N° d'ordre : **05/2008 - E/MT**

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE DES SCIENCES ET DE LA TECHNOLOGIE HOUARI BOUMEDIENE-ALGER

FACULTE DE MATHEMATIQUES

THÈSE

Présentée pour l'obtention du diplôme de Doctorat d'Etat

En : Mathématiques

Spécialité : Probabilités et Statistique

Par : TATACHAK Abdelkader

Sujet :

Prévision non paramétrique dans les modèles de troncature via l'estimation du mode

soutenue publiquement le 15 décembre 2008, devant le jury composé de :

M.	Kamal	BOUKHETALA	Professeur	U.S.T.H.B	Président
M.	Elias	OULD SAÏD	MCF-HDR	Univ. Littoral, France	Directeur de thèse
M.	Amar	AISSANI	Professeur	U.S.T.H.B	Examinateur
M.	Zaher	MOHDEB	Professeur	Univ. Constantine	Examinateur
M.	Abdelkader	MOKKADEM	Professeur	Univ. Versailles, France	Examinateur
Mlle.	Hafida	GUERBYENNE	Maître de Conférence	U.S.T.H.B	Examinatrice
M.	Ali	LAKSACI	Maître de Conférence	Univ. Sidi Bel-Abbès	Examinateur

Remerciements

J'aimerais tout d'abord, remercier vivement le Professeur Elias Ould-Saïd, qui a su diriger mes travaux de recherches. Je tiens à lui exprimer ma profonde gratitude pour ses encouragements, son soutien moral et son extrême patience. Toujours optimiste et positif, son expérience et savoir faire, son dynamisme et sa rigueur fûrent les facteurs principaux conduisant à l'aboutissement de cette thèse après trois années de travail.
Je suis reconnaissant au Professeur Kamal Boukhetala qui a accepté de présider le jury, malgré ses nombreuses responsabilités.
Je remercie particulièrement Madame Hafida Guerbyenne, pour l'intérêt qu'elle a porté à mon travail depuis mes premiers pas, et je suis très sensible à sa présence au sein du jury.
Je suis très honoré, que le Professeur Abdelkader Mokkadem, dont les nombreux travaux sur l'estimation et le mode sont plus qu'impressionnants, ait accepté immédiatement d'être rapporteur de ma thèse, bien qu'il soit extrêmement occupé.
Je suis très reconnaissant au Professeur Amar Aissani, dont l'aide et le soutien moral m'ont été d'une grande utilité. Il était le premier à me conseiller de continuer à travailler dans le domaine du non paramétrique après ma thèse de Magister. Je le remercie pour l'honneur qu'il me fait en acceptant de participer au jury de ma soutenance.
Je remercie vivement le Professeur Zaher Mohdeb, pour l'intérêt qu'il a porté à mon travail en acceptant sans hésiter de faire partie du jury.
A Monsieur, Ali Laksaci, spécialiste du mode, j'adresse mes vifs remerciements pour l'honneur qu'il me fait en acceptant de faire partie du jury.
Je tiens également, à adresser mes vifs remerciements, aux Professeurs Djamal Louani et Mohamed Lemdani, qui n'ont pas cessé de me conseiller et de m'encourager. Leur modestie, leur rigueur et honnêteté intellectuelle resteront pour moi un modèle.
Je remercie également mes amis du groupe de travail 'GTDI', Ourida, Zohra et Mostefa qui m'ont accompagné et épaulé au cours de ces trois dernières années. Aussi, je ne saurai remercier assez mon ami Yacine pour toute l'aide qu'il m'a apportée, ses conseils et son soutien m'ont été précieux. Je remercie mes amis, H'mida, Karim, K'rimo et Chakib pour leur amitié et disponibilité.
J'exprime ma reconnaissance à tous les membres du Laboratoire M.S.T.D, ainsi qu'à tous les collègues qui m'ont aidé de loin ou de près durant l'accomplissement de ce travail.
Mes pensées vont enfin à tous mes proches, ma famille et tout particulièrement mes parents, mes enfants 'mama et momo' et mon épouse, pour m'avoir aidé et supporté.

Résumé

Dans cette thèse, nous nous proposons d'étudier les propriétés asymptotiques d'estimateurs non paramétriques, du mode simple et conditionnel, pour des données tronquées aléatoirement à gauche. Les estimateurs proposés se positionnent comme alternative à certains estimateurs classiques, et en particulier à la fonction de régression, lorsque des prévisions non paramétriques sont envisagées, surtout lorsque les données observées présentent une forme fortement dissymétrique ou multimodale. Des simulations sont faites pour illustrer les résultats théoriques obtenus dans le cas i.i.d.

Cette thèse est organisée comme suit :

- **Chapitre 1 :** Ce chapitre est une introduction générale, introduisant des concepts généraux utiles pour faciliter la lecture des chapitres suivants.

- **Chapitre 2 :** Dans ce chapitre, nous présentons des travaux sur le mode simple dans le cas de données i.i.d., tronquées aléatoirement à gauche. Nous illustrons les résultats obtenus (consistance et normalité asymptotique) par des simulations pour montrer les performances des estimateurs proposés.

- **Chapitre 3 :** Ce chapitre est une extension des résultats du chapitre 2 au cas de données fortement mélangeantes. Seule la consistance est présentée, la normalité asymptotique fait l'objet d'un travail en cours de finalisation.

- **Chapitre 4 :** Cet avant dernier chapitre traite du mode conditionnel (consistance et normalité asymptotique) et les résultats sont illustrés par des simulations.

- **Chapitre 5 :** Ce dernier mini-chapitre est un rappel des résultats obtenus sous forme de conclusion, suivie de quelques perspectives.

Mots-clefs : $\alpha-$mélange, consistance forte, données tronquées, estimateur à noyau, estimateur de Lynden-Bell, mode, mode conditionnel, normalité asymptotique, simulation, V$-$C-classe.

Abstract

In this thesis, we are interested in the study of the asymptotic properties of non parametric estimators of the simple and the conditional mode, for randomly left-truncated data. The proposed estimators provide alterternative procedures to some classical estimators, in particular to the regression function, to make non parametric forecastings, especially when the observed data present a shape either greatly unsymetrical or has several modes. Simulation studies are made to enforce the theoretical results gotten in the i.i.d case.

This thesis is organized as follows :

- **Chapter 1 :** This chapter is a general introduction, introducing some general concepts, useful to facilitate the reading of the following chapters.

- **Chapter 2 :** In this chapter, we present works on the simple mode in the case of i.i.d. data, randomly truncated from the left. We illustrate the results gotten (consistence and asymptotic normality) by simulations to show the performances of the proposed estimators.

- **Chapter 3 :** This chapter is an extension of the results of chapter 2 to the case of stronly mixing data. Only the consistence is considered, the asymptotic normality is curerently under consideration.

- **Chapter 4 :** This next-to-last chapter deals with the conditional mode function (consistence and asymptotic normality) and the results are illustrated by simulations.

- **Chapter 5 :** This last mini-chapter recalls the results stated in this thesis as a conclusion, followed of some perspectives.

Keywords : $\alpha-$mixing, asymptotic normality, conditional mode, Kernel estimator, Lynden-Bell estimator, mode, rate of convergence, simulation, strong consistency, truncated data, V$-$C-class.

Articles parus et soumis

1. **Article paru** : OULD-SAÏD, E., TATACHAK, A. (2007) : *Asymptotic properties of the kernel estimator of the conditional mode for the left-truncated model, C. R. Acad. Sci. Paris, Ser. I* 344 :651−656.

2. **Article à paraître** : OULD-SAÏD, E., TATACHAK, A. : *Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left-truncation. "In Communications in Statistics-Theory and Methods, (2009)".*

3. **Article à paraître** : OULD-SAÏD, E., TATACHAK, A. : *On the nonparametric estimation of the simple mode under random left-truncation model. "Revue Roumaine de Mathématiques Pures at Appliquées, vol. 54, (2009)".*

4. **Article en révision** : OULD-SAÏD, E., TATACHAK, A. : *On the strong uniform consistency of the conditional mode estimator for randomly left truncated time series.*

5. **Article en finalisation** : BENRABAH, O., OULD-SAÏD, E., TATACHAK, A. : *Asymptotic Normality for the Kernel mode estimator under dependence and left-truncation.*

Table des matières

Chapitre 1

Introduction générale

1.1 Généralités

Il est très difficile de donner une définition précise à l'inférence non paramétrique. L'idée de base de l'inférence non paramétrique est d'utiliser les données pour inférer une quantité inconnue, en utilisant le moins possible d'hypothèses ; cela signifie l'utilisation de modèles statistiques appartenant à un espace de dimension infinie. De ce fait, il serait peut être plus judicieux de parler de l'inférence en dimension infinie.
En inférence en dimension infinie, le statisticien est souvent amené à supposer que la densité sous-jacente est unimodale. En effet, l'unimodalité est d'usage naturel, elle s'impose plus naturellement que la symétrie, par exemple ; car les vraies données présentent habituellement une dissymétrie, en particulier, les échantillons de petite taille.
Il est connu que "la tendance centrale" d'un échantillon, peut être quantifiée par la moyenne, la médiane, ou le mode ; la moyenne étant le paramètre le plus souvent utilisé, mais le mode, contrairement aux deux autres paramètres, possède la propriété de préserver sa définition face à une transformation de la variable d'intérêt du type $Quantitative \rightarrow Qualitative - Nominale$. Bien que ces trois mesures coïncident pour les distributions unimodales symétriques, elles peuvent être totalement différentes pour des données observées.
L'un des plus curieux résultats en statistique théorique classique, est la relation empirique liant les trois paramètres de position, mise en évidence pour des distributions continues, par Pearson (1895) est :

$$moyenne - mode \approx 3(moyenne - mediane).$$

Le résultat est trivial pour des distributions symétriques unimodales, et reste valide pour des distributions légèrement dissymétriques. A titre d'exemple, pour une distribution gamma de paramètre γ, on a : moyenne$=\gamma$, mode$=\gamma-1$ et médiane$=\gamma-1/3+O(1/\gamma)$ (lorsque $\gamma\to\infty$); (voir, Hall (1980) et Asselin de Beauville (1978) pour plus de détails). De son côté, Fréchet (1940) a donné une borne supérieure du mode, fonction de la borne supérieure de la moyenne, pour une variable discontinue non négative. Par exemple, pour une variable entière non négative, si sa moyenne est majorée par une constante positive a, alors son mode est majoré par $a(a+3)/2$.

1.2 Tester l'unimodalité

L'unimodalité est une caractéristique qui permet de révéler l'homogénèité de la population étudiée, ainsi l'exitence de plusieurs modes nous informe sur l'exitence de sous populations dans la composition de la population mère, et c'est pour cette raison que cette caractéristique est utilisée en classification automatique (non supervisée). Récemment, cette notion a été utilisée dans la classification des formes de la fonction de hasard pour des modèles mélange de Weibull à poids négatifs (Jiang et al. (2004), page 161).
Cependant, le nombre de travaux concernant la mise en oeuvre d'outils permettant de déceler l'unimodalité reste modeste. Parmi les tests d'hypothèses concernant le nombre de modes, on cite : Good and Gaskins (1980), Silverman (1981 ; 1983), Hartigan and Hartigan (1985), Muller et Sawitzki (1991), Mammen et al (1992), Minnotte (1997). Lemdani (1995) a généralisé les résultats de Hartigan and Hartigan (1985) pour étendre le pricipe à une classe beaucoup plus vaste que celle des lois à décroissance exponentielle, tout en montrant la robustesse de la statistique 'dip', alors que Hall et al. (2004) ont généralisé les travaux de Silverman (1981) en utilisant des noyaux autres que gaussiens. Il est utile de souligner que le noyau gaussien est réputé pour ses propriétés analytiques et en particulier celles des fonctions Pólya fréquence (voir, Schoenberg (1950)).

1.3 Estimation non paramétrique de densité

1.3.1 Les principes d'estimation de densités

Selon le Théorème de Lebesgue sur les densités,

$$\lim_{h\to 0}\int_{S_{yh}}\frac{f(x)}{\lambda(S_{yh})}dx=f(y)$$

pour presque tout y, où S_{yh} est la boule de centre y et de rayon h, et λ désigne la mesure de Lebesgue.

L'expression sous la limite, lorsqu'on observe un $n-$échantillon $Y_1, Y_2, ..., Y_n$, peut être estimée par

$$f(y) = \sum_{i=1}^{n} \frac{\mathbb{I}(Y_i \in S_{yh})}{n\lambda(S_{yh})}.$$

Cet estimateur a été proposé par Rosenblatt (1956) et développé par Parzen (1962). Depuis, cet estimateur est connu sous le nom d'estimateur à noyau de Parzen-Rosenblatt et parfois Akaike-Parzen-Rosenblatt, et s'écrit

$$f_n(y) = \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{y - Y_i}{h}\right), \tag{1.1}$$

où $f_n(y)$ est défini sur $\mathbb{R}$, $h > 0$ est un paramètre de lissage (appelé aussi fenêtre), et $K(y)$ est un noyau qui souvent satisfait, $K(y) \geq 0$ et d'intǵrale unité.
Comme exemples de noyaux, on cite : le noyau normal $\phi(y) = (2\pi)^{-1/2} \exp(-y^2/2)$, $\frac{1}{2}\exp(-|y|)$ (Piccard), $\pi^{-1}\sin^2 y/y^2$ (Fejér-de la Vallée Poussin), et le noyau d'Epanechnikov $K(y) = \frac{3}{4}(1 - y^2)\mathbb{I}(|y| \leq 1)$.
Cependant, la positivité du noyau et pour des raisons justifiées, est parfois violée, dans le sens que certaines valeurs de $K(y)$ peuvent être négatives. Cette dernière caractéristique est étroitement liée à la notion de noyau d'ordre supérieur.

Définition 1.1. *Un noyau $K(y)$ est dit d'ordre r, s'il satisfait :*

$$\omega_l(K) := \int u^l K(u)du = \begin{cases} 1, & si \quad l = 0 \\ 0, & si \quad 1 \leq l \leq r - 1, \\ K_r \neq 0, & si \quad l = r. \end{cases}$$

Pour $l > 2$, le noyau $K(y)$ et par conséquent l'estimateur $f_n(y)$ peut être négatif, alors, l'estimateur doit être normalisé comme suit :

$$(f_n(y))_+ \times \left(\int (f_n(y))_+ \, dy\right)^{-1}.$$

Plusieurs règles ont été proposées pour construire des noyaux d'ordre supérieur, dont on cite celle proposée par Jones et Foster (1993), (voir, Wand et Jones (1995), page 32), définie par :

$$K_{[r+2]}(y) = \frac{3}{2}K_{[r]}(y) + \frac{1}{2}yK'_{[r]}(y),$$

où $K_{[r]}(y)$ dńote le noyau d'ordre r, supposé différentiable. Par exemple, pour un noyau d'ordre 4, si on utilise le noyau normal $K_{[2]}(y) = \phi(y)$, on obtient

$$K_{[4]}(y) = \frac{1}{2}(3 - y^2)\phi(y).$$

La précision de l'estimateur dépend plus du choix de h que de celui de $K(y)$. La variance de tout estimateur à noyau décroit vers 0 avec une vitesse de l'ordre $O(1/nh)$, quand $nh \to +\infty$, alors que le biais est d'ordre $O(h^2)$ pour un noyau d'ordre 2, et d'ordre $O(h^4)$ pour un noyau d'ordre 4 (voir, Silverman (1986)). Par conséquent, la variance d'un estimateur à noyau augmente alors que son biais diminue quand h décroit vers 0, ainsi le choix de h doit faire l'objet d'un compromis entre les deux effets.
A titre d'exemple, la valeur optimale de h, lorsque le noyau est choisi gaussien et la densité sous-jacente est normale, est donnée par $h^{opt} = 1.06 \times \widehat{\sigma} \times n^{-1/5}$, où $\hat{\sigma}^2$ est un estimateur robuste de la variance de la loi en question. En pratique, il est recommandé de prendre (voir, Silverman (1986), page 47),

$$\hat{\sigma} = \min\left(s, \sigma_{IQR} := \frac{E_{[1/4,3/4]}}{\Phi^{-1}(3/4) - \Phi^{-1}(1/4)}\right),$$

s étant l'écart-type empirique, Φ^{-1} est la fonction quantile de la loi normale standard (le dénominateur de $\sigma_{IQR} \approx 1.349$) et $E_{[1/4,3/4]}$ désigne l'écart interquartile empirique. Il est bien connu que pour $d = 1$, en utilisant un noyau d'ordre 2 et une fenêtre invariable, pour estimer la densité et ses dérivées d'ordre k, le MSE est donné par :

$$\begin{aligned} MSE\left(f_n^{(k)}(y)\right) &= \mathbb{E}\left(f_n^{(k)}(y) - f^{(k)}(y)\right)^2 \\ &= \operatorname{var}\left(f_n^{(k)}(y)\right) + \left[\mathbb{E}\left(f_n^{(k)}(y)\right) - f^{(k)}(y)\right]^2 \\ &= \frac{h^4}{4}\left(f^{(k+2)}(y)\right)^2 K_2^2 + (nh^{2k+1})^{-1} f(y) R(K^{(k)}) + o\left\{(nh^{2k+1})^{-1} + h^4\right\}, \end{aligned}$$

losque $h \to 0$, où $R(K^{(k)}) = \int \left(K^{(k)}(u)\right)^2 du$, en imposant la continuité de $f^{(k+2)(y)}$.
Le choix optimal de h pour le critère MSE est de l'ordre de $n^{-1/(2k+5)}$, ce qui donne un choix de l'ordre $n^{-1/7}$ pour l'estimation de la dérivée première f'. De plus, l'optimalité de $MSE\left(f_n^{(k)}(y)\right)$ et $MISE\left(f_n^{(k)}(y)\right) := \mathbb{E}\int\left\{f_n^{(k)}(y) - f^{(k)}(y)\right\}^2 dy$, est de l'ordre $n^{-4/(2k+5)}$. En pratique, si le noyau d'Epanechnikov est optimal pour un tel critère, l'estimateur idéal ne peut pas être obtenu puisque f et $f^{(k)}$ sont inconnus. Pour contourner ce problème, la densité $f(y)$ est remplacée par

$$f_{n,h_0}(y) = \frac{1}{nh_0}\sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_0}\right).$$

Aussi, la dérivée d'ordre k de f est remplacée par celle de $f_{n,h_0}(y)$ et h_0 peut être choisi égal à $\sigma_0\left(84\sqrt{\pi}/(5n^2)\right)^{1/13}$, en utilisant le noyau gaussien $\phi(y)$, et où σ_0 est l'écart-type empirique.

L'estimateur à noyau présente des avantages et des inconvénients, parmi lesquels :

1. **Avantages**
   - Il peut être utilisé pour des densités multidimensionnelles
   - Comportement récursif : Il peut être adapté à une utilisation on-line.
2. **Inconvénients**
   - Il produit les effets aux bords pour des densités à supports compacts
   - Il n'est pas aussi bon pour des densités non suffisamment lisses que pour des densités lisses.

1.4 Estimer le mode

Le problème le plus lié à celui de l'estimation de la densité est l'estimation du mode. Il peut y avoir plusieurs raisons de vouloir estimer ce paramètre ; par exemple, comme une analyse exploratoire ou pour une confirmation secondaire des valeurs estimées d'autres paramètres de position. C'est aussi un problème difficile, car en inférence non paramétrique, n'importe quelle méthode utilisée pour estimer le mode, nécessite d'abord l'estimation d'une densité qui est un problème difficile, ce qui rend le problème d'estimation du mode intrinsèquement intéressant.

Deux approches principales ont été utilisées pour estimer le mode :

1.4.1 Estimation directe

Cette approche repose sur une estimation directe à partir des données ; c'est-à-dire sans nécessité d'estimation d'une densité. On peut citer l'estimateur de Chernoff (1964), Grenander (1965), Delenius (1965) et celui de Venter (1967), ce dernier a été repris et amélioré par Sager (1975).
Comme application du mode, Wegman (1971) a repris l'estimateur de Chernoff pour proposer l'utilisation du mode dans la calibration du flux sanguin.

1.4.2 Estimation indirecte

Cette approche repose sur l'estimation de la densité d'abord, et la localisation du mode comme étant un maximum global ou approximativement global, ensuite. Certaines méthodes d'estimation utilisent la totalité des données, alors que d'autres n'en utilisent qu'une partie, que ce soit pour le cas univarié ou multidimensionnel. Les références les plus importantes sont les premières à être adressées dans la littérature ; on cite : Parzen (1962), Chernoff (1964), Venter (1967), Wegman (1971) et Sager (1975), pour ne citer que ceux-là. Récemment, l'estimateur de Chernoff à fait l'objet d'un remarquable travail par Groeneboom et Wellner (2001).

Comme exemples d'estimateurs du mode, on peut citer les trois suivants :

1.4.2.1 Estimateur de Parzen

Soit $f_n(y)$ un estimateur de la densité $f(y)$, défini comme en (1.1), basé sur l'observation de $Y_1, Y_2, ..., Y_n$. L'estimateur du mode de Parzen, noté $\theta_{n,P}$ est défini comme étant une variable aléatoire satisfaisant

$$f_n(\theta_{n,P}) = \sup_{y \in \mathbb{R}^d} f_n(y) \tag{1.2}$$

1.4.2.2 Estimateur de Chernoff

Soit a_n une suite de réels positifs décroissant vers zéro avec n, l'estimateur de Chernoff noté $\theta_{n,C}$, est défini comme étant le milieu d'un quelconque intervalle de longueur $2a_n$ contenant le maximum des observations $Y_1, Y_2, ..., Y_n$.

1.4.2.3 Estimateur de Venter

Soit b_n une suite d'entiers naturels tendant vers ∞ quand $n \to \infty$. L'estimateur de Venter noté $\theta_{n,V}$, est défini comme étant le milieu du plus court intervalle contenant b_n observations parmi $Y_1, Y_2, ..., Y_n$.
Il est à noter que certains estimateurs du mode sont basés sur l'annulation de la dérivée. Pour une revue très détaillée concernant les propriétés asymptotiques (Consistance et loi limite) des trois estimateurs, on peut consulter (DasGupta (2008), Théorèmes 32.17 et 32.18).

Depuis l'article précurseur de Parzen (1962) traitant de l'estimation du mode, l'étude du mode a connu un manque d'intérêt, et cela est dû probablement aux difficultés rencontrées dans l'interprétation des résultats obtenus en utilisant ce paramètre. Aussi, comme l'a souligné Eddy (1980), ce manque d'intérêt est dû aux problèmes posés par l'estimation de la densité. Cette dernière décennie a connu un renouvellement de pari avec le mode, dû certainement à l'avancée technologique et à la disponibilité des moyens de calcul très sophistiqués. A titre d'exemple, Abraham et *al.* (2003, 2004) ont établi des résultats très utiles en pratique (traitement on-line, par exemple) pour contourner l'aspect gourmand en temps de calcul des processus utilisant le mode, en considérant le maillage le plus naturel qui soit pour estimer la densité, puisque les observations se concentrent naturellement autour de leur mode. Pratiquement, il est nécessaire de calculer au maximum n fois l'estimateur de la densité pour obtenir l'estimateur du mode. Ainsi, cette méthode d'estimation diminue considérablement le nombre d'opérations dans l'estimation du mode, tout en possédant d'aussi bonnes propriétés théoriques que ses concurrentes antérieures.

1.5 Conditionnement et prévision

Il arrive souvent que la variable d'intérêt notée Y ne soit pas libre, dans la mesure où sa réalisation dépend étroitement de la réalisation d'une seconde variable X. Dans ce cas, on parle de présence de covariables ou de variables explicatives. Dans la pratique, ce cas est souvent la situation la plus naturelle, raison ayant conduit à reprendre les travaux sur les trois pramètres de localisation. Bien que la médiane conditionnelle est devenue un cas particulier de l'étude du quantile conditionnel, la régression est plus ancienne et plus utilisée que le mode conditionnel.
Soit $Z = (X, Y)$, où Y est une variable aléatoire (v.a) et X un vecteur aléatoire de $\mathbb{R}^d$. Si X et Y ne sont pas indépendantes et X observabe, il est raisonnable de prédire Y à travers une transformation de X ; c'est-à-dire via une règle $h : \mathbb{R}^d \to \mathbb{R}$, qui associe à chaque réalisation x de X le réel $h(x)$. Comme le prédicteur $h(X)$ est maintenant aléatoire, le risque qui lui est associé est $r(h) = \mathbb{E}\ell(Y - h(X))$, où l'espérance est prise par rapport à la distribution conjointe de X et Y, et $\ell(.)$ est la fonction coût utilisée. Le meilleur prédicteur conditionnel de Y sachant X est une fonction h^* vérifiant $r(h^*) \leq r(h)$ pour toute fonction h. Si la fonction coût $\ell(.)$ est convexe et admet un minimum unique, alors on peut définir un estimateur $\hat{r}(.)$ de $r(.)$ comme étant solution du problème

$$\hat{r}(x) = \min_{c \in \mathbb{R}} \mathbb{E}\left(\ell(Y - c) | X = x\right) \tag{1.3}$$

Par exemple, si Y a une variance finie σ_Y^2, l'unique solution de (1.3) pour la fonction de coût $\ell(z) = z^2$ est :

$$\hat{r}(x) = \mathbb{E}\left(Y | X = x\right),$$

qui est la moyenne de la distribution conditionnelle de Y sachant $X = x$ (fonction de régression). Son étude est à l'origine de plusieurs publications, tant dans le domaine des statistiques qu'en économétrie.
Si $\ell(z) = |z| + 2p(z-1)z$, critère plus difficile à utiliser, la solution de (1.3) est la courbe

des quantiles de Y sachant $X = x$. Alors que si on considère la fonction de coût non convexe définie par :

$$\begin{cases} 0 & \text{si } u = 0 \\ 1 & \text{sinon,} \end{cases}$$

la solution de (1.3) est la fonction mode conditionnel. Ce dernier fut étudié pour la première fois par Collomb et al.(1987). Depuis, les modèles basés sur le mode conditionnel se sont imposés et positionnés comme alternative à la prévision basée sur la régression, lorsque la distribution d'intérêt présente une dissymétrie importante ou possède plusieurs modes. Il est important de rappeler qu'il y a des densités conditionnelles pour lesquelles la fonction de régression est nulle partout, par exemple :

$$g\left(y|x\right) = \frac{1}{2\sqrt{2\pi}} \left(\exp\left\{ -\frac{1}{2}\left(y-x\right)^2 \right\} + \exp\left\{ -\frac{1}{2}\left(y+x\right)^2 \right\} \right).$$

Par conséquent, il n'est pas raisonnable de faire des prévisions en utilisant la fonction de régression $\mathbb{E}\left(Y/X = x\right)$. Une autre situation critique, est lorsque la distribution d'intérêt est multimodale, avec un mode prédominant. Dans ce cas, la moyenne conditionnelle empirique peut fournir des valeurs totalement improbables (comprises entre deux modes) alors que le mode empirique fournit un des modes de la distribution. Pour illustrer la supériorité du mode conditionnel sur les autres méthodes de prévision, dans le cas de multimodalité des données, Matzner-Lϕber (1997) a comparé les performances de la moyenne conditionnelle, la médiane conditionnelle, la régression modale (Scott, 1992) et le mode conditionnel, en utilisant un jeu de données réelles représentant le temps d'attente relevé en minutes, entre deux éruptions du 'Old Faithful Geyser, Yellow Stone National Park, Wyoming', du premier au 15 Août 1985.
Une autre application intéressante du mode conditionnel dans le domaine des hautes technologies, a été proposée par Rossi (2004) dans un procédé de dépollution biologique.
Ainsi, le mode conditionnel, par son importance dans le domaine de la prévision non paramétrique, a motivé un certain nombre de chercheurs à l'investigation d'estimateurs du mode, comme en témoignent les travaux de Rosa (1993), Ould-Saïd (1993, 1997), Matzner-Lϕber et al.(1997) et Quintela-Del-Río et Vieu (1997). D'autres références et travaux sur le mode seront données dans les chapitres 2, 3 et 4.
Tous les travaux cités jusqu'à présent concernent des données recueillies par les méthodes classiques d'échantillonnage sans conditions restrictives, dans le sens à ne pas affecter un poids nul aux éléments de la population se trouvant dans des situations particulières. Dans ce cas, on parle de données complètes. Il n'est pas rare que les données à traiter ne soient pas complètes, dans ce cas les techniques statistiques classiques ne s'adaptent pas correctement aux données incomplètes, car, l'inférence faite sur l'échantillon observé ne

s'étend pas directement à la population mère.
Afin de rendre facile la lecture de cette thèse, nous donnons quelques définitions et exemples de données incomplètes.

1.6 Données incomplètes

1.6.1 Durée de vie

On entend par *durée de vie*, une variable aléatoire Y, souvent positive. En fait, cette variable est observée dans plusieurs domaines tels que l'économie, médecine, biologie, épidémiologie, astronomie et santé publique, de la durée de vie d'une ampoule à la durée de vie d'un malade sous observation thérapeutique. Cette notion de durée de vie est mesurée en fixant une origine, et sa valeur est quantifiée par le temps écoulé jusqu'à la survenue d'un évènement (première panne pour une machine, décès ou rechute pour un malade, embauche pour un chômeur,...). De manière générale, une durée de vie sera donc, le temps écoulé pour passer d'un état A à un état B. Lorsque les durées de vie sont observées dans leur totalité, on parle de données complètes, dans le cas contraire, les données sont incomplètes et nécessitent un traitement statistique particulier.
L'analyse des données incomplètes est complexe. Si la durée de vie d'un individu n'est observable que dans une période donnée, on parle de censure. Il s'agit de troncature, si l'individu doit survivre longtemps pour être observé, ou lorsque l'individu est retenu par l'étude uniquement s'il a connu l'évènement en question à une date t donnée.

1.6.1.1 Censure

Définition 1.2. *Soit Y la variable d'intérêt, si au lieu d'observer Y on observe une constante C et si on sait que $Y > C$, on dira alors qu'il y a censure à droite.*

Il arrive souvent que la constante de censure soit elle même une variable aléatoire, dans ce cas, on parle de censure aléatoire à droite et on a :

Définition 1.3. *Etant donnée une variable aléatoire Y, on dit qu'il y a censure aléatoire à droite s'il existe une variable aléatoire C, telle qu'au lieu d'observer Y on observe :*

$$\left\{(Z, \delta) / Z = \min(Y, C), \delta = \mathbb{1}_{\{Y \leq C\}}\right\}.$$

1.6.1.2 Troncature

Définition 1.4. *Soit Y la variable d'intérêt et T une autre variable aléatoire, si Y et T sont observables uniquement si $Y > T$, et rien sinon, on dira que la variable Y est aléatoirement tronquée à gauche par la variable de troncature T. Dans le cas contraire, on parle de troncature aléatoire à droite.*

1.6.1.3 Exemple de troncature à gauche

Un exemple de troncature à gauche concerne la durée de vie des retraités. On enregistre l'âge au décès et l'âge d'entrée au centre. Un individu donné, doit vivre suffisamment longtemps pour être parmi la communauté du centre de retraite, et s'il décède avant, il y a troncature à gauche.
Un autre exemple de troncature à gauche est le problème d'estimation de la distribution des diamètres de particules microscopiques. Selon la résolution du microscope, seules les particules suffisamment grandes sont retenues pour l'étude, alors que les autres restent invisibles pour l'investigateur et la troncature à gauche a lieu.

1.6.1.4 Exemple de troncature à droite

Supposons qu'une certaine maladie soit caractérisée par un évènement initial (*Etat A*) et un évènement final (*Etat B*). Un exemple est l'étude du Virus de l'Immunodéficience Humaine (HIV :(*Etat A*)) et le Syndrome d'ImmunoDéficience Acquise (SIDA :(*Etat A*)). Si W désigne le temps du calendrier de l'évènement initial et Y la durée de vie. Alors une observation $(W;Y)$ est enregistrée seulement si $W + Y \leq \tau$, où τ est la date de clôture de la collecte des données. C'est un exemple de troncature à droite : On observe la durée de vie si $Y \leq \tau - W =: T$; T est appelé le temps de troncature. Un autre exemple de troncature à droite survient dans l'estimation de la distribution des étoiles à partir de la terre. Les étoiles très lointaines ne sont pas visibles, et par conséquent, il y a troncature à droite.

Remarque 1.1. *A notre connaissance, les seuls travaux sur le mode en présence de données incomplètes (censure), sont ceux de Louani (1998), Gannoun et Saracco (2002), pour le mode simple, et Ould-Saïd et Cai (2005) pour le mode conditionnel.*

1.7 Estimation de la fonction de régression

1.7.1 un bref historique

Le problème de l'estimation en utilisant la régression a une longue histoire comme l'atteste la littérature. Déjà en 1632, Galileo Galilei utilisa une procédure qui peut être interprétée comme étant un ajustement linéaire de données contaminées. Un tel ajustement d'un nuage de points par une droite est le principe du problème classique de la régression linéaire. Une solution à ce problème est fournie par le célèbre principe des moindres carrés, mis en oeuvre indépendamment par A. M. Legendre et C. F. Gauss, et publié en (1805) et (1809), respectivement.
Le principe des moindres carrés peut être appliqué pour construire des estimateurs non paramétriques à condition de ne pas se réstreindre à une classe donnée de relations.
L'analyse par régression linéaire basée sur la fonction de régression (concept analytique) a été introduite par F. Galton (1889), alors que l'approche probabiliste dans le cas de lois normales multivariées, fut intrduite par A. Bravais (1946).
Le premier estimateur de la régression non paramétrique de type "local averaging" a été proposé par J. W. Tukey (1947).

