Quelques propriétés asymptotiques en estimation non paramétrique de fonctionnelles de processus stationnaires en temps continu

Sultana Didi

École Doctorale Paris Centre

Thèse de doctorat

Discipline : Statistiques

présentée par

Sultana DIDI

Quelques propriétés asymptotiques en estimation non paramétrique de fonctionnelles de processus stationnaires en temps continu.

---

dirigée par Djamal LOUANI

Soutenue le 15 septembre 2014 devant le jury composé de :

M. Paul DEHEVEULS	Paris 6	président
M Gérard BIAU	Paris 6	examinateur
Mme. Sophie DABO-NIANG	Lille 3	examinateur
M Jêrome SARACCO	Bordeaux 1	rapporteur
M. Djamal LOUANI	Paris 6	directeur
Mme. Céline VIAL	Polytech Lyon	rapporteur

Institut de Mathématiques de Jussieu
175, rue du chevaleret
75 013 Paris

École doctorale Paris centre Case 188
4 place Jussieu
75 252 Paris cedex 05

Au nom de Dieu, le Miséricordieux, le Clément.
Dis : " Oui, ma prière, mes actes de dévotion, ma vie et ma mort appartiennent à Dieu, le Seigneur des mondes ; Il n'a pas d'associé ! C'est cela qui m'a été ordonné, et je suis le premier de ceux qui se sont soumis (muslimûn)"
Le Bétail - Al-An'âm.

Cette thèse est dédiée à Maman et Grand Père,
ceux qui rêvaient de me voir obtenir mon doctorat et ne pouvaient pas terminer le long chemin avec moi.
Que Dieu ait vos âmes en paix.

Résumé

Les travaux de cette thèse portent sur les problèmes d'estimation non paramétrique des fonctions de densité, de régression et du mode conditionnel associés à des processus stationnaires à temps continu. La motivation essentielle est d'établir des propriétés asymptotiques tout en considérant un cadre de dépendance des données assez général qui puisse être facilement utilisé en pratique.

Cette contribution se compose de quatre parties. La première partie est consacrée à l'état de l'art relatif à la problématique qui situe bien notre contribution dans la littérature.

Dans le deuxième partie, nous nous intéressons à l'estimation, par la méthode du noyau, de la densité pour laquelle nous établissons des résultats de convergence presque sûre, ponctuelle et uniforme, avec des vitesses de convergence. Dans les parties suivantes, les données sont supposées stationnaires et ergodiques.

Dans la troisième partie, des propriétés asymptotiques similaires sont établies pour l'estimation à noyau de la fonction de régression.

Dans le même esprit, nous étudions dans la quatrième partie, l'estimation à noyau de la fonction mode conditionnel pour lequel nous établissons des propriétés de consistance avec des vitesses de convergence. L'estimateur proposé ici se positionne comme une alternative à celui de la fonction de régression dans les problèmes de prévision.

Mots-clefs : Estimation non-paramétrique, estimateur à noyau, consistance, convergence presque sûre et uniforme, ergodicité, stationarité, temps continu, densité, régression, mode conditionnel, vitesse de convergence.

Abstract

The work of this thesis focuses upon some nonparametric estimation problems. More precisely, considering kernel estimators of the density, the regression and the conditional mode functions associated to a stationary continuous-time process, we aim at establishing some asymptotic properties while taking a sufficiently general dependency framework for the data as to be easily used in practice.

The present manuscript includes four parts. The first one gives the state of the art related to the field of our concern and identifies well our contribution as compared to the existing results in the literature.

In the second part, we focus on the kernel density estimation. In a rather general dependency setting, where we use a martingale difference device and a technique based on a sequence of projections on σ-fields, we establish the almost sure pointwise and uniform consistencies with rates of our estimate.

In the third part, similar asymptotic properties are established for the kernel estimator of the regression function. Here and below, the processes are assumed to be ergodic

In the same spirit, we study in the fourth part, the kernel estimate of conditional mode function for which we establish consistency properties with rates of convergence. The proposed estimator may be viewed as an alternative in the prediction issues to the usual regression function.

Keywords : Non-parametric estimation, kernel estimator, consistency, uniform convergence, ergodicity, stationarity, continuous time, density function, regression function, conditional mode, rate of convergence.

