finite sample.

5.3.5.1 Consistency

The aim of the following simulations is to examine the performance of our estimator $\xi_{p,n}(x)$ in some particular model. In our simulation, we consider the following model : $T_i = X_i + \sigma\epsilon_i$, $i = 1, \dots, n$ where X_i and ϵ_i are two independent i.i.d. sequences distributed as $N(0, 1)$ and σ is a positive constant. The censoring times $\{C_i; i = 1, \dots, n\}$ are generated independently from $N(0, 1)$. Then we compute our estimator with the observed data (X_i, Y_i, δ_i) , where $Y_i = T_i \wedge C_i$ and $\delta_i = \mathbb{I}_{\{T_i \leq C_i\}}$.

We choose a gaussian kernel and it is well know that, in nonparametric estimation, optimality (in the MSE sense) is not seriously swayed by the choice of the kernel K but is affected by the choice of the bandwidth h_n . The bandwidth h_n is chosen according to the assumption **A2**, that is $h_n = C \left(\frac{\log \log n}{n}\right)^{\frac{1}{4}}$, with an appropriate choice of the constant C . Furthermore we choose H as the normal distribution function. We take several values of n and in each case, the conditional quantile estimator, along a grid of 160 equispaces points in $[-2, 2]$ has been calculated. We take several values of n and draw two curves corresponding to the conditional quantile estimator $\xi_{p,n}(x)$ for $x \in [-2, 2]$ and the theoretical conditional quantile $\xi_p(x)$ for $\sigma = 0.45$ and for the two values $p = 1/2$ and $p = 1/4$.

The curves show that our estimator perform well in particular when n increase.

5.3.5.2 Asymptotic normality

We now consider the problem of asymptotic normality. We show how good the normality is when dealing with samples of finite size which is the case in practice. The

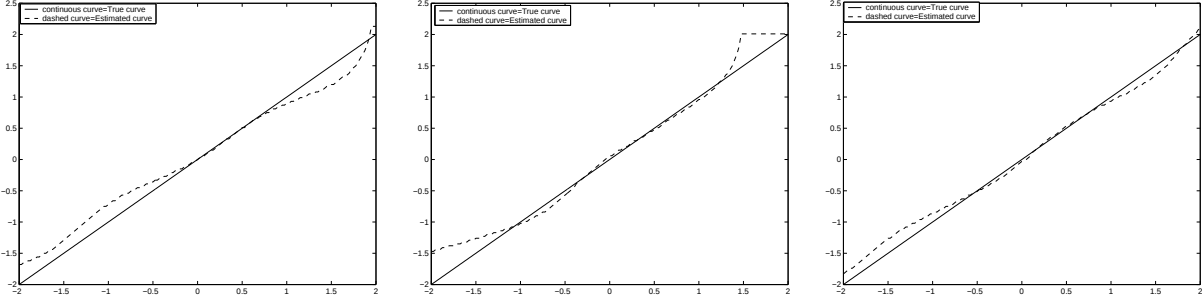


FIG. 5.2 – $p = 1/2$; $n = 50$, $n = 100$ and 500 respectively.

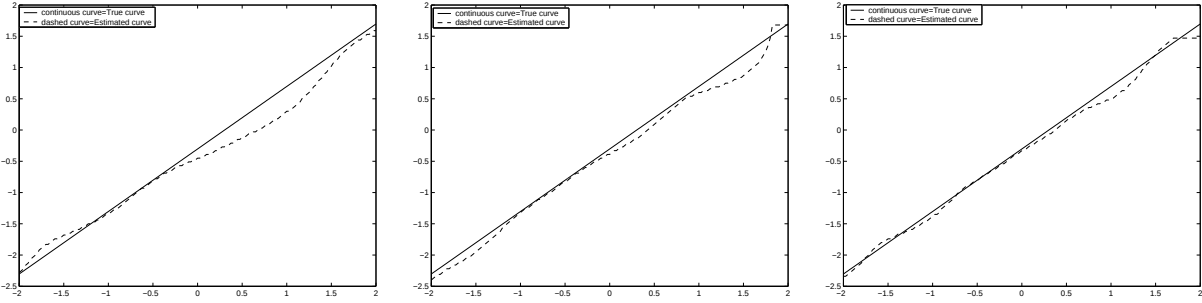


FIG. 5.3 – $p = 1/4$; $n = 50$, $n = 100$ and $n = 500$ respectively.

data arise from the same distribution as previously for a given size n , we estimate the conditional quantile function as before and calculate the normalized deviation between this estimate and the theoretical conditional quantile for $x = 0$ and $p = 1/2$;

$$\bar{\xi}_{p,n}(x) = \sqrt{nh_n} \left(\frac{\kappa p (1 - p \bar{G}_n(\xi_p(x)))}{f_n^2(\xi_p(x)|x) \ell_n(x) \bar{G}_n(\xi_p(x))} \right)^{-1/2} (\xi_{p,n}(x) - \xi_p(x)) \text{ which becomes}$$

$$\bar{\xi}_{p,n}(0) = \sqrt{nh_n} \left(\frac{\kappa \frac{1}{2} (1 - \frac{1}{2} \bar{G}_n(0))}{f_n^2(0|0) \ell_n(0) \bar{G}_n(0)} \right)^{-1/2} (\xi_{\frac{1}{2},n}(x) - 0).$$

We draw, using this scheme, B independent n -samples. The bandwidth h_n is chosen according to assumption **A2**. In order to estimate the density function of $\bar{\xi}_{p,n}(x)$ (by the kernel method), we make the classical bandwidth choice (see, e.g. Silverman (1986), p. 40) $h' = Cn^{-1/5}$, where the constant C is chosen appropriately.

We see that the normality of the r.v. $\bar{\xi}_{p,n}(\cdot)$ is better as n increase, which is clearly appear in the Q.Q. plot.

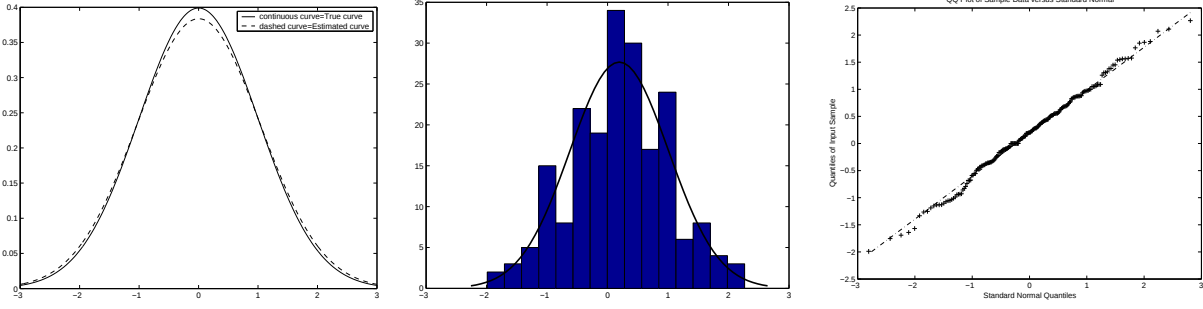


FIG. 5.4 – $n = 100$, $B = 200$.

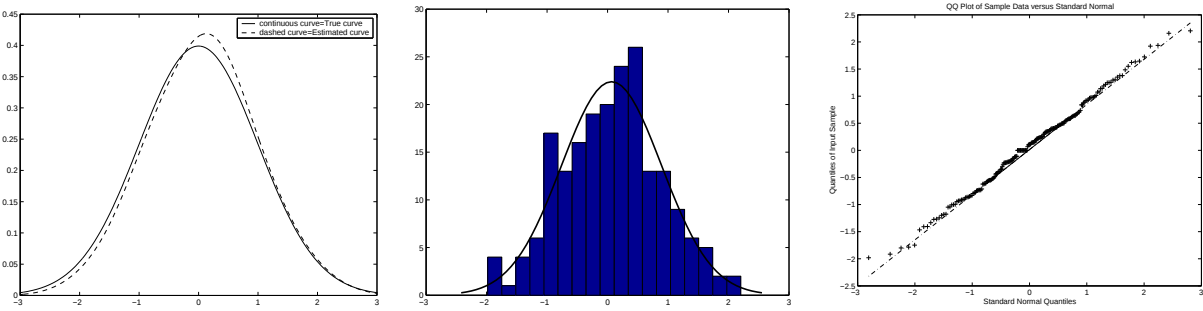


FIG. 5.5 – $n = 300$, $B = 200$.

5.3.5.3 Confidence curves

We construct approximate $(1 - \eta)\%$ confidence curves obtained from corollary 2, that is

$$\xi_{1/2,n}(x) \pm t_{1-\eta/2} \times \frac{\sigma_n(x, \xi_{1/2,n}(x))}{\sqrt{nh_n}}.$$

We draw in Figure 5, for several values of n , two curves corresponding to the approximate $(1 - \eta/2)\%$ lower and upper confidence curves for $p = 1/2$, $x \in [-2, 2]$, $\sigma = 0.45$ and $\eta = 5\%$. The bandwidth h_n is chosen as in paragraph 5.1.