1.7.2 La régression et le L_2-risque

En régression, on considère un vecteur aléatoire (X, Y), où $X \in \mathbb{R}^d$ et $Y \in \mathbb{R}$, et l'on s'intérèsse au degré de dépendance de la valeur de la variable Y, dite varaiable réponse, de la valeur observée du vecteur X. En d'autre terme, on cherche á trouver une fonction (mesurable) $f : \mathbb{R}^d \to \mathbb{R}$, telle que $f(X)$ soit une bonne approximation de Y, c'est à dire $f(X)$ proche de Y au sens d'un critère donné, ce qui équivaut à $|f(X) - Y|$ "petit". Comme X et Y sont aléatoires, $|f(X) - Y|$ l'est aussi. Cependant, la notion de "petit" n'est pas explicite. Pour résoudre ce problème de quantification, on peut introduire le L_2-risque, ou l'erreur quadratique moyenne de f

$$\mathbb{E}\,|f(X) - Y|^2\,,$$

que l'on veut avoir la plus petite possible. Deux raisons justifient ce choix, d'abord, la simplicité des calculs, ensuite, la minimisation du L_2-risque aboutit à des estimateurs qui sont calculables rapidement. Donc, on cherche à trouver une fonction $m^* : \mathbb{R}^d \to \mathbb{R}$ telle que

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

Une telle fonction peut être obtenue explicitement comme étant la fonction de régression $m(x) = \mathbb{E}\,\{Y|X = x\}$. Pour ce faire, il suffit d'appliquer la règle de l'espérance itérée pour aboutir à la relation $\mathbb{E}\left|f(X)-Y\right|^2 = \int_{\mathbb{R}^d} |f(x)-m(x)|^2\mu(dx) + \mathbb{E}\left|m(X)-Y\right|^2$, μ étant la mesure de Lebesgue.

1.7.3 Prévision via la fonction de régression

Il existe une littérature abondante en estimation non paramétrique de la fonction de régression, basée sur la méthode du noyau. Parmi les principaux estimateurs, on en rappelle les trois suivants :

1. Nadaraya-Watson

$$\hat{f}_n(x) := \frac{\sum_{l=1}^n Y_l K_d\left(\frac{x-X_l}{h_n}\right)}{\sum_{l=1}^n K_d\left(\frac{x-X_l}{h_n}\right)}$$

2. Priestly-Chao

$$\hat{f}_n(x) := \sum_{l=1}^n Y_{(l)}\left(X_{(l)}-X_{(l-1)}\right) K_d\left(\frac{x-X_{(l)}}{h_n}\right)$$

3. Gasser-Müller

$$\hat{f}_n(x) := \sum_{l=1}^n Y_{(l)} \int_{s_{l-1}}^{s_l} K_d\left(\frac{u-x}{h_n}\right) du,$$

où $s_{l-1} = \left(X_{(l)}+X_{(l-1)}\right)/2$, $\left(Y_{(l)}, X_{(l)}\right)$ est la l^{eme} statistique d'ordre suivant l'ordre de la seconde composante, avec $X_{(0)} = -\infty$ et $X_{(n+1)} = +\infty$.
Etant donnée une nouvelle variable X_{n+1}, la valeur prédite de la variable scalaire correspondante est donnée par :

$$\hat{y} = \hat{f}_n(X_{n+1}).$$

1.8 Estimation du mode conditionnel

L'estimation du mode conditionnel se fait en deux étapes :

1. Estimation de la densité conditionnelle : L'estimateur à noyau de la densité conditionnelle et ses dérivées $f^{(p)}(./x)$, d'ordre $p; p \geq 0$, de la variable Y sachant $X = x$ est donnée par :

$$f_n^{(p)}(y/x) := \frac{\partial^p f_n}{\partial y^p} = \frac{\sum_{i=1}^n K_d\left(\frac{x-X_i}{h_n}\right) K_0^{(p)}\left(\frac{y-Y_i}{h_n}\right)}{h_n^{(1+p)} \sum_{i=1}^n K_d\left(\frac{x-X_i}{h_n}\right)}, \forall y \in \mathbb{R}.$$

On note $f_n^{(0)}(./x) := f_n(./x)$, l'estimateur de $f^{(0)}(./x) := f(./x)$.

2. Estimation du mode : L'estimateur à noyau du mode conditionnel est donné par la solution de la relation

$$\Theta_n(x) = \arg\max_{y\in\mathbb{R}} f_n(y/x)$$

1.8.1 Prévision via le mode conditionnel

Le mode conditionnel est une alternative au problème de prévision lorsque les autres modèles ne conviennent pas. Supposons que l'on a observé X_i ; $i = 1, 2, ..., n$ une suite de v.a, et pour chaque valeur X_i, correspond une valeur Y_i. Etant donnée une nouvelle variable X_{n+1}, la valeur prédite de la variable scalaire est donnée par :

$$\hat{y} = \arg\max_{y\in\mathbb{R}} f_n(y/X_{n+1}).$$

1.9 Modèles de dépendance

Jusqu'ici, nous avons toujours supposé l'indépendance des données. Dans la réalité, il arrive souvent que les données traitées (durées de vies,...) présentent une certaine forme de dépendance. Par exemple, Voelkel et Crowley (1984) établissent un modèle pour les recherches cliniques sur le cancer pour lequel ils supposent que chaque patient peut, soit rester dans un état initial, soit voir sa maladie progresser, soit répondre positivement au traitement avant de connaître une éventuelle rechute. C'est notamment le cas en ce qui concerne les virus qui se transmettent d'homme à homme. En effet, un individu côtoyant des personnes atteintes par un virus donné, aura plus de chances d'être infecté que s'il résidait dans une zone relativement épargnée par ce virus. Les données ne sont alors pas indépendantes mais on peut considérer qu'elles le sont asymptotiquement. Si l'échantillon est composé d'individus que l'on ordonne selon leur provenance géographique, les données pour deux individus proches dans l'échantillon sont dépendantes, mais si les deux individus sont assez éloignés l'un de l'autre, il y a indépendance. Mathématiquement ceci est modélisé par les données mélangeantes. Il existe plusieurs formes de mélanges, exprimées par des coefficients, notés, selon les cas, α, β, ϕ, ψ ou ρ. Ces coefficients expriment l'indépendance des données selon leurs limites à l'infini (0 pour certains et 1 pour d'autres, du moins pour les coefficients qu'on présentera dans cette partie).
Parmi tous ces types de mélanges, l'$\alpha-$mélange est le plus faible et donc le moins restrictif, autrement dit, toute suite de variables aléatoires β, ϕ, ψ ou ρ-mélangeante sera donc forcément $\alpha-$mélangeante. Réciproquement, tout résultat énoncé pour des données

α-mélangeantes sera valable pour des données soumises à une autre forme de mélange. Les données mélangeantes ne sont pas rares, en particulier les processus ARMA, qui sont largement utilisés dans l'analyse des séries temporelles, sont $\alpha-$mélangeants avec un coefficient à décroissance exponentielle. Auestad et Tjøstheim (1990) fournissent des discussions intéressantes sur le rôle de l'$\alpha-$mélange pour l'identification du modèle dans le cadre de l'analyse des séries temporelles non linéaires. Le champ d'application de telles données est très large et il est important de pouvoir faire de l'estimation dans ce cadre de données dépendantes.
Dans cette thèse (chapitre 3), et en se basant sur les travaux de Cai (2001), Cai et Roussas (1992) et Sun et Zhou (2001), nous établissons un résultat très intéréssant pour l'estimateur de la probabilité de troncature, du type loi du logarithme itéré. Ce résultat permettra d'établir, la consistance forte de l'estimateur du mode simple, et sera sans doute utile pour d'autres perspectives.

1.9.1 Mesures de dépendance

1.9.1.1 Définitions

Dans toute la suite de cette partie, $(\Omega, \mathcal{F}, \mathbb{P})$ désigne l'espace de probabilité. Pour toute $\sigma-$algèbre $\mathcal{A} \subset \mathcal{F}$, soit $\mathcal{L}^2_{reel}(\mathcal{A})$ l'espace (classes d'équivalence) des variables aléatoires réèlles, $\mathcal{A}-$mesurables et de carré-intégrables.

Pour deux $\sigma-$algèbre $\mathcal{A}$ et $\mathcal{B} \subset \mathcal{F}$, on définit les huit mesures de dépendance suivantes :

1. $\alpha(\mathcal{A}, \mathcal{B}) := \sup\{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|; A \in \mathcal{A}, B \in \mathcal{B}\}$; (Rosenblatt(1956))
2. $\phi(\mathcal{A}, \mathcal{B}) := \sup\{|\mathbb{P}(B|A) - \mathbb{P}(B)|; A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A) > 0\}$; (Ibragimov(1959))
3. $\psi(\mathcal{A}, \mathcal{B}) := \sup\left\{\left|\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)\mathbb{P}(B)} - 1\right|; A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A) > 0, \mathbb{P}(B) > 0\right\}$; (Blum et al (1963))
4. $\rho(\mathcal{A}, \mathcal{B}) := \sup\{|corr(f, g)|; f \in \mathcal{L}^2_{reel}(\mathcal{A}), g \in \mathcal{L}^2_{reel}(\mathcal{B})\}$; (Kolmogorov et Rozanov(1960))
5. $\beta(\mathcal{A}, \mathcal{B}) := \sup\left\{\frac{1}{2}\sum_{i\in I}\sum_{j\in J}|\mathbb{P}(A_i \cap B_j) - \mathbb{P}(A_i)\mathbb{P}(B_j)|\right\}$, où le sup est pris sur l'ensemble des couples $(A_1, A_2, ..., A_I) \times (B_1, B_2, ..., B_J)$ tels que $A_i \in \mathcal{A}$ pour tout i et $B_j \in \mathcal{B}$ pour tout j ; (Volkonskii et Rozanov(1959))
6. $\psi^*(\mathcal{A}, \mathcal{B}) := \sup\left\{\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)\mathbb{P}(B)}; A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A) > 0, \mathbb{P}(B) > 0\right\}$; (Origine difficile à tracer)
7. $\psi'(\mathcal{A}, \mathcal{B}) := \inf\left\{\frac{\mathbb{P}(A\cap B)}{\mathbb{P}(A)\mathbb{P}(B)}; A \in \mathcal{A}, B \in \mathcal{B}, \mathbb{P}(A) > 0, \mathbb{P}(B) > 0\right\}$; (Origine difficile à tracer)

8. $I(\mathcal{A},\mathcal{B}) := \sup\left\{\sum_{i\in I}\sum_{j\in J}\mathbb{P}(A_i\cap B_j)\log\left(\frac{\mathbb{P}(A_i\cap B_j)}{\mathbb{P}(A_i)\mathbb{P}(B_j)}\right)\right\}$ (Introduit par Volkonskii et Rozanov(1961) et attribué à Pinsker(1964) !)

les notations sont les mêmes que pour la mesure β et $0/0 := 0$ ainsi que $0\log 0 = 0$.

1.9.1.2 Propriétés de base

P1. Les valeurs possibles des différents coefficients (en incluant la valeur ∞ dans certains cas), sont comme suit :

1. $0 \leq \alpha(\mathcal{A},\mathcal{B}) \leq 1/4$,
2. $0 \leq \phi(\mathcal{A},\mathcal{B}) \leq 1$,
3. $0 \leq \psi(\mathcal{A},\mathcal{B}) \leq \infty$,
4. $0 \leq \rho(\mathcal{A},\mathcal{B}) \leq 1$,
5. $0 \leq \beta(\mathcal{A},\mathcal{B}) \leq 1$,
6. $0 \leq \psi^*(\mathcal{A},\mathcal{B}) \leq \infty$,
7. $0 \leq \psi^{'}(\mathcal{A},\mathcal{B}) \leq 1$,
8. $0 \leq I(\mathcal{A},\mathcal{B}) \leq \infty$.

P2. Chacune des égalités suivantes est équivalente à l'indépendance des tribus $\mathcal{A}$ et $\mathcal{B}$. $\alpha(\mathcal{A},\mathcal{B}) = 0$, $\phi(\mathcal{A},\mathcal{B}) = 0$, $\psi(\mathcal{A},\mathcal{B}) = 0$, $\rho(\mathcal{A},\mathcal{B}) = 0$, $\beta(\mathcal{A},\mathcal{B}) = 0$, $\psi^*(\mathcal{A},\mathcal{B}) = 1$, $\psi^{'}(\mathcal{A},\mathcal{B}) = 1$, $I(\mathcal{A},\mathcal{B}) = 0$.

P3. Les différents coefficients satisfont les inégalités suivantes :

1. $2\alpha(\mathcal{A},\mathcal{B}) \leq \beta(\mathcal{A},\mathcal{B}) \leq \phi(\mathcal{A},\mathcal{B}) \leq \frac{1}{2}\psi(\mathcal{A},\mathcal{B})$;
2. $4\alpha(\mathcal{A},\mathcal{B}) \leq \rho(\mathcal{A},\mathcal{B}) \leq \psi(\mathcal{A},\mathcal{B})$;
3. $\rho(\mathcal{A},\mathcal{B}) \leq 2\left[\phi(\mathcal{A},\mathcal{B})\right]^{1/2}\left[\phi(\mathcal{B},\mathcal{A})\right]^{1/2} \leq 2\left[\phi(\mathcal{A},\mathcal{B})\right]^{1/2}$;
4. $\phi(\mathcal{A},\mathcal{B}) \leq 1 - \frac{1}{\psi^*(\mathcal{A},\mathcal{B})} \leq \psi^*(\mathcal{A},\mathcal{B}) - 1$;
5. $\phi(\mathcal{A},\mathcal{B}) \leq 1 - \psi^{'}(\mathcal{A},\mathcal{B})$;
6. $\psi(\mathcal{A},\mathcal{B}) = \max\left\{\psi^*(\mathcal{A},\mathcal{B}) - 1, 1 - \psi^{'}(\mathcal{A},\mathcal{B})\right\}$;
7. $I(\mathcal{A},\mathcal{B}) \leq \psi^*(\mathcal{A},\mathcal{B})\log\psi^*(\mathcal{A},\mathcal{B})$;
8. $\beta(\mathcal{A},\mathcal{B}) \leq \left[I(\mathcal{A},\mathcal{B})\right]^{1/2}$.

1.9.2 Conditions de mélangeance forte

1.9.2.1 Conditions de mélangeance forte basée sur le passé et le futur

Soit $Y := (Y_k, k \in \mathbb{Z})$ (pas nécessairement stationnaire), une suite de variables aléatoires. Pour $-\infty \leq J \leq L \leq \infty$, on définit la $\sigma-$algèbre $\mathcal{F}_J^L := \sigma\left(Y_k, J \leq k \leq L (k \in \mathbb{Z})\right)$.
Pour tout $n \geq 1$, définissons les coefficients de dépendance suivants :

1. $\alpha(n) := \sup\limits_{j\in\mathbb{Z}} \alpha\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
2. $\phi(n) := \sup\limits_{j\in\mathbb{Z}} \phi\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
3. $\psi(n) := \sup\limits_{j\in\mathbb{Z}} \psi\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
4. $\rho(n) := \sup\limits_{j\in\mathbb{Z}} \rho\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
5. $\beta(n) := \sup\limits_{j\in\mathbb{Z}} \beta\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
6. $\psi^*(n) := \sup\limits_{j\in\mathbb{Z}} \psi^*\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$;
7. $\psi^{'}(n) := \inf\limits_{j\in\mathbb{Z}} \psi^{'}\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$; et
8. $I(n) := \sup\limits_{j\in\mathbb{Z}} I\left(\mathcal{F}_{-\infty}^{j}, \mathcal{F}_{j+n}^{\infty}\right)$.

La variable aléatoire est dite :

1. "fortement mélangeante" ou $\alpha-$mélangeante si $\alpha(n) \to 0$ quand $n \to \infty$,
2. " $\phi-$mélangeante" si $\phi(n) \to 0$ quand $n \to \infty$,
3. "$\psi-$mélangeante" si $\psi(n) \to 0$ quand $n \to \infty$,
4. "$\rho-$mélangeante" si $\rho(n) \to 0$ quand $n \to \infty$,
5. "absolument régulière" ou $\beta-$mélangeante si $\beta(n) \to 0$ quand $n \to \infty$,
6. "ψ^*-mélangeante" si $\psi^*(n) \to 1$ quand $n \to \infty$,
7. "$\psi^{'}-$mélangeante" si $\psi^{'}(n) \to 1$ quand $n \to \infty$,
8. "information régulière" si $I(n) \to 0$ quand $n \to \infty$.

Remarque 1.2. *1. Il est utile de rappeler que les deux phrases en anglais "Strong-mixing condition" et "Strong-mixing conditions", qui ne se différencient que par le 's' du pluriel, alors que le sens est totalement différent.*
- La première phrase traduit le coefficient de l'$\alpha-$mélange ($\alpha(n) \to 0$).
- La deuxième phrase traduit les autres coefficients de mélange qui sont au moins aussi fort que l'$\alpha-$mélange (i.e qui impliquent l'$\alpha-$mélange).

2. *Aussi, comme on peut le constater, tous les coefficients sont symétriques à l'exception du coefficient ϕ, car si $\phi(\mathcal{A},\mathcal{B})$ est petit, ceci n'implique en aucun cas le même ordre de grandeur pour $\phi(\mathcal{B},\mathcal{A})$.*
3. *Dans le cas de la stricte stationnarité de Y, on a simplement,*
$$\alpha(n) := \alpha\left(\mathcal{F}_{-\infty}^{0}, \mathcal{F}_{n}^{\infty}\right)$$
et c'est aussi valable pour les sept autres modes de dépendance.
4. *La première inégalité dans (3), donnée plus haut concernant le coefficient ρ, a été établie par Peligrad (1992), qui s'est appuyée sur les arguments utilisés par Cogburn (1960) et Ibragimov (1962) pour établir la seconde inégalité. Indépendamment, Denker et Keller (1983) ont prouvé une inégalité similaire*
$$\rho(\mathcal{A},\mathcal{B}) \leq 2\max\left(\phi(\mathcal{A},\mathcal{B}), \phi(\mathcal{B},\mathcal{A})\right).$$

Pour plus de détails sur le sujet, on peut consulter par exemple, Doukhan (1994), Bosq (1998), Rio (2000), Bradley (2005) et Bertail et *al.* (2006).

Bibliographie

[1] Abraham, C., Biau, G., Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode. *ESAIM, Probab. Statist.* 8 :1−11.

[2] Abraham, C., Biau, G., Cadre, B.(2003). Simple estimation of the mode of a multivariate density. *Canadian Journal of Statistics* 31 :23−34.

[3] Asselin de Beauville, J. P. (1978). Estimation non paramétrique de la densité et du mode : Exemple de la distribution gamma. *Revue de Statistique Appliquée.* 26 :47−70

[4] Bertail, P., Doukhan, P., Soulier, P. (2006). *Dependence in Probability and Statistics.* Lectures Note in Statistics, Springer.

[5] Bickel, D.R., Frühwirth, R. (2006). On a fast, robust estimator of the mode : Comparaisons to other robust estimators with applications. *Computat. Statist. Data Anal.* 50 :3500−3530.

[6] Bickel, D.R.(2002). Robust estimators of the mode and skewness of continuous data. *Computat. Statist. Data Anal.* 38 :153−163.

[7] Blum, J.R., Hanson, D.L., Koopmans, L.H.(1963). On the strong law of large numbers for a class of stochastic processes. *Z. Wahrsch. verw. Gebiete* 2 :1−11.

[8] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction* (Second Edition). Lecture Notes in Statistics 110, Springer−Verlag, Berlin.

[9] Bradley, R.C. (2005). Introduction to Strong Mixing Conditions, Volume 3. Technical Report, Department of Mathematics, Indiana University, Bloomington. *Custom Publishing of I.U.*, Bloomington.

[10] Cai, Z. (2001). Estimating a Distribution Function for Censored Time Series Data, *J. Multivariate Anal.* 78 :299−318.

[11] Cai, Z., Roussas, G. G. (1992). Uniform strong estimation under α-mixing, with rates, *Statist. Probab. Lett.* 15 :47−55.

[12] Chernoff, H. (1964). Estimation of the mode. *Ann. Instit. Statist. Math.* 16 :31−41.

[13] Cogburn, R. (1960). Asymptotic properties of stationary sequences. *Univ. Calif. Publ. Statist.* 3 :99−146.

[14] Collomb ,G., Härdle, W. and Hassani, S. (1987). A note on prediction via estimation of the conditional mode function. *J. Statist. Plann. Inference.* 15 :227−236.

[15] DasGupta, A. (2008). *Asymptotic Theory of Statistics and Probability.* Springer-Verlag.

[16] Denker, M., Keller, G. (1983). On U-statistics and v. Mises statistics for weakly dependent processes. *Z. Wahrsch. verw. Gebiete* 64 :505−522.

[17] Doukhan, P. (1994). *Mixing : Properties and Examples.* Lecture Notes in Statistics, Springer−Verlag.

[18] Eddy, W.F. (1982). The asymptotic distributions of kernel estimators of the mode. *Z. Wahrsh. Verw. Gebiete.* 59 :279−290.

[19] Eddy, W.F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* 8 :870−882.

[20] Fréchet, M. (1947). The General Relation Between the Mean and the Mode for a Discontinuous Variate.*Ann. Math. Statist.* 18 :290−293.

[21] Gannoun, A., Saracco, J. (2002). A new proof of strong consistency of kernel estimation of density function and mode under random censorship. *Statist. Probab. Lett.* 59 :61−66.

[22] Good, I. J., Gaskins, R. A. (1980). Density estimation and bump-hunting by the penalized likelihood method exemplifed by scattering and meteorite data (with discussion). *J. Amer. Statist. Assoc.* 75 :42−73.

[23] Grenander, U. (1965). Some direct estimates of the mode. *Ann. Math. Stat.* 131−138.

[24] Groeneboom, P. and Wellner, J. (2001). Computing Chernoff's distribution. *J. Comput. Graph. Stat.* 10 :388−400.

[25] Hall, P. (1980). On the limiting behaviour of the mode and median of a sum of independent random variables. *Ann. Statist.* 8 :419−430

[26] Hall, P., Minotte, M. C., Zhang, C. (2004). Bump hunting with non-gaussian kernels. *Ann. Statist.* 13 :70−84.

[27] Hartigan, J. A., Hartigan, P. M. (1985). The DIP test of unimodality. *Ann. Statist.* 32 :2124−2141.

[28] Ibragimov, I.A. (1962). Some limit theorems for stationary processes. *Theor. Probab. Appl.* 7 :349−382.

[29] Ibragimov, I.A. (1959). Some limit theorems for stochastic processes stationary in the strict sense. *Dokl. Akad. Nauk SSSR* 125 :711−714.

[30] Ibragimov, I.A., Rozanov, Yu.A. (1978). *Gaussian Random Processes.* Springer-Verlag, New York.

[31] Jiang, R., Prabhakar Murthyd, D. N., Xie, M. (2004). *Weibull models.* John Wiley, Sons, INC.

[32] Kolmogorov, A.N., Rozanov, Yu.A.(1960). On strong mixing conditions for stationary Gaussian processes. *Theor. Probab. Appl.* 5 :204−208.

[33] Lemdani, M. (1995). *Tests dans le cas d'un mélange de lois dans les modèles paramétriques et non paramétriques.* PhD. Thesis, Ecole Polytechnique.

[34] Louani, D. (1998). On the asymptotic normality of the kernel estimators of the density function and its derivatives under random censoring. *Comm. Statist. Theory* & *Meth.* 27 :2909−2924.

[35] Lynden-Bell, D.(1971). A method of allowing for known observational selection in small samples applied to 3CR quasars. *Monthly Notices Roy. Astronom. Soc.* 155 :95−118.

[36] Mammen, E., Marron, J. S., Fisher, N. I. (1992). Some asymptotics for multimodality tests based on kernel density estimates. *Probab. Theory related fields.* 91 :115−132.

[37] Matzner-Lϕber, E. (1997). *Prévision nonparamétrique des processus stochastiques.* Thèse de Doctorat, Univ. Montpellier II.

[38] Minotte, M. C. (1997). Nonparametric testing of the existence of modes. *Ann. Statist.* 25 :21241646−1660.

[39] Müller, D. W., Sawitski, G. (1991). Excess mass estimates and tests for multimodality. *J. Amer. Statist. Assoc.* 86 :738−746

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

[41] Ould-Saïd, E. (1993). Estimation non paramétrique du mode conditionnel. Application à la prévision. *C. R. Acad. Sci. Paris. Ser. I* 316 :943−947.

[42] Ould-Saïd, E., Cai, Z. (2005). Strong uniform consistency of the nonparametric estimation of the censored conditional mode function, *Nonparametric Statist. J.* 17 :797−806.

[43] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33 :1065−1076.

[44] Peligrad, M.(1992). On the central limit theorem for weakly dependent sequences with a decomposed strong mixing coefficient.*Stochastic Process. Appl.* 42 :181-193.

[45] Peligrad, M. (1983). A note on two measures of dependence and mixing sequences. *Adv. Appl. Probab.* 15 :461−464.

[46] Philipp, W.(1969). The central limit problem for mixing sequences of random variables. *Z. Wahrsch. verw. Gebiete* 12 :155−171.

[47] Pinsker, M.S. (1964). *Information and Information Stability of Random Variables and Processes.* Holden-Day, San Francisco.

[48] Quintela-Del-Río, A, Vieu, P. (1997). A nonparametric conditional mode estimate. *J. Nonparametric Statist.* 8 :253−266.

[49] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants* (In French). Mathématiques et Applications 31, Springer-Verlag.

[50] Rosa, A. C. (1993). *Prévision robuste sous une hypothèse ergodique.* Thèse de Doctorat de l'Université de Toulouse I.

[51] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci. U.S.A.* 42 :43−47.

[52] Rossi, V. (2004). *Filtrage non linéaire par noyaux de convolution Application à un procédé de dépollution biologique.* Thèse de Doctorat, E.N.S.A.M.

[53] Sager, T. (1975). Consistency in nonparametric estimation of the mode. *Ann. Stat.* 3 :698−706.

[54] Samanta, M., and Thavaneswaran, A.(1990). Nonparametric estimation of the conditional mode. *Comm. Statist. Theory and Methods.* 16 :4515−4524.

[55] Schoenberg, I. J. (1950). On Pólya frequency functions. II. Variation-diminishing integral operators of the convolution type. *Acta Sci. Math.(Szeged)* 12 :97−106

[56] Scott, D.W. (1992). Multivariate Density Estimation : *Theory, Practice and Visualization.* John Wiley, New York.

[57] Silverman, B. W. (1983). Some properties of a test for multimodality based on kernel density estimates. In*Probability, Statistics and Analysis* (J. F. C. Kingman and G. E. H. Reuter, eds.) 248−259. Cambridge Univ. Press.

[58] Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *J. Roy. Statist. Soc. Ser.B* 43 :97−99.

[59] Sun, L., Zhou, X. (2001). Survival function and density estimation for truncated dependent data. *Statist. Probab. Lett.* 52 :47−57.

[60] Van Ryzin, J. (1969). On strong consistency of density estimates. *Ann. Math. Statist.* 40 :1765−1772.

[61] Venter, J. (1967). On estimation of the mode. *Ann. Math. Stat.* 38 :1446−1455.

[62] Volkonskii, V. A., Rozanov, Yu. A. (1959). Some limit theorems for random functions II. *Theor. Probab. Appl.* 4 :178−197.

[63] Wegman, E. (1971). A note on the estimation of the mode. *Ann. Math. Stat.* 42 :1909−1915.

Chapitre 2

Mode simple : Cas i.i.d.

On the nonparametric estimation of the simple mode under random left-truncation model†

Ould-Saïd, E., Tatachak, A.

Abstract

Let $(Y_N)_{N\geq 1}$ be a sequence of independent and identically distributed random variables of interest. Let θ be the mode of Y, in this paper, we define a new kernel estimator $\hat{\theta}_n$ of θ when Y is subject to random left-truncation by a variable T. We establish the strong consistency with a rate for the proposed kernel estimate of the simple mode, and state its asymptotic normality. Simulations are drawn to illustrate the results and to show how the estimator behaves for finite samples.

Keywords Almost sure convergence, asymptotic normality, Lynden-Bell estimator, kernel estimator, left-truncation, mode, rate of convergence, V−C-class.

Mathematics Subject Classification. 62G05 ; 62G07 ; 62G20.

†To appear, in Romanian Journal of Pure and Applied Mathematics. vol 54, (2009).

2.1 Introduction

Let Y and T be independent random variables with distribution functions F and G respectively, both assumed to be continuous, and let $(Y_1,T_1),(Y_2,T_2),...,(Y_N,T_N)$ be N independent and identically distributed (iid) copies of (Y,T), where the sample size N is deterministic, but unknown. In the random left-truncation model (RLTM), the random variable (rv) of interest Y is interfered by the truncation rv T, such that both quantities Y and T are observable only if $Y \geq T$ whereas nothing is observed if $Y < T$. Without possible confusion, we still denote $(Y_i,T_i)_{1\leq i\leq n}$, $(n \leq N)$, the observed iid pairs from the original N-sample. As a consequence of truncation, the size of the actually observed sample, n is a $Bin(N,\mu)$ random variable, with $\mu := \mathbb{P}(Y \geq T)$.
It is clear that if $\mu = 0$, no data can be observed and therefore, we suppose throughout this paper that $\mu \neq 0$.
By the strong law of large numbers, we have, as $N \longrightarrow +\infty$

$$\tilde{\mu}_n := \frac{n}{N} \longrightarrow \mu, \ \mathbb{P}-\text{a.s.} \tag{2.1}$$

Note that the iid property of the observed n-sample is implicitly assumed in the literature in the left-truncation (LT) framework, but it is not completely straightforward. Lemdani and Ould-Saïd [25, Proposition 1] proved that it is deduced from the N-sequence iid property.
Truncation frequently occurs in medical studies, when one wants to study the length of survival after the start of the disease : if Y denotes the elapsed time between the onset of the disease and death, and if the follow-up period starts T units of time after the onset of the disease, then clearly Y is left truncated by T.
Truncated data are also common in actuarial, astronomical, demographics, epidemiological, reliability testing and other studies [More examples and references dealing with truncated data can be found Woodroofe [49], Wang *et al.* [48], Tsai *et al.* [45], Anderson *et al.* [3], He and Yang [20] and Chen *et al.* [6]].
Denote by $f(\cdot)$ the probability density function of Y and assume that it has a unique mode θ, defined by

$$f(\theta) = \sup_{y\in\mathbb{R}} f(y).$$

The problem of estimating the unconditional/conditional mode of a probability density has a long history in statistics literature, and number of distinguished papers deal with this topic. To quote a few of them, Parzen [38] established weak convergence and asymptotic normality for the iid case, while the strong consistency was obtained by Nadaraya [33]. Using classical approaches of weak convergence, Eddy [12] derived the asymptotic normality, under weaker conditions than Parzen's [38]. Chernoff [7] studied the naive estimator of the mode, defined as the center of that interval which contains most of the observations (see also Eddy [13]). Romano [39] investigated the asymptotic behavior of the kernel estimate of the mode with data-dependent bandwidths, whereas Vieu [47] obtained a rate of convergence for both local and global estimates of the mode.

The multidimensional versions of these results were obtained by Samanta [41] and recently, Abraham et al. [1, 2] considered the estimate of the mode as an element of the sample points, and obtained the strong consistency as well as asymptotic properties. Mokkadem and Pelletier [30, 31] and Mokkadem et al. [32] studied the law of the iterated logarithm and the moderate deviations, as well as the large deviations upper bound for the kernel mode estimator. Hermann and Ziegler [22] obtained rates of nonparametric estimation of the mode in absence of smoothness assumptions. Bickel [4] and Bickel and Frühwirth [5] proposed with applications, a robust estimators of the mode based on densest half ranges and compared them to other robust estimators of measure of location. Bickel [4] concluded that, while the median is resistent to outliers, the mode is immune to them ; also, it is a safer measure of location when the data may suffer from the latter.
When conditioning by one of the coordinates of a bi-dimensional random vector in the iid random vector case, Samanta and Thavaneswaran [42] showed that under some regularity conditions, the kernel estimator of the conditional mode function is consistent, and asymptotically normal. Mehra *et al.* [29] established the law of iterated logarithm, the uniform almost sure convergence over a compact set and the asymptotic normality of smoothed rank nearest neighbor estimator of the conditional mode function.
For the dependent case, the strong consistency of the conditional mode estimator was established under a ϕ-mixing condition by Collomb et al. [8] and their results can be applied to process forecasting. In the α-mixing case, the strong consistency over a compact set and the asymptotic normality were obtained by Ould-Saïd [34] and, Louani and Ould-Saïd [27], respectively. In a general ergodic framework, a process prediction via the conditional mode estimation was described by Ould-Saïd [35] and its strong consistency was obtained. Lately, Ferraty et al. [16] established under $\alpha-$mixing condition, the almost complete convergence of the conditional mode of a scalar response given a functional random variable.
In the incomplete data case, Louani [27] studied the asymptotic normality of the kernel estimator of the mode for the random right censoring and recently, Ould-Saïd and Cai [37] established, in the iid case, the uniform strong consistency with rates of a nonparametric estimator of the censored conditional mode function.
When the conditioning variable takes values in a semi-normed vectorial space of possibly infinite dimension, Dabo-Niang et al. [10] established the strong consistency of the simple mode, while Dabo-Niang and Laksaci [9] established the consistency in L^p norm of the conditional mode estimator. The asymptotic normality have been obtained by Ezzahrioui and Ould-Saïd [14, 15], in both iid and α-mixing cases.
Recent years have witnessed a renewal of interest in mode estimation. The main reason that enables widespread use of the mode is the recent advent of powerful mathematical tools, and computational machinery that render some complex problems much more tractable. An important application of mode estimation is pattern recognition.
To the best of our knowledge, the problem of estimating the simple mode under truncation has not been addressed in the statistics literature. It is the goal of this paper.