Table des matières

Introduction **9**

0.1 Historique de l'estimation non-paramétrique 12

0.2 Théorie ergodique 18

0.3 Plan de la thèse 21

1 Consistency results for the kernel density estimate on continuous time stationary and dependent data **27**

1.1 Introduction 28

1.2 Results 29

1.3 Proofs 32

2 Asymptotic results for the regression function estimate on continuous time stationary and ergodic data **51**

2.1 Introduction 52

2.2 Results 54

2.3 Proofs 58

3 Consistency results for the conditional mode estimate on continuous time stationary and ergodic data **71**

3.1 Introduction 72

3.2 Assumptions and notation 73

3.3 Theoretical properties 75

3.4 Proofs 77

Annexe **93**

Perspectives de recherche **97**

Annexe aux perspectives **101**

Bibliographie **125**

Introduction

L'estimation statistique est un domaine très important de la statistique mathématique qui développe des techniques pour décrire certaines caractéristiques d'ensembles d'observations. On pourra distinguer deux composantes principales, à savoir, l'estimation paramétrique et l'estimation non paramétrique. Dans le cadre de l'estimation non-paramétrique, on est intéressé, essentiellement, par l'estimation d'une fonction inconnue appartenant à une certaine classe de fonctions. Notons que l'approche non-paramétrique est basée sur un procédé indépendant de la loi du processus observé. Dans ce contexte général, la procédure d'estimation ne se restreint pas, comme dans l'estimation paramétrique, à l'estimation d'un nombre fini de paramètres liés à la loi de l'échantillon. De ce fait, l'estimation non-paramétrique offre une très grande flexibilité de modélisation pour les applications réelles. Parmi les problèmes importants de la statistique mathématique, l'estimation de caractéristiques fonctionnelles associées à la loi des observations occupe une place prépondérante. À cet effet, de nombreux travaux ont été consacrés à l'estimation de fonctionnelles comme la fonction de densité et la fonction de régression.

Les contributions à l'étude des paramètres fonctionnels ont beaucoup enrichi la littérature ces dernières décennies. Nous citons à titre d'exemples les travaux de Rosenblatt (1956)[114], Parzen (1962)[107], Nadaraya (1965)[100], Watson (1964)[126], Banon (1978)[6] ainsi que celui de Castellana- Leadbetter (1986)[26], présentant une condition portant sur les densités jointes et marginales d'un processus strictement stationnaire et utilisée pour atteindre des vitesses suroptimal dans le cas d'un processus à temps continu (voir Bosq (1997)[15]). L'estimation non paramétrique des fonctions de la densité, de la régression et des argmax (fonction mode pour la densité et mode conditionnel pour la densité conditionnelle) constitue un sujet qui a suscité une intense activité de recherche et permis l'introduction de nombreux outils et méthodes considérant des contextes multiples et variés.

Le présent manuscrit est composé de quatre chapitres présentant des résultats de recherche originaux portant sur l'estimation non-paramétrique de fonctionnnelles associées à des processus à temps continu, chacun éclairant un aspect particulier de ce contexte général. Nous abordons des problèmes en rapport avec l'estimation de la fonction de densité, de la fonction de régression et de la fonction mode conditionnel.

Le chapitre introductif rappellera les principaux concepts ainsi que les propriétés les plus importantes qui nous permettent de mieux décrire les problèmes traités en situant au fur et à mesure les travaux antérieurs. Ensuite, nous présenterons brièvement les nouveaux résultats obtenus dans cette thèse. Pour être plus précis, nous avons caractérisé le comportement asymptotique, sous des hypothèses appropriées, relatif à la consistance avec des vitesse de convergence, pour des estimateurs à noyau construits à partir de l'observation d'une partie d'un processus stationnaire à temps continu défini dans un cadre de dépendance assez général.

Dans le chapitre 1, nous nous intéressons à la question de l'estimation de la fonction de densité d'un processus stationnaire à temps continu. Sous des conditions de dépendance assez générales, nous établissons les convergences presque sûres, ponctuelles et uniformes avec des vitesses de convergence de l'estimateur à noyau de la densité du processus.

L'objet du chapitre 2, est d'étudier le modèle de régression d'un processus réel $(Y_t)_{t\geq 0}$ sur un processus vectoriel $(\mathbf{X}_t)_{t\geq 0}$, tous les deux considérés comme stationnaires et ergodiques. En utilisant une version adaptée de l'estimateur à noyau de Nadaraya-Watson au cas du temps continu, nous établissons les convergences presque sûres, ponctuelles et uniformes avec des vitesses de convergence de cet estimateur. Les résultats sont, ensuite, appliqués à la prévision des séries temporelles.

Dans le dernier chapitre, nous étudions le mode conditionnel et nous établissons la consistance presque sûre, ponctuelle et uniforme, avec vitesse de convergence de l'estimateur à noyau proposé. Dans toute la suite de ce travail, nous supposons que le processus observé n'est pas assujetti à aucune hypothèse permettant d'utiliser des méthodes paramétriques.