It is clear, from Figure 5, that the confidence interval becomes more precise as the

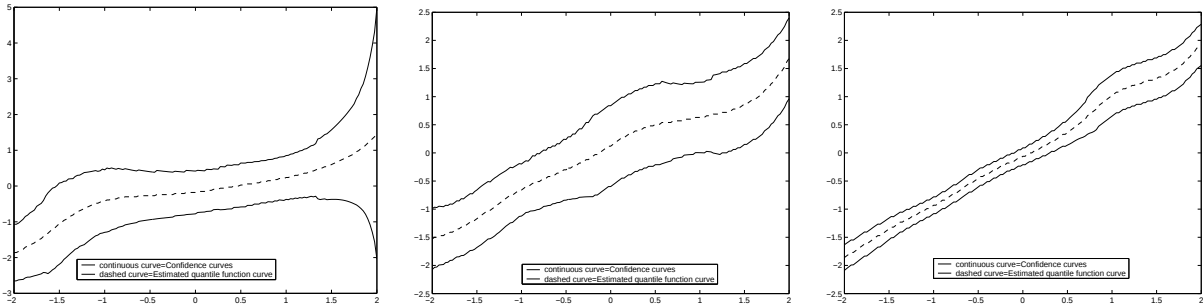


FIG. 5.6 – $p = 1/2$; $n = 50$, $n = 100$ and $n = 500$ respectively.

sample size increase.

5.3.6 Auxiliary results and proofs

The proof of our main results is split up into several Lemmas. The first Lemma deals with the behavior of the difference between $\mathbb{E} \left[\tilde{F}_{1,n}(t, x) \right]$ and $F_1(t, x)$.

Lemma 5.6. *Under assumptions **A1** (iii), **A3** (i), **A4** and **A5** then for n large enough, we have*

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \mathbb{E} \left[\tilde{F}_{1,n}(t, x) \right] - F_1(t, x) \right| = O(h_n) \quad a.s. \quad (5.30)$$

PROOF OF LEMMA 5.6

$$\begin{aligned} \mathbb{E} \left(\tilde{F}_{1,n}(t, x) \right) &= \frac{1}{h_n} \mathbb{E} \left[\frac{\delta_1}{\overline{G}(Y_1)} K \left(\frac{x - X_1}{h_n} \right) H \left(\frac{t - Y_1}{h_n} \right) \right] \\ &= \frac{1}{h_n} \mathbb{E} \left[K \left(\frac{x - X_1}{h_n} \right) \mathbb{E} \left[\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | X_1 \right] \right]. \end{aligned}$$

Moreover we have by integration by parts and changing variables,

$$\begin{aligned} \mathbb{E} \left[\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | X_1 \right] &= \mathbb{E} \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | T_1 \right) X_1 \right] \\ &= \mathbb{E} \left[H \left(\frac{t - T_1}{h_n} \right) | X_1 \right] \\ &= \int H \left(\frac{t - y}{h_n} \right) f(y | X_1) dy \\ &= \int H'(z) F(t - zh_n | X_1) dz \\ &= \int H'(z) [F(t - zh_n | X_1) - F(t | x)] dz + F(t | x), \end{aligned}$$

by the first part of the assumption **A4**.

Thus we have :

$$\begin{aligned} &\mathbb{E} \left[\tilde{F}_{1,n}(t, x) \right] \\ &= \frac{1}{h_n} \mathbb{E} \left[K \left(\frac{x - X_1}{h_n} \right) \int H'(z) [F(t - zh_n | X_1) - F(t | x)] dz \right] + \frac{F(t | x)}{h_n} \mathbb{E} \left[K \left(\frac{x - X_1}{h_n} \right) \right] \\ &=: \mathcal{I}_1 + \mathcal{I}_2. \end{aligned}$$

Making use the first part of the assumption **A5**, the second term $\mathcal{I}_2$ tends to $F_1(t, x)$ as n goes to infinity.

By the assumption **A3** (i) and **A4** (i)

$$\begin{aligned}\mathcal{I}_1 &< \int_{\mathbb{R}} H'(z) |F(t - zh_n | X_1) - F(t | x)| dz \\ &< C \int_{\mathbb{R}} H'(z) |h_n + zh_n| dz \\ &< Ch_n + Ch_n \int_{\mathbb{R}} |z| H'(z) dz.\end{aligned}\tag{5.31}$$

Making use **A4** (ii) and **A1** (iii), it clear that $\mathcal{I}_1 = O(h_n)$, this completes the proof of the Lemma 5.6. $\blacksquare$

Remark 5.2. *The last argument shows that*

$$\mathbb{E} \left[H \left(\frac{y - Y_1}{h} \right) | X_1 \right] - F(y|x) \longrightarrow 0, \quad \text{as } n \rightarrow \infty. \tag{5.32}$$

The second Lemma deals with the behavior of the difference between $F_{1,n}$ and $\tilde{F}_{1,n}$.

Lemma 5.7. *Under assumptions **A1** (i) and **A2** (i), then for n large enough, we have*

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_{1,n}(x, t) - \tilde{F}_{1,n}(t, x)| = O(h_n) \quad a.s. \tag{5.33}$$

PROOF OF LEMMA 5.7

$$\begin{aligned}|F_{1,n}(t, x) - \tilde{F}_{1,n}(t, x)| &\leq \frac{1}{nh_n} \sum_{i=1}^n \left| \delta_i K \left(\frac{x - X_i}{h_n} \right) H \left(\frac{t - Y_i}{h_n} \right) \right| \left| \frac{1}{\overline{G}_n(Y_i)} - \frac{1}{\overline{G}(Y_i)} \right| \\ &\leq \frac{1}{h_n} \frac{M^*}{\overline{G}_n(\tau_F) \overline{G}(\tau_F)} \sup_{t \in \mathcal{C}} |G_n(t) - G(t)| \frac{1}{n} \sum_{i=1}^n \delta_i.\end{aligned}$$

Since $\overline{G}(\tau_F) > 0$, in conjunction with the SLLN and the LIL on the censoring law (see formula 4.28 in Deheuvels and Einmahl (2000)), we have

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_{1,n}(x, t) - \tilde{F}_{1,n}(t, x)| \leq \frac{cM^*}{\overline{G}^2(\tau_F)} \frac{1}{h_n} \left(\frac{\log \log n}{n} \right)^{1/2} \quad a.s.$$

where c is a positive constant. Assumption **A2** (i) concludes the proof. $\blacksquare$

Lemma 5.8. *Under assumption **A1** (i) and (ii), **A2** (i), then we have*

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |\tilde{F}_{1,n}(t, x) - \mathbb{E} [\tilde{F}_{1,n}(t, x)]| \longrightarrow 0, \quad a.s. \quad \text{as } n \rightarrow \infty.$$

Furthermore, for all n large enough,

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |\tilde{F}_{1,n}(t, x) - \mathbb{E} [\tilde{F}_{1,n}(t, x)]| = O \left(\sqrt{\frac{\log \log n}{nh_n}} \right) \quad a.s. \tag{5.34}$$

PROOF OF LEMMA 5.8

As Ω and $\mathcal{C}$ are compact sets, then they can be covered by finite numbers s_n and d_n of intervals centered at x_k and t_j of length h_n^λ and h_n^μ respectively such $\lambda \geq \frac{\gamma+3}{\gamma}$ and $\mu \geq 4$. Since Ω and $\mathcal{C}$ are bounded, there exist two constant A_1 and A_2 such that $s_n \leq A_1 h_n^{-\lambda}$ and $d_n \leq A_2 h_n^{-\mu}$.

Now put $k(x) = \arg \min_{k=1,2,\dots,s_n} |x - x_k|$ and $j(t) = \arg \min_{j=1,2,\dots,d_n} |t - t_j|$. Thus we have the following decomposition

$$\begin{aligned}
& \left| \tilde{F}_{1,n}(t, x) - \mathbb{E} \left[\tilde{F}_{1,n}(t, x) \right] \right| \\
\leq & \underbrace{\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x) - \tilde{F}_{1,n}(t, x_{k(x)}) \right|}_{\mathcal{I}_1} + \underbrace{\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x) - \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right|}_{\mathcal{I}_2} \\
& + \underbrace{\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right|}_{\mathcal{I}_3} \\
& + \underbrace{\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] - \mathbb{E} \left[\tilde{F}_{1,n}(t, x_{k(x)}) \right] \right|}_{\mathcal{I}_4} \\
& + \underbrace{\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \mathbb{E} \left[\tilde{F}_{1,n}(t, x_{k(x)}) \right] - \mathbb{E} \left[\tilde{F}_{1,n}(t, x) \right] \right|}_{\mathcal{I}_5}.
\end{aligned}$$

Concerning $\mathcal{I}_1$ and $\mathcal{I}_5$. By **A1** *ii*) and the fact that is bounded we have

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x) - \tilde{F}_{1,n}(t, x_{k(x)}) \right| \leq \frac{c}{\overline{G}(\tau_F)} h_n^{\gamma\lambda - (\gamma+1)}.$$

Assumption **A2** *ii*) on h_n and the condition upon λ we get

$$\sqrt{\frac{nh_n}{\log \log n}} \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x) - \tilde{F}_{1,n}(t, x_{k(x)}) \right| = o(1). \quad (5.35)$$

Concerning $\mathcal{I}_1$ and $\mathcal{I}_5$. In the same way and by the fact K is bounded and **A4** (which implies that H is Lipschitzian) we have

$$\sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x_{k(x)}) - \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right| \leq \frac{c}{\overline{G}(\tau_F)} h_n^{\mu-2}.$$