Our main results establish, the strong uniform consistency with a rate, over a compact set, of the kernel density estimator which allows to deduce the strong consistency with a rate of the kernel mode estimator and state its asymptotic normality.
The paper is organized as follows : in Section 2 we recall some important and useful results in the random left-truncation model. In Section 3 we first define a kernel mode estimator in RLTM, then we give assumptions and main results. The asymptotic normality of the suggested estimator is illustrated in section 4, for finite samples via a simulation study. Finally, the proofs of the main results are postponed to Section 5, where some auxiliary results are also proved.

2.2 Preliminaries and notations

We first present some results from the literature for the univariate RLTM and which will be used to define our nonparametric kernel estimator of the mode.
Since N is unknown and n is known (although random), in the rest of this paper, our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the $N-$sample) but will involve the probability $\mathbf{P}$ (related to the $n-$sample). In the same way $\mathbb{E}$ and $\mathbf{E}$ denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively. In the sequel, as a convention, we denote by a superscript $(^*)$ any characteristic of the actually observed data (that is, conditionally on n).
Let Y be the variable of interest and T be the truncating variable, with continuous distribution functions F and G, respectively. Under LT sampling scheme, the conditional joint distribution of an observed (Y,T) (Stute [44] and Zhou [50]), is given by

$$\begin{aligned} H^*(y,t) &= \mathbf{P}\{Y \leq y, T \leq t\} = \mathbb{P}\{Y \leq y, T \leq t \mid Y \geq T\} \\ &= \mu^{-1}\int_{-\infty}^{y} G(t \wedge u)dF(u), \end{aligned}$$

where $t \wedge u = \min(t,u)$. The marginal distributions are defined by

$$F^*(y) := H^*(y,\infty) = \mu^{-1}\int_{-\infty}^{y} G(u)dF(u), \tag{2.2}$$

$$G^*(t) := H^*(\infty,t) = \mu^{-1}\int_{-\infty}^{+\infty} G(t \wedge u)dF(u),$$

which are estimated by

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{Y_i \leq y\}} \quad \text{and} \quad G_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{T_i \leq t\}}$$

where $\mathbb{1}_A$ denotes the indicator function of the set A.
Let us denote for any distribution function W

$$a_W = \inf\{x : W(x) > 0\} \quad \text{and} \quad b_W = \sup\{x : W(x) < 1\},$$

as the endpoints of the W support. Woodroofe [49] pointed out that F and G can be estimated completely only if

$$a_G \leq a_F, \;\; b_G \leq b_F \;\text{ and } \int_{a_F}^{\infty} \frac{dF}{G} < \infty. \tag{2.3}$$

Then under (2.3)

$$\mu := P(Y \geq T) = \int G(u)dF(u) > 0.$$

Now, let $C(\cdot)$ be defined by

$$C(y) := G^*(y) - F^*(y) = \mu^{-1}G(y)\left(1 - F(y)\right), \; y \in [a_F, +\infty[\tag{2.4}$$

with empirical estimator

$$C_n(y) = G_n^*(y) - F_n^*(y^-) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{T_i \leq y \leq Y_i\}}, \; y \in [a_F, +\infty[\,.$$

Then the nonparametric maximum likelihood estimates (NPMLE) of F and G, originally proposed by Lynden-Bell [28], are given by

$$F_n(y) = 1 - \prod_{i:Y_i \leq y}\left[\frac{nC_n(Y_i) - 1}{nC_n(Y_i)}\right] \;\text{ and }\; G_n(t) = \prod_{i:T_i > t}\left[\frac{nC_n(T_i) - 1}{nC_n(T_i)}\right], \tag{2.5}$$

respectively, assuming no ties in the data.
Asymptotic properties of (2.5) have been also studied by Woodroofe [49, Theorem 2] and established the uniform consistency results

$$\sup_{y \geq a_F} |F_n(y) - F(y)| \overset{\mathbf{P-a.s}}{\longrightarrow} 0 \quad \text{and} \quad \sup_{t \geq a_G} |G_n(t) - G(t)| \overset{\mathbf{P-a.s}}{\longrightarrow} 0\,.$$

Additional results were obtained by Keiding and Gill [24].
Note that the estimator $\tilde{\mu}_n$ defined in (2.1) cannot be calculated (since N is unknown). This situation requires the use of another estimator.
Indeed, according to (2.4) and replacing F and G by their respective NPMLE, we can consider the estimator of μ, namely

$$\mu_n = \frac{G_n(y)[(1 - F_n(y^-)]}{C_n(y)}, \tag{2.6}$$

for all y such that $C_n(y) > 0$. He and Yang [21, Theorem 2.5], showed that μ_n is equivalent to the more familiar estimate $\hat{\mu}_n = \int G_n(u)dF_n(u)$ studied in the literature, and that

$$\mu_n \overset{\mathbf{P-a.s}}{\longrightarrow} \mu, \text{ as } n \longrightarrow \infty.$$

The authors proved that μ_n does not depend on y and its value can then be obtained for any y such that $C_n(y) \neq 0$.
The fact that μ_n has a much simpler form than $\hat{\mu}_n$, makes it a useful instrument in the construction of new estimates as to be carried out below.

2.3 Assumptions and main results

First, we define our nonparametric estimator of the mode. Let us note that, if no truncation is present ($n = N$), it is well known that the kernel estimator of the mode θ is defined as the random variable θ_n maximizing the kernel estimator $f_n(\cdot)$ of $f(\cdot)$; that is,

$$f_n(\theta_n) = \sup_{y\in\mathbb{R}} f_n(y),$$

where

$$f_n(y) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{y-Y_i}{h_n}\right), \tag{2.7}$$

K is a probability kernel defined on $\mathbb{R}$, and (h_n) a positive bandwidth sequence, which decays to zero as n grows to infinity.
In RLTM, (2.7) is no longer adapted as an estimator of $f(\cdot)$.
Now, We define a new estimator $\tilde{f}_n(\cdot)$, based on the n actually observed pairs (X_i, Y_i).
By differentiating in the statement (2.2) and integrating the result over y, we get

$$F(y) = \mu \int_{a_G}^{y} \frac{1}{G(u)} dF^*(u).$$

A natural estimator of F is then given by

$$\tilde{F}_n(y) = \frac{\mu}{n}\sum_{i=1}^{n} \frac{1}{G(Y_i)} 1\!\!1_{\{Y_i \le y\}}, \tag{2.8}$$

which gives us an estimate of the density f

$$\begin{aligned} \tilde{f}_n(y) &= \frac{1}{h_n}\int K\left(\frac{y-v}{h_n}\right) d\tilde{F}_n(v) \\ &= \frac{\mu}{nh_n}\sum_{i=1}^{n} \frac{1}{G(Y_i)} K\left(\frac{y-Y_i}{h_n}\right). \end{aligned} \tag{2.9}$$

However, both (2.8) and (2.9) are not useful in practice since $G(\cdot)$ and μ are unknown. To overcome this difficulty and to make them usable, we propose the estimator

$$\hat{F}_n(y) = \frac{\mu_n}{n}\sum_{i=1}^{n} \frac{1}{G_n(Y_i)} 1\!\!1_{\{Y_i \le y\}}. \tag{2.10}$$

Note that in (2.10) and in the sequel, the sum is taken only for the $i's$ such that $G_n(Y_i) \neq 0$.
Finally (2.10) yields the density estimator

$$\begin{aligned} \hat{f}_n(y) &= \frac{1}{h_n}\int K\left(\frac{y-v}{h_n}\right) d\hat{F}_n(v) \\ &= \frac{\mu_n}{nh_n}\sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K\left(\frac{y-Y_i}{h_n}\right). \end{aligned} \tag{2.11}$$

Note that both estimators (2.10) and (2.11) are given in Ould-Saïd and Lemdani [36], and that the rate of convergence of the latter estimator is improved in this paper.
Now, we define our nonparametric kernel estimator of the mode, by

$$\hat{f}_n(\hat{\theta}_n) = \sup_{y \geq a_F} \hat{f}_n(y). \tag{2.12}$$

Remark that the estimate $\hat{\theta}_n$ is not necessarily unique, and if this is the case, all the remaining of our paper will concern any value $\hat{\theta}_n$ satisfying (2.12). We point out that we can specify our choice by taking

$$\hat{\theta}_n = \inf \left\{ t \geq a_F \text{ such that } \quad \hat{f}_n(t) = \sup_{y \geq a_F} \hat{f}_n(y) \right\}.$$

Suppose that K is chosen to be twice differentiable, we get the derivatives of $\hat{f}_n(\cdot)$ as

$$\hat{f}_n^{(j)}(y) = \frac{\mu_n}{nh_n^{j+1}} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K^{(j)}\left(\frac{y - Y_i}{h_n}\right), \quad \text{for } j = 1, 2.$$

The derivatives $\tilde{f}_n^{(j)}(\cdot)$ of $\tilde{f}_n(\cdot)$ are obtained analogously.
Throughout this paper, we suppose that $a_G < a_F, \;\; b_G \leq b_F$ and let Ω be a compact set such that $\Omega \subset \Omega_0 = \{y : \; y \geq a_F\}$.
We will make use of the following assumptions gathered here together for easy reference.

(A1) The bandwidth h_n satisfies for $n \longrightarrow \infty$, $h_n \longrightarrow 0$ and $\dfrac{nh_n}{\log n} \longrightarrow \infty$;

(A2) The kernel K is compactly supported, a $\mathcal{C}^1$-probability density, three times differentiable and such that K, $K^{(1)}$ and $K^{(2)}$ are of bounded variations ;

(A3) K is a second-order kernel ;

(A4) The density $f(\cdot)$ is bounded and three times continuously differentiable in a neighborhood of the mode θ with $f^{(2)}(\theta) \neq 0$. Furthermore, we suppose that $\theta \in \mathring{\Omega}$, where $\mathring{\Omega}$ denotes the interior of Ω ;

(A5) The unique mode θ satisfies, for any $\varepsilon > 0$ and y, there exists a $\beta > 0$ such that $|\theta - y| \geq \varepsilon$ implies that $|f(\theta) - f(y)| \geq \beta$;

(A6) The bandwidth h_n satisfies for $n \longrightarrow \infty, \;\; h_n \longrightarrow 0$;
i) $\frac{nh_n^6}{\log n} \longrightarrow \infty$;
ii) $nh_n^7 \longrightarrow 0$.

Discussion on the assumptions
1. Assumption **A1** is needed in the proof of the convergence of $\hat{f}_n(\cdot)$. For the convergence of $\hat{f}_n^{(2)}(\cdot)$, we need more than **(A1)** that is $\frac{nh_n^6}{\log n} \longrightarrow \infty$. Assumptions **(A2)-(A4)** are common in nonparametric estimation. Assumption **(A5)** stipulates the uniform uniqueness of the mode point. Finally, Assumption **(A6)** is used to prove the asymptotic normality result as well as the

convergence of $\hat{f}_n^{(2)}(\cdot)$.
2. Here, we point out that in view of the choice of Ω_0, many authors (see Stute [44]) used the milder conditions $a_G \leq a_F$ with additional integrability conditions. This choice is motivated by the rate of convergence of Lynden-Bell [28] estimator $G_n(\cdot)$ to $G(\cdot)$ where we have to consider a set of values of Y_i which do not include a_G (a uniform rate for $G_n(\cdot)$ is given by Woodroofe [49], only on $[a, b_G]$ with $a > a_G$), that is $a_G < a_F$.
3. Recall that since K has bounded variations, its first derivative $K^{(1)}$ is integrable. Furthermore $\left(K^{(1)}\right)^2$ is also integrable, which ensures the existence of the asymptotic variance term.

Remark 2.1. *Assumption (A2) implies condition (K_1) (Giné and Guillou [17]) under which*

$$\mathcal{G} = \left\{ K\left(\frac{y - \cdot}{h}\right); y \in \mathbb{R}, \ h \in \mathbb{R}\backslash \{0\} \right\}$$

is a bounded V–C-class of measurable functions. This is a consequence of (Dudley [11, Theorems 4.2.1 and 4.2.4]). This assumption is needed in order to use Talagrand's inequality.

Now, our first result is the uniform almost sure convergence with a rate of the kernel density estimator $\hat{f}_n(\cdot)$, as stated in Theorem 2.1. An immediate consequence is the almost sure convergence with a rate of the kernel mode estimator and is given in Corollary 2.1.

Theorem 2.1. *Under Assumptions (A1)-(A4), we have*

$$\sup_{y\in\Omega}\left|\hat{f}_n(y) - f(y)\right| = O\left\{\max\left(\sqrt{\frac{\log n}{nh_n}}, h_n^2\right)\right\}, \quad \mathbf{P}-a.s. \ as \ n \to +\infty.$$

Remark 2.2. *If we choose $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$, which is the optimal choice with respect to the almost sure uniform convergence criterion in the density estimation, we get*

$$\sup_{y\in\Omega}\left|\hat{f}_n(y) - f(y)\right| = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right), \quad \mathbf{P}-a.s.$$

which is the same optimal rate as the one established in the complete and iid case (Vieu [47, Lemma A-2]).

Corollary 2.1. *Under (A1)-(A5), then*

$$\left|\hat{\theta}_n - \theta\right| \longrightarrow 0, \ a.s. \ \ as \ \ n \longrightarrow +\infty,$$

whenever $\theta \in \Omega$. In addition, if (A4) holds, for n large enough, we have

$$\hat{\theta}_n - \theta = O\left(\max\left\{\left(\frac{\log n}{nh_n}\right)^{1/4}, h_n\right\}\right), \ \mathbf{P}-a.s. \ as \ n \to +\infty.$$

Now, suppose that the density function $f(\cdot)$ has a unique mode at θ. Under **(A4)**, we have

$$f^{(1)}(\theta) = 0 \text{ and } f^{(2)}(\theta) < 0.$$

Similarly, under (A2), with probability one, we have

$$\hat{f}_n^{(1)}(\hat{\theta}_n) = 0 \text{ and } \hat{f}_n^{(2)}(\hat{\theta}_n) < 0,$$

if $\hat{\theta}_n$ is the mode of $\hat{f}_n(\cdot)$.
A Taylor expansion of $\hat{f}_n^{(1)}(\cdot)$ in the neighborhood of θ, gives

$$0 = \hat{f}_n^{(1)}(\hat{\theta}_n) = \hat{f}_n^{(1)}(\theta) + \left(\hat{\theta}_n - \theta\right) \hat{f}_n^{(2)}(\theta_n^*),$$

where θ_n^* is between $\hat{\theta}_n$ and θ, which gives that

$$\hat{\theta}_n - \theta = -\frac{\hat{f}_n^{(1)}(\theta)}{\hat{f}_n^{(2)}(\theta_n^*)},$$

if the denominator does not vanish.
To state the asymptotic normality, we show that the numerator suitably normalized, is asymptotically normally distributed, and the denominator converges in probability to $f^{(2)}(\theta)$. The result is given in the following

Theorem 2.2. *Under Assumptions (A1)-(A4) and (A6), we have*

$$\left(\frac{nh_n^3 \left(f^{(2)}(\theta)\right)^2}{\sigma^2(\theta)}\right)^{\frac{1}{2}} \left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$$

where $\xrightarrow{\mathcal{D}}$ *denotes the convergence in distribution and* $\sigma^2(y) = \dfrac{\mu f(y)}{G(y)} \displaystyle\int \left(K^{(1)}(v)\right)^2 dv.$

Remark 2.3. *Using Theorem 2.2, it is possible to construct confidence intervals for θ. For that purpose, a plug-in estimate $\sigma_n^2(y) := \dfrac{\mu_n \hat{f}_n(y)}{G_n(y)} \displaystyle\int \left(K^{(1)}(v)\right)^2 dv$ for the asymptotic variance $\sigma^2(y)$ can be easily obtained using the estimators defined above ((2.6), (2.5) and (2.11)) which is a consistent estimator. This yields a confidence interval of asymptotic level $1-\zeta$ for θ.*

$$\left[\hat{\theta}_n - \frac{\sigma_n(\hat{\theta}_n)}{\sqrt{nh_n^3}\left|\hat{f}_n^{(2)}(\hat{\theta}_n)\right|} \times u_{1-\zeta/2}\ ,\ \hat{\theta}_n + \frac{\sigma_n(\hat{\theta}_n)}{\sqrt{nh_n^3}\left|\hat{f}_n^{(2)}(\hat{\theta}_n)\right|} \times u_{1-\zeta/2}\right]$$

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

2.4 Simulation study

The purpose of this section deals with asymptotic normality study. In situations involving asymptotic normality, it is important to know how good is this normality for finite sample.
First we present two simulated models which permitted to compute the estimator $\hat{\theta}_n$ defined in (3.12), and for each model, we simulated N iid rv (Y_i, T_i). Here it is assumed that the truncating variable T is distributed as an exponential with parameter λ. This parameter was adapted in order to obtain different values of the theoretical proprtion α as fixed bellow. We then kept the data (Y_i, T_i) such that $Y_i \geq T_i$.
Using this scheme, m independent samples of size n were generated. For each sample, using plug-in estimates $\sigma_n(\cdot)$ and $\hat{f}_n^{(2)}(\cdot)$ for $\sigma(\cdot)$ and $f^{(2)}(\cdot)$, respectively, we estimated the mode and computed the normalized deviation, $\mathring{\theta}_n := \left(\left(\sigma_n^2(\hat{\theta}_n)\right)^{-1} nh_n^3 \left(\hat{f}_n^{(2)}(\hat{\theta}_n)\right)^2\right)^{\frac{1}{2}} \left(\hat{\theta}_n - \theta\right)$. We estimated then, using the kernel method, the density function of the process $\left(\mathring{\theta}_n\right)$.

1. Model 1 (heavy tail case). The variable Y is distributed as a three-parameter Weibull with pdf,
$$f_Y(y) = \begin{cases} \frac{c}{b^c}(y-a)^{c-1}\exp(-(\frac{y-a}{b})^c) & y \geq a > 0 \\ 0 & y < a \end{cases}$$
which admits a mode θ at the point $a + b\left(\frac{c-1}{c}\right)$
2. Model 2 (exponential decreasing case). The variable Y is distributed as a normal $\mathcal{N}(3,1)$, with mode $\theta = 3$.

Recall 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. In this study, the bandwidth h_n is chosen to satisfy the assumptions above, and the kernel K is Gaussian. In estimating the value $\mathring{\theta}_n$, we choose $h_n \approx \hat{\sigma}_Y n^{-\delta}$, $\delta \in]\frac{1}{7}, \frac{1}{6}[$ and $\hat{\sigma}_Y$ is a robust estimate of the standart deviation parameter of Y. The bandwidth that we used in estimating the $(\mathring{\theta}_n)$-density, is the classical bandwidth choice (Silverman [43, p. 40]) which is $h_m^{opt} = 1.06\hat{\sigma}_Y m^{-1/5}$.
The normalized estimation biases under model 1, estimated by the mean values of the processes, corresponding to the different figures are given in Table 2.1.

n \ α	60%	75%	90%
50	.39	.40	.37
100	.26	.26	.25
500	-	.05	-

TAB. 2.1 – The normalized estimation biases.

It is clear that in any cases, except for $n = 500$, there is a substantial positive biases. We observe

n	α	$Bias$	$\hat{\sigma}_n^2$	MSE
50	0.45	−0.0100	0.0533	0.05 34
	0.60	0.027 7	0.0511	0.05 88
	0.75	−0.028	0.0452	0.0530
150	0.45	0.014 9	0.0245	0.02 67
	0.60	−0.005 1	0.0214	0.021 7
	0.75	−0.016 1	0.0285	0.02 87
300	0.45	−0.003 0	0.0159	0.0160
	0.60	0.009 2	0.0174	0.01 83
	0.75	0.016 3	0.0177	0.0204
500	0.75	−0.008 1	0.0087	0.0094

TAB. 2.2 – Average estimated Bias and the MSE, 100 replications.

that these biases decrease when the sample size increases but remained still relatively important for moderate sizes. In the case where $n = 500$, we give the histogram and the corresponding $Q - Q$-plot against the standard Gaussian distribution. Moreover, a Shapiro-Wilk normality test and the Kolmogorov-Smirnov (with Lilliefors correction) test, are used. The Kolmogorov-Smirnov significance level is greater than 0.05, and the Shapiro-Wilk test suggests a small disparity from normality.

Now, under Model 2, we compute the estimated bias of the estimator as well as the variance $\hat{\sigma}_n^2 := m^{-1} \sum_{k=1}^{m} \left(\hat{\theta}_n^k - \hat{\theta}_n\right)^2$, which give the MSE (Table 2.2).

As one can see it in Table 2.2, the quality of the estimator does not seem to be affected by α, and the MSE decreases when the sample size increases. For $n = 500$ and since the results do not depend on α, we studied only the case $\alpha \approx 75\%$).

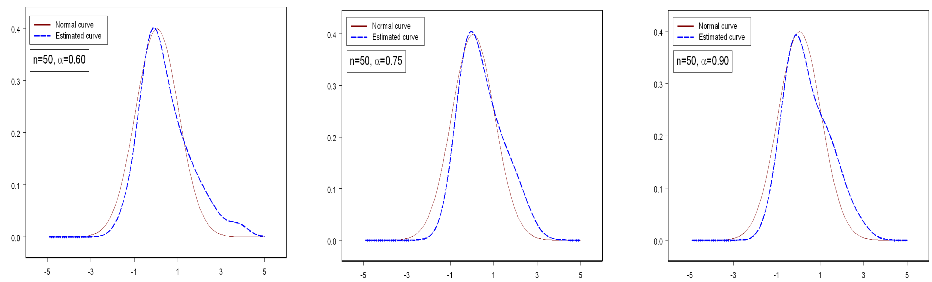


Figure 1: (Model 1);$n = 50$, $m = 100$, $\alpha =.60$, .75 and .90 respectively.

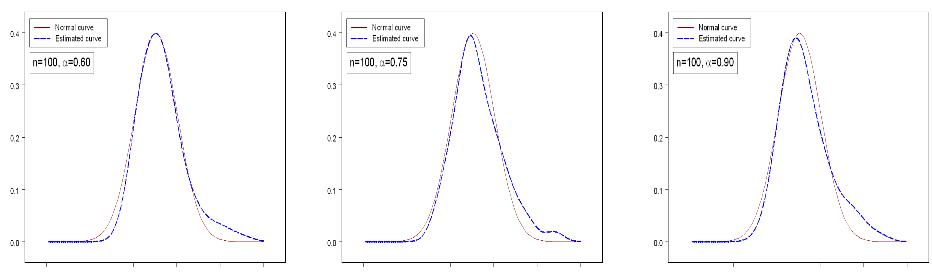


Figure 2: (Model 1);$n = 100$, $m = 100$, $\alpha =.60$, .75 and .90 respectively.

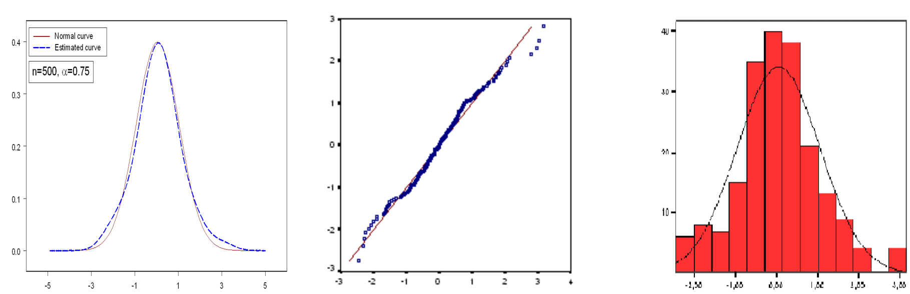


Figure 3: (Model 1);$n = 500$, $m = 200$ and $\alpha =.75$

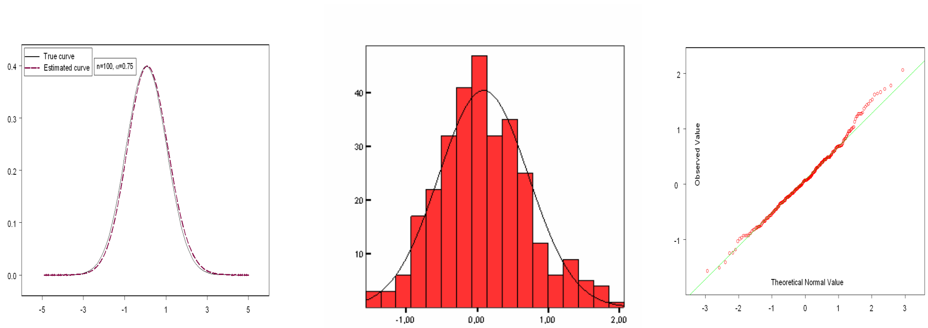


Figure 4: (Model 2);$n = 100$ and $m = 300$.

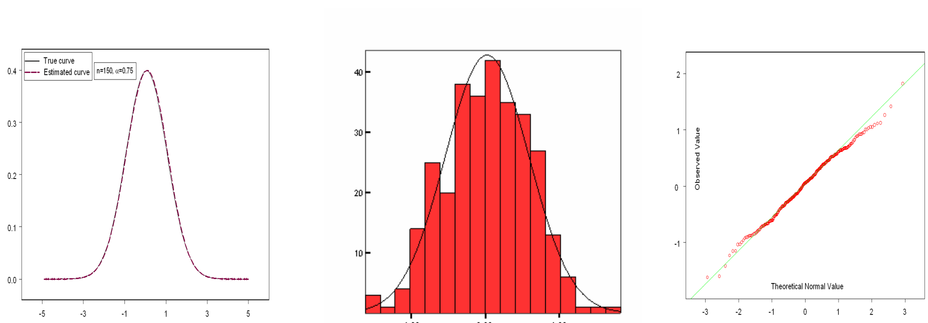


Figure 5: (Model 2);$n = 150$ and $m = 300$.

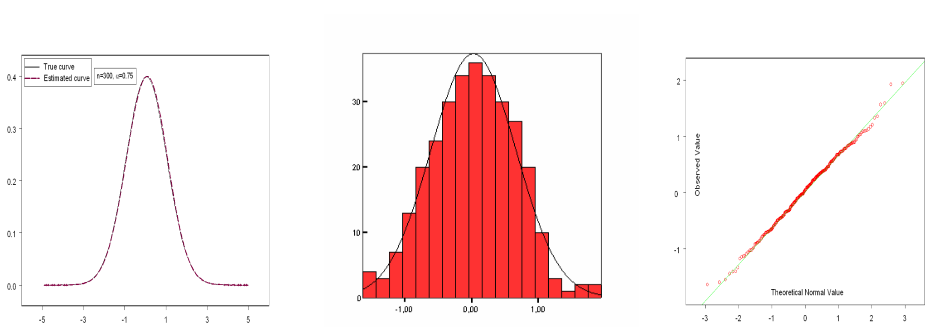


Figure 6: (Model 2);$n = 300$ and $m = 300$.

2.5 Proofs

In order to make the proofs easier, we need some auxiliary results and notations. The first Lemma gives the uniform consistency with rate of the estimator $\tilde{f}_n(y)$ defined in (2.9).

Lemma 2.1. *Under Assumptions (A1), (A2) and (A4), we have*

$$\sup_{y\in\Omega}\left|\tilde{f}_n(y)-\mathbf{E}\left[\tilde{f}_n(y)\right]\right| = O\left(\sqrt{\frac{\log n}{nh_n}}\right),\ \mathbf{P}-a.s.\ as\ n\to+\infty.$$

Proof. Under (A2), the class of functions

$$\mathcal{F}_n=\left\{\psi_y(\cdot)=\frac{\mu}{nh_n}K\left(\frac{y-\cdot}{h_n}\right);y\in\mathbb{R}\right\}$$

is a bounded V$-$C-class of measurable functions (see Remark 2.1) which are uniformly bounded with envelope $\mathbf{\Delta_1}=\frac{\|K\|_\infty}{nh_n}$. Then using Lemma 2.6.18 (vi) (Van der Vaart and Wellner [46, p. 147]) with $g(y)=1/G(y)$, the class of functions

$$\mathcal{H}_n=\left\{\Xi_y(u)=\frac{\mu}{nh_nG(y)}K\left(\frac{y-u}{h_n}\right);\ y\geq a_F\right\}$$

is a bounded V$-$C-class of measurable functions with respect to the envelope $\mathbf{\Delta_2}=\frac{\|K\|_\infty}{nh_nG(a_F)}$. Moreover

$$\mathbf{E}\left[\Xi_y(Y)\right]\leq\frac{\|f\|_\infty}{n}=:U_n\quad\text{and}\quad\mathbf{E}\left[\Xi^2_y(Y)\right]\leq\frac{\|f\|_\infty}{n^2h_nG(a_F)}=:\sigma_n^2$$

Applying Talagrand's inequality (Giné and Guillou [18, Proposition 2.2]) with $t=D\sqrt{\frac{\log n}{nh_n}}$, where D is a positive constant, there exist two positive constants c_1 and c_2 such that

$$\begin{aligned}
&\mathbf{P}\left\{\sup_{\Xi_y\in\mathcal{H}_n}\left|\sum_{i=1}^n\left\{\Xi_y(Y_i)-\mathbf{E}[\Xi_y(Y)]\right\}\right|\geq D\sqrt{\frac{\log n}{nh_n}}\right\}\\
\leq\ & c_1\exp\left\{-\frac{D\sqrt{\frac{n\log n}{h_n}}}{c_1\|f\|_\infty}\log\left[1+\frac{\frac{D\|f\|_\infty}{n}\sqrt{\frac{\log n}{nh_n}}}{c_1\left(\sqrt{\frac{\|f\|_\infty}{nh_nG(a_F)}}+\frac{\|f\|_\infty}{n}\sqrt{\log\left(c_2\sqrt{\|f\|_\infty h_nG(a_F)}\right)}\right)^2}\right]\right\}\\
\leq\ & c_1\exp\left\{-\frac{D\sqrt{\frac{n\log n}{h_n}}}{c_1\|f\|_\infty}\log\left[1+\frac{\frac{nh_nG(a_F)}{c_1}\frac{D}{n}\sqrt{\frac{\log n}{nh_n}}}{\left(1+\sqrt{\frac{1}{n}\|f\|_\infty G(a_F)h_n\log\left(c_2\sqrt{\|f\|_\infty G(a_F)h_n}\right)}\right)^2}\right]\right\}\\
\leq\ & c_1\exp\left\{-\frac{D}{c_1\|f\|_\infty}\sqrt{\frac{n\log n}{h_n}}\frac{G(a_F)}{c_1}D\sqrt{\frac{h_n\log n}{n}}\right\}\\
\leq\ & c_1 n^{-\frac{D^2G(a_F)}{c_1^2\|f\|_\infty}}
\end{aligned}$$

which, for n large enough and by an appropriate choice of D, can be made $O\left(n^{-3/2}\right)$. The latter being a general term of a summable series, we then have by Borel-Cantelli's Lemma and (A1), $\mathcal{I}_2 = O\left(\sqrt{\frac{\log n}{nh_n}}\right)$, $\mathbf{P} - a.s.$ ■

Lemma 2.2. *Under Assumption (A4), we have*

$$\sup_{y\in\Omega}\left|\hat{f}_n(y) - \tilde{f}_n(y)\right| = O_{\mathbf{P}}\left(n^{-1/2}\right) \quad a.s. \quad as\ n \to +\infty.$$

Proof. We observe that one can have the following decomposition

$$\begin{aligned}
\left|\hat{f}_n(y) - \tilde{f}_n(y)\right| &\leq \frac{|\mu_n - \mu|}{nh_n}\sum_{i=1}^{n}\frac{1}{G_n(Y_i)}K\left(\frac{y - Y_i}{h_n}\right) + \frac{\mu}{nh_n}\sum_{i=1}^{n}\left|\frac{1}{G_n(Y_i)} - \frac{1}{G(Y_i)}\right|K\left(\frac{y - Y_i}{h_n}\right) \\
&\leq \left\{\underbrace{|\mu_n - \mu|}_{V_{1n}}\frac{1}{G_n(a_F)} + \mu\frac{1}{G_n(a_F)G(a_F)}\underbrace{\sup_{y\in\mathcal{D}}\ |G_n(y) - G(y)|}_{V_{2n}}\right\}|f_n^*(y)|.
\end{aligned}$$

From (He and Yang [21, Theorem 3.2]), $|\mu_n - \mu| = O_{\mathbf{P}}\left(n^{-1/2}\right)$, and $G_n(a_F) \overset{\mathbf{P}-a.s.}{\longrightarrow} G(a_F) > 0$ we get, uniformly on y, $V_{1n} = O_{\mathbf{P}}\left(n^{-1/2}\right)$. In the same way, and using (Woodroofe [49, Remark 6]) we get $V_{2n} = O_{\mathbf{P}}\left(n^{-1/2}\right)$. Finally, using the fact that $f_n^*(y)$ goes to $f^*(y) = \frac{G(y)}{\mu}f(y)$ (by differentiating in the statement (2.2)) which is bounded by (A4). Then we get the result. ■

Proof of Theorem 2.1. Using the triangular inequality, we have

$$\begin{aligned}
\sup_{y\in\Omega}\left|\hat{f}_n(y) - f(y)\right| &\leq \sup_{y\in\Omega}\left|\tilde{f}_n(y) - \mathbf{E}\left[\tilde{f}_n(y)\right] + \right|\sup_{y\in\Omega}\left|\hat{f}_n(y) - \tilde{f}_n(y)\right| + \sup_{y\in\Omega}\left|\mathbf{E}\left[\tilde{f}_n(y)\right] - f(y)\right| \\
&=: \mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3.
\end{aligned}$$

For $\mathcal{I}_1$ and $\mathcal{I}_2$ we use Lemma 2.1 and Lemma 2.2 respectively. It remains to examine the term $\mathcal{I}_3$. Using a change of variable and a Taylor expansion, we get under (A2)-(A4), $\mathcal{I}_3 = O\left(h_n^2\right)$. Then we end the proof of Theorem 2.1.