La prise en compte de structure de dépendance pour l'estimation de paramètres rend le cadre de travail plus proche de ce qui est observé dans la réalité. L'étude des propriétés relatives aux processus à temps continu ne peut être envisagée en dehors d'une structure de dépendance des données. Les conditions de mélange sont les plus utilisées, à travers la littérature, dans l'analyse des propriétés de ce type de processus. Pour de nombreux processus bien connus, les propriétés de mélange ont déjà été établies. Cependant, la structure de dépendance de multiple processus utiles reste un problème ouvert. À titre d'exemple, Ibragimov et Linnik (1971)[74], Chernick (1981)[31] et Andrews (1984)[2] ont montré que dans certains cas le processus autorégressif linéaire du premier ordre en temps discret n'est pas fortement mélangeant. En particulier, le processus stationnaire $(X_t)_{t\in\mathbb{Z}}$ vérifiantle modèle AR(1) défini par $X_t = \theta X_{t-1} + \varepsilon_t$, où les $(\epsilon_t)_{t\in\mathbb{Z}}$ est une suite d'innovation de Bernoulli i.i.d, n'est pas fortement mélangeant. Toutefois, l'ergodicité est conservée en prenant des fonctions mesurables d'un processus ergodique. Si $(\varepsilon_t)_{t\in\mathbb{Z}}$ est un processus strictement stationnaire et ergodique, $Y_t = \vartheta((\ldots,\varepsilon_{t-1},\varepsilon_t),(\varepsilon_t,\varepsilon_{t+2},\ldots))$ pour une certaine fonction borélienne $\vartheta(\cdot)$, alors $(Y_t)_{t\in\mathbb{Z}}$ est également ergodique ; voir Proposition 2.10 à la page 54 de Bradley (2007)[24]. Comme le processus autorégressif, ci-dessus, peut être représenté comme une fonction linéaire des ε_t, il s'ensuit alors qu'il est aussi ergodique. Nous construisons maintenant un autre exemple de processus ergodique non mélangeant et cette fois-ci en temps continu. Il est bien connu que le mouvement brownien fractionnaire $(W_t^H; t \geq 0)$ de paramètre de Hurst $H \in (0,1)$ admet des incréments strictement stationnaires. Par ailleurs, le bruit gaussien fractionnaire, défini pour tout $s > 0$ par $(G_t^H)_{t\geq 0} = (W_{t+s}^H - W_t^H)_{t\geq 0}$, est un processus centré strictement stationnaire à longue mémoire lorsque $H \in (\frac{1}{2}, 1)$ (voir, par exemple, Beran (1994)[11] p. 55, Lu (2009)[93] p. 17). De ce fait, il contrevient alors à la condition de mélange fort. Pour un processus réel gaussien strictement stationnaire et centré $(G_t)_{t\geq 0}$, considérons la fonction de corrélation $R(t) = \mathbb{E}[G_0 G_t]$. En se reportant au travail de Maslowski et Pospíšil (2008)[96], Lemme 4.2, il ressort que le processus $(G_t)_{t\geq 0}$ est ergodique lorsque $\lim_{t\to\infty} R(t) = 0$. Un calcul facile permet alors de voir que le processus $(G_t^H)_{t\geq 0}$ vérifie bien cette condition.

Par ailleurs, il est bien connu que les conditions de mélange sont difficiles à vérifier en pratique. Il est donc préférable de définir un cadre de travail général en considérant des processus satisfaisant des conditions d'ergodicité qui incluent à la fois des processus mélangeants et non-mélangeants. Il est bien connu que l'ergodicité est satisfaite par les processus mélangeants et elle est, en outre, plus facile à vérifier en pratique (Krengel (1985)[77], p. 24, Ash et Gardner (1975)[3], p. 120). La difficulté principale de l'ergodicité est qu'elle ne nous permet pas, sans conditions supplémentaires, d'obtenir des vitesses de convergence (Krengel (1985)[77]). Notre objectif est de construire un carde de travail relatif à l'estimation non paramétrique de fonctionnelles, d'un processus à temps continu, vérifiant une condition d'ergodicité et d'établir des résultats de consistances ponctuelles et uniformes, avec des vitesses de convergence des différents estimateurs.

Dans la suite de cette introduction, nous donnerons successivement :

– une introduction à l'estimation non-paramétrique ;

– une introduction à la théorie ergodique ;

– un plan de la thèse ;

0.1 Historique de l'estimation non-paramétrique

L'approche non-paramétrique, permettant de surmonter les restrictions imposées par les méthodes paramétriques sur le modèle, est fréquement adoptée par les statisticiens pour estimer certaines fonctionnelles liées aux données.

Dans la suite, nous introduisons les estimateurs qui seront utilisés dans cette thèse.

Estimation de la densité de probabilité

Soit $(X_i)_{i\geq 1}$ une suite stationnaire des variables aléatoires. On note $\mathbb{P}$ la loi de probabilité sur $\mathbb{R}$ régissant ce processus et f sa densité par rapport à la mesure de Lebesgue.