The condition upon μ imply that

$$\sqrt{\frac{nh_n}{\log \log n}} \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t, x_{k(x)}) - \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right| = o(1). \quad (5.36)$$

Concerning $\mathcal{I}_3$, for all $\varepsilon > 0$ we have

$$\begin{aligned}
& \mathbb{P} \left\{ \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| > \varepsilon \right\} \\
&= \mathbb{P} \left\{ \max_{k=1,2,\dots,s_n} \max_{j=1,2,\dots,d_n} \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| > \varepsilon \right\} \\
&\leq s_n d_n \mathbb{P} \left\{ \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| > \varepsilon \right\}. \tag{5.37}
\end{aligned}$$

Now, we have

$$\mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right] = 0$$

and

$$\left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| \leq \frac{2M^*}{nh_n \overline{G}(\tau)}.$$

Hoeffding's inequality (cf. Shorack and Wellner, 1986, p. 855) yields

$$\mathbb{P} \left\{ \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| > \varepsilon \right\} \leq \exp \{ -\varepsilon^2 c n^2 h_n^2 \}$$

where c is a universal constant. Then, (5.37) becomes

$$\begin{aligned}
& \mathbb{P} \left\{ \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| > \varepsilon \right\} \\
&\leq A_1 A_2 h_n^{-(\lambda+\mu)} \exp \{ -c \varepsilon^2 n^2 h_n^2 \} \\
&= A_1 A_2 (nh_n)^{-(\lambda+\mu)} n^{\lambda+\mu} c \varepsilon^2 n^2 h_n^2 / \log n \tag{5.38}
\end{aligned}$$

where c is a positive constant depending only on M^* and $\overline{G}(\tau)$.

A2 (i) implies that $\frac{n^2 h_n^2}{\log n} \rightarrow +\infty$, which yields that the last term of (5.38) is the general term of a convergent series; then by Borel-Cantelli's Lemma, the first term of (5.37) goes to zero almost surely. Otherwise, if we replace ε by $\varepsilon_0 \sqrt{\frac{\log \log n}{nh_n}}$, for some $\varepsilon_0 > 0$ in all steps of the Lemma, we have

$$\sqrt{\frac{nh_n}{\log \log n}} \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} \left| \tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) - \mathbb{E} \left[\tilde{F}_{1,n}(t_{j(t)}, x_{k(x)}) \right] \right| = o(1) \text{ a.s.}$$

in conjunction with (5.35) and (5.36) we complete the proof of the Lemma. ■

Lemma 5.9. *Under assumptions **A1** (iii), **A2** (i) and **A4** (i) we have :*

$$\text{var} \left(\tilde{F}'_{1,n}(t, x) \right) \longrightarrow 0, \quad \text{as } n \rightarrow \infty.$$

PROOF OF LEMMA 5.9

Recall that

$$\tilde{F}'_{1,n}(x, y) = \frac{1}{nh_n^2} \sum_{i=1}^n \frac{\delta_i}{\bar{G}(Y_i)} K\left(\frac{x - X_i}{h}\right) H'\left(\frac{y - Y_i}{h}\right),$$

then

$$\begin{aligned} \text{var}\left(\tilde{F}'_{1,n}(t, x)\right) &= \frac{1}{nh_n^4} \text{var}\left[\frac{\delta_1}{\bar{G}(Y_1)} K\left(\frac{x - X_1}{h_n}\right) H'\left(\frac{t - Y_1}{h_n}\right)\right] \\ &= \frac{1}{nh_n^4} \mathbb{E}\left[\frac{\delta_1}{\bar{G}^2(Y_1)} K^2\left(\frac{x - X_1}{h_n}\right) H'^2\left(\frac{t - Y_1}{h_n}\right)\right] \\ &\quad - \frac{1}{nh_n^4} \left[\mathbb{E}\left[\frac{\delta_1}{\bar{G}(Y_1)} K\left(\frac{x - X_1}{h_n}\right) H'\left(\frac{t - Y_1}{h_n}\right)\right]\right]^2 \\ &= \frac{1}{nh_n^4} \mathbb{E}\left[K^2\left(\frac{x - X_1}{h_n}\right) \mathbb{E}\left(H'^2\left(\frac{t - Y_1}{h_n}\right) \frac{\delta_1}{\bar{G}(Y_1)} \middle| X_1\right)\right] \\ &\quad - \frac{1}{nh_n^4} \left[\mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right) \mathbb{E}\left(H'\left(\frac{t - Y_1}{h_n}\right) \frac{\delta_1}{\bar{G}(Y_1)} \middle| X_1\right)\right]\right]^2 \\ &=: J_1 - J_2. \end{aligned}$$

By the assumptions **A1** (iii) and **A4** (i), we have

$$\begin{aligned} J_1 &\leq \frac{1}{nh_n^4} \frac{M^2}{\bar{G}^2(\tau)} \mathbb{E}\left[K^2\left(\frac{x - X_1}{h_n}\right)\right] \\ &\leq \frac{1}{nh_n^3} \frac{M^2}{\bar{G}^2(\tau)} \kappa \ell(x) + O\left(\frac{1}{nh_n^3}\right). \end{aligned}$$

Then by **A2** (i), $J_1 \rightarrow 0$ as $n \rightarrow +\infty$.

For the same arguments as before, we have

$$\begin{aligned} J_2 &\leq \frac{1}{nh_n^4} \frac{M^2}{\bar{G}^2(\tau)} \left[\mathbb{E}\left[K\left(\frac{x - X_1}{h_n}\right)\right]\right]^2 \\ &= \frac{1}{nh_n^2} \frac{M^2}{\bar{G}^2(\tau)} \ell^2(x) + O\left(\frac{1}{nh_n^2}\right) \end{aligned}$$

and J_2 tends again to zero as n goes to infinity. This completes the proof of the Lemma 3. $\blacksquare$

Remark 5.3. Under the assumptions of Lemma 3, and making use of Tchebychev's inequality we have,

$$\tilde{F}'_{1,n}(t, x) \rightarrow F'_1(t, x), \text{ in probability, as } n \rightarrow \infty$$

and thus if we add assumption **A5** then

$$\tilde{f}_n(t|x) = \frac{\tilde{F}'_{1,n}(t, x)}{\ell_n(x)} \rightarrow f(t|x) \text{ in probability, as } n \rightarrow \infty. \quad (5.39)$$

The following Lemma deals with the convergence of the conditional probability density estimator $f_n(t|x)$ to $f(t|x)$

Lemma 5.10. *Under assumptions **A1** (iii), **A2** (i), **A4** and **A5**, we have*

$$f_n(t|x) \longrightarrow f(t|x) \text{ in probability, as } n \rightarrow \infty. \quad (5.40)$$

PROOF OF LEMMA 5.10

We have

$$f_n(t|x) - f(t|x) = \left(f_n(t|x) - \tilde{f}_n(t|x) \right) + \left(\tilde{f}_n(t|x) - f(t|x) \right).$$

$$\begin{aligned} \left| f_n(t|x) - \tilde{f}_n(t|x) \right| &= \frac{1}{\ell_n(x)} \left| \left(F'_{1,n}(t, x) - \tilde{F}'_{1,n}(t, x) \right) \right| \\ &= \frac{1}{\ell_n(x)} \frac{1}{nh_n^2} \left| \sum_{i=1}^n \delta_i K \left(\frac{x - X_i}{h_n} \right) H' \left(\frac{t - Y_i}{h_n} \right) \left(\frac{1}{\overline{G}_n(Y_i)} - \frac{1}{\overline{G}(Y_i)} \right) \right| \\ &\leq \frac{1}{\ell_n(x)} \frac{1}{h_n^2} \frac{M^* M}{\overline{G}_n(\tau) \overline{G}(\tau)} \sup_{t \leq \tau} |G_n(t) - G(t)| \frac{1}{n} \sum_{i=1}^n \delta_i. \end{aligned}$$

Since $\ell_n(x) \geq \inf_x \ell_n(x) \geq \gamma > 0$ and $\overline{G}(\tau) > 0$, in conjunction with the SLLN and the LIL on the censoring law, we have

$$\left| f_n(t|x) - \tilde{f}_n(t|x) \right| \leq \frac{cM^*M}{\overline{G}^2(\tau)} \frac{1}{h_n^2} \left(\frac{\log \log n}{n} \right)^{1/2} \text{ a.s.}$$

where c is a positive constant. Assumption **A2** (i) give us that

$$\left| f_n(t|x) - \tilde{f}_n(t|x) \right| \longrightarrow 0 \text{ a.s. as } n \rightarrow \infty.$$

The proof can be concluded by the remark 6 . ■

PROOF OF PROPOSITION 5.2

In view of (5.20), we have

$$\begin{aligned} \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_n(t|x) - F(t|x)| &\leq \frac{1}{\gamma - \sup_{x \in \Omega} |\ell_n(x) - \ell(x)|} \left\{ \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_{1,n}(x, t) - F_1(x, t)| \right. \\ &\quad \left. + \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F(t|x)| \sup_{x \in \Omega} |\ell_n(x) - \ell(x)| \right\}. \end{aligned} \quad (5.41)$$

The kernel estimator $\ell_n(x)$ is almost surely bounded away from 0 on Ω because of the second part of assumption **A5**.