Proof of Corollary 2.1. Standard arguments give us

$$\left|f(\hat{\theta}_n) - f(\theta)\right| \leq 2\sup_{y\in\Omega}\left|\hat{f}_n(y) - f(y)\right|. \tag{2.13}$$

Then, an application of Theorem 2.1 and **(A5)**, we get the result of part 1. For the second part, a Taylor expansion of $f(.)$ in a neighborhood of θ gives

$$f(\hat{\theta}_n) - f(\theta) = \frac{1}{2}(\hat{\theta}_n - \theta)^2 f^{(2)}(\theta^*),$$

where θ^* is between $\hat{\theta}_n$ and θ. Then, by (2.13) and (A4), we have

$$\left(\hat{\theta}_n - \theta\right)^2\left|f^{(2)}(\theta^*)\right| \leq 4\sup_{y\in\Omega}\left|\hat{f}_n(y) - f(y)\right|.$$

Hence, by Theorem 2.1, we complete the proof of Corollary 2.1. ■

The following Lemmas are needed to state the asymptotic normality of the kernel mode estimator.

Lemma 2.3. *Under Assumptions (A1), we have*

$$\sup_{y\in\Omega}\left|\tilde{F}_n(y)-F(y)\right| = O\left(\sqrt{\frac{\log n}{n}}\right), \quad \mathbf{P}-a.s. \text{ as } n\to+\infty$$

Proof. This proof is slightly different from that given in (Ould-Saïd and Lemdani [36, Lemma 1]). Consider, the iid sequence $Y_1, Y_2, ..., Y_n$ and the class of functions

$$\mathcal{L}_n = \left\{\Psi_y : \Omega \longrightarrow \mathbb{R}^+ /\ \Psi_y(z) = \frac{\mu.\mathbb{1}_{(z\le y)}}{nG(z)},\ y\in\Omega\right\}.$$

According to (Giné and Guillou [19, Lemma 3(b)]), $\mathcal{L}_n$ is a Vapnik-Ćervonenkis class (V−C-class) of non negative measurable functions which are uniformly bounded with respect to the envelope $\mathbf{\Delta} = [nG(a_F)]^{-1}$. Moreover

$$\mathbf{E}\left[\Psi_y(Y)\right] \le \frac{1}{nG(a_F)} \quad \text{and} \quad \mathbf{E}\left[\Psi_y^2(Y)\right] \le \frac{1}{n^2G^2(a_F)}.$$

Applying again Talagrand's inequality as before, with $t = D\sqrt{\dfrac{\log n}{n}}$, one gets the result.

Lemma 2.4. *As $n \longrightarrow +\infty$, we have*

$$\sup_{y\in\Omega}\left|\hat{F}_n(y)-F(y)\right| = O\left(\sqrt{\frac{\log n}{n}}\right), \quad \mathbf{P}-a.s.$$

Proof. We have

$$\begin{aligned}
\left|\hat{F}_n(y)-F(y)\right| &\le \underbrace{\left|\hat{F}_n(y)-\tilde{F}_n(y)\right|}_{\mathcal{T}_1} + \underbrace{\left|\tilde{F}_n(y)-F(y)\right|}_{\mathcal{T}_2} \\
&\le \frac{|\mu_n-\mu|}{n}\sum_{i=1}^n \frac{1}{G_n(Y_i)}\mathbb{1}_{\{Y_i\le y\}} + \frac{\mu}{n}\sum_{i=1}^n\left|\frac{1}{G_n(Y_i)}-\frac{1}{G(Y_i)}\right|\mathbb{1}_{\{Y_i\le y\}} + \mathcal{T}_2 \\
&\le \frac{|\mu_n-\mu|}{G_n(a_F)} + \frac{\mu}{G_n(a_F)G(a_F)}\sup_{y\ge a_F}|G_n(y)-G(y)| + \mathcal{T}_2 \\
&=: (I)+(II)+(\mathcal{T}_2). \qquad (2.14)
\end{aligned}$$

Using (He and Yang [21, Theorem 3.2]), $|\mu_n-\mu| = O_{\mathbf{P}}\left(n^{-1/2}\right)$, and $G_n(a_F) \overset{\mathbf{P}-a.s.}{\longrightarrow} G(a_F) > 0$ we get, uniformly on y, $(I) = O_{\mathbf{P}}\left(n^{-1/2}\right)$. In the same way, and using (Woodroofe [49, Remark 6]), we get $(II) = O_{\mathbf{P}}\left(n^{-1/2}\right)$. Finally, using Lemma 2.1 and Remark 2.4 below, we get the result. ■

Remark 2.4. *Recall that, as noted in formula (2.10), all the sums involving the* $(G_n(Y_i))^{-1}$ *are taken for the i's such that* $G_n(Y_i) \neq 0$. *It follows that an additional term*

$$(IV) = \frac{\mu}{n}\sum_{i=1}^{n}\frac{1}{G(Y_i)}\mathbb{1}_{\{Y_i\le y\}}\mathbb{1}_{\{G_n(Y_i)=0\}},$$

should be added in (2.14). This term is clearly negligible by the law of large numbers. The same remark can be made for all similar quantities in what follows.

Proof of Theorem 2.2. From (2.9) and (2.11) we have the following decomposition

$$\begin{aligned}
\sqrt{nh_n^3}\frac{\hat{f}_n^{(1)}(\theta)}{\hat{f}_n^{(2)}(\theta_n^*)} &= \sqrt{nh_n^3}\frac{\hat{f}_n^{(1)}(\theta)-\tilde{f}_n^{(1)}(\theta)}{\hat{f}_n^{(2)}(\theta_n^*)}\\
&+ \sqrt{nh_n^3}\frac{\tilde{f}_n^{(1)}(\theta)-\mathbf{E}\left[\tilde{f}_n^{(1)}(\theta)\right]}{\hat{f}_n^{(2)}(\theta_n^*)}\\
&+ \sqrt{nh_n^3}\frac{\mathbf{E}\left[\tilde{f}_n^{(1)}(\theta)\right]}{\hat{f}_n^{(2)}(\theta_n^*)}\\
&=: \mathcal{J}_1+\mathcal{J}_2+\mathcal{J}_3.
\end{aligned}$$

To prove the result, we establish that the numerators of the terms $\mathcal{J}_1$ and $\mathcal{J}_3$ are negligible and that of $\mathcal{J}_2$ is normally distributed. Then, in conjunction with the fact that the denominator converges in probability, we get the result by using Slutsky's theorem.
For the first term $\mathcal{J}_1$, we have

$$\begin{aligned}
\sqrt{nh_n^3}\left(\hat{f}_n^{(1)}(\theta)-\tilde{f}_n^{(1)}(\theta)\right) &\le \Bigg\{|\mu_n-\mu|\sqrt{nh_n^3}\times\frac{\sup_{y\ge a_F}|G_n(y)-G(y)|}{G(a_F)G_n(a_F)}\\
&+\mu\sqrt{nh_n^3}\times\frac{\sup_{y\ge a_F}|G_n(y)-G(y)|}{G(a_F)G_n(a_F)}\\
&+\sqrt{nh_n^3}\times\frac{|\mu_n-\mu|}{G(a_F)}\Bigg\}\times\frac{1}{nh_n^2}\left|\sum_{i=1}^{n}K^{(1)}\left(\frac{\theta-Y_i}{h_n}\right)\right|\\
&=: \mathcal{U}_1+\mathcal{U}_2+\mathcal{U}_3
\end{aligned}$$

Since $|\mu_n-\mu| = O_{\mathbf{P}}\left(n^{-1/2}\right)$, $\sup\limits_{y\ge a_F}|G_n(y)-G(y)| = O_{\mathbf{P}}\left(n^{-1/2}\right)$, $G_n(a_F)\overset{\mathbf{P}\ a.s.}{\longrightarrow} G(a_F)$(see proof of Lemma 2.4) and the fact that $\frac{1}{nh_n^2}\sum_{i=1}^{n}K^{(1)}\left(\frac{\theta-Y_i}{h_n}\right) =: (f_n^*)^{(1)}(\theta)$, goes to $(f^*)^{(1)}(\theta)$ (where $(f^*)^{(1)}(\cdot)$ denotes the first derivative of the density function $f^*(\cdot)$ of the observed data), which is not necessarily equal to zero at the point θ, we get $\mathcal{U}_1 = O_{\mathbf{P}}\left(\sqrt{\frac{h_n^3}{n}}\right) = o_{\mathbf{P}}(h_n)$. In the same way, we get $\mathcal{U}_2 = O_{\mathbf{P}}\left(\sqrt{h_n^3}\right)$ and $\mathcal{U}_3 = O_{\mathbf{P}}\left(\sqrt{h_n^3}\right)$ which permit us to conclude that $\mathcal{J}_1$ is negligible.
Now, we turn out to $\mathcal{J}_3$ and state the following result

Lemma 2.5. *Under Assumptions (A2)-(A4) and (A6) ii), we have*

$$\left(nh_n^3\right)^{\frac{1}{2}} \mathbf{E}\left[\tilde{f}_n^{(1)}(\theta)\right] \longrightarrow 0, \quad as\ n \to \infty.$$

Proof. We have

$$\begin{aligned}
\mathbf{E}\left[\tilde{f}_n^{(1)}(\theta)\right] &= \frac{\mu}{h_n^2} \int \frac{K^{(1)}\left(\frac{\theta-u}{h_n}\right)}{G(u)} dF^*(u) \\
&= \frac{1}{h_n^2} \int K^{(1)}\left(\frac{\theta-u}{h_n}\right) f(u)du \\
&= \frac{1}{h_n} \int K^{(1)}(v)\, f(\theta - vh_n)dv.
\end{aligned}$$

By part integrating, we have

$$\mathbf{E}\left[\tilde{f}_n^{(1)}(\theta)\right] = \int K(v)\, f^{(1)}(\theta - vh_n)\, dv.$$

By Taylor's expansion of $f^{(1)}(\cdot)$ around θ, (A2)-(A3) and the definition of the mode, we get

$$\left(nh_n^3\right)^{\frac{1}{2}} \mathbf{E}\left[f_n^{(1)}(\theta)\right] = \sqrt{nh_n^7} \int v^2 K(v) f^{(3)}(\theta^*)\, dv,$$

where θ^* is between θ and $\theta - vh_n$.
By **(A2)**, **(A4)** and **(A6)** *ii*), we get the result. ∎
Finally, to state the asymptotic normality of $\mathcal{J}_2$, we need some auxiliary results.

Lemma 2.6. *Under Assumptions (A2), (A4) and (A6) ii), we have*

$$nh_n^3 Var\left[\tilde{f}_n^{(1)}(\theta)\right] \longrightarrow \frac{\mu f(\theta)}{G(\theta)} \int \left(K^{(1)}(v)\right)^2 dv, \quad as \quad n \to \infty.$$

Proof. First, start with calculating $Var\left(\tilde{f}_n^{(1)}(\theta)\right)$.

$$\begin{aligned}
Var\left(\tilde{f}_n^{(1)}(\theta)\right) &= Var\left(\frac{\mu}{nh_n^2}\sum_{i=1}^{n}\frac{1}{G(Y_i)}K^{(1)}\left(\frac{\theta-Y_i}{h_n}\right)\right)\\
&= \frac{\mu^2}{nh_n^4}\left\{\mathbf{E}\left[\frac{1}{G(Y_1)}K^{(1)}\left(\frac{\theta-Y_1}{h_n}\right)\right]^2-\mathbf{E}^2\left[\frac{1}{G(Y_1)}K^{(1)}\left(\frac{\theta-Y_1}{h_n}\right)\right]\right\}\\
&= \frac{\mu^2}{nh_n^4}\int\frac{G(v)}{\mu}\left[\frac{1}{G(v)}K^{(1)}\left(\frac{\theta-v}{h_n}\right)\right]^2 f(v)dv\\
&\quad -\frac{1}{nh_n^4}\left[\int K^{(1)}\left(\frac{\theta-v}{h_n}\right)f(v)dv\right]^2\\
&= \frac{\mu}{nh_n^3}\int\frac{1}{G(\theta-vh_n)}\left(K^{(1)}(u)\right)^2 f(\theta-uh_n)du\\
&\quad -\left[\frac{1}{\sqrt{nh_n^2}}\int K^{(1)}(u)\,f(\theta-uh_n)du\right]^2\\
&=: \mathcal{V}_1-\mathcal{V}_2.
\end{aligned}$$

In analogous manner as done before and under **(A1)-(A4)** and the definition of the mode, we get $nh_n^3\mathcal{V}_2\longrightarrow 0$.
Again, a Taylor expansion gives us

$$\mathcal{V}_1=\frac{\mu f(\theta)}{nh_n^3}\int_{a_F}^{\frac{\theta-a_F}{h_n}}\frac{1}{G(\theta-uh_n)}\left(K^{(1)}(u)\right)^2du+o\left(\frac{1}{nh_n^2}\right).$$

Since $\theta>a_F$, observe that there exists n_0 such that for all fixed $n\geq n_0$,

$$\int_{a_F}^{\frac{\theta-a_F}{h_n}}\frac{1}{G(\theta-uh_n)}\left(K^{(1)}(u)\right)^2du\leq\frac{1}{G(a_F)}\int_{a_F}^{+\infty}\left(K^{(1)}(u)\right)^2du,$$

whence, by the dominated convergence Theorem and for $n\to\infty$

$$\int\frac{1}{G(\theta-uh_n)}\left(K^{(1)}(u)\right)^2du\longrightarrow\frac{1}{G(\theta)}\int\left(K^{(1)}(u)\right)^2du,$$

which is finite by **(A2)**. Finally, we obtain

$$nh_n^3Var\left(\tilde{f}_n^{(1)}(\theta)\right)\longrightarrow\frac{\mu f(\theta)}{G(\theta)}\int\left(K^{(1)}(u)\right)^2du.$$

Now, all what is left to be shown is that, the numerator of $\mathcal{J}_2$ is a sum of iid rv's which satisfies Lindeberg's Theorem. For this purpose, let us to consider

$$Z_{in}(y)=\frac{\mu}{\sqrt{nh_n}}\left(\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)-\mathbf{E}\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]\right).$$

Simple algebra shows that

$$\sum_{i=1}^{n} Z_{in}(y) = \sqrt{nh_n^3}\left(\tilde{f}_n^{(1)}(y) - \mathbf{E}\left[\tilde{f}_n^{(1)}(y)\right]\right).$$

Hence

$$Var\left(\sum_{i=1}^{n} Z_{in}(y)\right) = nh_n^3 Var\left(\tilde{f}_n^{(1)}(y)\right).$$

Then we have the following

Lemma 2.7. *Under Assumptions (A2), (A4), (A6) i), we have*

$$\forall\varepsilon > 0,\ \sum_{i=1}^{n}\int_{\left\{Z_{in}^2(y)>\varepsilon^2 Var\left(\sum_{i=1}^{n} Z_{in}(y)\right)\right\}} Z_{in}^2(y)dF^*(y) \longrightarrow 0, \quad as \quad n\to\infty.$$

Proof. On the one hand we have

$$\begin{aligned} Z_{in}^2(y) &\leq \frac{2\mu^2}{nh_n}\left\{\frac{1}{G^2(Y_i)}\left(K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 + \mathbf{E}^2\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]\right\} \\ &= \frac{2\mu^2}{nh_n}\frac{1}{G^2(Y_i)}\left(K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 + \frac{2\mu^2}{nh_n}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]. \end{aligned} \tag{2.15}$$

Note that

$$\frac{2\mu^2}{nh_n}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] = \frac{2h_n}{n}\mathbf{E}^2\left[\frac{\mu}{h_nG(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right],$$

which gives

$$\frac{2\mu^2}{nh_n}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] \longrightarrow 0, \text{ as } n\to\infty. \tag{2.16}$$

On the other hand, by Lemma 2.4 we have

$$Var\left(\sum_{i=1}^{n} Z_{in}(y)\right) \longrightarrow \frac{\mu f(\theta)}{G(\theta)}\int\left(K^{(1)}(u)\right)^2 du,$$

then for $\varepsilon = \frac{\mu f(\theta)}{2G(\theta)}\int\left(K^{(1)}(u)\right)^2 du > 0$, $\exists n_0 \in \mathbb{N}^*$ such that : $\forall n \geq n_0$, we have

$$Var\left(\sum_{i=1}^{n} Z_{in}(y)\right) \geq \frac{\mu f(\theta)}{2G(\theta)}\int\left(K^{(1)}(u)\right)^2 du.$$

Now denote by

$$V(Y_i) = \frac{1}{G^2(Y_i)}\left(K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 + \mathbf{E}^2\left[\frac{1}{G(Y_i)}K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right];$$

from (4.6), we have

$$Z_{in}^2(y) \leq \frac{2\mu^2 V(Y_i)}{nh_n},$$

making use of (2.16), for $n \geq n_0$ we have

$$\left\{Z_{in}^2(y) > \varepsilon^2 Var\left(\sum_{i=1}^n Z_{in}(y)\right)\right\} \subset \left\{Z_{in}^2(y) > \varepsilon^2 \frac{\mu f(\theta)}{2G(\theta)} \int \left(K^{(1)}(u)\right)^2 du\right\}.$$

Now, set $\varepsilon' = \varepsilon^2 \frac{\mu f(\theta)}{4G(\theta)} \int \left(K^{(1)}(u)\right)^2 du$, we have

$$\begin{aligned}
\left\{Z_{in}^2(y) > \varepsilon^2 Var\left(\sum_{i=1}^n Z_{in}(y)\right)\right\} &\subset \left\{Z_{in}^2(y) > 2\varepsilon'\right\} \\
&= \left\{\frac{nh_n}{2\mu^2} Z_{in}^2(y) > \varepsilon' nh_n\right\} \\
&\subset \left\{V(Y_i) > \varepsilon' nh_n\right\} \\
&\subset \left\{\frac{1}{G^2(Y_i)}\left(K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 > \frac{\varepsilon' nh_n}{2}\right\} \\
&\bigcup \left\{\mathbf{E}^2\left[\frac{1}{G(Y_i)} K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] > \frac{\varepsilon' nh_n}{2}\right\}.
\end{aligned}$$

From (2.16) and for n large enough we have

$$\left\{\mathbf{E}^2\left[\frac{1}{G(Y_i)} K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] > \frac{\varepsilon' nh_n}{2}\right\} = \emptyset. \tag{2.17}$$

Furthermore, by the fact that G is lower bounded and $K^{(1)}$ is bounded, then by (A6) $i)$ and for n large enough

$$\left\{\frac{1}{G^2(Y_i)}\left(K^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 > \frac{\varepsilon' nh_n}{2}\right\} = \emptyset. \tag{2.18}$$

Therefore, from (2.17) and (2.18) and for n large enough, we get

$$\left\{Z_{in}^2(y) > \varepsilon^2 Var\left(\sum_{i=1}^n Z_{in}(y)\right)\right\} = \emptyset,$$

which completes the proof. ■

Now, to end the proof of Theorem 2.2, it suffices to prove that $\hat{f}_n^{(2)}(\theta_n^*) \xrightarrow{\mathbf{P}} f^{(2)}(\theta)$. Observe that one can write the following

$$\begin{aligned}
\hat{f}_n^{(2)}(y) - f^{(2)}(y) &= \frac{1}{h_n^3} \int K^{(2)}\left(\frac{y-v}{h_n}\right)\left\{d\hat{F}_n(v) - dF(v)\right\} \\
&\quad + \left\{\frac{1}{h_n^3} \int K^{(2)}\left(\frac{y-v}{h_n}\right) f(v)dv - \frac{1}{h_n} \int K\left(\frac{y-v}{h_n}\right) f^{(2)}(y)dv\right\} \\
&=: \mathcal{S}_1 + \mathcal{S}_2.
\end{aligned}$$

That $\mathcal{S}_1$ and $\mathcal{S}_2$ tend to zero uniformly as n grows to infinity. Indeed, on the one hand, by Taylor's expansion, integration by part and assumptions (A1), (A2), (A4) and (A6) $i)$ we have

$$\begin{aligned} |\mathcal{S}_1| &\leq \frac{1}{h_n^3}\int \left|K^{(3)}(v)\right| \left|\hat{F}_n(y-vh_n) - F(y-vh_n)\right| dv \\ &\leq \sup_{y\in\Omega}\left|\hat{F}_n(y) - F(y)\right| \frac{1}{h_n^3}\int \left|K^{(3)}(v)\right| dv. \end{aligned}$$

Then, by Lemma 2.4, $\mathcal{S}_1 = O_{\mathbf{P}}\left(\sqrt{\frac{\log n}{nh_n^6}}\right)$.

On the other hand, under assumptions (A2), (A4) and by part integrating, we have

$$\mathcal{S}_2 = \frac{1}{h_n}\left\{\int K\left(\frac{y-v}{h_n}\right) f^{(2)}(v)dv - \int K\left(\frac{y-v}{h_n}\right) f^{(2)}(y)dv\right\},$$

then, using a change of variable and a Taylor expansion, we have

$$\begin{aligned} |\mathcal{S}_2| &\leq \int K(v)\left|f^{(2)}(y-vh_n) - f^{(2)}(y)\right| dv \\ &\leq h_n \int |v|\, K(v)\left|f^{(3)}(v^*)\right| dv \\ &= O(h_n) \end{aligned}$$

where v^* is between $y - vh_n$ and y. Now, we have almost surely, for n large enough,

$$\left|\hat{f}_n^{(2)}(\theta_n^*) - f^{(2)}(\theta)\right| \leq \sup_{y\in\Omega}\left|\hat{f}_n^{(2)}(y) - f^{(2)}(y)\right| + \left|f^{(2)}(\theta_n^*) - f^{(2)}(\theta)\right|.$$

As θ_n^* is between θ and $\hat{\theta}_n$, Corollary 2.1 entails the almost sure convergence of θ_n^* to θ. Hence, the continuity of $f^{(2)}(\cdot)$ ensures the almost sure convergence of $f^{(2)}(\theta_n^*)$ to $f^{(2)}(\theta)$. Furthermore, because of the convergence in probability of $\hat{f}_n^{(2)}(\cdot)$ to $f^{(2)}(\cdot)$ uniformly on Ω, we conclude that $\hat{f}_n^{(2)}(\theta_n^*)$ converges in probability to the real number $f^{(2)}(\theta) \neq 0$.

Bibliographie

[1] C. Abraham, G. Biau and B. Cadre, *On the asymptotic properties of a simple estimate of the mode.* ESAIM, Probab. Statist. **8** (2004), 1-11.

[2] C. Abraham, G. Biau and B. Cadre, *Simple estimation of the mode of a multivariate density.* Canadian Journal of Statistics **31** (2003), 23-34.

[3] P.K. Anderson, O. Borgan, R.D. Gill and N. Keiding, *Statistical models based on counting processes.* Springer, New York, 1993.

[4] D.R. Bickel, *Robust estimators of the mode and skewness of continuous data.* Computat. Statist. Data Anal. **38** (2002), 153-163.

[5] D.R. Bickel and R. Frühwirth, *On a fast, robust estimator of the mode : Comparaisons to other robust estimators with applications.* Computat. Statist. Data Anal. **50** (2006), 3500-3530.

[6] K. Chen, M.T. Chao and S.W. Lo, *On strong uniform consistency of the Lynden-Bell estimator for truncated data.* Ann. Statist. **23** (1995), 440-449.

[7] H .Chernoff, *Estimation of the mode.* Ann. Instit. Statist. Math. **16** (1964), 31-41.

[8] G. Collomb, W. Hardle and S. Hassani, *A note on prediction via estimation of the conditional mode function.* J. Statist. Plann. Inference **15** (1987), 227-236.

[9] S. Dabo-Niang and A. Laksaci, *Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle.* C. R. Acad. Sci. Paris, Ser. I **344** (2007), 49-52.

[10] S. Dabo-Niang, F. Ferraty and P. Vieu, *Estimation du mode dans un espace vectoriel semi-normé.* C. R. Acad. Sci. Paris, Ser. I **339** (2004), 659-662.

[11] R. M. Dudley, *Uniform Central Limit Theorems.* Cambridge University Press, Cambridge, UK, 1999.

[12] W.F. Eddy, *Optimum kernel estimators of the mode.* Ann. Statist. **8** (1980), 870-882.

[13] W.F. Eddy, *The asymptotic distributions of kernel estimators of the mode.* Z. Wahrsh. Verw. Gebiete. **59** (1982), 279-290

[14] M. Ezzahrioui and E. Ould-Saïd, *Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data.* Nonparametric Statist. J. (In press).

[15] M. Ezzahrioui and E. Ould-Saïd, *Some asymptotic results of a nonparametric conditional mode estimator for functional time series data.* Technical Report, LMPA, Janvier 2007, Univ. Littoral Côte d'Opale (Submitted).

[16] F. Ferraty, A. Laksaci and P. Vieu, *Functional times series prediction via conditional mode.* C. R. Math. Acad. Sci. Paris, Ser. I **340** (2005), 389-392.

[17] E. Giné and A. Guillou, *Rates of strong uniform consistency for multivariate kernel density estimators.* Ann. I. H. Poincaré **38** (2002), 907-921.

[18] E. Giné and A. Guillou, *On consistency of kernel density estimators for randomly censored data : rates holding over adaptive intervals.* Ann. inst. Henry poincaré **37** (2001), 503-522.

[19] E. Giné and A. Guillou, *Laws of iterated logarithm for censored data.* Ann. Probab. **27** (1999), 2042-2067.

[20] S. He and G.L. Yang, *Estimating a lifetime distribution under different sampling plan.* In : S.S Gupta and J.O. Berger (Ed.),*Statistical Decision theory and related topics*, **5** pp. 73-85. Berlin Heidelberg, Springer, New York, 1994.

[21] S. He and G.L. Yang, *Estimation of the truncation probability in the random truncation model.* Ann. Statist. **26** (1998), 1011-1028.

[22] E. Hermann and K. Ziegler, *Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions.* Statist. Probab. Lett. **72** (2004), 359-368.

[23] E.M. Kaplan and P. Meier, *Nonparametric estimation from incomplete observations.* J. Amer. Statist. Assoc. **53** (1958), 457-481.

[24] N. Keiding and R.D. Gill, *Random truncation models and Markov processes.* Ann. Statist., **18** (1990), 582-602.

[25] M. Lemdani and E. Ould-Saïd, *Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data.* Comm. Statist. Theory Methods **36** (2007), 155-174.

[26] D. Louani, *On the asymptotic normality of the kernel estimators of the density function and its derivatives under random censoring.* Comm. Statist. Theory Methods **27** (1998), 2909-2924.

[27] D. Louani and E. Ould-Saïd, *Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis.* J. Nonparametric Statist. **11** (1999), 413-442.

[28] D. Lynden-Bell, *A method of allowing for known observational selection in small samples applied to 3CR quasars.* Monthly Notices Roy. Astronom. Soc. **155** (1971), 95-118.

[29] K.L. Mehra, Y.S. Ramakrishnaiah and P. Sashikala, *Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode.* Ann. Inst. Statist. Math. **52** (2000), 630-645.

[30] A. Mokkaddem and M. Pelletier, *Moderate deviations for the kernel mode estimator and some applications.* J. Statist. Plann. Inference **135** (2005), 276-299.

[31] A. Mokkaddem and M. Pelletier, *The law of the iterated logarithm for the multivariate kernel mode estimator*. ESAIM : Probability and statistics **7** (2003), 1-21.

[32] A. Mokkaddem, M. Pelletier, and J. Worms, *A large deviations upper bound for the kernel mode estimator*. Theory Probab. Appl. **50** (2006), 153-165.

[33] E.N. Nadaraya, *On nonparametric estimates of density functions and regression curves.* Theory Probab. Appl. **10** (1965), 186-190.

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

[35] E. Ould-Saïd, *Estimation non paramétrique du mode conditionnel. Application à la prévision.C. R. Acad. Sci. Paris, Ser. I* **316** (1993), 943-947.

[36] E. Ould-Saïd and M. Lemdani, *Asymptotic properties of the kernel estimators of the regression under random left truncation*. Ann. Inst. Statist. Math. **58** (2006), 357-378.

[37] E. Ould-Saïd, and Z. Cai, *Strong uniform consistency of the nonparametric estimation of the censored conditional mode function*. J. Nonparametric Statist. **17** (2005), 797-806.

[38] E. Parzen, *On estimation of a probability density function and mode*. Ann. Math. Statist. **33** (1962), 1065-1076.

[39] J. P. Romano, *On weak convergence and optimality of kernel density estimates of the mode*. Ann. Statist. **16** (1988), 629-647.

[40] M. Rosenblatt, *Remarks on some nonparametric estimates of a density function*. Ann. Math. Statist. **27** (1956), 832-837.

[41] M. Samanta, *Nonparametric estimation of the mode of a multivariate density*. South African Statist. J. **7** (1973), 109-117.

[42] M. Samanta and A. Thavaneswaran, *Nonparametric estimation of the conditional mode* . Comm. Statist. Theory Methods **16** (1990), 4515-4524.

[43] B.W. Silverman, *Density estimation for statistics and data analysis*. Chapman and Hall, London, 1986.

[44] W. Stute, *Almost sure representation of the product-limit estimator for truncated data*. Ann. Statist. **21** (1993), 146-156.

[45] W.Y. Tsai, N.P. Jewell and M.C. Wang, *A note on the product-limit estimator under right censoring and left truncation*. Biometrika **74** (1987), 883 -886.

[46] A.W. van der Vaart and J.A. Wellner, *Weak convergence and empirical processes*. Springer, New York, 1996.

[47] P. Vieu, *A note on density mode estimation*. Statist. Probab. Lett. **26** (1996), 297-307.

[48] M.C. Wang, N.P. Jewell and W.Y. Tsai, *Asymptotic properties of the limit-product estimate under random truncation*. Ann. Statist. **14** (1986), 1597-1805.

[49] M. Woodroofe, *Estimating a distribution function with truncated data.* Ann. Statist. **13** (1985), 163-177.

[50] Y. Zhou, *A note on the TJW product-limit estimator for truncated and censored data.* Statist. Probab. Lett. **26** (1996), 381-387.

Chapitre 3

Mode simple : Cas $\alpha-$mélange

Strong consistency rate for the kernel mode estimator under strong mixing hypothesis and left-truncation[‡]

Ould-Saïd, E., Tatachak, A.

Abstract

Let $(Y_N)_{N\geq 1}$ be a sequence of copies of a random variable of interest Y. In this paper, we study the kernel estimator, say $\hat{\theta}_n$, of the mode of Y when Y is subject to the random left-truncation. While based on n ($n \leq N$) actual observations fulfilling the well-known $\alpha-$mixing condition, we establish the strong consistency with a rate of the proposed estimator $\hat{\theta}_n$.

Keywords Kernel estimator ; Lynden$-$Bell estimator ; mode ; random left-truncation ; rate of convergence ; strong consistency ; strong mixing.

Mathematics Subject Classification. 62G05 ; 62G07 ; 62G20.

[‡]To appear, in Comunications in Statistics$-$Theory and Methods, (2009).