La méthode du noyau a d'abord été décrite en 1951 dans un rapport non publié par Fix et Hodges (voir Silverman and Jones (1989)[123]). Un peu plus tard, la première forme de l'estimateur à noyau était introduite dans les travaux de Rosenblatt (1956)[114] et Parzen (1962) [107] qui ont défini une application réelle K vérifiant la condition $\int_{\mathbb{R}} K(u)du = 1$ et connue depuis sous le nom du noyau de Parzen-Rosenblatt. L'estimateur de Parzen-Rosenblatt est défini, pour tout $x \in \mathbb{R}$, par

$$\hat{f}_n(x) = \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

où h_n, la fenêtre de lissage, est un paramètre tendant avec une certaine vitesse vers 0 lorsque n tend vers l'infini.

La première version de cet estimateur fut donnée par Rosenblatt(1956), en choisissant le noyau uniforme $K = \frac{\mathbb{I}_{[-1,1]}}{2}$. Il donna l'erreur quadratique moyenne de l'estimateur de la densité pour des observations univariées indépendantes et identiquement distribuées. La généralisation de ce résultat fut obtenue par Parzen(1962) pour une vaste classe de noyaux qui établit aussi la normalité asymptotique. Calcoullos (1966)[25] traita le cas multivarié. Résultats de la section 6.3 du Bosq et Blanke (2008)[21] sont des améliorations de résultats parus dans Bosq (1998)[17]. Concernant la convergence presque sûre, ils ont délibérément choisi une condition de mélange assez fort (nommément une décroissance exponentielle du coefficient) de présenter des résultats optimales (c'est-à constantes asymptotiques similaires à ceux du cas i.i.d.). Pour les conditions de dépendance affaiblis, le lecteur intéressé pourra se référer aux ouvrages suivants (et les références qui y sont) : Liebscher (2001)[82] pour les processus arithmétique fortement mélangeants, Doukhan et Louhichi (2001)[53] pour un nouveau concept de dépendance y compris les processus qui ne sont pas mélangeants en général. Pour des choix pratiques de h_n dans le cas dépendant, nous renvoyons également à : Hall et al. (1995)[68] ; Hart et Vieu (1990)[72] et Kim (1997)[76].

Banon (1978)[6], fut le premier à s'intéresser à l'estimation de la densité en temps continu à partir de l'observation d'une partie $(X_t)_{0\leq t\leq T}$ d'un processus stationnaire $(X_t)_{t\geq 0}$. S'inspirant du cas discret, la forme de cet estimateur est donnée, pour tout $x \in \mathbb{R}$, par

$$\hat{f}_T(x) = \frac{1}{Th_T}\int_0^T K\left(\frac{x - X_t}{h_T}\right) dt.$$

Les premiers résultats de l'estimation de la fonction de densité pour les processus de diffusion sont apparus dans Banon (1978)[6] suivie de Banon et Nguyen ((1978)[7], (1981)[8]) ; Nguyen (1979)[102] et Nguyen et Pham (1980)[103].

Par la suite, Delecroix (1979)[43], étudia la convergence presque sûre et la convergence en moyenne quadratique des estimateurs de la densité et de la densité conditionnelle pour des processus strictement stationnaires et fortement mélangeants. Quelques années plus tard, Castellana et Leadbetter (1986)[26] démontrèrent qu'il était possible, sous certaines conditions portant sur la densité jointe du processus, d'atteindre une vitesse de convergence en moyenne quadratique de l'estimateur de la densité en T^{-1}, qu'on appelle vitesse "suroptimale" ou "paramétrique" puisqu'elle est identique à la vitesse obtenue usuellement en statistique paramétrique. Ce résultat historique a par la suite inspiré de nombreux travaux en temps continu. On pourra se référer par exemple aux ouvrages de Bosq (1998) [20] et Bosq et Blanke (2007)[17] pour une étude complète des estimateurs à noyau et par projection, et de Kutoyants (2004)[79] pour l'inférence des processus de diffusion ergodiques, ainsi qu'à leurs bibliographies respectives. Le cas particulier des ondelettes a également été examiné par Leblanc (1995)[91]. Enfin, Kutoyants (1995)[78] a prouvé que $(1/T)$ est la vitesse minimax lorsque les processus de diffusion sont observés. Bosq (1997)[15] et Bosq *et al* (1999)[18] ont obtenu la normalité asymptotique pour les estimateurs à noyaux basés sur des échantillons des processus $\alpha-$mélangeants. Didi & Louani (2013a)[50] ont obtenu des résultats de convergence presque sûre, ponctuelle et uniforme, avec des vitesses de convergence non paramétrique dans un cadre de convergence assez général où des méthodes de preuve basées sur des différences de martingale et des projections successives sur une famille de $\sigma-$algèbre emboitées, comparables à celle définies dans Wu *& al* (2010)[128] dans le cas discret, ont été utilisées (voir, aussi Chapitre 1). Wu *& al* (2010)[128] établit, dans le cas discret, la normalité asymptotique ainsi que la consistance avec des vitesses de convergence pour des estimateurs non paramétriques de la densité et de la fonction de régressioon pour une large classe de processus linéaires et non-linéaires utilisés dans les series temporelles. Finalement, le choix de la fenêtre de lissage est un problème qui a suscité beaucoup d'intérêt, Yondjé *et al* (1994)[132] a proposé une méthode pour sélectionner une fenêtre qui soit asymptotiquement "bonne". Cette méthode s'inspire des idées de validation croisée qui ont été proposées dans d'autres problèmes d'estimation fonctionnelle, citons par exemple, Marron (1987)[95] pour la densité, Härdle et Marron (1985)[71] pour la régression, Sarda et Vieu, (1991)[122] pour la fonction de hasard et Sarda (1991)[121] pour le cas de la fonction de répartition. Nous pouvons aussi nous référer aux travaux de Youndé *et al* (1993a)[131], Youndjé et Wells (2008)[133], Absava (1999)[1], Chacón et Tenreiro (2012)[28] qui traitent de la même question.