Now, under assumptions **A5** and **A1** (i) – (iii), we have

$$\sup_{x \in \Omega} |\mathbb{E}(\ell_n(x)) - \ell(x)| \longrightarrow 0 \text{ as } n \rightarrow \infty.$$

Using Taylor's expansion up to the first order to $\ell(\cdot)$, for n large enough, we have

$$\sup_{x \in \Omega} |\mathbb{E}(\ell_n(x)) - \ell(x)| = O(h_n). \quad (5.42)$$

Furthermore under **A1** (i), (ii), (iii) and **A2** (i), an analogous proof as in the Lemma 3 (by taking $\Delta_i(x) = \frac{1}{nh_n} K\left(\frac{x - X_i}{h_n}\right) - \mathbb{E}\left[\frac{1}{nh_n} K\left(\frac{x - X_i}{h_n}\right)\right]$ and by replacing ϵ by $\epsilon_0 \left(\frac{\log \log n}{nh_n}\right)^{1/2}$, for some $\epsilon_0 > 0$ yields to

$$\sup_{x \in \Omega} |\mathbb{E}(\ell_n(x)) - \ell_n(x)| = O\left(\left(\frac{\log \log n}{nh_n}\right)^{1/2}\right) \quad a.s. \quad (5.43)$$

Then the proposition can be obtained straightforwardly from (5.41)- (5.43), Lemmas 5.6, 5.7 and 5.7 ■

PROOF OF THEOREM 5.2

Let $x \in \Omega$. As $F(\cdot|x)$ and $F_n(\cdot|x)$ are continuous, we have

$F(\xi_p(x)|x) = F_n(\xi_{p,n}(x)|x) = p$. Then

$$\begin{aligned} |F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x)| &= |F(\xi_{p,n}(x)|x) - F_n(\xi_{p,n}(x)|x)| \\ &\leq \sup_{t \in \mathcal{C}} |F_n(t|x) - F(t|x)|. \end{aligned} \quad (5.44)$$

The consistency of $\xi_{p,n}(x)$ follows then immediately from Proposition 1 and continuity of $F(\cdot|x)$. Now, a Taylor expansion of $F(\cdot|x)$ in a neighborhood of $\xi_p(x)$, implies that

$$F(\xi_{p,n}(x)|x) - F(\xi_p(x)|x) = (\xi_{p,n}(x) - \xi_p(x)) f(\xi_p^*(x)|x)$$

where $\xi_p^*(x)$ is between $\xi_p(x)$ and $\xi_{p,n}(x)$. Then, by (5.44), we have

$$\sup_{x \in \Omega} |\xi_{p,n}(x) - \xi_p(x)| |f(\xi_p^*(x)|x)| \leq \sup_{x \in \Omega} \sup_{t \in \mathcal{C}} |F_n(t|x) - F(t|x)|.$$

The result is then a consequence of Proposition 1 and the assumption of $f(\xi_p(\cdot)|\cdot)$ being uniformly bounded away from zero. ■

In order to proof Proposition 5.7, we will use the following decomposition and the Lemmas below

$$\begin{aligned} \tilde{F}_n(t|x) - F(t|x) &= \frac{\ell(x)}{\ell_n(x)} \left[\frac{\tilde{F}_{1,n}(t, x) - \mathbb{E}(\tilde{F}_{1,n}(t, x)) - F(t|x) [\ell_n(x) - \mathbb{E}(\ell_n(x))]}{\ell(x)} \right] \\ &- \frac{1}{\ell_n(x)} \left[F_1(t, x) - \mathbb{E}(\tilde{F}_n(t|x)) - F(t|x) [\ell(x) - \mathbb{E}(\ell_n(x))] \right]. \end{aligned}$$

Note that

$$\begin{aligned} (nh_n)^{1/2} (F_n(t|x) - F(t|x)) &= \frac{\ell(x)}{\ell_n(x)} \left((nh_n)^{1/2} A_n(t, x) \right) \\ &+ (nh_n)^{1/2} C_n(t, x), \end{aligned} \quad (5.45)$$

where

$$\begin{aligned} A_n(t, x) &= \frac{\tilde{F}_{1,n}(t, x) - \mathbb{E} \left(\tilde{F}_{1,n}(t, x) \right) - F(t|x) [\ell_n(x) - \mathbb{E}(\ell_n(x))]}{\ell(x)} \\ B_n(t, x) &= \frac{F_1(t, x) - \mathbb{E} \left(\tilde{F}_n(t|x) \right) - F(t|x) [\ell(x) - \mathbb{E}(\ell_n(x))]}{\ell_n(x)} \\ \text{and } C_n(t, x) &= \left(F_n(t|x) - \tilde{F}_n(t|x) \right) - B_n(t, x). \end{aligned}$$

The next Lemmas show the asymptotic normality of $(nh_n)^{1/2} A_n(x, y)$ and the convergence in probability of $(nh_n)^{1/2} C_n(x, y)$ to zero. We begin to show the second part :

Lemma 5.11. *Under the assumptions **A1** (i) – (iii), **A2**, **A3** (i), **A4** and **A5** we have*

$$(nh_n)^{1/2} C_n(x, y) \longrightarrow 0 \text{ in probability as } n \rightarrow \infty.$$

PROOF OF LEMMA 5.11

Let use the following decomposition

$$\begin{aligned} &(nh_n)^{1/2} C_n(x, y) \\ &= \frac{(nh_n)^{1/2}}{\ell_n(x)} \left[\left(F_{1,n}(t, x) - \tilde{F}_{1,n}(t, x) \right) - \left(F_1(t, x) - \mathbb{E}(\tilde{F}_{1,n}(t, x)) \right) + F(t|x) (\ell(x) - \ell_n(x)) \right] \end{aligned}$$

On the one hand, making use of Lemmas 5.6 and 5.7 we have that each term in brackets goes to zero in probability as n goes to infinity.

On the other hand,

$$\frac{(nh_n)^{1/2}}{\ell_n(x)} = \frac{(nh_n)^{1/2} \tilde{f}_n(t|x)}{\ell_n(x) \tilde{f}_n(t|x)} = \frac{(nh_n)^{1/2} \tilde{f}_n(t|x)}{\tilde{F}'_{1,n}(t, x)}$$

where

$$\tilde{F}'_{1,n}(t, x) = \frac{1}{nh_n^2} \sum_{i=1}^n K \left(\frac{x - X_i}{h_n} \right) H' \left(\frac{t - Y_i}{h_n} \right) \frac{\delta_i}{\overline{G}(Y_i)}.$$

By assumptions **A1** (i) and **A4** (i), we obtain that $\tilde{F}'_{1,n}(t, x) \geq \frac{m^* m}{h_n^2} \frac{1}{n} \sum_{i=1}^n \delta_i$.

Therefore

$$\frac{(nh_n)^{1/2}}{\ell_n(x)} \leq \frac{(nh_n^5)^{1/2}}{m^* m} \frac{\tilde{f}_n(t|x)}{\frac{1}{n} \sum_{i=1}^n \delta_i}.$$

By **A2** (ii), Lemma 3 and in conjunction with the SLLN the right hand side of the last inequality goes to zero in probability as n goes to infinity, which completes the proof of Lemma 5.11 $\blacksquare$

Lemma 5.12. *Under assumptions **A1** (iii), **A3**, **A4**, **A5** and **A6** we have*

$$nh_n \text{Var} [A_n(t, x)] \longrightarrow (\sigma(t, x))^2 \text{ as } n \rightarrow \infty.$$

PROOF OF LEMMA 5.12

Clearly, we have

$$\begin{aligned} A_n(t, x) &= \frac{1}{nh_n \ell(x)} \sum_{i=1}^n \left[K \left(\frac{x - X_i}{h_n} \right) H \left(\frac{t - Y_i}{h_n} \right) \frac{\delta_i}{\overline{G}(Y_i)} - \mathbb{E} \left[K \left(\frac{x - X_i}{h_n} \right) H \left(\frac{t - Y_i}{h_n} \right) \frac{\delta_i}{\overline{G}(Y_i)} \right] \right] \\ &\quad - F(t|x) \sum_{i=1}^n \left[K \left(\frac{x - X_i}{h_n} \right) - \mathbb{E} \left(K \left(\frac{x - X_i}{h_n} \right) \right) \right] \\ &=: \frac{1}{nh_n \ell(x)} \sum_{i=1}^n N_i(t, x) \end{aligned}$$

where

$$\begin{aligned} N_i(t, x) &= K \left(\frac{x - X_i}{h_n} \right) \left[H \left(\frac{t - Y_i}{h_n} \right) \frac{\delta_i}{\overline{G}(Y_i)} - F(t|x) \right] \\ &\quad - \mathbb{E} \left[K \left(\frac{x - X_i}{h_n} \right) \left(H \left(\frac{t - Y_i}{h_n} \right) \frac{\delta_i}{\overline{G}(Y_i)} - F(t|x) \right) \right]. \end{aligned}$$

Then

$$\begin{aligned} &nh_n \text{Var} (A_n(t, x)) \\ &= \frac{1}{h_n \ell^2(x)} \text{Var}(N_1(t, x)) \\ &= \frac{1}{h_n \ell^2(x)} \mathbb{E}(N_1^2(t, x)) \\ &= \frac{1}{h_n \ell^2(x)} \mathbb{E} \left[K^2 \left(\frac{x - X_1}{h_n} \right) \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) - F(t|x) \right)^2 \right] \\ &\quad - \frac{1}{h_n \ell^2(x)} \left[\mathbb{E} \left[K \left(\frac{x - X_1}{h_n} \right) \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) - F(t|x) \right) \right] \right]^2. \quad (5.46) \end{aligned}$$

We can write the second term of the the right hand side of (5.46) as :

$$\begin{aligned} &\frac{1}{h_n \ell^2(x)} \left[\mathbb{E} \left[K \left(\frac{x - X_1}{h_n} \right) \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) - F(t|x) \right) \right] \right]^2 \\ &= \frac{h_n}{\ell^2(x)} \left[\mathbb{E} \left(\widetilde{F}_{1,n}(t, x) \right) - F(t|x) \mathbb{E}(\ell_n(x)) \right]^2. \end{aligned}$$

Using Lemma 5.6, we can conclude that the second term of (5.46) tends to zero as n goes to infinity.