3.1 Introduction

Let Y and T be two real random variables with unknown cumulative distribution functions (df) F and G respectively. Denote by f the probability density of Y with respect to Lebesgue measure. Let $(Y_1, T_1), (Y_2, T_2), ..., (Y_N, T_N)$ be N copies of (Y, T), where the sample size N is fixed, but unknown. In the random left-truncation (RLT) model, the random variable (rv) of interest Y is interfered by the truncation rv T, such that both quantities Y and T are observable only if $Y \geq T$, whereas nothing is observed if $Y < T$. Without possible confusion, we still denote (Y_i, T_i), $i = 1, 2, ..., n, (n \leq N)$ the observed pairs from the original $N-$sample.
As a consequence of truncation, the size of the actually observed sample, n, is a $Bin(N, \mu)$ random variable, with $\mu := \mathbb{P}(Y \geq T)$. It is clear that if $\mu = 0$, no data can be observed and therefore, we suppose throughout this paper that $\mu \neq 0$. By the strong law of large numbers (SLLN) we have, as $N \longrightarrow \infty$

$$\mu_n := \frac{n}{N} \longrightarrow \mu, \ \mathbb{P}-\text{a.s.} \tag{3.1}$$

In the complete data case (no truncation), the mode estimator for the independent and identically distributed (iid) random variables has been studied by many authors among whom we quote Parzen (1962), Nadaraya (1965), Van Ryzin (1969), Samanta (1973), Eddy (1980, 1982), Romano (1988), Samanta and Thavaneswaran (1990) and Vieu (1996).
The last quinquennial knew an increased interest and numerous are the papers dealing with the problem of the mode. To quote only a few of them, Abraham *et al.* (2004) considered the estimate of the mode as an element of the sample points, and obtained the strong consistency. Mokkadem and Pelletier (2005) and Mokkadem *et al.* (2006), studied the moderate and large deviations upper bound for the kernel mode estimator. Hermann and Ziegler (2004) obtained rates of nonparametric estimation of the mode in absence of smoothness assumptions and Bickel and Frühwirth (2006) proposed a robust estimator of the mode based on the densest half ranges. When the conditioning variable takes values in a semi-normed vectorial space of possibly infinite dimension, Dabo$-$Niang *et al.* (2004) established the strong consistency of the simple mode, while Dabo$-$Niang and Laksaci (2007) established the consistency in L^p norm of the conditional mode function estimator.
For the complete dependent data case, the strong consistency of the conditional mode estimator was established under a $\phi-$mixing condition by Collomb *et. al.* (1987) and their results can be applied to process forecasting. In the $\alpha-$mixing setting, the strong consistency over a compact set and the asymptotic normality were obtained by Ould-Saïd (1993) and Louani and Ould-Saïd (1999) respectively. In a general ergodic framework, a process prediction via the conditional mode estimation was described by Ould-Saïd (1997) and its strong consistency was established. Recently, Ferraty *et al.* (2005) established under $\alpha-$mixing condition, the almost complete convergence of the conditional mode of a scalar response given a functional random variable.
In the incomplete data case, for iid random variables under random right censoring, Louani (1998) studied the asymptotic normality of the kernel estimator of the mode, while Ould-Saïd and Cai

(2005) established, the uniform strong consistency with rates of a nonparametric estimator of the conditional mode function.
Under RLT model, in the iid case, the kernel mode estimate has been studied by Ould-Saïd and Tatachak (2005) and their results were extended to the conditional mode estimator (see Ould-Saïd and Tatachak (2007)).
The present work deals with nonparametric estimation of the mode for the RLT model, when the underlying variables of interest are assumed to fulfill the $\alpha-$mixing condition. The strong mixing condition was introduced by Rosenblatt (1956) whose definition is given below.
Our main result establishes the strong uniform consistency with a rate, over a compact set, of the kernel density estimator which allows to deduce the strong consistency with a rate of the kernel mode estimator. The paper is organized as follows : In Section 2, we recall some important and useful results in the RLT model. Section 3 is devoted to notations and definitions, while assumptions and main results are given in Section 4. Finally, some auxiliary results and proofs are relegated to the last Section.

3.2 The Model and the Estimates

Throughout this paper, the letters C and C' are used indiscriminately as generic constants, and all the limits are taken as $n \to \infty$ unless otherwise specified.
We first present some results from the literature for the univariate truncation model, which will be used in the sequel to define our nonparametric kernel estimators.
Since N is unknown and n is known (although random), our results will not be stated with respect to the probability measure $\mathbb{P}$ (related to the $N-$sample) but will involve the probability $\mathbf{P}$ (related to the $n-$sample). In the same way $\mathbb{E}$ and $\mathbf{E}$ denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively. From now on, as a convention, we denote by a superscript $(^*)$ any distribution function that is associated to the observed sample.
Under the RLT sampling scheme, the conditional joint distribution of an observed (Y,T) (see Stute (1993) and Zhou (1996)) is given by

$$\begin{aligned} H^*(y,t) = \mathbf{P}\{Y \le y, T \le t\} &= \mathbb{P}\{Y \le y, T \le t \,|Y \ge T\} \\ &= \mu^{-1}\int_{-\infty}^{y} G(t\wedge u)dF(u) \end{aligned}$$

where $t \wedge u = \min(t,u)$. The marginal distributions are defined by

$$F^*(y) := H^*(y,\infty) = \mu^{-1}\int_{-\infty}^{y} G(u)dF(u), \tag{3.2}$$

$$\begin{aligned} G^*(t) := H^*(\infty,t) &= \mu^{-1}\int_{-\infty}^{\infty} G(t\wedge u)dF(u) \\ &= \mu^{-1}\int_{-\infty}^{t} (1-F(u))\,dG(u) \end{aligned}$$

which are estimated by

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{Y_i \leq y\}} \quad \text{and} \quad G_n^*(t) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{T_i \leq t\}}$$

respectively, where $1\!\!1_A$ denotes the indicator function of the set A. Let f^* denote the density function of the observed variables. Then the kernel estimator is defined by

$$f_n^*(y) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{y - Y_i}{h_n}\right), \tag{3.3}$$

where K is a probability density function (called kernel) defined on $\mathbb{R}$, and $h_n =: h$ a positive bandwidth tending to zero as n grows to infinity. Now, for any df W, define :

$$a_W = \inf\{x : W(x) > 0\} \text{ and } b_W = \sup\{x : W(x) < 1\},$$

as the endpoints of the W support. Woodroofe (1985) pointed out that F and G can be estimated completely only if

$$a_G \leq a_F, \ \ b_G \leq b_F \ \text{ and } \int_{a_F}^{\infty} \frac{dF}{G} < \infty. \tag{3.4}$$

Then under (3.4)

$$\mu := P(Y \geq T) = \int G(u)dF(u) > 0.$$

Let $R(\cdot)$ be a function defined by

$$R(y) = G^*(y) - F^*(y) = \mu^{-1}G(y)\left[1 - F(y)\right], \tag{3.5}$$

with the empirical estimator

$$R_n(y) = G_n^*(y) - F_n^*(y^-) = \frac{1}{n}\sum_{i=1}^{n} 1\!\!1_{\{T_i \leq y \leq Y_i\}}, \quad a_F \leq y < +\infty. \tag{3.6}$$

Then the nonparametric maximum likelihood estimates (NPMLE) of F and G, originally proposed by Lynden$-$Bell (1971), are given by

$$F_n(y) = 1 - \prod_{i:Y_i \leq y}\left[\frac{nR_n(Y_i) - 1}{nR_n(Y_i)}\right] \text{ and } G_n(t) = \prod_{i:T_i > t}\left[\frac{nR_n(T_i) - 1}{nR_n(T_i)}\right], \tag{3.7}$$

respectively, assuming no ties in the data.
Asymptotic properties of (3.7) have been also studied by Woodroofe (1985), in his Theorem 2, he established the following uniform consistency results

$$\sup_{y \geq a_F} |F_n(y) - F(y)| \overset{\mathbf{P-a.s}}{\longrightarrow} 0 \quad \text{and} \quad \sup_{t \geq a_G} |G_n(t) - G(t)| \overset{\mathbf{P-a.s}}{\longrightarrow} 0\ .$$

Under, $\alpha-$mixing structure, Sun & Zhou (2001) expressed the product limit estimator F_n as an average of a sequence of bounded random variables plus a remainder term of order $O(n^{-1/2}\log^{-\varsigma} n)$ for some $\varsigma > 0$, and obtained similar results as those obtained for the Kaplan$-$Meier estimator for censored dependent data (see Cai (1998)).
Note that the estimator μ_n defined in (3.1) cannot be calculated (since N is unknown). Another estimator, namely :

$$\hat{\mu}_n = \frac{G_n(y)[(1-F_n(y^-)]}{R_n(y)},$$

is used. He and Yang (1998) proved that $\hat{\mu}_n$ does not depend on y and its value can then be obtained for any y such that $R_n(y) \neq 0$. Furthermore, they showed (see their Corollary 2.5), that

$$\hat{\mu}_n \longrightarrow \mu \quad \mathbf{P}-a.s.$$

3.3 Notations and Definitions

First, let $\mathcal{F}_i^k(Z)$ denotes the σ-field of events generated by $\{Z_j,\ i \leq j \leq k\}$. For easy reference, let us recall the following definition.
Definition. *Let $\{Z_i,\ i \geq 1\}$ denotes a sequence of random variables. Given a positive integer n, set :*

$$\alpha(n) = \sup\left\{|\mathbf{P}(A\cap B) - \mathbf{P}(A)\mathbf{P}(B)| : A \in \mathcal{F}_1^k(Z),\ B \in \mathcal{F}_{k+n}^{\infty}(Z),\ k \in \mathbb{N}\right\}.$$

The sequence is said to be $\alpha-$mixing (strongly mixing) if the mixing coefficient $\alpha(n) \to 0$.
Among various mixing conditions used in the literature, $\alpha-$mixing is reasonably weak and has many practical applications (see, e.g. Doukhan (1994) or Cai (1998, 2001) for more details).
Now, we come back to our main problem. We would like to estimate the mode under RLT model. Note that, if no truncation is present ($n = N$), it is well known that the kernel estimator of the mode θ is defined as the random value θ_n maximizing the kernel estimator $f_n(\cdot)$ of $f(\cdot)$; that is,

$$\theta_n = \arg\max_{-\infty<y<\infty} f_n(y),$$

where $f_n(\cdot)$ is defined similarly as in (3.3).
In the RLT model, (3.3) is no longer adapted as an estimator of $f(\cdot)$. We define a new estimator $\tilde{f}_n(\cdot)$, based on the n actually observed pairs (Y_i, T_i).
By differentiating and integrating in the statement (3.2), we get

$$F(y) = \mu \int_{a_G}^{y} \frac{1}{G(u)} F^*(du).$$

A natural estimator of F is then given by

$$\tilde{F}_n(y) = \frac{\mu}{n}\sum_{i=1}^{n} \frac{1}{G(Y_i)} \mathbb{I}_{\{Y_i \leq y\}} \tag{3.8}$$

which allows us to define an estimate of the density f as follows

$$\begin{aligned}\tilde{f}_n(y) &= \frac{1}{h}\int K\left(\frac{y-v}{h}\right) d\tilde{F}_n(v) \\ &= \frac{\mu}{nh}\sum_{i=1}^{n}\frac{1}{G(Y_i)}K\left(\frac{y-Y_i}{h}\right). \qquad (3.9)\end{aligned}$$

However, both (3.8) and (3.9) are useless in practice since $G(\cdot)$ and μ are unknown. By adapting (3.8), we propose the estimator

$$\hat{F}_n(y) = \frac{\hat{\mu}_n}{n}\sum_{i=1}^{n}\frac{1}{G_n(Y_i)}\mathbb{1}_{\{Y_i\le y\}}. \qquad (3.10)$$

Note that in (3.10) and in the sequel, the sum is taken only over the $i's$ such that $G_n(Y_i) \neq 0$. Finally (3.10) yields the density estimator

$$\begin{aligned}\hat{f}_n(y) &= \frac{1}{h}\int K\left(\frac{y-v}{h}\right) d\hat{F}_n(v) \\ &= \frac{\hat{\mu}_n}{nh}\sum_{i=1}^{n}\frac{1}{G_n(Y_i)}K\left(\frac{y-Y_i}{h}\right). \qquad (3.11)\end{aligned}$$

Recall that both estimators (3.10) and (3.11) are given in Ould-Saïd and Lemdani (2006) and used in Ould-Saïd and Tatachak (2007). Now, we define our nonparametric kernel estimator of the mode by

$$\hat{f}_n(\hat{\theta}_n) = \sup_{-\infty<y<\infty}\hat{f}_n(y). \qquad (3.12)$$

Remark 3.1. *The estimate $\hat{\theta}_n$ is not necessarily unique, and therefore, all the results of our paper will concern any sequence of random variables $\hat{\theta}_n$ satisfying (3.12). We point out that we can specify our choice by taking*

$$\hat{\theta}_n = \inf\left\{t\in\mathbb{R} \text{ such that } \hat{f}_n(t) = \sup_{-\infty<y<\infty}\hat{f}_n(y)\right\}.$$

3.4 Assumptions and Asymptotic Study

In what follows, we suppose that $a_G < a_F,\ b_G \le b_F$ and let $\mathcal{D} = [a,b]$ be a compact set such that $\mathcal{D} \subset \Omega = \{y:\ y\in[a_F; b_F[\}$. Our assumptions are gathered here for easy reference.

(K1) The kernel function K is a bounded probability density, Hölder continuous with exponent $\beta > 0$ and satisfies $|u|\,K(u) \to 0$ as $|u| \to \infty$;

(K2) K is a second-order kernel;

(K3) The kernel K is differentiable.

(M1) $\{Y_j;\ j \geq 1\}$ is a sequence of stationary α-mixing random variables with coefficient $\alpha(n)$;

(M2) $\{T_j;\ j \geq 1\}$ is a sequence of iid truncating variables with common continuous df G, and are independent of $\{Y_j;\ j \geq 1\}$;

(M3) There exist $\nu > 5 + 1/\beta$, and $C > 0$, such that $\forall\, n$, $\alpha(n) \leq Cn^{-\nu}$.

(D1) The mode θ satisfies the following property : for any $\varepsilon > 0$ and y, there exists a $\tau > 0$ such that $|\theta - y| \geq \varepsilon$ implies that $|f(\theta) - f(y)| \geq \tau$;

(D2) The probability density $f(\cdot)$ is twice continuously differentiable on $\mathcal{D}$ with second derivative $f^{(2)}(\theta) \neq 0$. Furthermore, we suppose that $\theta \in \mathring{\mathcal{D}}$, where $\mathring{\mathcal{D}}$ denotes the interior of $\mathcal{D}$;

(D3) The joint density f^*_{ij} of (Y_i, Y_j) exists and satisfies : $\sup\limits_{u,v} \left| f^*_{ij}(u,v) - f^*(u)f^*(v) \right| \leq C < \infty$ for some constant C not depending on (i,j).

(H) The bandwidth h satisfies :

i) $h \to 0$ and $\frac{\log\log n}{nh^2} \longrightarrow 0$;

ii) $Cn^{\frac{(3-\nu)\beta}{\beta(\nu+1)+2\beta+1}+\eta} < h < C'n^{\frac{1}{1-\nu}}$

where η is such that

$\frac{1}{\beta(\nu+1)+2\beta+1} < \eta < \frac{(\nu-3)\beta}{\beta(\nu+1)+2\beta+1} + \frac{1}{1-\nu}$.

Remark 3.2. *(Discussion of the Assumptions)*
*Assumptions (**K1**)–(**K3**) are usual in density estimation and are not very restrictive. Note that the Hölder property of the kernel is often needed when dealing with probabilities of small balls. Condition (**D3**) is needed for covariance calculus and takes a similar form than that used in complete data under dependence. Hypothesis (**H**) i) allows us to establish Lemma 3.3, while (**H**) ii) was used in Ferraty et al. (2005) and is needed to establish Lemma 3.4.*

Remark 3.3. *Here, we point out that many authors (see Stute (1993)) used the milder condition $a_G \leq a_F$ with additional integrability conditions. Our choice of strict inequality $a_G < a_F$ is motivated by the rate of convergence of the Lynden-Bell estimator $G_n(\cdot)$ to $G(\cdot)$, where we have to consider a set of values of Y_i which do not include a_G (a uniform rate for G_n with $a > a_G$ has been stated by Woodroofe (1985)).*

Now, our main result is the strong uniform consistency with rates of the kernel density estimator $\hat{f}_n(\cdot)$. A direct application to the mode estimate is given in Corollary 3.1.

Theorem 3.1. *Assume that hypotheses (**K1**)–(**K3**), (**M1**)–(**M3**), (**D2**), (**D3**), and (**H**) hold, then we have*

$$\sup_{y\in\mathcal{D}} \left| \hat{f}_n(y) - f(y) \right| = O\left\{ \max\left(\left(\frac{\log n}{nh}\right)^{1/2}, h^2 \right) \right\}, \quad \mathbf{P} - a.s.$$

Remark 3.4. *We point out that there is no result dealing with the bandwidth optimal choice, with respect to the almost sure uniform convergence criterion, for the density estimation, under random left-truncation model.*

Remark 3.5. *If we choose* $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$*, which is the optimal choice with respect to the almost sure uniform convergence criterion in the density estimation (see Stute (1982) and Stone (1983)), we get*

$$\sup_{y\in\mathcal{D}}\left|\hat{f}_n(y) - f(y)\right| = O\left(\left(\frac{\log n}{n}\right)^{2/5}\right) \quad , \quad \mathbf{P}-a.s.$$

which is the same optimal rate as the one established in the complete and iid case (see Vieu (1996), Lemma A-2).

As an application of Theorem 3.1, we obtain

Corollary 3.1. *Under the hypotheses of Theorem 3.1 and (**D1**), we have*

$$\hat{\theta}_n - \theta = O\left\{\max\left(\left(\frac{\log n}{nh}\right)^{1/4}, h\right)\right\} \quad , \quad \mathbf{P}-a.s.$$

Furthermore, if $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$*, then*

$$\hat{\theta}_n - \theta = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right), \quad \mathbf{P}-a.s.$$

Remark 3.6. *The choice* $h = O\left(\left(\frac{\log n}{n}\right)^{1/5}\right)$ *satisfies the condition (**H**) ii).*

3.5 Proofs

Proof of Theorem 3.1. The proof is based on decomposition (3.13) hereafter, and is broken into proofs of the following Lemmas.
First, observe that

$$\begin{aligned}\left|\hat{f}_n(y) - f(y)\right| &\leq \left|\mathbf{E}\left[\tilde{f}_n(y)\right] - f(y)\right| + \left|\hat{f}_n(y) - \tilde{f}_n(y)\right| + \left|\tilde{f}_n(y) - \mathbf{E}\left[\tilde{f}_n(y)\right]\right| \\ &=: \mathcal{J}_{1n} + \mathcal{J}_{2n} + \mathcal{J}_{3n}.\end{aligned} \tag{3.13}$$

Lemma 3.1. *Assume that Hypotheses (**K2**) and (**D2**) hold, then*

$$\sup_{y\in\mathcal{D}}\left|\mathbf{E}\left[\tilde{f}_n(y)\right] - f(y)\right| = O(h^2).$$

Proof. The asymptotic behavior of $\mathcal{J}_{1n}$ is standard, in the sense that it is not affected by the dependence structure. Indeed, using a change of variable and a Taylor expansion, we have

$$\begin{aligned}\mathbf{E}\left[\tilde{f}_n(y)\right] - f(y) &= \frac{1}{h}\int \frac{\mu}{G(u)} K\left(\frac{y-u}{h}\right) f^*(u)du - f(y)\\ &= \int K(u)\frac{(hu)^2}{2} f^{(2)}(\varsigma)du,\end{aligned}$$

with ς is between $y-uh$ and y. Thus

$$\left|\mathbf{E}\left[\tilde{f}_n(y)\right] - f(y)\right| \le \frac{h^2}{2}\sup_{y\in\mathcal{D}}\left|f^{(2)}(y)\right|\int u^2K(u)du.$$

Under the given conditions, the result holds. ■

To bound $\mathcal{J}_{2n}$, we state a rate of convergence for $\hat{\mu}_n$ under $\alpha-$mixing hypothesis, which is interesting in itself, similar to the one established for the iid case in He and Yang (1998). We have the following :

Lemma 3.2. *Under Conditions (**M1**)–(**M3**), we have*

$$|\hat{\mu}_n - \mu| = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{2}}\right).$$

Proof. First, observe that

$$\begin{aligned}\hat{\mu}_n - \mu &= \frac{G_n(y)[(1-F_n(y^-)]}{R_n(y)} - \frac{G(y)[(1-F(y)]}{R(y)}\\ &= \frac{1}{R_n(y)R(y)}\left\{R(y)\Big(\underbrace{G_n(y)-G(y)}_{\mathcal{I}_{1n}}\Big)\left[1-F_n(y^-)\right] + R(y)G(y)\Big(\underbrace{F(y)-F_n(y^-)}_{\mathcal{I}_{2n}}\Big)\right.\\ &+ \left. G(y)\Big(\underbrace{R_n(y)-R(y)}_{\mathcal{I}_{3n}}\Big)\left[F(y)-1\right]\right\}\end{aligned}$$

and from (3.5) and (3.6), we have

$$\begin{aligned}|\mathcal{I}_{3n}| &\le |F_n^*(y) - F^*(y)| + |G_n^*(y) - G^*(y)|\\ &=: \mathcal{I}_{31n} + \mathcal{I}_{32n}.\end{aligned}$$

On the one hand, from Theorem 3.2 in Cai and Roussas (1992), (**M2**) and the classical Law of the Iterated Logarithm, we have,

$$\mathcal{I}_{31n} = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{2}}\right) \quad \text{and} \quad \mathcal{I}_{32n} = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{\frac{1}{2}}\right). \tag{3.14}$$

Thus

$$\sup_{y\in D} |\mathcal{I}_{3n}| = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right). \tag{3.15}$$

On the other hand, using Remark 6 in Woodroofe (1985), we have

$$\sup_{y\in D} |\mathcal{I}_{1n}| = O_{\mathbf{P}}(n^{-1/2}). \tag{3.16}$$

To end the proof, it suffices to apply result $A-13$ in Sun and Zhou (2001) (see the Appendix), which gives

$$\sup_{y\in \mathcal{D}} |\mathcal{I}_{2n}| = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right). \tag{3.17}$$

Hence, from the continuity of F and equations (3.15)$-$(3.17), we obtain

$$|\hat{\mu}_n - \mu| = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right). \tag{3.18}$$

For the terms $\mathcal{J}_{2n}$ and $\mathcal{J}_{3n}$, we have the following results :

Lemma 3.3. *Assume that Assumptions (**K1**)$-$(**K3**), (**D2**) and (**H**) i) hold, then*

$$\sup_{y\in\mathcal{D}} \left|\hat{f}_n(y) - \tilde{f}_n(y)\right| = O_{\mathbf{P}}\left(\left(\frac{\log\log n}{n}\right)^{1/2}\right).$$

Proof. We write

$$\begin{aligned}\left|\hat{f}_n(y) - \tilde{f}_n(y)\right| &\le \frac{|\hat{\mu}_n - \mu|}{nh}\sum_{i=1}^{n}\frac{1}{G_n(Y_i)}K\left(\frac{y-Y_i}{h}\right) + \frac{\mu}{nh}\sum_{i=1}^{n}\left|\frac{1}{G_n(Y_i)} - \frac{1}{G(Y_i)}\right| K\left(\frac{y-Y_i}{h}\right)\\ &\le \left\{\frac{|\hat{\mu}_n - \mu|}{G_n(a_F)} + \frac{\mu}{G_n(a_F)G(a_F)}\sup_{y\in\mathcal{D}}\ |G_n(y) - G(y)|\right\} f_n^*(y).\end{aligned}$$

Since $G_n(a_F) \overset{\mathbf{P}-a.s.}{\longrightarrow} G(a_F) > 0$, and using (3.16) and (3.18), the proof is achieved provided that $f_n^*(y) = O_{\mathbf{P}}(1)$. Moreover, using integration by parts, condition (**K3**) (this implies that the derivative $K^{(1)}$ is integrable) and a change of variable, we have

$$\begin{aligned}|f_n^*(y)| &\le |f_n^*(y) - \mathbf{E}\left[f_n^*(y)\right]| + |\mathbf{E}\left[f_n^*(y)\right]|\\ &= \frac{1}{h}\left|\int K\left(\frac{y-u}{h}\right)\{dF_n^*(u) - dF^*(u)\}\right| + \frac{1}{h}\int K\left(\frac{y-u}{h}\right) f^*(u)du\\ &\le \frac{1}{h}\sup_{y\in\mathcal{D}}|F_n^*(y) - F^*(y)| \int \left|K^{(1)}(u)\right| du + \frac{1}{\mu h}\int G(u)K\left(\frac{y-u}{h}\right) f(u)du\\ &\le \frac{C}{h}\sup_{y\in\mathcal{D}}|F_n^*(y) - F^*(y)| + C'\int K(u)\, f(y-uh)du.\end{aligned}$$

The result follows by using (3.14), a Taylor expansion and Assumptions (**K1**), (**K2**), (**D**2) and (**H**) i). ■

Lemma 3.4. *Assume that Hypotheses (**K1**), (**M1**), (**M3**), (**D2**), (**D3**) and (**H**) hold, then*

$$\sup_{y\in\mathcal{D}}\left|\tilde{f}_n(y)-\mathbf{E}\left[\tilde{f}_n(y)\right]\right| = O_{\mathbf{P}}\left(\left(\frac{\log n}{nh}\right)^{1/2}\right).$$

Proof. The proof is based on the following observation : The compact set $\mathcal{D}$ can be covered by a finite number q_n of intervals of half length $\omega_n = \left(n^{-1}h^{1+2\beta}\right)^{\frac{1}{2\beta}}$, where β is the Hölder exponent. Let $I_k := I(y_k, \omega_n); 1 \le k \le q_n$, denote each interval centered at some point y_k.
Since $\mathcal{D}$ is bounded, one can find $M > 0$ such that $\omega_n q_n \le M$. For any $y \in \mathcal{D}$, there exists an interval I_k which contains y such that :

$$|y - y_k| \le \omega_n. \tag{3.19}$$

Set

$$\Delta_i(y) = \frac{\mu}{nh}\left\{\frac{1}{G(Y_i)}K\left(\frac{y-Y_i}{h}\right) - \mathbf{E}\left[\frac{1}{G(Y_i)}K\left(\frac{y-Y_i}{h}\right)\right]\right\}.$$

Obviously, we have

$$\begin{aligned}\sum_{i=1}^n \Delta_i(y) &= \left(\left(\tilde{f}_n(y)-\tilde{f}_n(y_k)\right) - \left(\mathbf{E}\left[\tilde{f}_n(y)\right]-\mathbf{E}\left[\tilde{f}_n(y_k)\right]\right)\right) + \left(\tilde{f}_n(y_k)-\mathbf{E}\left[\tilde{f}_n(y_k)\right]\right)\\ &=: \sum_{i=1}^n \tilde{\Delta}_i(y) + \sum_{i=1}^n \Delta_i(y_k).\end{aligned}$$

Hence

$$\sup_{y\in\mathcal{D}}\left|\sum_{i=1}^n \Delta_i(y)\right| \le \max_{1\le k\le q_n}\ \sup_{y\in I(y_k,\omega_n)}\underbrace{\left|\sum_{i=1}^n \tilde{\Delta}_i(y)\right|}_{\mathcal{S}_1} + \max_{1\le k\le q_n}\underbrace{\left|\sum_{i=1}^n \Delta_i(y_k)\right|}_{\mathcal{S}_2}.$$

Observe now that

$$\begin{aligned}\mathcal{S}_1 &\le \frac{1}{nh}\sum_{i=1}^n \frac{\mu}{G(Y_i)}\left|K\left(\frac{y-Y_i}{h}\right)-K\left(\frac{y_k-Y_i}{h}\right)\right|\\ &\quad + \mathbf{E}\left[\frac{\mu}{G(Y_1)}\frac{1}{h}\left|K\left(\frac{y-Y_1}{h}\right)-K\left(\frac{y_k-Y_1}{h}\right)\right|\right]\\ &=: \mathcal{S}_{1n}+\mathcal{S}_{2n}.\end{aligned}$$

Assumption (**K1**) and (3.19), yield

$$\begin{aligned}\mathcal{S}_{1n} &\leq \frac{\mu\,|y-y_k|^{\beta}}{G(a_F)h^{\beta+1}}\\ &\leq C\omega_n^{\beta}h^{-(\beta+1)} = O((nh)^{-1/2}).\end{aligned}$$

Similar arguments as above lead to the same bound for $\mathcal{S}_{2n}$. Hence, by (**H**) $i)$ and for n large enough, we get $\mathcal{S}_1 = o_{\mathbf{P}}(1)$.
In order to study the term $\mathcal{S}_2$, we use an exponential inequality involving the $\alpha-$mixing structure as the main argument. Set $\xi_i = nh\Delta_i(y_k)$. Using the condition (**K1**), we clearly obtain

$$|\xi_i| \leq \frac{2\mu\,\|K\|_{\infty}}{G(a_F)} < \infty.$$

The use of the well known Fuk-Nagaev's inequality (see Rio (2000), formula 6.19b, page 87), slightly modified in Ferraty and Vieu (2006) (see Proposition A.11$-ii$), page 237), allows to have, for all $\varepsilon > 0$ and $r > 1$:

$$\mathbf{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(y_k)\right| > \varepsilon\right\} \leq C\left\{\left(1+\frac{\varepsilon^2n^2h^2}{rs_n}\right)^{-\frac{r}{2}} + nr^{-1}\left(\frac{r}{\varepsilon nh}\right)^{\nu+1}\right\}, \tag{3.20}$$

where

$$s_n = \sum_{1\leq i\leq n}\sum_{1\leq j\leq n}|Cov(\xi_i,\xi_j)|\,.$$

Remark now that we can write

$$s_n = \underbrace{\sum_{i\neq j}|Cov(\xi_i,\xi_j)|}_{s_n^{cov}} + \underbrace{\sum_{i=1}^{n}Var\,(\xi_i)}_{s_n^{var}}.$$

On the one hand, by (**K1**), (**D2**) and a change of variable, we obtain

$$\begin{aligned}s_n^{var} &= nVar(\xi_1)\\ &= n\left\{\mathbf{E}\left[\frac{\mu^2}{G^2(Y_1)}K^2\left(\frac{y_k-Y_1}{h}\right)\right] - \mathbf{E}^2\left[\frac{\mu}{G(Y_1)}\left(\frac{y_k-Y_1}{h}\right)\right]\right\}\\ &= O(nh).\end{aligned} \tag{3.21}$$

Furthermore, by a change of variable, (**K1**), (**M1**) and (**D3**) we obtain

$$\begin{aligned}|Cov(\xi_i,\xi_j)| &= |\,\mathbf{E}\,[\xi_i\xi_j]|\\ &\leq \frac{\mu^2}{G^2(a_F)}\int\int K(\frac{y_k-u}{h})K(\frac{y_k-v}{h})\left|f_{ij}^{*}\,(u,v) - f^{*}(u)f^{*}(v)\right|dudv\\ &= O(h^2).\end{aligned} \tag{3.22}$$

Note that these covariances can be controlled by the mean of the usual Davydov covariance inequality stated for mixing processes (see Rio (2000), formula 1.12a, page 10 ; or Bosq (1998), formula 1.11, page 22). Clearly, we have

$$|Cov(\xi_i,\xi_j)| \leq C\alpha\left(|i-j|\right). \tag{3.23}$$

To evaluate the term s_n^{cov}, we use a technique developed by Masry (1986). Taking $\varphi_n = \left\lceil \frac{1}{h\log n}\right\rceil$ (where $\lceil\cdot\rceil$ denotes the smallest integer greater than the argument), we can write

$$s_n^{cov} = \sum_{0<|i-j|\leq\varphi_n} |Cov(\xi_i,\xi_j)| + \sum_{|i-j|>\varphi_n} |Cov(\xi_i,\xi_j)|. \tag{3.24}$$

Applying the upper bound in (3.22) to the first term in (3.24), we have

$$\sum_{0<|i-j|\leq\varphi_n} |Cov(\xi_i,\xi_j)| \leq Cnh^2\varphi_n. \tag{3.25}$$

In order to control the second term, thanks to (**M3**) and use (3.23) to obtain

$$\begin{aligned}\sum_{|i-j|>\varphi_n} |Cov(\xi_i,\xi_j)| &\leq C\sum_{|i-j|>\varphi_n}\alpha\left(|i-j|\right)\\ &\leq Cn^2\alpha\left(\varphi_n\right).\end{aligned} \tag{3.26}$$

According to the right-hand side of (**H**) *ii*), there exists $\zeta>0$, such that $\log^\nu n = o(n^\zeta)$ and

$$h^{\nu-1} = O(n^{-1-\zeta}). \tag{3.27}$$

Using (**M3**) and (3.25)$-$(3.27), we get

$$s_n^{cov} = O(nh). \tag{3.28}$$

Thus, by (3.21) and (3.28), it arises that

$$s_n = O(nh). \tag{3.29}$$

Therefore, putting $\varepsilon = \varepsilon_0\sqrt{\frac{\log n}{nh}}$ and replacing (3.29) in (3.20), we obtain

$$\begin{aligned}&\mathbf{P}\left\{\left|\sum_{i=1}^n \Delta_i(y_k)\right| > \varepsilon_0\sqrt{\frac{\log n}{nh}}\right\}\\ \leq\ & C\left\{\left(1+\frac{C'\varepsilon_0^2\log n}{r}\right)^{-\frac{r}{2}} + nr^{-1}\left(\frac{r}{\varepsilon_0\sqrt{nh\log n}}\right)^{\nu+1}\right\}\\ \leq\ & C\exp\left[-\frac{r}{2}\log\left(1+\frac{C'\varepsilon_0^2\log n}{r}\right)\right] + nr^{-1}\left(\frac{r}{\varepsilon_0}\right)^{\nu+1}(nh\log n)^{-\frac{\nu+1}{2}}.\end{aligned} \tag{3.30}$$

By taking $r = (\log n)^{1+\delta}$, $(\delta > 0)$ and using Taylor series expansion of $\log(1 + x)$, inequality (3.30) becomes

$$\mathbf{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(y_k)\right| > \varepsilon_0\sqrt{\frac{\log n}{nh}}\right\} \leq Cn^{-C'\varepsilon_0^2} + C\varepsilon_0^{-(\nu+1)}(\log n)^{\nu(1+\delta)}n^{1-\frac{\nu+1}{2}}h^{-\frac{\nu+1}{2}}.$$

Consequently

$$\begin{aligned}
\mathbf{P}\left\{\max_{1\leq k\leq q_n}\left|\sum_{i=1}^{n}\Delta_i(y_k)\right| > \varepsilon\right\} &\leq \sum_{i=1}^{q_n}\mathbf{P}\left\{\left|\sum_{i=1}^{n}\Delta_i(y_k)\right| > \varepsilon\right\}\\
&\leq M\omega_n^{-1}\left(Cn^{-C'\varepsilon_0^2} + C\varepsilon_0^{-(\nu+1)}(\log n)^{\nu(1+\delta)}n^{1-\frac{\nu+1}{2}}h^{-\frac{\nu+1}{2}}\right)\\
&\leq C\left\{\underbrace{\frac{n^{\frac{1}{2\beta}-C'\varepsilon_0^2}}{h^{\left(1+\frac{1}{2\beta}\right)}}}_{\mathcal{S}_{21n}} + \underbrace{\frac{n^{1-\frac{\nu+1}{2}+\frac{1}{2\beta}}(\log n)^{\nu(1+\delta)}}{\varepsilon_0^{(\nu+1)}h^{1+\frac{\nu+1}{2}+\frac{1}{2\beta}}}}_{\mathcal{S}_{22n}}\right\}.
\end{aligned}$$

From the left-hand side of the condition $(\mathbf{H})$ $ii)$, it follows then that

$$\begin{aligned}
\mathcal{S}_{22n} &\leq C(\log n)^{\nu(1+\delta)}n^{1-\frac{\nu+1}{2}+\frac{1}{2\beta}}n^{-\frac{3-\nu}{2}-\eta\frac{\beta(\nu+1)+2\beta+1}{2\beta}}\\
&\leq C(\log n)^{\nu(1+\delta)}n^{-1-\frac{\eta}{2\beta}(\beta(\nu+1)+2\beta+1-\frac{1}{\eta})}.
\end{aligned}$$

So, for any η as in $(\mathbf{H})$ $ii)$, $\mathcal{S}_{22n}$ is bounded by the term of a finite-sum series. In the same way for $\mathcal{S}_{21n}$, which by an appropriate choice of ε_0 can be made $O(n^{-3/2})$, which in turn is the general term of convergent series. Hence, $\sum_{n\geq 1}(\mathcal{S}_{21n} + \mathcal{S}_{22n}) < \infty$, and therefore the result follows by applying Borel-Cantelli Lemma. ■

Proof of Corollary 3.1. Observe that

$$\begin{aligned}
\left|f(\hat{\theta}_n) - f(\theta)\right| &\leq \left|f(\hat{\theta}_n) - \hat{f}_n(\hat{\theta}_n)\right| + \left|\hat{f}_n(\hat{\theta}_n) - f(\theta)\right|\\
&\leq \sup_{y\in\mathcal{D}}\left|\hat{f}_n(y) - f(y)\right| + \left|\hat{f}_n(\hat{\theta}_n) - f(\theta)\right|.
\end{aligned}$$

Since

$$\left|\hat{f}_n(\hat{\theta}_n) - f(\theta)\right| = \left|\sup_{y\in\mathcal{D}}\hat{f}_n(y) - \sup_{y\in\mathcal{D}}f(y)\right| \leq \sup_{y\in\mathcal{D}}\left|\hat{f}_n(y) - f(y)\right|,$$

then we have

$$\left|f(\hat{\theta}_n) - f(\theta)\right| \leq 2\sup_{y\in\mathcal{D}}\left|\hat{f}_n(y) - f(y)\right|. \tag{3.31}$$

Making use of Theorem 3.1 and $(\mathbf{D1})$, we obtain the convergence of $\hat{\theta}_n$ to θ almost surely.