Estimation de la fonction de régression

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

Cette fonction de régression est définie pour tout $\mathbf{x} \in \mathbb{R}^d$ par :

$$m(\mathbf{x}) = \mathbb{E}[l(Y) \mid \mathbf{X} = \mathbf{x}],$$

qui est la moyenne de la distribution conditionnelle de Y sachant $\mathbf{X} = \mathbf{x}$.

Soit $((\mathbf{X}_t, Y_t), t \in \mathcal{J})$ un processus stationnaire mesurable sur $\mathbb{R}^d \times \mathbb{R}$ définie sur un espace de probabilité $(\Omega, \mathcal{A}, P)$. L'ensemble des indices $\mathcal{J}$ peut être discret $\mathbb{N}$, ou continu $\mathbb{R}^+$ et soit l une fonction borélienne réelle dans $\mathbb{R}$ tel que $\mathbb{E}\left[|l(Y_0)|\right] < \infty$. Soient K un noyau positif et h_n une suite de réels positifs. L'estimateur à noyau de type Nadaraja-Watson $\widehat{m}_n$ de la fonction de régression m, basé sur un échantillon $(\mathbf{X}_1, Y_1), \cdots, (\mathbf{X}_n, Y_n)$ du couple (X, Y), est défini par

$$\widehat{m}_n(\mathbf{x}) = \frac{\sum_{i=1}^n l(Y_i) K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}.$$

La méthode du noyau introduite par Rozenblatt (1956)[114] pour estimer la fonction de densité a été reprise par Nadaraja (1965)[100] et Watson (1964)[126] pour estimer la fonction de régression. En raison de la forme de l'estimateur à noyau de la fonction de régression, qui est un rapport, les propriétés de convergence sont obtenues sous des conditions plus restrictives que pour l'estimateur de la densité. Notamment, la convergence uniforme n'est obtenue que sur des ensembles bornés. Cet estimateur a été largement étudié par plusieurs auteurs, nous citons entre autres, Collomb (1985)[34] qui a donné une condition nécessaire et suffisante pour la convergence, uniforme presque sûre, sur un ensemble borné, généralisant ainsi les résultats antérieurs de Nadaraja (1964,1965)[99],[100], Devroye (1978)[48], Mack et Silverman (1982)[94], Gasser et Müller (1984)[64]. Pour plus de détails sur ce sujet, nous renvoyons le lecteur au livre de Bosq (1998, Chapitre 3)[17].

L'étude de l'estimation de la régression non paramétrique dans le cadre des données ergodiques en temps discret a motivé un certain nombre de travaux. Nous nous référons, entre autres, aux travaux de Delecroix et Rosa (1996), Laib (1999,2005), Morvai *et al.* (1996)[47] et Yakowitz *et al.* (1999)[129] où certaines propriétés asymptotiques sont obtenues. Notons également que la consistance avec vitesse ainsi que la normalité asymptotique ont été obtenues par Laïb et Louani (2010,2011)[88],[89] pour l'estimateur de régression à partir des données fonctionnelles en temps discret.