Now let us turn to the first term of (5.46), we have

$$\begin{aligned} & \frac{1}{h_n \ell^2(x)} \mathbb{E} \left[K^2 \left(\frac{x - X_1}{h_n} \right) \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) - F(t|x) \right)^2 \right] \\ &= \frac{1}{h_n \ell^2(x)} \mathbb{E} \left[K^2 \left(\frac{x - X_i}{h_n} \right) \mathbb{E} \left[\left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_i}{h_n} \right) - F(t|x) \right)^2 \middle| X_1 \right] \right]. \end{aligned}$$

Using the definition of the conditional variance we have :

$$\begin{aligned} \mathbb{E} \left[\left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) - F(t|x) \right)^2 \middle| X_1 \right] &= \text{Var} \left[\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) \middle| X_1 \right] \\ &+ \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) \middle| X_1 \right) - F(t|x) \right]^2 \\ &=: \mathcal{V}_1 + \mathcal{V}_2. \end{aligned}$$

Using Lemma 5.6, we have $\mathcal{V}_2 = O(h_n^2)$ as $n \rightarrow +\infty$.

Let's now examine the term $\mathcal{V}_1$

$$\begin{aligned} \mathcal{V}_1 &= \mathbb{E} \left[\frac{\delta_1}{\overline{G}^2(Y_1)} H^2 \left(\frac{t - Y_1}{h_n} \right) \middle| X_1 \right] - \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) \middle| X_1 \right) \right]^2 \\ &=: \mathcal{J}_1 + \mathcal{J}_2. \end{aligned}$$

The first term of the last equality can be developed as follow

$$\begin{aligned} \mathcal{J}_1 &= \mathbb{E} \left[\mathbb{E} \left(H^2 \left(\frac{t - Y_1}{h_n} \right) \frac{\delta_1}{\overline{G}^2(Y_1)} \middle| T_1 \right) \middle| X_1 \right] \\ &= \mathbb{E} \left[H^2 \left(\frac{t - T_1}{h_n} \right) \frac{1}{\overline{G}(T_1)} \middle| X_1 \right] \\ &= \int_{\mathbb{R}} H^2 \left(\frac{t - y}{h_n} \right) \frac{1}{\overline{G}(y)} f(y|X_1) dy \\ &= \int_{\mathbb{R}} H^2(z) \frac{1}{\overline{G}(t - h_n z)} dF(t - h_n z | X_1). \end{aligned}$$

By Taylor's expansion, we have

$$\begin{aligned} \mathcal{J}_1 &= \int_{\mathbb{R}} H^2(z) \frac{1}{\overline{G}(t)} dF(t - h_n z | X_1) + \frac{h_n}{\overline{G}^2(t)} \int_{\mathbb{R}} z H(z) \overline{G}'(t^*) dF(t - h_n z | X_1) + o(1), \\ &=: \mathcal{J}_1' + \mathcal{J}_2' \end{aligned} \tag{5.47}$$

where t^* is between t and $t - h_n z$.

Under assumption **A6**, the second term of (5.47) can be bounded by

$$\mathcal{J}_2' \leq \frac{h_n^2 \sup_{x \in \mathbb{R}} |g(x)|}{\overline{G}^2(t)} \int_{\mathbb{R}} z f(t - h_n z | X_1) dz,$$

then under assumption **A3** (ii), we get that $\mathcal{J}'_2 = O(h_n^2)$.
On the other hand, by integrating by parts we have

$$\begin{aligned}\mathcal{J}'_1 &= \frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z)F(t - h_n z | X_1) dz \\ &= \frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z) (F(t - h_n z | X_1) - F(t|x)) dz + \frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z)F(t|x) dz.\end{aligned}$$

Remark that $\int_{\mathbb{R}} 2H(z)H'(z)F(t|x) dz = F(t|x)$ and using the same idea as in (5.31), we get

$$\int_{\mathbb{R}} 2H'(z)H(z) (F(t - h_n z | X_1) - F(t|x)) dz = O(h_n) \quad \text{as } n \rightarrow +\infty$$

$$\text{and } \int_{\mathbb{R}} H^2(z) \frac{1}{\overline{G}(t)} dF(t - h_n z | X_1) \longrightarrow \frac{F(t|x)}{\overline{G}(t)} \quad \text{as } n \rightarrow +\infty.$$

Therefore

$$\begin{aligned}\mathcal{V}_1 &= \frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z)F(t|x) dz + O(h_n^2) + O(h_n) - \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | X_1 \right) \right]^2 \\ &= \frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z)F(t|x) dz + O(h_n) - \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | X_1 \right) \right]^2.\end{aligned}$$

We proved that

$$\begin{aligned}\lim_{n \rightarrow \infty} n h_n \text{Var} (A_n(t, x)) &= \lim_{n \rightarrow \infty} \frac{1}{h_n \ell^2(x)} \mathbb{E} \left[K^2 \left(\frac{x - X_1}{h_n} \right) \left(\frac{1}{\overline{G}(t)} \int_{\mathbb{R}} 2H'(z)H(z)F(t|x) dz + O(h_n) \right) \right] \\ &\quad - \lim_{n \rightarrow \infty} \frac{1}{h_n \ell^2(x)} \mathbb{E} \left[K^2 \left(\frac{x - X_1}{h_n} \right) \left[\mathbb{E} \left(\frac{\delta_1}{\overline{G}(Y_1)} H \left(\frac{t - Y_1}{h_n} \right) | X_1 \right) \right]^2 \right] \\ &= \frac{\kappa}{\ell(x)} \left[\frac{F(t|x)}{\overline{G}(t)} - (F(t|x))^2 \right] \\ &= (\sigma(t, x))^2.\end{aligned}$$

■

Let $Z_{ni}(t, x) = \frac{N_i(t, x)}{(n h_n \ell^2(x))^{1/2}}$. Remark that

$$(n h_n)^{1/2} A_n(x, y) = \sum_{i=1}^n Z_{ni}(x, y),$$

so to prove the asymptotic normality of $(n h_n)^{1/2} A_n(t, x)$, it suffices to shown that $\sum_{i=1}^n Z_{ni}(t, x)$ satisfies the Lindberg-Feller condition, which is given in the following Lemma.

Lemma 5.13. *Under the assumptions **A1** (i) and (iii), **A2** (i), **A3**, **A4**, **A5** and **A6** for all $\epsilon > 0$ we have*

$$\sum_{i=1}^n \int_{\{Z_{ni}^2(t,x) > \epsilon^2 \text{Var}(\sum_{i=1}^n Z_{ni}(t,x))\}} Z_{ni}^2(t,x) d\mathbb{P}_{(X_i, Y_i)} \longrightarrow 0 \quad \text{as } n \rightarrow \infty.$$

PROOF OF LEMMA 5.13

By Lemma 5.12 we have clearly,

$$\text{Var}\left(\sum_{i=1}^n Z_{ni}(t,x)\right) \longrightarrow (\sigma(t,x))^2 \text{ as } n \rightarrow \infty,$$

therefore for n large enough

$$\left\{ Z_{ni}^2(t,x) > \epsilon^2 \text{Var}\left(\sum_{i=1}^n Z_{ni}(t,x)\right) \right\} \subset \left\{ Z_{ni}^2(t,x) > \frac{\epsilon^2}{2} (\sigma(t,x))^2 \right\}.$$

On the one hand, using the fact that $|H\left(\frac{t-Y_i}{h_n}\right) \frac{\delta_i}{\bar{G}(t)} - F(t|x)| \leq 2$ and the condition **A1** (i), we have

$$\begin{aligned} Z_{ni}^2(x,y) &= \frac{N_i^2(t,x)}{nh_n} \\ &< \frac{2}{nh_n \ell^2(x)} \left[K\left(\frac{x-X_n}{h_n}\right) \left(H\left(\frac{t-Y_i}{h_n}\right) \frac{\delta_i}{\bar{G}(t)} - F(t|x) \right) \right]^2 \\ &+ \frac{2}{nh_n \ell^2(x)} \mathbb{E}^2 \left[K\left(\frac{x-X_n}{h_n}\right) \left(H\left(\frac{t-Y_i}{h_n}\right) \frac{\delta_i}{\bar{G}(t)} - F(t|x) \right) \right] \\ &\leq \frac{8}{nh_n} \left[K^2\left(\frac{x-X_n}{h_n}\right) + \mathbb{E}^2 \left(K\left(\frac{x-X_n}{h_n}\right) \right) \right] \\ &\leq \frac{16M^*}{nh_n \ell^2(x)}. \end{aligned}$$

On the other hand, the second part of assumption **A2** (i) implies that $nh_n \rightarrow +\infty$ as $n \rightarrow +\infty$, therefore

$$Z_{ni}^2(t,x) \longrightarrow 0 \quad \text{as } n \rightarrow \infty,$$

then for n large enough, we have

$$Z_{ni}^2(t,x) \leq \frac{\epsilon^2}{2} (\sigma(t,x))^2.$$

Finally this shows clearly, that for n large enough, the set

$$\left\{ Z_{ni}^2(t,x) > \epsilon^2 \text{Var}\left(\sum_{i=1}^n Z_{ni}(t,x)\right) \right\}$$

is empty, then this completes the proof of the Lemma 5.13 and therefore Proposition 5.6. ■

PROOF OF THEOREM 5.3

Using a Taylor expansion we have

$$\xi_{p,n}(x) - \xi_p(x) = \frac{F_n(\xi_{p,n}(x)|x) - F_n(\xi_p(x)|x)}{f_n(\xi_{p,n}^*(x)|x)}$$

where $\xi_{p,n}^*(x)$ lies between $\xi_p(x)$ and $\xi_{p,n}(x)$.