A Taylor series expansion of $f(.)$ in a neighborhood of θ gives

$$f(\hat{\theta}_n) - f(\theta) = \frac{1}{2}(\hat{\theta}_n - \theta)^2 f^{(2)}(\theta^*),$$

where θ^* is between $\hat{\theta}_n$ and θ. Then, by (3.31), we have

$$\left(\hat{\theta}_n - \theta\right)^2 \left|f^{(2)}(\theta^*)\right| \leq 4 \sup_{y \in \mathcal{D}} \left|\hat{f}_n(y) - f(y)\right|.$$

Hence, by Theorem 3.1, the convergence of $\hat{\theta}_n$ to θ and (**D2**), we complete the proof of Corollary 3.1. ■

Bibliographie

[1] Abraham, C., Biau, G., Cadre, B. (2004). On the asymptotic properties of a simple estimate of the mode. *ESAIM, Probab. Statist.* 8 :1−11.

[2] Bickel, D.R., Frühwirth, R. (2006). On a fast, robust estimator of the mode : Comparaisons to other robust estimators with applications. *Computat. Statist. Data Anal.* 50 :3500−3530.

[3] Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes : Estimation and Prediction* (Second Edition). Lecture Notes in Statistics 110, Springer−Verlag, Berlin.

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

[5] Cai, Z. (2001). Estimating a Distribution Function for Censored Time Series Data, *J. Multivariate Anal.* 78 :299−318.

[6] Cai, Z., Roussas, G. G. (1992). Uniform strong estimation under α-mixing, with rates, *Statist. Probab. Lett.* 15 :47−55.

[7] Collomb, G., Härdle, W., Hassani, S. (1987). A note on prediction via conditional mode estimation, *J. Statist. Plann. Inference* 15 :227−236.

[8] Dabo-Niang, S., Laksaci, A. (2007). Estimation non paramétrique du mode conditionnel pour variable explicative fonctionnelle, *C. R. Acad. Sci. Paris, Ser. I* 344 :49−52.

[9] Dabo-Niang, S., Ferraty, F., Vieu, P. (2004). Estimation du mode dans un espace vectoriel semi−normé, *C. R. Acad. Sci. Paris, Ser. I* 339 :659–662.

[10] Davydov, A. (1970). The invariance principle for stationary processes. *Theory Probab. Appl.* 15 :487−498.

[11] Doukhan, P. (1994). *Mixing : Properties and Examples.* Lecture Notes in Statistics, Springer−Verlag.

[12] Eddy, W.F. (1980). Optimum kernel estimators of the mode. *Ann. Statist.* 8 :870−882.

[13] Eddy, W.F. (1982). The asymptotic distributions of kernel estimators of the mode. *Z. Wahrsh. Verw. Gebiete.* 59 :279−290.

[14] Ferraty, F., Vieu, P. (2006). *Nonparametric Functional Data Analysis, Theory and Practice.* Springer−Verlag.

[15] Ferraty, F., Laksaci, A., Vieu, P. (2005), Functional times series prediction via conditional mode, *C. R. Math. Acad. Sci. Paris, Ser. I* 340 :389−392.

[16] He, S., Yang, G.L. (1998). Estimation of the truncation probability in the random truncation model. *Ann. Statist.* 26 :1011−1028.

[17] Hermann, E., Ziegler, K. (2004). Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions. *Statist. Probab. Lett.* 72 :359−368.

[18] Louani, D. (1998). On the asymptotic normality of the kernel estimators of the density function and its derivatives under random censoring. *Comm. Statist. Theory* & *Meth.* 27 :2909−2924.

[19] Louani, D., Ould-Saïd, E. (1999). Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis. *Nonparametric Statist. J.* 11 :413−442.

[20] Lynden-Bell, D. (1971). A method of allowing for known observational selection in small samples applied to 3CR quasars, *Monthly Notices Roy. Astronom. Soc.* 155 :95−118.

[21] Masry, E. (1986). Recursive probability density estimation for weakly dependent processes. *IEEE Transactions on Information Theory* 32 :254−267.

[22] Mokkaddem, A., Pelletier, M. (2005). Moderate deviations for the kernel mode estimator and some applications. *J. Statist. Plann. Inference* 135 :276−299.

[23] Mokkaddem, A., Pelletier, M., Worms, J. (2006). A large deviations upper bound for the kernel mode estimator. *Theory Probab. Appl.* 50 :153−165.

[24] Nadaraya, E.N. (1965). On nonparametric estimates of density functions and regression curves. *Theory Probab. Appl.* 10 :186−190.

[25] Ould-Saïd, E. (1993). Estimation non paramétrique du mode conditionnel. Application à la prévision. *C. R. Acad. Sci. Paris. Ser. I* 316 :943−947.

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

[27] Ould-Saïd, E., Cai, Z. (2005). Strong uniform consistency of the nonparametric estimation of the censored conditional mode function, *Nonparametric Statist. J.* 17 :797−806.

[28] Ould-Saïd, E., Lemdani, M. (2006). Asymptotic properties of the kernel estimators of the regression under random left truncation. *Ann. Inst. Statist. Math.* 58 :357−378.

[29] Ould-Saïd, E., Tatachak, A. (2005). On the nonparametric estimation of the simple mode under left−truncation model. Technical Report L.M.P.A. n^o 271. Univ. du Littoral Cote d'Opale.

[30] Ould-Saïd, E., Tatachak, A. (2007). Asymptotic properties of the kernel estimator of the conditional mode for the left-truncated model, *C. R. Acad. Sci. Paris, Ser. I* 344 :651−656.

[31] Parzen, E. (1962). On estimation of a probability density function and mode. *Ann. Math. Statist.* 33 :1065−1076.

[32] Rio, E. (2000). *Théorie asymptotique des processus aléatoires faiblement dépendants* (In French). Mathématiques et Applications 31, Springer-Verlag.

[33] Romano, J. P. (1988). On weak convergence and optimality of kernel density estimates of the mode. *Ann. Statist.* 16 :629−647.

[34] Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proc. Nat. Acad. Sci.* U.S.A. 42 :43−47.

[35] Samanta, M. (1973). Nonparametric estimation of the mode of a multivariate density. *South African Statist. J.* 7 :109−117.

[36] Samanta, M., Thavaneswaran, A. (1990). Nonparametric estimation of the conditional mode. *Comm. Statist. Theory* & *Meth.* 16 :4515−4524.

[37] Stone, C.J. (1983). Optimal uniform rate of convergence for nonparametric estimators of a density function or its derivatives, *Recent Advances in Statistics : Papers presented in Honor of Herman Chernoff's Sixtieth Birthday, M.H. Rizvi, J.S. Rustagi and D. Siegmund, (eds.)* 393−406, Academic Press, New York.

[38] Stute, W. (1982). A law of the logarithm for kernel density estimators, *Ann. Probab.* 10 :414−422.

[39] Stute, W. (1993). Almost sure representation of the product-limit estimator for truncated data. *Ann. Statist.* 21 :146−156.

[40] Sun, L., Zhou, X. (2001). Survival function and density estimation for truncated dependent data. *Statist. Probab. Lett.* 52 :47−57.

[41] Van Ryzin, J. (1969). On strong consistency of density estimates, *Ann. Math. statist.* 40 :1765−1772.

[42] Vieu, P. (1996). A note on density mode estimation. *Statist. Probab. Lett.* 26 :297−307.

[43] Woodroofe, M. (1985). Estimating a distribution function with truncated data. *Ann. Statist.* 13 :163−177.

[44] Zhou, Y. (1996). A note on the TJW product limit estimator for truncated and censored data. *Statist. Probab. Lett.* 26 :381−387.

Chapitre 4

Mode conditionnel : Cas i.i.d.

Prediction from randomly left-truncated model via estimation of the conditional mode function[§]

Ould-Saïd, E., Tatachak, A.

Abstract

Let $(Y_N)_{N\geq 1}$ be a sequence of independent and identically distributed random variables of interest and $(\mathbf{X}_N)_{N\geq 1}$ be a sequence of covariates taking values in $\mathbb{R}^d$. Let $\Theta(\mathbf{x})$ be the conditional mode of Y given $\mathbf{X} = \mathbf{x}$, in this paper, we define a new kernel estimator $\hat{\Theta}_n(\mathbf{x})$ of $\Theta(\mathbf{x})$ when Y is subject to random left-truncation, then we establish the strong uniform consistency of the estimate with rates and state its asymptotic normality. Some simulations are given to show how the estimator behaves for finite samples.

Keywords Asymptotic normality, conditional mode function, kernel estimator, Lynden-Bell estimator, random left-truncation model, rate of convergence, uniform almost sure convergence, V-C class.

Mathematics Subject Classification. 62G05 ; 62G07 ; 62G20.

[§]A condensed version of this document has been published in the CRAS (2007).

4.1 Introduction

Let Y and T be independent random variables (r.v.) with distribution functions F and G respectively, both assumed to be continuous. Let $(Y_1, T_1), (Y_2, T_2), ..., (Y_N, T_N)$ be independent and identically distributed (i.i.d.) as (Y, T), where the population size N is deterministic, but unknown. In the random left-truncation model, the random variable (rv) of interest Y is interfered by the truncation rv T, such that both quantities Y_i and T_i are observable only if $Y_i \geq T_i$ whereas neither is observed if $Y_i < T_i$. Without possible confusion, we still denote (Y_i, T_i), $i = 1, 2, ..., n; (n \leq N)$ the actually observed r.v. These observed r.v. are still i.i.d. (see Proposition 1 in [14]). As a consequence of truncation, the size of the actually observed sample, n, is a $Bin(N, \mu)$ random variable, with $\mu := \mathbb{P}(Y \geq T)$. It is clear that if $\mu = 0$, no data can be observed and therefore, we suppose through this paper that $\mu \neq 0$. By the strong law of large numbers (SLLN) we have, as $N \longrightarrow +\infty$

$$\tilde{\mu}_n := \frac{n}{N} \longrightarrow \mu, \ \mathbb{P}\text{−a.s.} \tag{4.1}$$

Truncation frequently occurs in medical studies, when one wants to study the length of survival after the start of the disease : If Y denotes the elapsed time between the onset of the disease and death, and if the follow-up period starts T units of time after the onset of the disease, then clearly Y is left truncated by T. For examples, an important feature of AIDS development is the induction period between infection with the AIDS virus and the onset of clinical AIDS. The data collected from persons infected from contaminated blood transfusions provide a unique source of information for the induction period. Of person infected in this way, only those who have developed AIDS can be identified. Truncated data are also common in actuarial, astronomic, demographic, epidemiologic, reliability testing and other studies (More examples and references of truncated data can be found in Woodroofe [37], Wang et al. [36], Tsai et al. [32], Wang [36], Anderson et al. [2], He [9] and Chen et al. [3]).
Let $\mathbf{X}$ be a random vector $\mathbb{R}^d$-valued and for any $\mathbf{x} = {}^t(x_1, x_2, ..., x_d)$, we denote by $g(\cdot|\mathbf{x})$ the conditional probability density function of Y given $\mathbf{X} = \mathbf{x}$, which can be written as $g(y|\mathbf{x}) = \frac{f(\mathbf{x},y)}{\ell(\mathbf{x})}$, where $f(\cdot, \cdot)$ is the joint probability density function of $(\mathbf{X}, Y)$ and $\ell(\cdot)$ is the marginal density of $\mathbf{X}$. Assuming that $g(\cdot|\mathbf{x})$ has a unique mode at $\Theta(\mathbf{x})$, the conditional mode function of Y given $\mathbf{X} = \mathbf{x}$, is defined by the following equation

$$g(\Theta(\mathbf{x})|\mathbf{x}) = \max_{-\infty<y<\infty} g(y|\mathbf{x}). \tag{4.2}$$

The problem of estimating the unconditional/conditional mode of a probability density has a long history in the statistics literature, and number of distinguished papers deal with this topic. There are among others, for example, Parzen [26] established weak convergence and asymptotic normality for the i.i.d. case, and the strong consistency was obtained by Nadaraya [21] and Van Ryzin [33]. Using classical techniques of weak convergence, Eddy [6] derived the asymptotic normality, under weaker conditions than those imposed by Parzen [26], Chernoff [4] studied the naive estimator of the mode, defined as the center of that interval which contains the most observations ; see also Eddy [7]. The multidimensional versions of these results were obtained by Samanta [28], Konakov [13], and recently Abraham et al. [1] considered the estimate of the mode as an element of the sample points, and obtained the strong consistency. Mokkadem and Pelletier [20] studied the law of the iterated logarithm of the kernel mode estimator, and Hermann and Ziegler [12] obtained rates of consistency in absence of smoothness.

The recent developments focus on the nonparametric estimation of the conditional mode. For example, Romano [27] investigated the asymptotic behavior of the kernel estimate of the conditional mode, with data-dependent bandwidths, and obtained results under weaker smoothness assumptions on the underlying density $\ell(\cdot)$. Vieu [34] obtained a rate of convergence for both local and global estimates of the mode function. For the right random censoring, Louani [16] studied the asymptotic normality of the kernel estimator of the mode.

When conditioning by one of the coordinates of a bi-dimensional random vector in the i.i.d. random vector case, Samanta and Thavaneswaran [29] showed that under some regularity conditions, the kernel estimator of the conditional mode function is consistent, and asymptotically normally distributed. Mehra et al. [19] established the law of iterated logarithm, uniform almost sure convergence over a compact set, and asymptotic normality of smoothed rank nearest neighbor estimator of the conditional mode function.

Recently, Ould-Saïd and Cai [25] established in the i.i.d. case, the uniform strong consistency of nonparametric estimation of the censored conditional mode function.

For the dependent case, the strong consistency of conditional mode estimator was established under ϕ-mixing condition by Collomb et al. [5] and their results can be applied to process forecasting. In the α-mixing case, the strong consistency over a compact set and the asymptotic normality were obtained by Ould-Saïd [22] and Louani and Ould-Saïd [17] respectively. In a general ergodic framework, a process prediction via the conditional mode estimation was described by Ould-Saïd [23], and its strong consistency was obtained.

It is well known that the conditional mode function, provides an alternative procedure to prediction in contract to classical regression function. More especially, there exist conditional densities for which the regression function is anywhere vanished, so that is not suitable in problems involving process prediction. Hence, it is the interest to have alternative methods at one's disposal.

At the best of our knowledge, the problem of estimating the conditional mode function under random left-truncation has not been addressed in the statistics literature. This is the object of central interest of this paper. We define a new estimator of our interest parameter $\Theta(\mathbf{x})$ for that, we establish over a compact set, the uniform almost sure convergence based on a conditional density, and then we settle its asymptotic normality.

The paper is organized as follows. In section 2 we recall some important and useful results in relation with the aim of this article. The nonparametric conditional density estimate is exposed in section 3. In section 4 we first introduce the proposed estimator, followed regularity assumptions and main results. Some simulations are given in section 5 to illustrate the estimate's asymptotic behavior. Finally, the proofs of the main results are postponed to section 6 with some auxiliary results with their proofs.

4.2 Preliminaries

We first present some results existing in the literature for univariate left-truncation model, that allow the introduction and the definition of a new nonparametric estimator.
Since N is unknown and n is known (although random), our results won't be stated with respect to the probability measure $\mathbb{P}$ (related to the $N-$sample) but will involve the conditional probability $\mathbf{P}$ related to the $n-$sample). Note also that $\mathbb{E}$ and $\mathbf{E}$ denote the expectation operators related to $\mathbb{P}$ and $\mathbf{P}$, respectively. In the sequel, as a convention, we denote by a superscript $(^*)$ any distribution function that is associated with the truncated random variables.
Let Y be the variable of interest and T be the truncating variable, with distribution functions F and G respectively. Under the left-truncation sampling scheme, the conditional joint distribution of (Y, T) (Stute [31]) becomes :

$$\begin{aligned}\mathbf{V}^*(y,t) &= \mathbf{P}\{Y \leq y, T \leq t\} = \mathbb{P}\{Y \leq y, T \leq t \mid Y \geq T\} \\ &= \mu^{-1}\int_{-\infty}^{y} G(t\wedge v)dF(v),\end{aligned}$$

where $t \wedge v = \min(t, v)$. The marginal distributions are defined by

$$F^*(y) := \mathbf{V}^*(y,\infty) = \mu^{-1}\int_{-\infty}^{y} G(v)dF(v),$$

$$G^*(t) := \mathbf{V}^*(\infty,t) = \mu^{-1}\int_{-\infty}^{+\infty} G(t\wedge v)dF(v),$$

and let

$$R(y) = G^*(y) - F^*(y) = \mu^{-1}G(y)\left[1-F(y)\right]. \tag{4.3}$$

The corresponding empirical distributions are defined by

$$F_n^*(y) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(Y_i\leq y)},\quad G_n^*(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(T_i\leq t)},$$

and

$$R_n(y) = G_n^*(y) - F_n^*(y^-) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(T_i\leq y\leq Y_i)},\qquad -\infty<y<+\infty$$

respectively, where $\mathbb{1}_A$ denotes the indicator function of the set A.
Let, for any distribution function W, $a_W = \inf\{x : W(x) > 0\}$, $b_W = \sup\{x : W(x) < 1\}$, be the endpoints of the support of W. Assuming that $\{a_G \leq a_F; b_G \leq b_F\}$, we have

$$\mu := P(Y \geq T) = \int_{\mathbb{R}} G(v) dF(v) > 0. \tag{4.4}$$

Now, it is well known that the product-limit estimators for $F(y)$ and $G(y)$ are defined by

$$\bar{F}_n(y) = \prod_{s \leq y} \left[1 - \frac{F_n^*(s)}{R_n(s)}\right] \quad \text{and} \quad G_n(y) = \prod_{s > y} \left[1 - \frac{G_n^*(s)}{R_n(s)}\right], \tag{4.5}$$

respectively (see Lynden-Bell [18]).
Note that the estimator $\tilde{\mu}_n$ defined in (4.1) cannot serve to estimate μ defined in (4.4), since N is unknown. This situation requires to use the empirical estimator of μ. According to (4.3) and replacing F and G by their product-limit estimators and R by its empirical version, we have

$$\mu_n = \frac{G_n(y) \bar{F}_n(y)}{R_n(y)}, \tag{4.6}$$

for all y such that $R_n(y) > 0$. He and Yang [11], showed that μ_n is equivalent to the more familiar estimate $\tilde{\mu}_n = \int_{\mathbb{R}} G_n(u) dF_n(u)$ studied in the literature, and that (their Corollary 2.5)

$$\mu_n \overset{\mathbf{p-a.s}}{\longrightarrow} \mu \quad \text{as } n \longrightarrow \infty. \tag{4.7}$$

The fact that μ_n is independent of y and has a much simpler form than $\tilde{\mu}_n$, it will be useful as instrument in the construction of new estimates as to be carried out below.

4.3 Conditional densities of observed data.

Now, in addition to the considered previously variables Y and T, we consider a random vector $\mathbf{X} \in \mathbb{R}^d$ of covariates, assumed to be absolutely continuous with distribution function L, having density ℓ . Suppose that one observes the n triplets $(\mathbf{X}_i, Y_i, T_i)$ among N such that $(Y_i \geq T_i)$. From now on, T is assumed to be independent of $(\mathbf{X}, Y)$, and $(\mathbf{X} \leq \mathbf{x})$ stands for $(X_1 \leq x_1, X_2 \leq x_2, ..., X_d \leq x_d)$.
Given this model, the observed triplets can be considered to arise from the following trivariate conditional distribution H^* :

$$\begin{aligned} H^*(x, y, t) &= \mathbf{P}\{\mathbf{X} \leq \mathbf{x}, Y \leq y, T \leq t\} = \mathbb{P}\{\mathbf{X} \leq \mathbf{x}, Y \leq y, T \leq t \mid Y \geq T\} \\ &= \mu^{-1} \int_{\mathbf{u} \leq \mathbf{x}} \int_{a_G \leq v \leq y} G(v \wedge t) \mathbf{F}(\mathbf{du}, dv), \end{aligned} \tag{4.8}$$

where μ is as defined before, and $\mathbf{F}(\cdot, \cdot)$ is the joint distribution function of $(\mathbf{X}, Y)$. Taking $t = +\infty$, the observed pair then has the following distribution $\mathbf{F}^*(\cdot, \cdot)$:

$$\mathbf{F}^*(\mathbf{x}, y) = H^*(\mathbf{x}, y, \infty) = \mu^{-1} \int_{\mathbf{u} \leq \mathbf{x}} \int_{a_G \leq v \leq y} G(v) \mathbf{F}(\mathbf{du}, dv). \tag{4.9}$$

By differentiating (4.9), we get

$$\mathbf{F}(\mathbf{dx}, dy) = [\mu^{-1}G(y)]^{-1}\mathbf{F}^*(\mathbf{dx}, dy) \quad \text{for } y \geq a_G. \tag{4.10}$$

Hence

$$f(\mathbf{x}, y) = \left[\mu^{-1}G(y)\right]^{-1} f^*(\mathbf{x}, y). \tag{4.11}$$

Integrating (4.10) over y, we obtain the df of $\mathbf{X}$

$$L(\mathbf{x}) = \mu \int_{\mathbf{u}\leq\mathbf{x}} \int_{a_G \leq y} \frac{1}{G(y)} \mathbf{F}^*(\mathbf{du}, dy).$$

A natural estimator of $L(\mathbf{x})$ is then given by

$$\tilde{L}_n(\mathbf{x}) = \frac{\mu}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} 1\!\!1_{(\mathbf{X}_i \leq \mathbf{x})}. \tag{4.12}$$

Note that in (4.12) and in the sequel, the sum is taken only for i such that $G(Y_i) \neq 0$. Finally (4.12) yields the density estimator of $\mathbf{X}$ as

$$\tilde{\ell}_n(\mathbf{x}) := \frac{1}{a_n^d} \int_{\mathbb{R}^d} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{a_n}\right) \tilde{L}_n(\mathbf{du}) = \frac{\mu}{n a_n^d} \sum_{i=1}^{n} \frac{1}{G(Y_i)} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{a_n}\right), \tag{4.13}$$

where $K_d : \mathbb{R}^d \longrightarrow \mathbb{R}$ is a fixed kernel with $\int_{\mathbb{R}^d} K_d = 1$, and $(a_n)_{n\geq 1}$ a positive bandwidth sequence tending to zero as n grows to infinity.
Adopting the same methodology, and while observing (4.11), we get an estimator of $\mathbf{F}(\mathbf{x}, y)$ as follows.

$$\tilde{\mathbf{F}}_n(\mathbf{x}, y) = \frac{\mu}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} 1\!\!1_{(\mathbf{X}_i \leq \mathbf{x}, Y_i \leq y)}.$$

Now, according to (4.11), and in order to construct our estimator of the conditional density, we define the kernel estimate of the joint probability density function $f(\cdot, \cdot)$ as follows

$$\begin{aligned}
\tilde{f}_n(\mathbf{x}, y) &= \frac{1}{a_n^d b_n} \int_{\mathbb{R}^d \times \mathbb{R}} K_d\left(\frac{\mathbf{x}-\mathbf{u}}{a_n}\right) K_0\left(\frac{y-v}{b_n}\right) \tilde{\mathbf{F}}_n(\mathbf{du}, dv) \\
&= \frac{\mu}{n a_n^d b_n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{a_n}\right) K_0\left(\frac{y-Y_i}{b_n}\right),
\end{aligned} \tag{4.14}$$

where $K_0 : \mathbb{R} \longrightarrow \mathbb{R}$ is a fixed kernel with $\int_{\mathbb{R}} K_0 = 1$, and $(b_n)_{n\geq 1}$ is defined as $(a_n)_{n\geq 1}$ above.
Finally, if we assume that $\tilde{\ell}_n(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbb{R}^d$, and while dividing (4.14) by (4.13), we get the definition of a new estimator of the conditional density $g(y|\mathbf{x})$. Indeed, we have

$$\tilde{g}_n(y|\mathbf{x}) := \frac{\tilde{f}_n(\mathbf{x}, y)}{\tilde{\ell}_n(\mathbf{x})} = \frac{\frac{1}{b_n} \sum_{i=1}^{n} (G(Y_i))^{-1} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{a_n}\right) K_0\left(\frac{y-Y_i}{b_n}\right)}{\sum_{i=1}^{n} (G(Y_i))^{-1} K_d\left(\frac{\mathbf{x}-\mathbf{X}_i}{a_n}\right)}. \tag{4.15}$$

4.4 Assumptions and main results

In order to define our nonparametric estimator of the conditional mode function, throughout this paper and without loss of generality, we assume that the covariates are univariate, that is $d = 1$. Also it is assumed that $a_n = b_n = h_n$.
If no truncation is present ($n = N$), it is well known (see Collomb et al. [5]) that the kernel estimator of the conditional mode function $\Theta(x)$ is defined as a random variable $\Theta_n(x)$ maximizing the kernel estimator $g_n(y|x)$ of $g(y|x)$; that is

$$g_n(\Theta_n(x)|x) = \max_{-\infty<y<\infty} g_n(y|x),$$

with

$$g_n(y|x) = \frac{f_n(x,y)}{\ell_n(x)},$$

where

$$f_n(x,y) = \frac{1}{nh_n^2}\sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_n}\right) K_0\left(\frac{y-Y_i}{h_n}\right),$$

and

$$\ell_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K_1\left(\frac{x-X_i}{h_n}\right).$$

In the left-truncation model, we define $\tilde{g}_n(y|x)$ as an estimate of $g(y|x)$, based on the n observed i.i.d. pairs (X_i, Y_i), and given by

$$\tilde{g}_n(y|x) = \frac{\tilde{f}_n(x,y)}{\tilde{\ell}_n(x)}, \tag{4.16}$$

where

$$\tilde{f}_n(x,y) = \frac{\mu}{nh_n^2}\sum_{i=1}^{n} \frac{1}{G(Y_i)} K_1\left(\frac{x-X_i}{h_n}\right) K_0\left(\frac{y-Y_i}{h_n}\right), \tag{4.17}$$

and

$$\tilde{\ell}_n(x) = \frac{\mu}{nh_n}\sum_{i=1}^{n} \frac{1}{G(Y_i)} K_1\left(\frac{x-X_i}{h_n}\right).$$

Then a natural estimator of $\Theta(x)$ is defined as a random variable $\hat{\Theta}_n(x)$ maximizing $\tilde{g}_n(y|x)$, that is

$$\tilde{g}_n(\hat{\Theta}_n(x)|x) = \max_{-\infty<y<\infty} \tilde{g}_n(y|x). \tag{4.18}$$

However, the computation of $\hat{\Theta}_n(x)$ from (4.18) is not possible since both μ and $G(\cdot)$ are usually unknown, and therefore (4.16) is not useful in practice. One way to overcome this difficulty, is to replace $G(\cdot)$ by its product-limit estimator $G_n(\cdot)$ defined in (4.5) and μ by μ_n defined in (4.6), which allows us to define our nonparametric kernel estimator of the mode. Therefore, the feasible estimator of the conditional mode is then given as the solution of the equation

$$\hat{g}_n(\hat{\Theta}_n(x)|x) = \max_{-\infty<y<\infty} \hat{g}_n(y|x), \tag{4.19}$$

where

$$\hat{g}_n(y|x) = \frac{\hat{f}_n(x,y)}{\hat{\ell}_n(x)} = \frac{\sum_{i=1}^n (G_n(Y_i))^{-1} K_1\left((x-X_i)\diagup h_n\right) K_0\left((y-Y_i)\diagup h_n\right)}{h_n \sum_{i=1}^n (G_n(Y_i))^{-1} K_1\left((x-X_i)\diagup h_n\right)}, \tag{4.20}$$

with

$$\begin{aligned}
\hat{f}_n(x,y) &= \frac{1}{h_n^2}\int_{\mathbb{R}^2} K_1\left(\frac{x-u}{a_n}\right) K_0\left(\frac{y-v}{b_n}\right) \hat{\mathbf{F}}_n(du,dv) \\
&= \frac{\mu_n}{nh_n^2}\sum_{i=1}^n \frac{1}{G_n(Y_i)} K_1\left(\frac{x-X_i}{h_n}\right) K_0\left(\frac{y-Y_i}{h_n}\right),
\end{aligned} \tag{4.21}$$

and

$$\hat{\ell}_n(x) := \frac{1}{h}\int_{\mathbb{R}} K_1\left(\frac{x-u}{a_n}\right) \hat{L}_n(du) = \frac{\mu_n}{nh_n}\sum_{i=1}^n \frac{1}{G_n(Y_i)} K_1\left(\frac{x-X_i}{a_n}\right). \tag{4.22}$$

The corresponding d.f's of (4.21) and (4.22) are respectively given by

$$\hat{\mathbf{F}}_n(x,y) = \frac{\mu_n}{n}\sum_{i=1}^n \frac{1}{G_n(Y_i)} 1\!\!1_{(X_i\leq x, Y_i\leq y)}, \tag{4.23}$$

and

$$\hat{L}_n(x) = \frac{\mu_n}{n}\sum_{i=1}^n \frac{1}{G_n(Y_i)} 1\!\!1_{(X_i\leq x)}. \tag{4.24}$$

Note that (4.22) has been proposed lately by Ould-Saïd and Lemdani [24] in another context and its consistency has been stated.

Remark 4.1. *It can easily be verified that (4.23) is a bivariate distribution function since* $\frac{1}{n}\sum_{i=1}^n \frac{1}{G_n(Y_i)}$ *and* μ_n^{-1} *are both consistent estimates of* μ^{-1}.

It is useful to mention that $\hat{\Theta}_n(x)$ is not necessarily unique, and if this is the case, all the remaining of our paper will concern any value $\hat{\Theta}_n(x)$ satisfying (4.19). We point out that we can specify our choice by taking

$$\hat{\Theta}_n(x) = \inf\left\{ z\in\mathbb{R} \text{ such that } \hat{g}_n(z|x) = \sup_{-\infty<y<\infty} \hat{g}_n(y|x)\right\}.$$

To formulate our results, some additional notations are required. From now on, for all kernel K; $K^{(j)}$ denotes the $j-$order derivative of K.