L'adaptation de l'estimateur de Nadaraja-Watson ([100]-[126]) en temps continu a fait sa première apparition en 1978, Banon [6],

$$\widehat{m}_T(\mathbf{x}) = \frac{\int_0^T l(Y_t)K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt}{\int_0^T K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)}.$$

Les vitesses de convergence de l'estimateur à noyau de la fonction de régression relatif aux processus à temps continu ont été relativement peu étudiées. Pour les travaux qui ont été l'objet de cette problématique, on se réfère ici à Cheze- Payaud (1994)[32] ainsi que Bosq (1998, chapitre 5)[17] dans les cas dits optimaux et suroptimaux. Récemment, Lejeune (2007)[92] a donné des vitesses de convergence de l'erreur quadratique moyenne du régressogramme (analogue à l'histogramme pour l'estimation de la régression) en temps continu, citons aussi Bosq et Blanke (2007 Chapitre 7, 2008)[20],[21]. Plus récemment encore, Didi et Louani (2013b)[51] (voir aussi Chapitre 2) ont obtenu des vitesses non-paramétriques pour les processus stationnaires ergodiques et ont donné une application à la prévision.

Par ailleurs, les applications pour le modèle de la régression ont pris une place importante dans la prévision des séries chronologiques issues de différentes disciplines telles que la communication, les systèmes de contrôle, la climatologie, les études biomédicales ainsi que l'économétrie. Il s'agit donc des domaines de prévision sur lesquels les premiers résultats furent établis. On peut citer, pour plus de détails sur ce sujet, Collomb (1983, 1984)[27], [33] et Robinson (1983)[113]. Ces domaines ont connu des développements continus, citons par exemple, les livres généraux de Devroye *et al.* (1986)[49], Györfi et al. (1989)[66], Yoshihara (1994)[134]. Les travaux de Ramsay et Silverman (2002, 2005)[111],[112] constituent une collection importante de méthodes statistiques, en particulier pour la régression. Le livre de Bosq (2000)[19], aborde le traitement des modèles fonctionnels linéaires, en particulier le processus autorégressif, on cite aussi Bosq & Blanke (2008)[21].

Estimation du mode conditionnel

Soit $(\mathbf{X}, Y)$ un couple de variables aléatoires à valeurs dans $\mathbb{R}^d \times \mathbb{R}$. On suppose que le couple $(\mathbf{X}_t, Y_t)_{t\geq 0}$ est un processus strictement stationnaire et ergodique, admettant une densité jointe $r_{\mathbf{X},Y}$ et que le vecteur aléatoire $\mathbf{X}$ admette une densité marginale $g_{\mathbf{X}}$. Nous pouvons alors définir le mode conditionnel $\Theta(\mathbf{x})$, supposé unique sur $\mathbb{R}$, comme étant la valeur maximisant la densité conditionnelle de Y sachant $\mathbf{X} = \mathbf{x}$,

$$\Theta(\mathbf{x}) = \arg\max_{y\in\mathbb{R}} f(y \mid \mathbf{x}),$$

avec

$$f(y \mid \mathbf{x}) = \frac{r_{\mathbf{X},Y}(\mathbf{x}, y)}{g_{\mathbf{X}}(\mathbf{x})}.$$

Soient $(\mathbf{X}_1, Y_1), \ldots, (\mathbf{X}_n, Y_n)$ un échantillon d'observations du couple $(\mathbf{X}, Y)$. Pour estimer le mode conditionnel, on se donne deux noyaux de Parzen-Rosenblatt K_1 et K_2 et une fenêtre de lissage h_n, et on pose

$$f_n(y \mid \mathbf{x}) = \frac{\sum_{i=1}^n K_1\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right) K_2\left(\frac{y-Y_i}{h_n}\right)}{\sum_{i=1}^n K_1\left(\frac{\mathbf{x}-\mathbf{X}_i}{h_n}\right)}.$$

Des propriétés de convergence sur des estimateurs de la densité conditionnelle ont été obtenues entre autres par Bosq ((1971)[12],(1973)[13]), Chahboun (1984)[29] et Delecroix (1975)[42]. Pour des reférences plus récentes sur le sujet, on pourra consulter Efromovich (2007)[55]. Dans cet article, l'auteur expose la théorie minimax appliquée à la densité conditionnelle. Il montre en outre que l'estimateur par la méthode des séries orthogonales atteint la vitesse minimax et que ce résultat reste vrai même si les variables X et Y sont indépendantes. Hyndman *et al* (1996)[73] étudie les propriétés quadratiques de l'estimateur à noyau. L'article propose une variante de l'estimateur à noyau qui permet de réduire le biais de l'estimation. Ferraty *et al.* (2006)[62] et Laksaci (2007)[80] étudient les propriétés quadratiques d'estimateurs à noyau de la densité conditionnelle pour des variables fonctionnelles. Les travaux de Hall *et al* (2004)[69] et Fan et Yim (2004)[58] représentent des reférences récentes dans lesquelles sont étudiées les propriétés de certains estimateurs de la densité conditionnelle.