The continuity of $f(\cdot|x)$, Theorem 5.2 and Lemma 5.10 imply the convergence in probability of the above denominator to $f(\xi_p(x)|x)$. Proposition 5.3 is used to finish the proof. ■

Chapitre 6

Conclusion

6.1 Commentaires

La littérature sur les données censurées est très riche mais reste encore d'actualité, notamment en ce qui concerne l'estimation des paramètres fonctionnels présents dans ce modèle. Ce type de données a des caractéristiques très spécifiques dont on doit tenir compte dans leur modélisation.

L'introduction présentée au début de cette thèse, nous a permis de situer l'importance du modèle de censure dans l'analyse des durées de vie. Cette constatation nous a amené à réaliser une étude bibliographique sur le modèle censuré. Nous avons également dégagé l'intérêt de l'estimation de la fonction des quantiles simples et conditionnels ; ces derniers restent un sujet très abordé de nos jours.

La réalité pratique de l'étude des durées de vie, nous a amené à considérer le cas où les données présentent une forme de dépendance, nous avons opté pour le mélange fort ou l' α -mélange, car ce cas de dépendance est le moins restrictif et est très répandu ; par exemple les processus AR et ARMA sont α -mélangeants.

Nous avons présenté quelques notions fondamentales sur l' α -mélange et principalement le résultat de Cai (2001) sur les propriétés de l'estimateur de Kaplan-Meier pour des données α -mélangeantes.

Nous nous sommes intéressés, dans un premier temps, à la fonction des quantiles simples, nous avons rappelé les principales propriétés des quantiles simples lorsque les observations sont complètes pour les deux cas i.i.d. et α -mélange.

Nous avons considéré les données censurées et α -mélangeantes pour établir, sous une hypothèse sur le coefficient de mélange, la consistance, la représentation forte et une représentation de type Bahadur de l'estimateur $\hat{F}_n^{-1}(\cdot)$ du quantile, $\hat{F}_n$ étant l'estimateur de Kaplan-Meier de la f.d.r. F ; en travaillant sur la base de la représentation forte de l'estimateur de Kaplan-Meier dans le cas de mélange fort (voir le théorème 2 dans Cai 1998 et le théorème 1 dans Cai 2001)

Nous avons aussi montré, grâce à la version d'Ibragimov et Linnick (1981) du T.C.L.,

la normalité asymptotique du processus $\sqrt{n} \left(\hat{F}_n^{-1}(p) - F^{-1}(p) \right)$ tout en précisant le terme de variance. Nous avons ainsi généralisé le résultat de Cheng (1983) à un cas de dépendance.

Au regard des champs d'applications et pour obtenir une modélisation plus fine, nous avons, dans un second temps, envisagé de travailler sur l'estimation de la fonction quantile lorsqu'on dispose d'informations supplémentaires regroupées dans une variable explicative X .

A notre connaissance, le cas de données censurées et α -mélangeantes lorsqu'on dispose d'une variable explicative corrélée avec la variable d'intérêt n'a pas encore été envisagé.

Nous donnons, une extension des résultats d'Ould-Saïd (2006) qui construit un estimateur lissé $F_n(t|x)$ de la f.d.r. conditionnelle $F(t|x)$ de T sachant $X = x$ et par conséquent du quantile conditionnelle $F^{-1}(p|x)$, au cas du modèle censuré et α -mélangeant. Nous établissons la convergence uniforme presque sûre de $F_n^{-1}(p|x)$, tout en précisant la vitesse de convergence.

Nous nous sommes penchés sur la construction d'estimateurs de la fonction quantile conditionnelle asymptotiquement normaux, ce qui ouvre un champ plus grand aux applications des données censurées, en particulier pour faire de la prévision. Nous sommes revenus au modèle censuré dans le cas i.i.d., sous ce modèle nous construisons un nouvel estimateur lissé de la f.d.r. conditionnelle $F(.|.)$. Nous établissons sa convergence uniforme presque sûre et sa normalité asymptotique et nous déduisons la normalité asymptotique de l'estimateur du quantile conditionnel correspondant. Nous donnons quelques applications de ces résultats notamment à la prévision et la construction d'intervalles de confiance.

Les simulations, que nous avons réalisé, nous permettent de montrer la performance de nos estimateurs à taille finie des échantillons .

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

6.2 Perspectives de recherche

Pour conclure, nous pouvons soulever des perspectives à court terme et d'autres à plus long terme de nos travaux ; nous en citons quelques-uns :

- Les résultats établis au paragraphe 5.3 peuvent être étendus au cas dépendant (cas d' α -mélange). Dans ce cas les résultats intéressant à établir sont :
La convergence presque sûre et la normalité asymptotique des estimateurs lissés de la f.d.r. conditionnelle et du quantile conditionnel.
- S'inspirant des travaux de Xiang (1995) sur l'efficacité relative de l'estimateur empirique du quantile simple et de l'estimateur lissé et des résultats de Lemdani et Ould Saïd (2001) sur la comparaison entre l'E.K.M. et sa version lissée, il

nous semble possible d'étudier l'efficacité relative entre les estimateurs (5.10) et (5.20). Ce qui nous permettra de les comparer avec un choix optimal de la fenêtre et donnera, en plus de la normalité asymptotique, une preuve théorique que (5.20) est meilleur que (5.10).

- Utiliser d'autres méthodes d'estimation non paramétrique des quantiles conditionnels dans le modèle censuré, nous pensons en particulier à la méthode des polynômes locaux. Cette méthode nous donnera des versions robustes des quantiles conditionnels locaux.

Le cas du modèle censuré i.i.d. est traité par Gannoun *et al.* (2005), nous pensons améliorer leurs résultats, en pénalisant par l'estimateur de Kaplan-Meier de C , la variable de censure afin de tenir compte des premières valeurs censurées. Ceci en modifiant l'estimateur de la fonction de répartition conjointe de T et X , $F(.,.)$ proposé par Stute (1993). Puis, nous regarderons le cas du modèle censuré dépendant.

- Grâce à la performance accrue des appareils de mesure et l'augmentation des capacités de stockage informatique, de nombreuses données sont collectées et sauvegardées sur des échelles temporelles ou des grilles spatiales de plus en plus fines (courbe d'évolution de température, courbes spectrométriques, images satellites,...). On est ainsi amené à traiter des données assimilables à des courbes ou à des fonctions de variables continues, ce qu'on appelle données fonctionnelles. Nous envisageons de traiter l'estimation du quantile conditionnel dans le cas où la variable explicative est fonctionnelle. Le cas où les données sont complètes est traité par Ezzahrioui et Ould Saïd (2006), nous nous pencherons sur le cas des données censurées.

Bibliographie

- [1] AALEN, O. (1978). Nonparametric inference for a family of count process. *Ann. Statist.* **6**, 701-726.
- [2] ALY, E. A., CSÖRGŐ, M. AND HORVÁRTH, L. (1985). Strong Approximations of the uantile Process of the Product-Limit Estimator. *J. Multivariate Anal.* **16**, 185-210.
- [3] AUESTAD, B., TJØSTHEIM, D., (1990). Identification of nonlinear time series : First order characterisation and order determination, *Biometrika*, **77**, 669-687.
- [4] BABU, G. J. AND SINGH, K. (1978). On deviations between empirical ans quan- tile processes for mixing random variables. *J. of Multivariate Anal.* **8**, 532-549.
- [5] BAHADUR, R.R.(1967). Rates of convergence of estimates and test statistics, *Ann. Math. Statist.* **38**; 303-324.
- [6] BAHADUR, R.R. (1966).A note on quantiles in large samples. *Ann. Math. Statist.* **37**, 577-580.
- [7] BERAN, R. (1981). nonparametric regression with randomly censored survival data, *Technical Report university of California*. Berkley.
- [8] BERLINET, A., GANNOUN, A. AND MATZNER-LØBER, E. (2001). Asymptotic normality of convergent estimates of conditional quantiles, *Statistics*, **35**, 139-168.
- [9] BHATTACHARYA, P.K., GANGOPADHYAY, A. (1990). Kernel and nearest neigh- bor estimation of conditional quantile, *Ann. Statist.* **18**, 1400-1415.
- [10] BILINGSLEY, P. (1968). Convergence of probability measures. Wiley & Sons.
- [11] BLUM, J.R., HANSON, D. L. AND KOOPSMANS L.H. (1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. Verw. Gebiete* **2** 1-11.
- [12] BOSQ, D. (1998). *Nomparametric Statistics for Stochastic Processes, Estimation and Prediction*. Lect. Notes in Statistics, **110**, Springer Verlag.
- [13] BOSQ, D. ET LECOUTRE, J.J. (1987). *Théorie de l'estimaion fonctionnelle*, Economica .
- [14] BRESLOW, N. AND CROWLEY, J. (1974). A large sample study of the life table and product-limit estimates under random censorship. *Ann. Statist.* **2**, 437-443
- [15] BURKE, M.D., CSÖRGŐ, S. AND HORVÁRTH, L. (1981),Strong approximations of some biometric estimates under random censorship. *Z. Wahrsh. view. gebiete*, **56**, 87-112.