For $(i,j) \in \mathbb{N}^2$, set

$$f^{(i,j)}(x,y) = \frac{\partial^{(i+j)}}{\partial x^i \partial y^j} f(x,y),$$

and for $j \geq 1$,

$$\hat{f}_n^{(0,j)}(x,y) = \frac{\partial^j}{\partial y^j} \hat{f}_n(x,y) = \frac{\mu_n}{n h_n^{(2+j)}} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} K_1\left(\frac{x - X_i}{h_n}\right) K_0^{(j)}\left(\frac{y - Y_i}{h_n}\right).$$

The derivatives $\tilde{f}_n^{(0,j)}(\cdot,\cdot)$ of $\tilde{f}_n(\cdot,\cdot)$ are obtained analogously.
Define $\Xi_0 = \{x \in \mathbb{R} \ |\ell(x) > 0\}$, let $\Xi \subset \Xi_0$ be a compact set of $\mathbb{R}$ and $\mathcal{C}(\ell)$ be the set of continuity points of ℓ. In the sequel a_F will denote the left endpoint of F. We will make use of the following assumptions gathered together here for easy reference.

(A) **(a)** The joint density $f(\cdot,\cdot)$ is differentiable up to order 4 and $\sup_{x,y} \left| f^{(i,j)}(x,y) \right| < \infty$, for $i + j \leq 4$;

(b) $f^{(0,2)}(\cdot,\cdot)$ is continuous and does not vanish ;

(c) The marginal density $\ell(\cdot)$ satisfies the Lipschitz condition and $\ell(x) \geq \gamma_0$ for all $x \in \Xi$, for some $\gamma_0 > 0$;

(d) $g(y|x)$ is twice differentiable, uniformly continuous in y and the second derivative $g^{(2)}(\cdot|x)$ is continuous ;

(e) The function $\Theta(\cdot)$ satisfies for any $\varepsilon > 0$ and $\eta(x)$, there exists a $\beta > 0$ such that $\sup_{x \in \Xi} |\Theta(x) - \eta(x)| \geq \varepsilon$, implies $\sup_{x \in \Xi} |g(\Theta(x)|x) - g(\eta(x)|x)| \geq \beta$.

(B) **(a)** K is a continuous function satisfying $\lim_{|u| \longrightarrow \infty} K(u) = 0$ and $\int_{\mathbb{R}} K(u)du = 1$;

(b) $\int_{\mathbb{R}} u^j K(u)du = 0$; for $j = 1,\ 2$;

(c) $\int_{\mathbb{R}} |u|^3 K(u)du < \infty$;

(d) K_1 is differentiable and of bounded variations ;

(e) K_0 is three times differentiable and K_0 and its two first derivatives are functions of bounded variations ;

(f) $K_1(\cdot)K_0^{(1)}(\cdot)$ is bounded.

(H) **(a)** The bandwidth h_n satisfies for $n \to \infty$, $h_n \to 0$ and $\dfrac{n h_n^8}{(\log n)} \to \infty$;

(b) $n h_n^9 \to 0$.

Remark 4.2 (discussion of the assumptions). *Conditions (**B :a-f**) are very common in functional estimation, they allow us to evaluate the asymptotic variance term of $\tilde{f}_n^{(0,1)}$. Usually, derivatives functions estimation uses finite order conditions, for example (**B :b,c**) are of this type. Note that condition (**B :c**) and the continuity of K imply that $\int_{\mathbb{R}} |u| K(u)du < \infty$. The condition (**B :a**) implies that $\hat{g}_n(\cdot|x)$ is a continuous function of y and tends to zero as y tends to $\pm\infty$, which ensures the existence of $\hat{\Theta}_n(x)$ satisfying (4.19). The set of conditions (**H**) gives* *the behavior of h_n for which the asymptotic results obtained below are valid. Assumptions (**A : a,b**) intervene in the convergence of $\hat{f}_n^{(0,2)}$ and (**A :c**) in that of $\hat{\ell}_n(\cdot)$. As a consequence of the boundedness of $f^{(1,0)}$ and $f^{(0,1)}$, the function $f(x, y)$ is uniformly continuous. Finally, conditions (**A :d,e**) stipulate the existence and the uniform uniqueness of the conditional mode function.*

Remark 4.3. *Recall that since K_0 has bounded variations, its first derivative $K_0^{(1)}$ is integrable and hence $K_1K_0^{(1)}$ is integrable. Furthermore $\left(K_1K_0^{(1)}\right)^2$ is also integrable, which ensures the existence of the asymptotic variance term.*

An example of a kernel K which satisfies the conditions **(B)** is given by

$$K(x) = \frac{1}{2\sqrt{2\pi}}\left(3 - x^2\right)\exp\left(-\frac{x^2}{2}\right).$$

Now, our first result is the uniform almost sure convergence with a rate of the conditional probability density function $\hat{g}_n(\cdot|x)$ as stated in Theorem 4.1. An immediate consequence is the strong uniform convergence with a rate of the kernel conditional mode estimator and given in Corollary 4.1.

Theorem 4.1. *Suppose that Assumptions (**A :a,c**), (**B :c,d,e**) and (**H :a**) hold. Then, as $n \longrightarrow \infty$*

$$\sup_{y\in\mathbb{R}} |\hat{g}_n(y|x) - g(y|x)| \longrightarrow 0 \quad \textit{almost surely.}$$

Moreover, we have for any compact set $\Xi \subset \Xi_0$

$$\sup_{x\in\Xi}\sup_{y\in\mathbb{R}} |\hat{g}_n(y|x) - g(y|x)| = O\left[\max\left\{\left(\frac{\log n}{nh_n^4}\right)^{\frac{1}{2}}, h_n\right\}\right], \ \mathbf{P} - a.s.$$

Corollary 4.1. *Under Assumptions (**A :a,b,e**),(**B :d,e**) and (**H :a**) we have, as $n \longrightarrow \infty$,*

$$\lim_{n\longrightarrow\infty}\sup_{x\in\Xi}\left|\hat{\Theta}_n(x) - \Theta(x)\right| = 0, \ \textit{almost surely.}$$

*In addition, if the assumption (**A :d**) holds, then for n large enough, we have*

$$\sup_{x\in\Xi}\left|\hat{\Theta}_n(x) - \Theta(x)\right| = O\left[\max\left\{\left(\frac{\log n}{nh_n^4}\right)^{\frac{1}{4}}, h_n^{\frac{1}{2}}\right\}\right], \ \mathbf{P} - a.s.$$

Now, suppose that the density function $g\left(\cdot|x\right)$ is unimodal at $\Theta(x)$, then by assumption (**A :b,d**), we have

$$g^{(1)}\left(\Theta(x)|x\right) = 0 \text{ and } g^{(2)}\left(\Theta(x)|x\right) < 0.$$

Similarly, under (**B :d,e**), with a great probability we have

$$\hat{g}_n^{(1)}\left(\hat{\Theta}_n(x)|x\right) = 0 \text{ and } \hat{g}_n^{(2)}\left(\hat{\Theta}_n(x)|x\right) < 0$$

if $\hat{\Theta}_n(x)$ is the mode of $\hat{g}_n(\cdot|x)$.
A Taylor series expansion of $\hat{g}_n^{(1)}(\cdot|x)$ in the neighborhood of $\Theta(x)$, gives

$$0 = \hat{g}_n^{(1)}\left(\hat{\Theta}_n(x)|x\right) = \hat{g}_n^{(1)}(\Theta(x)|x) + \left(\hat{\Theta}_n(x) - \Theta(x)\right)\hat{g}_n^{(2)}\left(\bar{\Theta}_n(x)|x\right),$$

where $\bar{\Theta}_n(x)$ is between $\hat{\Theta}_n(x)$ and $\Theta(x)$. Therefore,

$$\hat{\Theta}_n(x) - \Theta(x) = -\frac{\hat{g}_n^{(1)}(\Theta(x)|x)}{\hat{g}_n^{(2)}\left(\bar{\Theta}_n(x)|x\right)}.$$

Using (4.20), one may write

$$\hat{\Theta}_n(x) - \Theta(x) = -\frac{\hat{f}_n^{(0,1)}(x, \Theta(x))}{\hat{f}_n^{(0,2)}\left(x, \bar{\Theta}_n(x)\right)}, \tag{4.25}$$

if the denominator does not vanish.
To state the asymptotic normality of $\hat{\Theta}_n(x)$, we show that the numerator in (4.25), suitably normalized is asymptotically normally distributed, and that the denominator converges in probability to $f^{(0,2)}(x, \Theta(x))$. The result is given in the following Theorem.

Theorem 4.2. *Suppose that Assumptions (**A :b**), (**B :b,d,f**) and (**H :a,b**) holds, we have for any $x \in \mathcal{A}$*

$$\left(\frac{nh_n^4\left(f^{(0,2)}(x,\Theta(x))\right)^2}{Var(x,\Theta(x))}\right)^{\frac{1}{2}}\left(\hat{\Theta}_n(x) - \Theta(x)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1),$$

where $\xrightarrow{\mathcal{D}}$ denotes the convergence in distribution, $\mathcal{A} = \{x : x \in \mathcal{C}(f), Var(x, \Theta(x)) \neq 0\}$ and

$$Var(x,\Theta(x)) = \frac{\mu f(x,\Theta(x))}{G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(r)K_0^{(1)}(s)\right]^2 drds.$$

4.5 Simulations

The purpose of this section is to show how good is the behavior of our estimator for some particular functions $\Theta(\cdot)$. First we present the simulated model which permits to compute the estimator $\hat{\Theta}_n(\cdot)$.
It is well known that if $\mathbf{Z}^t = (Z^1, Z^2) \hookrightarrow \mathcal{N}(\mathbf{O}, \mathbf{I})$, then $\mathbf{V} = \mathbf{TZ} + \mu \hookrightarrow \mathcal{N}(\mu, \sum)$, where $\mathbf{T} = \mathbf{\Sigma}^{1/2} = \sqrt{\beta}D\sqrt{A}D^t$ with $\mathbb{E}(\mathbf{V}) = \mu^{\mathbf{t}} = (\mu_1, \mu_2)$ and $\sum = var(\mathbf{V})$. In our case we take $\mathbf{V}^t = (X, Y)$, $\mu^t = (0, 3)$, $\beta = \frac{9}{15}$, $D = \begin{pmatrix} \cos(\pi/4) & \sin(\pi/4) \\ -\sin(\pi/4) & \cos(\pi/4) \end{pmatrix}$ and $A = \begin{pmatrix} 1/3 & 0 \\ 0 & 3 \end{pmatrix}$.

A simple algebra gives $\sum = \begin{pmatrix} 1 & 4/5 \\ 4/5 & 1 \end{pmatrix}$ and hence the density of $\mathbf{V}$ is given by

$$f(x,y) = \frac{5}{6\pi} \exp\left(\frac{-25}{18}\left(x^2 - \frac{8}{5}x(y-3) + (y-3)^2\right)\right).$$

Under this model, the conditional mode function is the value of y which maximizes $f(x,y)$, we have

$$\Theta(x) = \frac{4}{5}x + 3.$$

The kernel estimator $\hat{\Theta}_n(x)$ is then computed according to the model

$$\begin{aligned} X_i &= \frac{2Z_i^1}{\sqrt{5}} + \frac{Z_i^2}{\sqrt{5}} \\ Y_i &= 2X_i + 3 - \frac{3}{\sqrt{5}} Z_i^1;\ i = 1, ..., N \end{aligned} \tag{4.26}$$

where Z_i^1 and Z_i^2 are two independent i.i.d. sequences both distributed as $\mathcal{N}(0,1)$. We also simulate N i.i.d. rv $T_i \hookrightarrow \mathcal{E}(\lambda)$, where λ is adapted in order to get different values of μ. We then keep the data $(X_i, Y_i),\ i = 1, \ldots, n$ such that $Y_i \geq T_i$. We do it in a way to obtain a given n (which means that in this case n is not random whereas N is), then we compute our estimator with the observed data $(Y_i, T_i, X_i),\ i = 1, \ldots, n$. We choose a Gaussian kernel and it is well known 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.

4.5.1 Consistency

The aim of the following simulations deals with the consistency (Theorem 4.1) and are made according to the methodology described previously . The bandwidth h_n is chosen as so as to satisfy conditions (H) and several values x_k are taken in $[-1, 1[$ to draw the graphs $\left\{(\Theta(x_k), \hat{\Theta}_n(x_k))\right\}$, and we show of how the diagrams become more close to the first bisector (see figure 1), and then we plot in the same figure $\left\{(x_k, \hat{\Theta}_n(x_k)\right\}$ and $\{(x_k, \Theta(x_k))\}$ respectively, when n increases. We notice that the estimator is bad for small n, but has a good behavior for n large enough (see figure 2). We then try to see if the quality depends on the truncation proportion μ. We take $n = 500$ and choose three values of the percentage of truncated data : $\mu \approx 45\%$, $\approx 60\%$ and $\approx 75\%$. The estimator's quality does not seem to be affected by μ as shown in figure 3 (though higher values of N are needed for small μ to achieve $n = 500$).

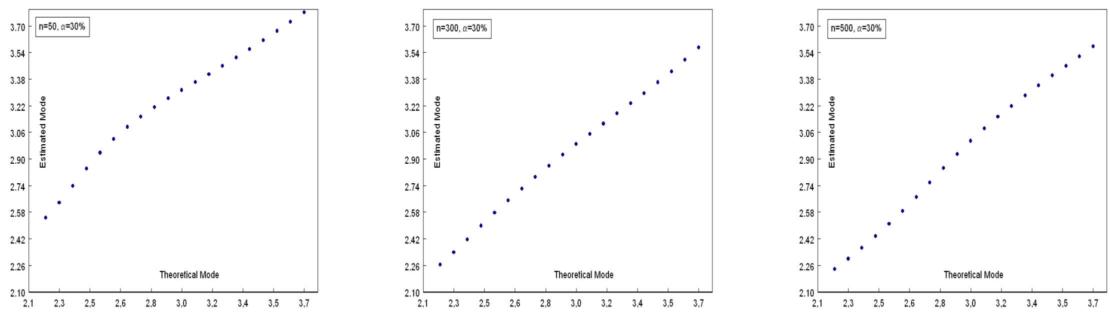


Figure 1: $\alpha = 0.30$ with $n = 50, 300$ and 500, respectively.

Figure 2: $\alpha = 0.30$ with $n = 50, 300$ and 500, respectively.

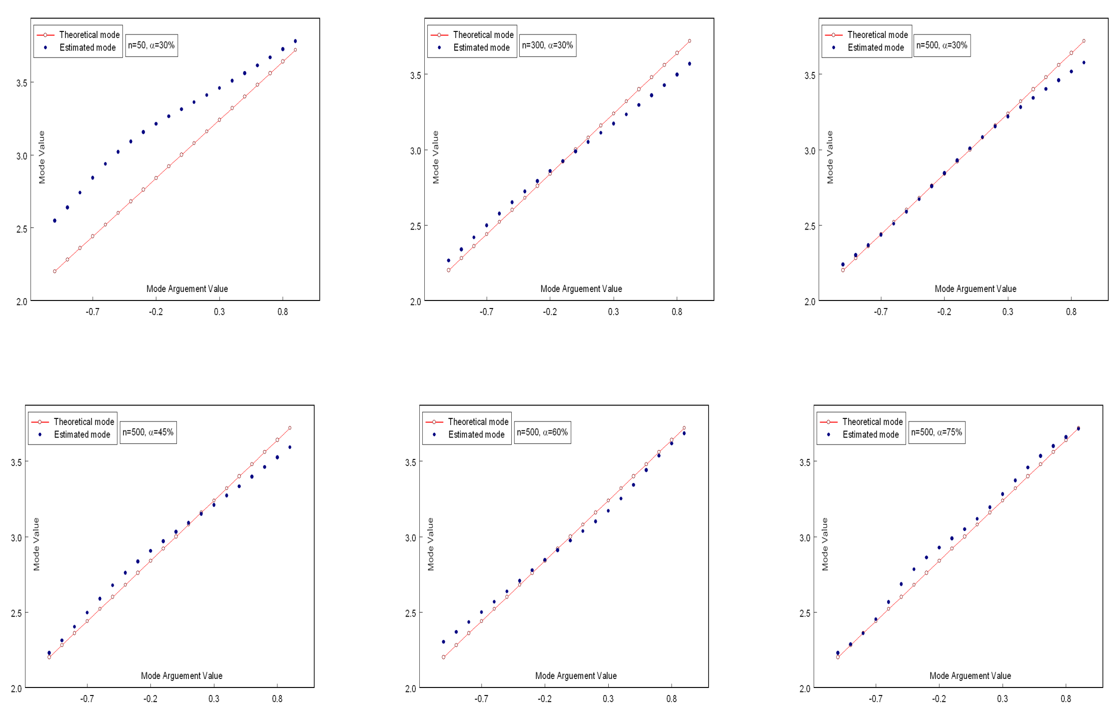


Figure 3: $n = 500$ with $\alpha = 0.45, 0.60$ and 0.75, respectively.

4.5.2 Asymptotic normality

In situation involving asymptotic normality, it is important to show how good the normality is when dealing with samples of finite sizes. There are questions arise and require answers in order to use correctly this normality in practice. Furthermore, how the shape of the estimated curve fit the standard normal density ?. The crucial problem in nonparametric estimation is, how to choose the bandwidth parameter ?. Some responses to these questions are brought out by studying the following example.

We consider the same model (4.26) as before, and the data has the same law as in the last simulations. Also we take the same law for the truncated r.v. T_i. For each sample of size n, we estimate the conditional mode function as before and calculate the normalized deviation between this estimate and the theoretical conditional mode function given in Theorem 4.2, that is

$$\mathring{\Theta}_n(x) = \sqrt{nh_n^4}\hat{\sigma}^{-1}(x, \hat{\Theta}_n(x))\hat{f}_n^{(0,2)}(x, \hat{\Theta}_n(x))\left\{\hat{\Theta}_n(x) - \Theta(x)\right\}.$$

We draw, using this scheme, B independent samples of $\mathring{\Theta}_n(x)$. Here, x is taken to be equal to 0. In estimating the value $\hat{\Theta}_n(0)$, the value of the bandwidth h_n is chosen as before (i.e. $h_n = Cn^{-\delta}$, $\delta \in]\frac{1}{9}, \frac{1}{8}[$), but, in estimating the density function of the processes $(\mathring{\Theta}_n(0))$, we use the classical bandwidth choice (see, e.g., Silverman [30], p. 40) which is $h'_B = C'\,B^{-1/5}$, where the constants C and C' are appropriately chosen.

We have worked with several values of $n = 50, 100$ and 200, $B = 300$ and we draw the corresponding curves (see figure 4, 5 and 6). It is clear that quality is better when n grows. Furthermore, in each case, we give the corresponding $Q-Q$-plot against a normal distribution, which puts in evidence the adjustment of better quality for n=200.

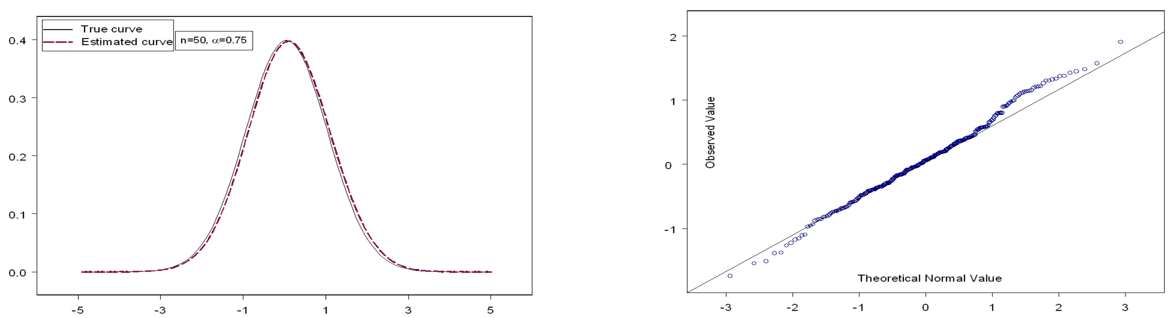
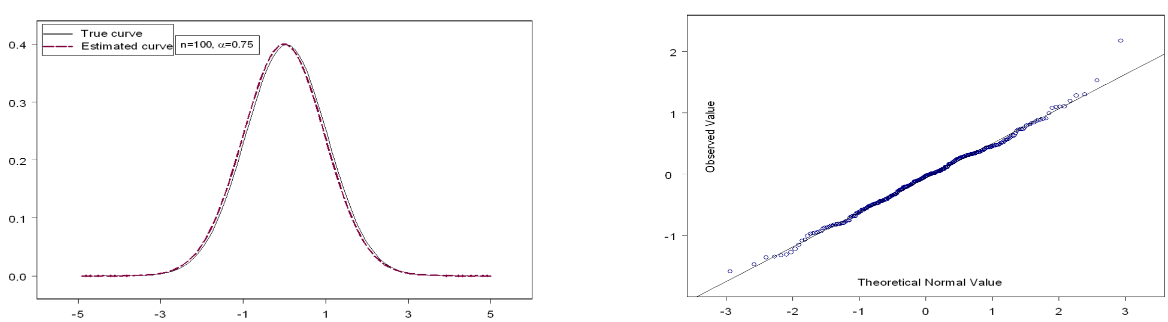


Figure 4: $n = 50$ and $B = 300$.

Figure 5: $n = 100$ and $B = 300$.



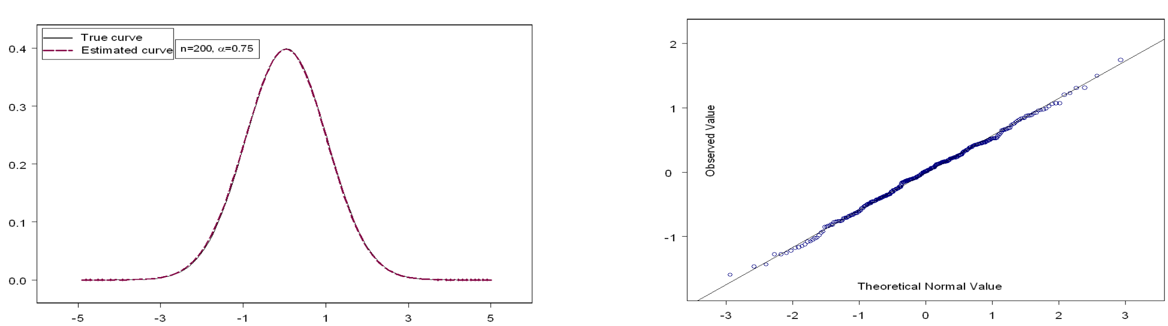


Figure 6: $n = 200$ and $B = 300$.

4.6 Proofs

To prove the main results, we need some auxiliary results.

Lemma 4.1. *As* $n \longrightarrow \infty$

$$\tilde{L}_n(x) \longrightarrow L(x), \ \mathbf{P}-a.s.$$

Furthermore, for any set $\Xi \subset \Xi_0$, *we have*

$$\sup_{x\in\Xi}\left|\tilde{L}_n(x) - L(x) \ \right| = O\left(\sqrt{\frac{\log n}{n}}\right), \ \mathbf{P}-a.s.$$

Proof. This proof is slightly different from the one given in Lemma 6.1 in Ould-Saïd and Lemdani [24]. The first part of Lemma 4.1 is a consequence of the strong law of large numbers. For the second part, we consider the i.i.d. sequence $(X_{i,}Y_i)_{1\leq i\leq n}$ and consider the class of functions

$$\mathbf{\Psi}_n = \left\{\psi_x : \Xi\times\mathbb{R} \longrightarrow \mathbb{R}^+ / \psi_x(u,y) = \frac{\mathbb{1}_{(u\leq x)}}{n\mu^{-1}G(y)}, \ x\in\Xi\right\},$$

$\mathbf{\Psi}_n$ is Vapnik-Ćervonenkis (V-C) class of non negative measurable functions which are uniformly bounded with respect to the envelope $\mathbf{\Omega} = [nG(a_F)]^{-1}$. Moreover

$$\mathbf{E}\left[\psi_x(X,Y)\right] = \mathbf{E}\left[\frac{\mathbb{1}_{(X\leq x)}}{n\mu^{-1}G(Y)}\right] \leq \mathbf{\Omega} =: U_n,$$

and

$$\mathbf{E}\left[\psi_x^2(X,Y)\right] = \mathbf{E}\left[\frac{\mathbb{1}_{(X\leq x)}}{(n\mu^{-1}G(Y))^2}\right] \leq \mathbf{\Omega}^2 =: \sigma_n^2.$$

Applying Talagrand's inequality (see : proposition 2.2 in Giné and Guillou [8]), with $\tau_n = D\sqrt{\frac{\log n}{n}}$ and D is a constant, we have

$$\begin{aligned}\mathbf{P}\left\{\sup_{x\in\Xi}\left|\tilde{L}_n(x) - L(x)\right| \geq \tau_n\right\} &= \mathbf{P}\left\{\sup_{\psi_x\in\Psi_n}\left|\sum_{i=1}^n \left\{\psi_x(X_i,Y_i) - \mathbf{E}[\psi_x(X,Y)]\right\}\right| \geq \tau_n\right\} \\ &\leq B_1\exp\left\{-\frac{G(a_F)}{B_1}\times n\tau_n\log\left[1+\frac{\tau_n}{n\delta_n B_1 G(a_F)}\right]\right\} \quad (4.27)\end{aligned}$$

where $\delta_n = \left(\frac{1}{n^{1/2}G(a_F)} + \frac{\sqrt{\log B_2}}{nG(a_F)}\right)^2$ and B_1, B_2 are positive constants.

For n large enough, the right hand side of (4.27) becomes of order $B_1 n^{-\left(\frac{DG(a_F)}{B_1}\right)^2}$ which by an appropriate choice of D can be made $O(n^{-3/2})$, which in turn is the term of a finite-sum series. Hence by Borel Cantelli's Lemma we have the second part of the result. ∎

Lemma 4.2. *As $n \longrightarrow \infty$, we have for any x*

$$\hat{L}_n(x) \longrightarrow L(x), \ \mathbf{P} - a.s.$$

Moreover, this convergence is achieved with a $O_{\mathbf{P}}\left(\sqrt{\frac{\log n}{n}}\right)$ rate and is uniform on any compact set $\Xi \subset \Xi_0$.

Proof. We have

$$\begin{aligned}
\left|\hat{L}_n(x) - L(x)\right| &\leq \left|\hat{L}_n(x) - \tilde{L}_n(x)\right| + \left|\tilde{L}_n(x) - L(x)\right| \\
&\leq \frac{|\mu_n - \mu|}{n} \sum_{i=1}^{n} \frac{1}{G_n(Y_i)} 1\!\!1_{(X_i \leq x)} + \frac{\mu}{n} \sum_{i=1}^{n} \left| \frac{1}{G_n(Y_i)} - \frac{1}{G(Y_i)} \right| 1\!\!1_{(X_i \leq x)} + \left|\tilde{L}_n(x) - L(x)\right| \\
&\leq \frac{|\mu_n - \mu|}{G_n(a_F)} + \frac{\mu}{G_n(a_F)G(a_F)} \sup_{y \in \mathbb{R}} |G_n(y) - G(y)| + \left|\tilde{L}_n(x) - L(x)\right| \\
&=: \mathcal{T}_1 + \mathcal{T}_2 + \mathcal{T}_3.
\end{aligned} \tag{4.28}$$

From Theorem 3.2 in He and Yang [11], $|\mu_n - \mu| = O_{\mathbf{P}}\left(n^{-1/2}\right)$, and $G_n(a_F) \stackrel{\mathbf{P}-a.s.}{\longrightarrow} G(a_F) > 0$, we get uniformly on x, $\mathcal{T}_1 = O_{\mathbf{P}}\left(n^{-1/2}\right)$. In the same way, we get $\mathcal{T}_2 = O_{\mathbf{P}}\left(n^{-1/2}\right)$, by using Remark 6 in Woodroofe [37]. Finally, the proof is achieved by using the result of Lemma 4.1 and Remark 4.4 below.

Remark 4.4. *Recall that, as noted in formula (4.12), all the sums involving the $(G_n(Y_i))^{-1}$ are taken for the i's such that $G_n(Y_i) \neq 0$. It follows that an additional term*

$$(IV) = \frac{\mu}{n} \sum_{i=1}^{n} \frac{1}{G(Y_i)} 1\!\!1_{(X_i \leq x)} 1\!\!1_{(G_n(Y_i)=0)},$$

should be added to the rhs of (4.28). this term is clearly negligible by the law of large numbers. The same remark can be made for all similar quantities in what follows. ■

Lemma 4.3. *As $n \longrightarrow \infty$, we have*

$$\tilde{\mathbf{F}}_n(x, y) \longrightarrow \mathbf{F}(x, y), \ \mathbf{P} - a.s.$$

Furthermore, for any compact set $\Xi \subset \Xi_0$

$$\sup_{x \in \Xi} \sup_{y \in \mathbb{R}} \left|\tilde{\mathbf{F}}_n(x, y) - \mathbf{F}(x, y)\right| = O\left(\sqrt{\frac{\log n}{n}}\right), \ \mathbf{P} - a.s.$$

Proof. The first part is the consequence of the SLLN. For the second part, we follow the same gait as in Lemma 4.1 by considering the following V-C class

$$\mathbf{\Phi}_n = \left\{ \varphi_{x,y} : \Xi \times \mathbb{R} \longrightarrow \mathbb{R}^+ \diagup \varphi_{x,y}(u,v) = \frac{1\!\!1_{(u \leq x, v \leq y)}}{n\mu^{-1}G(v)},\ x \in \Xi \right\}$$

with envelope $\mathbf{\Omega} = [nG(a_F)]^{-1}$. The upper bounds for the first and second moment of $\varphi_{x,y}$ are the same as in Lemma 4.1. ∎

Lemma 4.4. *As $n \longrightarrow \infty$, we have*

$$\hat{\mathbf{F}}_n(x,y) \longrightarrow \mathbf{F}(x,y),\ \mathbf{P} - a.s.$$

Moreover, for any compact set $\Xi \subset \Xi_0$

$$\sup_{x\in\Xi}\sup_{y\in\mathbb{R}} \left| \hat{\mathbf{F}}_n(x,y) - \mathbf{F}(x,y) \right| = O\left(\sqrt{\frac{\log n}{n}} \right),\ \mathbf{P} - a.s.$$

Proof. Analogously to Lemma 4.2, we have

$$\begin{aligned}
\left| \hat{\mathbf{F}}_n(x,y) - \mathbf{F}(x,y) \right| \quad &\leq \quad \frac{|\mu_n - \mu|}{G_n(a_F)} \frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{(X_i \leq x, Y_i \leq y)} \\
&+ \quad \frac{\mu}{G_n(a_F)G(a_F)} \sup_{y\in\mathbb{R}} |G_n(y) - G(y)| \frac{1}{n} \sum_{i=1}^{n} 1\!\!1_{(X_i \leq x, Y_i \leq y)} \\
&+ \quad \left| \tilde{\mathbf{F}}_n(x,y) - \mathbf{F}(x,y) \right|.
\end{aligned}$$

Similar arguments as in Lemma 4.2 and using Lemma 4.3 give the result. ∎

Lemma 4.5. *As $n \longrightarrow \infty$, we have for any fixed x*

$$\hat{\ell}_n(x) \longrightarrow \ell(x),\ \mathbf{P} - a.s.$$

Moreover, for any compact set $\Xi \subset \Xi_0$,

$$\sup_{x\in\Xi} \left| \hat{\ell}_n(x) - \ell(x) \right| = O\left[\max\left\{ \left(\frac{\log n}{nh_n^2} \right)^{\frac{1}{2}}, h_n \right\} \right],\ \mathbf{P} - a.s.$$

Proof. we have

$$\begin{aligned}
\hat{\ell}_n(x) - \ell(x) \quad &= \quad \frac{1}{h_n} \int_{\mathbb{R}} K_1\left(\frac{x-u}{h_n} \right) \left(\hat{L}_n(du) - dL(u) \right) \\
&+ \quad \frac{1}{h_n} \int_{\mathbb{R}} K_1\left(\frac{x-u}{h_n} \right) (\ell(u) - \ell(x))\, du \\
&=: \quad \mathcal{I}_1 + \mathcal{I}_2.
\end{aligned}$$

By part integrating, we have

$$\begin{aligned}|\mathcal{I}_1| &\le \frac{1}{h_n^2}\int_{\mathbb{R}}\left|\hat{L}_n(u)-L(u)\right|\left|K_1^{(1)}\left(\frac{x-u}{h_n}\right)\right|du\\ &\le \sup_{x\in\Xi}\left|\hat{L}_n(x)-L(x)\right|\frac{1}{h_n}\int_{\mathbb{R}}\left|K_1^{(1)}(r)\right|dr,\end{aligned}$$

which by Lemma 4.2 and (**B :d**) is $O_{\mathbf{P}}\left(\left(\frac{\log n}{nh_n^2}\right)^{1/2}\right)$ for any x. Moreover this rate is uniform on any compact set $\Xi\subset\Xi_0$. On the other hand, by using a change of variable, (**A :c**) and (**B :c**) the bias term $\mathcal{I}_2$ is $O(h_n)$, which yields the result. ■

Lemma 4.6. *Under Assumptions (**A :a**), (**B :c,d,e**) and (**H :a**), we have, as* $n\longrightarrow\infty$

$$\hat{f}_n(x,y)\longrightarrow f(x,y),\ \mathbf{P}-a.s.$$

Moreover, for any compact set $\Xi\subset\Xi_0$

$$\sup_{x\in\Xi}\sup_{y\in\mathbb{R}}\left|\hat{f}_n(x,y)-f(x,y)\right| = O\left[\max\left\{\left(\frac{\log n}{nh_n^4}\right)^{\frac{1}{2}},h_n\right\}\right],\ \mathbf{P}-a.s.$$

Proof. we have

$$\begin{aligned}\hat{f}_n(x,y)-f(x,y) &= \frac{1}{h_n^2}\int_{\mathbb{R}^2}K_1\left(\frac{x-u}{h_n}\right)K_0\left(\frac{y-v}{h_n}\right)\left(\hat{\mathbf{F}}_n(du,dv)-\mathbf{F}(du,dv)\right)\\ &+ \frac{1}{h_n^2}\int_{\mathbb{R}^2}K_1\left(\frac{x-u}{h_n}\right)K_0\left(\frac{y-v}{h_n}\right)(f(u,v)-f(x,y))\,dudv\\ &=: \mathcal{J}_1+\mathcal{J}_2.\end{aligned}$$

Using Fubini's theorem, we obtain by multiple integration by parts

$$\mathcal{J}_1=\frac{1}{h_n^4}\int_{\mathbb{R}^2}K_1^{(1)}\left(\frac{x-u}{h_n}\right)K_0^{(1)}\left(\frac{y-v}{h_n}\right)\left(\hat{\mathbf{F}}_n(u,v)-\mathbf{F}(u,v)\right)dudv.$$

Hence

$$|\mathcal{J}_1|\le h_n^{-2}\sup_{x\in\Xi}\sup_{y\in\mathbb{R}}\left|\hat{\mathbf{F}}_n(x,y)-\mathbf{F}(x,y)\right|\left[\int_{\mathbb{R}^2}\left|K_1^{(1)}(r)K_0^{(1)}(s)\right|drds\right],\tag{4.29}$$

which by Lemma 4.4 and assumptions (**B :d,e**) is $O_{\mathbf{P}}\left(\sqrt{\frac{\log n}{nh_n^4}}\right)$.