Un estimateur du mode conditionnel $\Theta(\mathbf{x})$ est défini comme la v.a. $\Theta_n(\mathbf{x})$ maximisant un estimateur de la densité conditionnelle $f_n(\cdot \mid \mathbf{x})$ de $f(\cdot \mid \mathbf{x})$, à savoir,

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

En temps continu, nous considérons l'estimateur de la fonction de densité conditionnelle suivant

$$f_T(y \mid \mathbf{x}) = \frac{h_T(\mathbf{x}, y)}{g_T(\mathbf{x})},$$

où

$$r_T(\mathbf{x}, y) = \frac{1}{Th_T^{d+1}} \int_0^T K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) K_2\left(\frac{y - Y_t}{h_T}\right) dt,$$

et

$$g_T(\mathbf{x}) = \frac{1}{Th_T^{d}} \int_0^T K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt.$$

Ici K_1 et K_2 sont deux noyaux et h_T est le paramètre de lissage. L'estimateur du mode conditionnel est alors défini par l'expression

$$\Theta_T(\mathbf{x}) = \arg\max_{y\in\mathbb{R}} f_T(y \mid \mathbf{x}).$$

Le problème de l'estimation du mode conditionnel et de la fonction de densité conditionnelle a fait l'objet de beaucoup d'attention dans le passé. Pour les données dépendantes ou indépendantes, Parzen [107](1962) établit la convergence faible et la normalité asymptotique dans le cas d'observations indépendantes et identiquement distribuées (i.i.d). Roussas (1969)[118] fut le premier à établir ces propriétés asymptotiques pour des données markoviennes, ainsi que la convergence en probabilité. Youndjé (1993)[130] s'est intéressé à l'étude de la densité conditionnelle pour des données indépendantes identiquement distribuées, il a proposé un critère de selection de la largeur de la fenêtre de lissage basé sur la validation croisée. Laksasi et Yousfate (2002)[90] ont établi, pour un processus markovien stationnaire, la convergence en norme $\mathcal{L}^p$ pour l'estimateur à noyau de la densité conditionnelle. Comme application, Rossi(2004)[115] a utilisé le mode conditionnel dans le domaine des hautes technologies pour décrire un procédé de dé-pollution biologique. Ainsi, le mode conditionnel, par son importance dans le domaine de la prévision non paramétrique était à l'origine de la motivation de nombreux d'auteurs. Sans prétendre à l'exhaustivité, citons les travaux de Samanta et Thavasneswaran (1990)[120] qui ont donné les propriétés de consistance et de normalité asymptotique dans le cadre i.i.d alors que les conditions de convergence dans le cas de données $\phi-$mélangeantes avaient été établies par Collomb et al. (1987)[35]. Dans le cas des données ergodiques des résultats sont obtenus par Rosa (1993)[116] et Ould Saïd (1997)[105], alors que dans le cas de données $\alpha-$mélangantes, nous citons Ould Saïd (1993)[104] et Berlinet *et al* (1998)[10]. Enfin, Quintela-Del-Ráo et Vieu (1997)[110] ont estimé le mode conditionnel comme étant le point annulant la dérivée d'ordre un de l'estimateur de la densité conditionnelle et établi la convergence presque complète de cet estimateur sous la condition d'$\alpha-$mélange. Louani et Ould Saïd (1999)[84] ont établi la normalité asymptotique dans le cas des données fortement mélangeantes. Plus récémment, Ezzahrioui et Ould Saïd (2008, 2010)[56],[57] ont étudié la normalité asymptotique et la consistance dans le cadre des données fonctionnelles.

En temps continu, nous renvoyons au chapitre 4 de cette thèse dans lequel nous démontrons des résultats de convergence presque sûre, ponctuelle et uniforme, avec des vitesses de convergence sous l'hypothèse d'ergodicité de l'estimateur à noyau proposé.

0.2 Théorie ergodique

La théorie ergodique est apparue dans la mécanique statistique, notamment dans la théorie de Maxwell et dans la théorie de Gibbs. Il est nécessaire de faire une sorte de transition logique entre le comportement moyen de l'ensemble des systèmes dynamiques et la moyenne temporelle des comportements d'un système dynamique unique. Elle découle d'une hypothèse ingénieuse dont on s'est servi pendant une longue période sans la justifier, et sous des formes variées. Le premier résultat rigoureux pour le mathématicien, est le célèbre théorème de récurrence de Poincaré datant de l'année 1890 (voir Peskir (2000)[109], théorème 1.3, p.6) mais l'essor de la théorie date de 1931 avec les théorèmes de Neumann et de Birkhoff, elle entre alors définitivement dans le cadre de l'analyse fonctionnelle. Les théorèmes fondamentaux sont dus à Koopman, von Neumann et Carlman, d'une part, et à Birkhoff d'autre part, nous nous référons au Peskir (2000)[109] pour le théorème ergodique ponctuelle de Birkhoff (théorème 1.6, p.8), le théorème ergodique de Von Neumann (théorème 1.7, p.10) et le théorème ergodique sous-additif de Kingman (théorème 1.8, p.11).