- [16] BURKE, M.D., CSÖRGŐ, S. AND HORVARTH, L. (1986). A correction to an improvement of strong approximations of some biometric estimates under random censorship, *Probab. Theory Related fields* **79**, 51-57.
- [17] CAI, Z. (2001). Estimating a distribution function for censored times series data, *J. of Multivariate Anal.* **78**, 299-318.
- [18] CAI, Z., (1998). Asymptotic properties of Kaplan-Meier estimator for censored dependent data, *Statist. Probab. Lett.*, **37**, 381-389.
- [19] CAI, Z. AND ROUSSAS, G.G. (1998). Kaplan-Meier estimator under association. *J. of Multivariate anal.* **67**, 318-348.
- [20] CAI, Z. AND ROUSSAS, G.G. (1992). Uniform strong estimation under α mixing with rates. *Statist. Probab. Letters* **15**, 47-55.
- [21] CAO, R., LOPEZ-DE-ULLIBARRI, I., JENSEN, P. AND VERAVERBEKE, N. (2005). Persmoothed Kaplan-Meier and Neslson-Aalen estimators, **17**, 31-56. *J. of Multivariate Anal.* **17**, 31-56.
- [22] CARBONEZ, A., GYÖRFI, L. AND VAN DER MEULIN, E. C. (1995); Partitioning estimates of a regression function under random censoring, *Statist. Decisions* **13**, 21-37.
- [23] CHAUDHURI, P. (1991a). Nonparametric estimates of regression quantiles and their local Bahadur representation, *Ann. Statist.*, **19**, 760-777.
- [24] CHAUDHURI, P. (1991b). Global nonparametric estimation of conditional quantile functions and their derivatives, *J. Multivariate Anal.*, **39**, 246-269.
- [25] CHAUDHURI, P., DOKSUM, K. AND SAMAROV, A. (1997). On average derivative quantile regression, *Ann. Statist.*, **25**, 715-744.
- [26] CHENG, K.F., (1984). On almost sure representation for quantiles of the product limit estimator with applications, *Sankhyā*. **46**, Series A, 426-443.
- [27] CSÖRGŐ, M., (1983). *Quantile processes with Statistical applications*, SIAM, Philadelphia PA. 1983
- [28] DABROWSKA, D. (1995) Nonparametric regression with censored covariates, *J. Multivariate Anal.*, **54** 253-283.
- [29] DABROWSKA, D. (1992). Nonparametric quantile regression with censored data. *Sankhyā*. **54**, Series A, 252-259.
- [30] DABROWSKA, D. (1989). Uniform consistency of the kernel conditional kaplan-Meier estimate, *Ann. Statist.* **17**, 1157-1167.
- [31] DABROWSKA, D. (1987). (Nonparametric regression with censored survival data, *Scand. J. Statist.* **14** , 181-197.
- [32] DEHEUVELS, P. AND EINMAHL, J.H.J. (2000). Functional limit laws for the increments of Kaplan-Meier product limit processes and applications. *Ann. Probab.* **28**, 1301-1335.
- [33] DEVROYE L., GYÖRFI L. AND LUGOSI G. (1996). *A Probability Theory of Pattern Recognition*. Springer, New York.

- [34] DHOMPONGSA, S. (1984). *A note on the almost sure approximation of the empirical process of weakly dependent random variables*. *Yokohama Math. J.* **32**, 113-121.
- [35] DOUKHAN, P., (1994). *Mixing : Properties and Examples*, *Lect. Notes in Statistics.*, **61**, Springer Verlag.
- [36] DROESBEKE, J.J., FICHET, B. ET TASSI, P. (1989). *Analyse Statistique des Durées de Vie*, *Economica*.
- [37] ENGLE, R.F., (1982). *Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflation*. *Econometrics*, **50**, 4, 987-1007.
- [38] M.EZZAHRIOUI, E. OULD-SAÏD.(2007) *Asymptotic normality of the kernel estimator of the conditional quantile in the normed space*. *Far East journal of Theoretical Statistics*.
- [39] FALK, M. (1983). *Relative efficiency and deficiency of kernel type estimators of smooth functions*. *Statist. neerlandica*, **37**, 642-669.
- [40] FAN, J., HU, T.C. AND TRUONG, Y.K., (1994). *Robust nonparametric function estimation*, *Scand. J. Statist.*, **21**, 433-446.
- [41] FOTOPOULOS, S., AHN, S.K. AND CHO, S., (1994). *Strong approximation of the quantile processes and its applications under strong mixing properties*, *J. Multivariate Anal.* **51**, 17-47.
- [42] GANNOUN, A., SARRACCO, J. , YUAN, A. & BONNET, G. E. (2005). *Nonparametric quantile regression with censored data*. *Scand. J. Statist.* **32**, 527-550.
- [43] GANNOUN, A., SARRACCO, J. AND YU, K.(2003). *Nonparametric prediction by conditional median and quantiles*, *J. Statist. Plann. Inference* **117**, 207-223.
- [44] GANZÁLEZ MANTEIGA, W. & CADARSO SUÁREZ, C. (1994). *Asymptotic properties of generalized Kaplan-Meier estimator with some applications*. *J. Nonpar. Statist.* **4**, 65-78.
- [45] GEBELEIN H.(1941). *Das Statistische Problem der Korrelation als Variation und Eigenwertproblem und sein Zusammenhang mit Ausgleichungsrechnung*. *Z. Angew. Math. Mech.* **21**, 364-379.
- [46] GINNÉ, E. AND GILLOU, A. (2002). *Rates of strong uniform consistency for multivariate kernel density estimators*. *Ann. I. H. Poincaré* **38**, 907-921.
- [47] GILL, R. D. (1980). *Censoring and stochastic integrals*. *Mathematical center tracts*, **124**, Mathematics centrum, Amsterdam.
- [48] HANSEN, K.W. (1996). *ambulatory blood pressure in insulin-dependent diabetes*. *J. Diabetes Complications*, **10** (6), 331-351
- [49] HIRSCHFELD, H. O. (1935). *A connection between correlation and contingency*. *Math. Proc. Camb. Phil. Soc.*, **31**, 520-524.
- [50] Hoeffding, W.(1963). *Probability inequalities for sums of bounded random variables*. *J. Amer. Statist. Assoc.*, **58**, 13-30.

- [51] HONDA, T. (2000). *Nonparametric estimation of a conditional quantile for α -mixing processes*, *Ann. Inst. Statist. Math.* **52** : 459-470.
- [52] HORVÁRTH, L.(1982) *Strong approximations of quantiles of the product-limit estimator. In Carleton Math. lecture Note* **38**, Carleton University, Ottawa.
- [53] IBRAGIMOV, I. A. AND LINNICK, Y.V. (1981). *Independent and stationary processes or random variables*. wiley.
- [54] IBRAGIMOV, I. A. (1962) *Some limit theorems for stationary processes. Theory. Probab. Appl.*, **7**, 349–382.
- [55] JONES, M.C., AND HALL, P. (1990). *Mean squared error properties of kernel estimates of regression quantiles*, *Statist. Probab. Letters*, **10**, 283-289.
- [56] KANG, S. AND KOEHLER, K.J., (1997). *Modification of the Greenwood formula for correlated response times*, *Biometrics*, **53**, 885-899.
- [57] KAPLAN, E.M. AND MEIER, P. (1958). *Nonparametric estimation from incomplete observations*, *J. Amer. Statist. Assoc.* **53**, 457-481.
- [58] KIEFER, J. (1967); *On Bahadur's representation of sample quantiles*, *Ann. Math. Statist.*, **38**, 1323-1342.
- [59] KLEIN, J. P. (1991). *Small sample moments of some estimator of the variance of the Kaplan-Meier and Nelson-Aalen estimators*. *Scand. J. Statist.*, **18**, 333-340.
- [60] KOEHLER, K.J. (1995). *An analysis of temperature effects on bean leaf egg hatch times. In " Proceedings of the 1994 KSU Conference on Applied Statistics in Agriculture", Vol 6, 215-229, Kansas State Univ. Press.*
- [61] KOEHLER, K.J. AND SYMANOWSKI, J., (1995). *Constructing multivariate distributions with specific marginal distributions*, *J. Multivariate Anal.*, **55**, 261-282.
- [62] KOHLER, M., MÁTHÉ, K. AND PINTÉR, M. (2002). *Prediction from randomly right censored data. J. Multivariate Anal.* **80**, 73-100.
- [63] KOMLÓS, J., MAJOR, P. AND TUSNÁDY, G. (1975). *Weak convergence and embedding limit theorems of probability theory. Colloq. Math. Soc. Janos Bolyai, North Holland* **11** 149-165.
- [64] LECOUTRE, J.P. AND OULD-SAÏD, E. (1995a). *Convergence of the conditional Kaplan-Meier under strong mixing. J. Statist. Plann. Inf.* **44**, 359-369.
- [65] LECOUTRE, J.P. AND OULD-SAÏD, E., (1995b). *Hazard rate estimation for strong-mixing and censored processes.it J. Nonparametric Statist.*, **5**, 83-89.
- [66] LEMDANI, M. AND OULD-SAÏD, E. (2001). *Relative deficiency of Kaplan-Meier estimator with respect to asmoothed estimator. Mathematical Methods of Statistics*, **10**, 215-231.
- [67] LEMDANI, M., OULD-SAÏD, E. AND POULIN, N. (2005). *Strong representation of the quantile function for left truncated and dependent data. Math. Meth. Statist.* **14**