On the other hand we have

$$\mathcal{J}_2=\int_{\mathbb{R}^2}K_1(r)\,K_0(s)\left\{f(x-rh_n,y-sh_n)-f(x,y)\right\}drds.\tag{4.30}$$

Taylor expansion of $f(x-rh_n,y-sh_n)$ around (x,y) to order 2 yields

$$|\mathcal{J}_2|\le h_n\left\{\int_{\mathbb{R}}\left|rK_1(r)\,f^{(1,0)}(\bar{x},\bar{y})\right|dr+\int_{\mathbb{R}}\left|sK_0(s)\,f^{(0,1)}(\bar{x},\bar{y})\right|ds\right\},$$

where $(\bar{x}, \bar{y})$ is between (x, y) and $(x - rh_n, y - sh_n)$. Hence under assumptions (**A :a**), (**B :c** see Remark 4.2) and (**H :a)**) yields the result. ■

Proof of Theorem 4.1. By applying a classical decomposition, it is easy to see that with probability one

$$\sup_{x\in\Xi}\sup_{y\in\mathbb{R}}|\hat{g}_n(y|x) - g(y|x)| \leq \frac{1}{\inf_{x\in\Xi}\hat{\ell}_n(x)}\left\{\sup_{x\in\Xi}\sup_{y\in\mathbb{R}}\left|\hat{f}_n(x,y) - f(x,y)\right| + \kappa.\sup_{x\in\Xi}\left|\hat{\ell}_n(x) - \ell(x)\right|\right\},$$

where κ is the upper bound of $g(y|x)$. An application of Lemma 4.5, Lemma 4.6 and (**A :a,c**) complete the proof. ■

Proof of Corollary 4.1. We have

$$\begin{aligned}\left|g\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right)\right| &\leq \left|\hat{g}_n\left(\hat{\Theta}_n(x)|x\right) - g\left(\hat{\Theta}_n(x)|x\right)\right| + \left|\hat{g}_n\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right)\right| \\ &\leq \sup_{x\in\Xi}\sup_{y\in\mathbb{R}}|\hat{g}_n\left(y|x\right) - g\left(y|x\right)| + \left|\hat{g}_n\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right)\right|.\end{aligned}$$

Since

$$\left|\hat{g}_n\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right)\right| = \left|\sup_{y\in\mathbb{R}}\hat{g}_n\left(y|x\right) - \sup_{y\in\mathbb{R}}g\left(y|x\right)\right| \leq \sup_{x\in\Xi}\sup_{y\in\mathbb{R}}|\hat{g}_n\left(y|x\right) - g\left(y|x\right)|.$$

Then

$$\left|g\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right)\right| \leq 2\sup_{x\in\Omega}\sup_{y\in\mathbb{R}}|\hat{g}_n\left(y|x\right) - g\left(y|x\right)|. \tag{4.31}$$

An application of Theorem 4.1 and (**A :e**), we get the result of part 1. For the second part, a Taylor expansion of $g\left(\cdot\ |x\right)$ in a neighborhood of $\Theta(x)$ gives

$$g\left(\hat{\Theta}_n(x)|x\right) - g\left(\Theta(x)|x\right) = \frac{1}{2}\left(\hat{\Theta}_n(x) - \Theta(x)\right)^2 g^{(2)}\left(\bar{\Theta}(x)|x\right),$$

where $\bar{\Theta}(x)$ is between $\hat{\Theta}_n(x)$ and $\Theta(x)$. Then, by (4.31) and (**A :d**), we have

$$\left(\hat{\Theta}_n(x) - \Theta(x)\right)^2\left|g^{(2)}\left(\bar{\Theta}(x)|x\right)\right| \leq 4\sup_{x\in\Xi}\sup_{y\in\mathbb{R}}|\hat{g}_n\left(y|x\right) - g\left(y|x\right)|.$$

Hence, by Theorem 4.1, we complete the proof of Corollary 4.1. ■

Proof of Theorem 4.2. From (4.17) and (4.21) we have the following decomposition

$$
\begin{aligned}
\sqrt{nh_n^4}\frac{\hat{f}_n^{(0,1)}(x,\Theta(x))}{\hat{f}_n^{(0,2)}(x,\bar{\Theta}_n(x))} &= \sqrt{nh_n^4}\frac{\hat{f}_n^{(0,1)}(x,\Theta(x))-\tilde{f}_n^{(0,1)}(x,\Theta(x))}{\hat{f}_n^{(0,2)}(x,\bar{\Theta}_n(x))}\\
&+ \sqrt{nh_n^4}\frac{\tilde{f}_n^{(0,1)}(x,\Theta(x))-\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right]}{\hat{f}_n^{(0,2)}(x,\bar{\Theta}_n(x))}\\
&+ \sqrt{nh_n^4}\frac{\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right]}{\hat{f}_n^{(0,2)}(x,\bar{\Theta}_n(x))}\\
&=: \mathcal{S}_1+\mathcal{S}_2+\mathcal{S}_3.
\end{aligned}
$$

To prove the result, we establish that the numerators of the terms $\mathcal{S}_1$ and $\mathcal{S}_3$ are negligible, and that of $\mathcal{S}_2$ is normally distributed. Then, in conjunction with the fact that the denominator converges in probability and Slutsky's theorem we get the result.
For the first term and by setting

$$
S_{h,i}=\sum_{i=1}^{n}\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-Y_i}{h_n}\right),
$$

we have

$$
\begin{aligned}
\sqrt{nh_n^4}\left(\hat{f}_n^{(0,1)}(x,\Theta(x))-\tilde{f}_n^{(0,1)}(x,\Theta(x))\right) &\le \frac{|\mu_n-\mu|\sqrt{nh_n^4}}{G(a_F)G_n(a_F)}\times\sup_y|G_n(y)-G(y)|\times\frac{S_{h,i}}{nh_n^3}\\
&+ \frac{\mu\sqrt{nh_n^4}}{G(a_F)G_n(a_F)}\times\sup_y|G_n(y)-G(y)|\times\frac{S_{h,i}}{nh_n^3}\\
&+ \frac{|\mu_n-\mu|\sqrt{nh_n^4}}{G(a_F)}\times\frac{S_{h,i}}{nh_n^3}\\
&=: \mathcal{U}_1+\mathcal{U}_2+\mathcal{U}_3.
\end{aligned}
$$

By the fact that $|\mu_n-\mu|=O_{\mathbf{P}}\left(n^{-1/2}\right)$ (from Theorem 3.2 in He and Yang [11]), $\sup_y|G_n(y)-G(y)|=O_{\mathbf{P}}\left(n^{-1/2}\right)$ (remark 6 in Woodroofe [37]), $G_n(a_F)\overset{\mathbf{P}\ a.s.}{\longrightarrow}G(a_F)$ and by using the usual kernel's method, by adding and subtracting the expectation of $\left(nh_n^3\right)^{-1}S_{h,i}$, we can prove that this term converges to $f^*(x,\Theta(x))$ which is bounded. Then we get, $\mathcal{U}_1=o_{\mathbf{P}}\left(h_n^2\right)$. In the same way, we get $\mathcal{U}_2=O_{\mathbf{P}}\left(h_n^2\right)$ and $\mathcal{U}_3=O_{\mathbf{P}}\left(h_n^2\right)$ which permit us to conclude that $\mathcal{S}_1$ is negligible.
Now, we turn out to $\mathcal{S}_3$ which is given in the following :

Lemma 4.7. *Under Assumptions (**A** :**a**), (**B** :**a-c**) and (**H** :**b**), we have*

$$
\left(nh_n^4\right)^{\frac{1}{2}}\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right]\longrightarrow 0,\quad \text{as } n\to\infty.
$$

Proof. We have

$$\begin{aligned}
\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right] &= \frac{\mu}{h_n^3}\int_{\mathbb{R}^2}\frac{1}{G(v)}K_1\left(\frac{x-u}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-v}{h_n}\right)d\mathbf{F}^*(u,v)\\
&= \frac{1}{h_n^3}\int_{\mathbb{R}^2}K_1\left(\frac{x-u}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-v}{h_n}\right)f(u,v)dudv.
\end{aligned}$$

Integrating by parts with respect to v, using conditions (**A :a**) and (**B :a**), we obtain

$$\begin{aligned}
\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right] &= \frac{1}{h_n^2}\int_{\mathbb{R}^2}K_1\left(\frac{x-u}{h_n}\right)K_0\left(\frac{\Theta(x)-v}{h_n}\right)f^{(0,1)}(u,v)dudv\\
&= \int_{\mathbb{R}^2}K_1(r)\,K_0(s)\,f^{(0,1)}(x-rh_n,\Theta(x)-sh_n)drds.
\end{aligned}$$

Expanding $f^{(0,1)}(x-rh_n,\Theta(x)-sh_n)$ around $(x,\Theta(x))$ to the order of h_n^3 , we get

$$\begin{aligned}
\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right] &= \sum_{j=0}^{2}\frac{(-h_n)^j}{j!}\sum_{i=0}^{j}\binom{j}{i}f^{(i,j+1)}(x,\Theta(x))\int_{\mathbb{R}^2}r^{j-i}s^iK_1(r)\,K_0(s)\,drds\\
&+ \frac{h_n^3}{3!}\sum_{i=0}^{3}\int_{\mathbb{R}^2}\binom{3}{i}r^{3-i}s^if^{(3-i,i+1)}(\bar{x},\bar{\Theta}(x))K_1(r)\,K_0(s)\,drds\\
&=: \Re_1+\Re_2.
\end{aligned}$$

Assumptions (**A :a**), (**B :b,c**) and the definition of the mode imply that $\Re_1 = 0$ and $\Re_2 = O(h_n^3)$. Thus by (**H :b**) we conclude that $\left(nh_n^4\right)^{\frac{1}{2}}\mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right] = o(1)$. ■

Lemma 4.8. *Under Assumptions (**A :a**) and (**B :a,d,f**), we have*

$$nh_n^4 Var\left[\tilde{f}_n^{(0,1)}(x,\Theta(x))\right] \longrightarrow \frac{\mu f(x,\Theta(x))}{G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2du\,dv, \quad as\ n\to\infty.$$

Proof. First, start with calculating $Var\left(\tilde{f}_n^{(0,1)}(x,\Theta(x))\right)$.

$$\begin{aligned}
Var\left(\tilde{f}_n^{(0,1)}(x,\Theta(x))\right) &= Var\left(\frac{\mu}{nh_n^3}\sum_{i=1}^{n}\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-Y_i}{h_n}\right)\right)\\
&= \frac{\mu^2}{nh_n^6}\mathbf{E}\left[\frac{1}{G(Y_1)}K_1\left(\frac{x-X_1}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-Y_1}{h_n}\right)\right]^2\\
&\quad -\frac{\mu^2}{nh_n^6}\mathbf{E}^2\left[\frac{1}{G(Y_1)}K_1\left(\frac{x-X_1}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-Y_1}{h_n}\right)\right]\\
&= \frac{\mu^2}{nh_n^6}\int_{\mathbb{R}^2}\frac{G(v)}{\mu}\left[\frac{1}{G(v)}K_1\left(\frac{x-u}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-v}{h_n}\right)\right]^2 f(u,v)dudv\\
&\quad -\frac{1}{nh_n^6}\left[\int_{\mathbb{R}^2}K_1\left(\frac{x-u}{h_n}\right)K_0^{(1)}\left(\frac{\Theta(x)-v}{h_n}\right)f(u,v)dudv\right]^2\\
&= \frac{\mu}{nh_n^4}\int_{\mathbb{R}^2}\frac{\left[K_1(r)K_0^{(1)}(s)\right]^2}{G(\Theta(x)-sh_n)}f(x-rh_n,\Theta(x)-sh_n)drds\\
&\quad -\left[\frac{1}{\sqrt{nh_n^4}}\int_{\mathbb{R}^2}K_1(r)K_0^{(1)}(s)f(x-rh_n,\Theta(x)-sh_n)drds\right]^2\\
&=: \mathcal{V}_1-\mathcal{V}_2.
\end{aligned}$$

In analogous manner as before and under (**A :a**) and (**B :a,d,f**), we get

$$nh_n^4\mathcal{V}_2 \longrightarrow 0.$$

Again, a Taylor expansion gives us

$$\mathcal{V}_1 = \frac{\mu f(x,\Theta(x))}{nh_n^4}\int_{a_F}^{\frac{\Theta(x)-a_F}{h_n}}\frac{\left[K_1(r)K_0^{(1)}(s)\right]^2}{G(\Theta(x)-sh_n)}drds + o\left(\frac{1}{nh_n^2}\right).$$

Since $\Theta(x) > a_F$, observe that there exists n_0 such that for all $n \geq n_0$,

$$\int_{\mathbb{R}}(K_1(r))^2\left(\int_{a_F}^{\frac{\Theta(x)-a_F}{h_n}}\frac{\left(K_0^{(1)}(s)\right)^2}{G(\Theta(x)-sh_n)}ds\right)dr \leq \frac{1}{G(a_F)}\int_{a_F}^{+\infty}\left[K_1(r)K_0^{(1)}(s)\right]^2 drds,$$

whence, by the Dominated Convergence Theorem and for $n\to\infty$

$$\int_{\mathbb{R}^2}\frac{\left(K_1(r)K_0^{(1)}(s)\right)^2}{G(\theta-sh_n)}drds \longrightarrow \frac{1}{G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(r)K_0^{(1)}(s)\right]^2 drds,$$

which is finite by (**B :f**) (see Remark 4.3). Finally, we obtain

$$nh_n^4 Var\left(\tilde{f}_n^{(0,1)}(x,\Theta(x))\right) \longrightarrow \frac{\mu f(x,\Theta(x))}{G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(r)K_0^{(1)}(s)\right]^2 drds.$$

Now, all what is left to be shown is that, the numerator of $\mathcal{J}_2$ is a sum of i.i.d. rv's which must satisfies a Lindeberg's Theorem. For this purpose, let us consider

$$\Delta_{in}(x,y) = \frac{\mu}{\sqrt{nh_n^2}}\left\{\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right) - \mathbf{E}\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]\right\}.$$

It is obvious that

$$\sum_{i=1}^{n}\Delta_{in}(x,y) = \sqrt{nh_n^4}\left(\tilde{f}_n^{(0,1)}(x,y) - \mathbf{E}\left[\tilde{f}_n^{(0,1)}(x,y)\right]\right).$$

Hence

$$Var\left(\sum_{i=1}^{n}\Delta_{in}(x,y)\right) = nh_n^4 Var\left(\tilde{f}_n^{(0,1)}(x,y)\right).$$

Then we have the following

Lemma 4.9. *Under Assumptions **(A :a)**, **(B :d)** and **(H :a)**, we have*

$$\forall\varepsilon > 0,\ \sum_{i=1}^{n}\int_{\left\{\Delta_{in}^2(x,y)>\varepsilon^2 Var\left(\sum_{i=1}^{n}\Delta_{in}(x,y)\right)\right\}}\Delta_{in}^2(x,y)d\mathbf{F}^*(x,y) \longrightarrow 0,\ \text{as } n\to\infty.$$

Proof. On the one hand we have

$$\begin{aligned}\Delta_{in}^2(x,y) &\leq \frac{2\mu^2}{nh_n^2}\frac{1}{G^2(Y_i)}\left(K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2 &(4.32)\\ &+ \frac{2\mu^2}{nh_n^2}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]. &(4.33)\end{aligned}$$

Note that

$$\frac{2\mu^2}{nh_n^2}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] = \frac{2h_n^4}{n}\left[\mathbf{E}\left\{\frac{\mu}{h_n^3 G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right\}\right]^2,$$

which gives

$$\frac{2\mu^2}{nh_n^2}\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right] \longrightarrow 0,\ \text{as } n\to\infty. \tag{4.34}$$

On the other hand, by Lemma 4.8, we have

$$Var\left(\sum_{i=1}^{n}\Delta_{in}(x,y)\right) \longrightarrow \frac{\mu f(x,\Theta(x))}{G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2 dudv,$$

then for $\varepsilon = \dfrac{\mu f(x,\Theta(x))}{2G(\Theta(x))}\displaystyle\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2 dudv > 0$, $\exists n_0\in\mathbb{N}^*$ such that : $\forall n\geq n_0$, we have

$$Var\left(\sum_{i=1}^{n}\Delta_{in}(x,y)\right) \geq \frac{\mu f(x,\Theta(x))}{2G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2 dudv.$$

Now denote by

$$V(X_i,Y_i)=\frac{1}{G^2(Y_i)}\left(K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2+\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right];$$

clearly we have

$$\Delta_{in}^2(x,y)\leq\frac{2\mu^2V(X_i,Y_i)}{nh_n^2},$$

making use of (4.34), for $n\geq n_0$ we have

$$\left\{\Delta_{in}^2(x,y)>\varepsilon^2Var\left(\sum_{i=1}^n\Delta_{in}(x,y)\right)\right\}\subset\left\{\Delta_{in}^2(x,y)>\varepsilon^2\frac{\mu f(x,\Theta(x))}{2G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2dudv\right\}.$$

Now let $\varepsilon'=\varepsilon^2\frac{\mu f(x,\Theta(x))}{4G(\Theta(x))}\int_{\mathbb{R}^2}\left[K_1(u)K_0^{(1)}(v)\right]^2dudv$, we have

$$\begin{aligned}\left\{\Delta_{in}^2(x,y)>\varepsilon^2Var\left(\sum_{i=1}^n\Delta_{in}(x,y)\right)\right\}&\subset\quad\{\Delta_{in}^2(x,y)>2\varepsilon'\}\\&=\quad\left\{\frac{nh_n^2}{2\mu^2}\Delta_{in}^2(x,y)>\varepsilon'\frac{nh_n^2}{\mu^2}\right\}\\&\subset\quad\left\{V(X_i,Y_i)>\varepsilon'\frac{nh_n^2}{\mu^2}\right\}\\&\subset\quad\left\{\frac{1}{G^2(Y_i)}\left(K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2>\frac{\varepsilon'nh_n^2}{2\mu^2}\right\}\\&\bigcup\quad\left\{\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]>\frac{\varepsilon'nh_n^2}{2\mu^2}\right\}.\end{aligned}$$

From (4.34) and for n large enough we have

$$\left\{\mathbf{E}^2\left[\frac{1}{G(Y_i)}K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right]>\frac{\varepsilon'nh_n^2}{2\mu^2}\right\}=\emptyset. \tag{4.35}$$

Furthermore, by the fact that G is lower bounded and $K_1.K_0^{(1)}$ is bounded, then for n large enough, the set

$$\left\{\frac{1}{G^2(Y_i)}\left(K_1\left(\frac{x-X_i}{h_n}\right)K_0^{(1)}\left(\frac{y-Y_i}{h_n}\right)\right)^2>\frac{\varepsilon'nh_n^2}{2\mu^2}\right\}, \tag{4.36}$$

is empty by (**H :a**).
Therefore, from (4.35) and (4.36) and for n large enough, we get

$$\left\{\Delta_{in}^2(x,y)>\varepsilon^2Var\left(\sum_{i=1}^n\Delta_{in}(x,y)\right)\right\},$$

is empty, which completes the proof. ∎

Now, to end the proof of Theorem 4.2, it suffices to prove that $\hat{f}_n^{(0,2)}(x, \bar{\Theta}_n(x))) \xrightarrow{\mathbf{P}} f^{(0,2)}(x, \Theta(x))$. Indeed, we have the following.

Lemma 4.10. *Under Assumptions (**A :a,b**), (**B :c,d,e**) and (**H :a**), we have*

$$\lim_{n\to+\infty} \sup_{y\in\mathbb{R}} \left|\hat{f}_n^{(0,2)}(x,y) - f^{(0,2)}(x,y)\right| = 0, \ \mathbf{P} - a.s.$$

Proof. We have

$$\begin{aligned}
\hat{f}_n^{(0,2)}(x,y) - f^{(0,2)}(x,y) &= \frac{1}{h_n^4}\left\{\int_{\mathbb{R}^2} K_1\left(\frac{x-u}{h_n}\right) K_0^{(2)}\left(\frac{y-v}{h_n}\right)\left\{\hat{\mathbf{F}}_n(du,dv) - d\mathbf{F}(u,v)\right\}\right\} \\
&\quad + \left\{\frac{1}{h_n^4}\int_{\mathbb{R}^2} K_1\left(\frac{x-u}{h_n}\right) K_0^{(2)}\left(\frac{y-v}{h_n}\right) f(u,v)dudv - f^{(0,2)}(x,y)\right\} \\
&=: \mathcal{J}_1 + \mathcal{J}_2.
\end{aligned}$$

By multiple integrating by parts, and using assumptions (**B :d,e**), we obtain

$$|\mathcal{J}_1| \le h_n^{-4} \sup_{x\in\Xi}\sup_{y\in\mathbb{R}} \left|\hat{\mathbf{F}}_n(x,y) - \mathbf{F}(x,y)\right| .\xi$$

where $\xi = \displaystyle\int_{\mathbb{R}^2} \left|K_1^{(1)}(u) K_0^{(3)}(v)\right| dudv < \infty$. Applying the result of Lemma 4.4, we have $\mathcal{J}_1 = O_{\mathbf{P}}\left(\sqrt{\dfrac{\log n}{nh_n^8}}\right)$.

Integrating by parts two times with respect to the second component and using a change of variable, it follows that

$$\mathcal{J}_2 = \int_{\mathbb{R}^2} K_1(r) K_0(s)\left\{f^{(0,2)}(x - rh_n, y - sh_n) - f^{(0,2)}(x,y)\right\} drds.$$

Taylor expansion of $f^{(0,2)}(x - rh_n, y - sh_n)$ in the neighborhood of (x,y) yields

$$|\mathcal{J}_2| \le h_n \int_{\mathbb{R}^2} \left|K_1(r) K_0(s)\left\{rf^{(1,2)}(\bar{x},\bar{y}) + sf^{(0,3)}(\bar{x},\bar{y})\right\}\right| drds,$$

where $(\bar{x},\bar{y})$ is between (x,y) and $(x - rh_n, y - sh_n)$. Making use of assumptions (**A :a**) and (**B :c**), it follows that $\mathcal{J}_2 = O(h_n)$.
In order to achieve the proof, we observe that we have, for n large enough

$$\begin{aligned}
\left|\hat{f}_n^{(0,2)}(x, \bar{\Theta}_n(x)) - f^{(0,2)}(x, \Theta(x))\right| &\le \sup_{y\in\mathbb{R}} \left|\hat{f}_n^{(0,2)}(x,y) - f^{(0,2)}(x,y)\right| \\
&\quad + \left|f^{(0,2)}(x, \bar{\Theta}_n(x)) - f^{(0,2)}(x, \Theta(x))\right|.
\end{aligned}$$

By Lemma 4.10 and assumption (**A :b**), we get the result. ∎

Remark 4.5. *A plug-in-type estimate $\hat{\sigma}^2(x, \hat{\Theta}_n(x))$ for the asymptotic variance $\sigma^2(x, \Theta(x))$ can easily be obtained using the estimators ((4.6) and (4.21)) of $\dfrac{\mu}{G(\cdot)}$ and $f(\cdot,\cdot)$, respectively. This yields a confidence interval of asymptotic level $(1-\varsigma)$ for $\Theta(x)$.*

$$\hat{\Theta}_n(x) \pm t_{1-\varsigma/2} \times \left(\frac{\hat{\sigma}^2(x, \hat{\Theta}_n(x))}{nh_n^4 \left(\hat{f}_n^{(0,2)}(x, \hat{\Theta}_n(x)) \right)^2} \right)^{1/2}$$

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

Bibliographie

[1] C. Abraham, G. Biau, B. Cadre, On the asymptotic properties of a simple estimate of the mode. ESAIM, Probab. Statist. 8 (2004) 1-11.

[2] P.K. Andersen, O. Borgan, R.D. Gill, N. Keiding, Statistical models based on counting processes, Springer Verlag New-York, 1993.

[3] K. Chen, M.T. Chao, S.W. Lo, On strong uniform consistency of the Lynden-Bell estimator for truncated data, Ann. Statist. 23, (1995) 440-449.

[4] H. Chernoff, Estimation of the mode, Ann. Inst. Statist. Math. 16 (1964) 31-41.

[5] G. Collomb, W. Hardle, S. Hassani, A note on prediction via estimation of the conditional mode function, J. Statist. Plann and Inference 15 (1987) 227-236.

[6] W.F. Eddy, Optimum kernel estimators of the mode, Ann. Statist. 8 (1980) 870-882.

[7] W.F. Eddy, The asymptotic distributions of kernel estimators of the mode. Z. Wahrsh. Verw. Gebiete 59 (1982) 279-290.

[8] E. Giné, A. Guillou, On consistency of kernel density estimators for randomly censored data : rates holdng unifomly over adaptive intervals. Ann. Inst. H. Poincarré, 37 (2001) 503-522.

[9] S. He, Estimating a distribution function with truncated data, Acta. Math. Appl. Sin., Eng. Ser. 10 (1994) 12-33.

[10] S. He, G.L. Yang, The strong law under random truncation, Ann. Statist. 26 (1998a) 992-1010.

[11] S. He, G.L. Yang, Estimation of the truncation probability in the random truncation model, Ann. Statist. 26 (1998b) 1011-1027.

[12] E. Hermann, K. Ziegler, Rates on consistency for nonparametric estimation of the mode in absence of smoothness assumptions. Statist. Probab. Lett. 68 (2004) 359-368.

[13] V.D. Konakov, On the asymptotic normality of the mode of multidimensional distributions, Th. Probab. Appl. 19 (1974) 794-799.

[14] M. Lemdani, E. Ould-Saïd, Relative deficiency of the Kaplan-Meier estimator with respect to a smoothed estimator, Math. Method. Statist. 10 (2001) 215-231.

[15] M. Lemdani, E. Ould-Saïd, Asymptotic behavior of the hazard rate kernel estimator under truncated and censored data, to appear in Comm. Statist. Theory and Metods (2007).

[16] D. Louani, On the asymptotic normality of the kernel estimators of the density function and its derivatives under random censoring. Comm. Statist. Theory & Meth. 27 (1998) 2909-2924.

[17] D. Louani, E. Ould-Saïd, Asymptotic normality of kernel estimators of the conditional mode under strong mixing hypothesis, J. Nonparametric Statist. 11 (1999) 413-442.

[18] D. Lynden-Bell, A method of allowing for known observational selection in small samples applied to 3CR quasars, Monthly Notices Roy. Astronom. Soc. 155 (1971) 95-118.

[19] K.L. Mehra, Y.S. Ramakrishnaiah, P. Sashikala, Laws of iterated logarithm and related asymptotics for estimators of conditional density and mode, Ann. Inst. Statist. Math. 52 (2000) 630-645.

[20] A. Mokkaddem, M. Pelletier, A law of the iterated logarithm for the kernel mode estimator, ESAIM : Probability and statistics 7 (2003) 1-21.

[21] E.N. Nadaraya, On nonparametric estimates of density functions and regression curves, Theory Probab. Appl. 10 (1965) 186-190.

[22] E. Ould-Saïd, Estimation non paramétrique du mode conditionnel. Application à la prévision. C. R. Acad. Sci. Paris. Série I, 316 (1993) 943-947.

[23] E. Ould-Saïd, A note on ergodic processes prediction via estimation of the conditional mode function, Scandinavian. J. statist. 24 (1997) 231-239.

[24] E. Ould-Saïd, Z. Cai, Strong uniform consistency of the nonparametric estimation of the censored conditional mode function, J. Nonparametric Statist. 17 (2005) 797-806.

[25] E. Ould-Saïd, M. Lemdani, Asymptotic properties of a nonparametric regression function estimator with randomly truncated data, In press in Ann. Inst. Statist. Math. (2006).

[26] E. Parzen, On estimation of a probability density function and mode, Ann. Math. Statist. 33 (1962) 1065-1076.

[27] J.P. Romano, On weak convergence and optimality of kernel density estimates of the mode, Ann. Statist. 16 (1988) 629-647.

[28] M. Samanta, Nonparametric estimation of the mode of a multivariate density. South African Statist. J. 7 (1973) 109-117.

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

[30] B.W. Silverman, Density estimation for statistics and data analysis, Chapman and Hall, London (1986).

[31] W. Stute, Almost sure representations of the product-limit estimator for truncated data, Ann. Statist. 21 (1993) 146-156.

[32] W.Y. Tsai, N.P. Jewel, M.C. Wang, A note on the product limit estimator under right censoring and left truncation, Biometrika 74 (1987) 883-886.

[33] J. Van Ryzin, On strong consistency of density estimates, Ann. Math. Statist. 40 (1969) 1765-1772.

[34] P. Vieu, A note on density mode estimation, Statist. Probab. Lett. 26 (1996) 297-307.

[35] M.C. Wang, A semiparametric model for randomly truncated data, J. Amer. Statist. Assoc. 84 (1989) 742-748.

[36] M.C. Wang, N.P. Jewell, W.Y. Tsai, Asymptotic properties of the limit-product estimate under random truncation, Ann. Statist 14 (1986) 1597-1805.

[37] M. Woodroofe, Estimating a distribution function with truncated data, Ann. Statist. 13 (1985) 163-177.

Chapitre 5

Conclusion et Perspectives

Dans cette thèse, nous avons proposé des réponses à certaines questions, et contribuer à compléter des travaux récents établis pour la régression dans le cas tronqué (voir, Ould-Saïd et Lemdani (2006)), en construisant des estimateurs pour les densités ainsi que pour le mode, simple et conditionnel. Les réponses apportées sont illustrées par des simulations pour montrer les performances des estimateurs. Cependant, certaines vitesses obtenues sont jugées optimales et d'autres restent à améliorer en se basant, par exemple, sur l'utilisation des VC-classes pour les densités (pour le mode conditionnel). Les outils utilisés dans les preuves sont d'actualité, ayant permi d'écourter les démonstrations en évitant ainsi le recours aux techniques basées sur le recouvrement des compacts par les petites boules, dans le cas i.i.d.
Losque les données présentent une forme de dépendance, nous avons traité le cas de l'α-mélange. Le choix de l'inégalité de Fuk-Nagaev (voir, Rio (2000)) et son utilisation dans le cas de données dépendantes, est justifié par le fait qu'elle s'adapte mieux et pose moins de problèmes techniques que celle de Bosq (1998), par exemple.

Ci-dessous, certaines questions qui vont guider nos recherches futures.

1. **Choix du paramètre de lissage**
   Le critère sur lequel nous nous sommes basés pour évaluer l'optimalité de la vitesse de convergence est celui du critère de la convergence uniforme presque sûre, établi dans le cas i.i.d. par Stute (1982). Il serait intéressant de regarder si de tels résultats demeurent vrais pour des données tronquées.

2. **Grandes déviations**
   Les travaux sur la densité et le mode établis par, Mokkadem et Pelletier (2003, 2005), Mokkadem et al. (2006) et, Louani et Ould Maouloud (2007), sont des documents de référence et pourraient, à notre avis être étendus au cas de données incomplètes.

3. **Récursivité**
   Au meilleur de nos connaissances, il n'existe pas d'estimateurs récursifs pour des données tronquées, que ce soit pour la densité ou la fonction de régression. Nous pensons qu'il serait intéressant de proposer des estimateurs pour le traitement on-line, dans le cas de données incomplètes.

4. **Test d'additivité en présence de troncature**
   Depuis les travaux de Bhattacharya et al. (1983) sur l'estimation paramétrique de la fonction de régression pour des données tronquées (troncature déterministe), les rares travaux sur l'estimation non paramétriques et paramétriques, concernant la fonction de régression, en présence de troncature aléatoire, sont ceux de Ould-Saïd et Lemdani (2006) et, He et Yang (2003), respectivement. Le fléau de la dimension est un inconvénient en estimation non paramétrique, une manière de contourner ce problème en régression, est de supposer l'additivité. A notre connaissance, une telle approche lorsque les données sont aléatoirement tronquées, n'a pas été abordée et il serait intéressant d'envisager des travaux sur les tests d'addtivité.

5. **Cas de variables fonctionnelles**
   Il nous semble possible d'étendre les travaux sur la régression et le mode conditionnel, en présence de variable fonctionnelle, au cas de données incomplètes.