- [68] LIEBSCHER, E. (1996). *Strong convergence of sums of α mixing random variables with applications to density estimation*. *Stoch. Processes and their Appl.* **56**, 69-80.
- [69] LIEBSCHER, E. (2002). *Kernel density and hazard rate estimation for censored data under α mixing condition*. *Ann. Inst. Statist. Math.* **54**, 19-28.
- [70] LYNDALL-BELL, D. (1971). *A method of allowing for known observational selection in small samples applied to 3CR quasars*. *Monthly Notices Roy. Astron. Soc.* **155**, 95-118.
- [71] LIO, Y. L. AND PADGETT, W. J. (1986). *Some convergence results for kernel-type quantile estimators under censoring*. *Statist. & Probab. letters* **53**, 457-481.
- [72] LIO, Y. L. AND PADGETT, W. J. (1987). *On the mean squared error of nonparametric quantile estimators under random right-censorship*. *Commun. Statist. Theory Meth.* **16**, 1617-1628.
- [73] LO, H.S., MACK, Y.P. AND WANG, J.L., (1989). *Density and hazard rate estimation for censored data via strong representation of the Kaplan-Meier estimator*, *Probab. Th. Rel. fields*, **80**, 461-473.
- [74] LO, H.S. AND SINGH, (1986), *The product limit estimator and bootstrap : some asymptotic representations*, *Probab. TH. Rel. Fields* **71**, 455-465.
- [75] MAJOR, P. AND REJTŐ, L. (1988). *Strong embedding of the estimator of the distribution function under random censorship*. *Ann. Statist.* **16**, 1113-1132.
- [76] MASRY, E. AND TJOSTHEIM, D., (1995). *Nonparametric estimation and identification of nonlinear ARCH time series : Strong convergence and asymptotic normality*. *Econ. Theory*, **11**, 258-289.
- [77] MASRY, E. AND TJOSTHEIM, D., (1997). *Additive nonlinear ARX times series and rpojection estimates*. *Econ. Theory*, **13**, 214-256.
- [78] MEHRA, K.L., RAO, M.S. AND UPADRASTA, S.P. (1991). *A smooth conditional quantile estimator and related applications of conditional empirical processes*, *J. Multivariate Anal.*, **37**, 151-179.
- [79] MILLER, R. AND HALPEN, J. (1982). *Regression with censored data*. *Biometrika* **69**, 521-531.
- [80] NADARAYA, E. A. (1964). *On estimating regression*. *Theor. Probability Appl.* **9** 141-142.
- [81] NELSON, W. (1972). *Theory and applications of hazard plotting for censored failure data*. *Technometrics* **14** 945-965.
- [82] OULD-SAÏD, E. (2006). *A strong uniform convergence rate of kernel conditional quantile estimator under random censorship*. *Statist. Probab. Letters.* **76**, 579-586.
- [83] OULD-SAÏD, E. AND CAI, Z. (2005). *Strong uniform consistency of nonparametric estimation of the conditional mode function ; J. Nonparametric Statist.* **17** 797-806.

- [84] OULD-SAÏD, E. AND SADKI, O. (2005). *Strong approximation of quantile function for strong mixing and censored processes*, *Comm. Statist. Theory and Methods* **34** 1449-1459.
- [85] OULD-SAÏD, E. (2002), *Sur l'estimation fonctionnelle pour des observations complète ou censurées*. Thès H.D.R., Université du Littoral Côte d'Opale.
- [86] PADGETT, W. T., THOMBS, L. A., (1989). *A smooth nonparametric quantile estimation from right censored data.*, *Statist. Probab. Lett.*, **7**, 113-121.
- [87] PARZEN, E. (1962). *On estimation a probability density function and mode*. *Ann. Math. Statist.* **33** 832-837.
- [88] QIN, Y. S. AND WU, Y. (2001). *An estimator of a conditional quantile in the presence of auxiliary information*. *J. Statist. Plann. and Inf.* **99**, 59-70.
- [89] REID, N. (1981). *Estimating the median survival time*. *Bimetrika* **68**, 601-608.
- [90] REISS, R. D. (1981). *Nonparametric estimation of smooth distribution functions*. *Scand. J. Statist.* **8**, 116-119.
- [91] RIO, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants*, Springer ISMAI.
- [92] ROSENBLATT, M. (1971). *Curve estimates*. *Ann. Math. Statist.* **42**, 1815-1842.
- [93] ROSENBLATT, M. (1956). *A central limit theorem and strong mixing condition*, *Proc. Nat. Acad. Sci. USA*, **42**, 43-47.
- [94] ROUSSAS, G. G. (1988). *Nonparametric estimation in mixing sequences of random variables*. *J. Statist. Palnn. Inference*, **18** 135-149.
- [95] ROUSSAS, G. G. (1969). *Nonparametric estimation of the transition distribution function of a Markov process*. *Ann. Math. Statist.* **40**, 1386-1400.
- [96] SAMANTA, M. (1989). *Non-parametric estimation of conditional quantiles*, *Statist. Probab. Lett.*, **7**, 407-412.
- [97] SANDER, J., (1975). *The weak convergence of quantiles of the product limit estimator*, Technical report Number 5, Stanford University. Department of Statistics.
- [98] SEN, P.K., (1972). *On the Bahadur representation of samples quantiles for sequences of ϕ -mixing random variables*, *J. Multivariate Anal.*, **2**, 77-95.
- [99] SERFLING, R. J. (1980). *Approximations theorems of mathematical statistics*; Wiley, New York.
- [100] SHORACK, G. R. AND WELLNER, J.A . (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [101] STUTE, W. (1993). *Consistent estimation under random censorship when co-variables are present*. *J. Multivariate Anal.* **45**, 89-103. *Ann. Statist.* **21**, 14-44.
- [102] STUTE, W. (1984). *The oscillation behavior of empirical processes : the multivariate case*. *Ann. Probab.* **12**, 361-379.
- [103] SUN, S. (2006). *The Bahadur representation for sample quantiles under weak dependance*. *Statist. Probab. Lett.*, **76** 1238-1244.

- [104] VAN KEILEGOM, I. AND VERAVERBEKE, N. (1998). *Bootstrapping quantiles in a fixed design regression model with censored data. J. Statist. Plan. Inference* **69** 115- 131.
- [105] VAN KEILEGOM, I. AND VERAVERBEKE, N. (1997). *Weak convergence of the bootstrapped conditional Kaplan-Meier process and its quantile process. Commun. Statist. Theory Meth.* **26** 853-869.
- [106] VAN KEILEGOM, I. AND VERAVERBEKE, N. (1996). *Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. Comm. Statist. Theory Methods*, **25**, 2251-2265.
- [107] VAN DER VAART, A.W. (1998). *Asymptotic statistique, Cambridge series in statistical and probabilistic mathematics.*
- [108] VOEKEL, J. AND CROWLY, J. (1984). *Nomparametric inference for a class of semi-Markov process with censored observation. Ann. Satist.* **12** 142-160.
- [109] WATSON, G. S. (1964). *Smooth regression analysis. Sankhyā, Series A*, **26**, 359-372.
- [110] WEI, L.J., LIN, D.Y., WEISSFELD, L. (1989). *Regression analysis of multivariate incomplete failure times data by modeling marginal distributions. J. Amer. Statist. Assoc.*, **84**, 1064-1072.
- [111] WOODROOFE, M. (1985). *Estimating a distribution function with truncated data. Ann. Statist.* **13**, 163-177.
- [112] XIANG, X. (1995). *Deficiency of the sample quantile estimator whith respect to kernel quantile estimators for censored data. Ann. Statist.* **23**, 836-854.
- [113] XIANG, X. (1996). *A kernel estimator of a conditional quantile. J. Multivariate Anal.* **59**, 206-216.
- [114] YING, Z. WEI, L.J. (1994). *The Kaplan-Meier estimate for dependent failure time observations, J. Multivariate Anal.* **50** (1994), 17-29.
- [115] YOSHIHARA, K.I., (1995). *The bahadur representation of sample quantiles for sequences of strongly mixing random variables, Statist. Probab. Lett.*, **24**, 299-304.
- [116] YU, K. AND JONES, M. C. (1998). *Local linear quantile regression. J. Amer. Statist. Assoc.* **93**, 228-237.
- [117] YU, H. (1996). *A note on strong approximation for quantile processes of strong mixing sequences, Statist. Probab. lett.*, **30**, 1-7.