La question naturelle qui se pose est : quand les moyennes des grandeurs générés de façon stationnaire convergent ? Dans la situation classique la stationnarité est décrite par une mesure préservant la transformation τ, et l'on considère les moyennes prises sur une séquence $f, f \circ \tau, f \circ \tau^2, \ldots$ pour une fonction f intégrable. Cela correspond au concept probabiliste de la stationnarité.

Nous nous sommes intéressés au théorème ergodique, qui s'applique aux processus stationnaires et ergodiques. Le théorème ergodique de Birkhoff, défini pour les processus stationnaires au lieu d'endomorphismes de τ, a maintenant la forme suivante (voir Krengel (1985)[77], p. 26, théorème 4.4.)

Théorème 0.1. Si $Y_0, Y_1, \ldots$ est un processus réel stationnaire et Y_0 est intégrable, alors

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} Y_k = \mathbb{E}[Y_0 \mid \mathcal{F}], \quad \text{p.s.} \tag{1}$$

où $\mathcal{F}$ est le $\sigma-$algèbre des ensembles invariants. Si, de plus, le processus est ergodique la limite coïncide avec l'espérance de la variable Y_0.

Pour la convergence en L^2 de (1) et de la sa version uniforme (en temps discret et continu) nous renvoyons au papier du Peskir (1998)[108]

Nous nous servirons d'une version de ce théorème pour les processus strictement stationnaires et ergodiques à temps continu.

Par souci de clarté et de compréhension, nous présentons quelques détails qui définissent la propriété ergodique des processus en temps continu.

Soit $X = \{X_t\}_{t\geq 0}$ un processus à temps continu prenant ses valeurs dans un espace mesurable $(E, \mathcal{X})$ sur lequel est définie une mesure de probabilité μ. Pour $\delta > 0$, soit Υ^δ une transformation δ, i.e., $(\Upsilon^\delta(x))_s = x_{s+\delta}$. Un ensemble mesurable $\mathcal{A}$ est dit $\delta-$invariant s'il ne change pas sous une transformation $\delta-$shift, i.e., $\Upsilon^\delta(\mathcal{A}) = \mathcal{A}$.

Définition 0.1. ($\delta-$ergodicité) Un processus en temps continu $X = \{X_t\}_{t\geq 0}$ est dit δ-ergodique, si chaque ensemble mesurable δ-invariant lié au processus X a une probabilité soit 1 ou 0. Autrement dit, pour tout ensemble δ-invariant $\mathcal{A}$, $\mu(\mathcal{A}) = (\mu(\mathcal{A}))^2$.

Cette définition signifie que si nous discrétisons le processus X en blocs de temps de longueur δ, le processus discrétisé $(X_0^\delta, X_\delta^{2\delta}, X_{2\delta}^{3\delta}, X_{3\delta}^{4\delta}, \ldots)$ obtenu est ergodique.

Définition 0.2. (Ergodicité) Un processus en temps continu $X = \{X_t\}_{t\geq 0}$ est ergodique s'il est δ-ergodique pour tout $\delta > 0$.

Alors le théorème ergodique ponctuel est donné sous la forme

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T Y_t dt = \mathbb{E}[Y_0], \quad \text{p.s.} \tag{2}$$

Dans notre travail nous utilisons une version fonctionnelle de l'équation (2). Delecroix (1987)[45] fut le premier à introduire cette expression sous la forme d'une hypothèse vérifiée par les processus stationnaires, on pourra consulter à ce sujet aussi Gyorfi *et al* (1989)([66], théorème 3.5.1), Delecroix *et al* (1991)[46] imposent l'existence des densités conditionnelles, $f_{X_i}^{\mathcal{F}_{i-1}}$, des variables de X_i par rapport aux tribu ($\mathcal{F}_{i-1} = \sigma(X_{i-j}, j \geq 1)$ et posant $U_n = \frac{1}{n}\sum_{i=1}^n f_{X_i}^{\mathcal{F}_{i-1}}$, ils utilisent les hypothèses suivantes pour établir les convergences souhaitées

(1) $f \in \mathcal{C}^0(\mathbb{R}^d), \quad f_{X_i}^{\mathcal{F}_{i-1}} \in \mathcal{C}^0(\mathbb{R}^d)$,

(2) $\sup_{\mathbf{x}\in\mathbb{R}^d} |U_n(\mathbf{x}) - f(\mathbf{x})| \xrightarrow[n\to\infty]{a.s} 0$.

L'hypothèse (2) implique, pour tout $\mathbf{x} \in \mathbb{R}^d$, que

$(2)^\star$ $U_n(\mathbf{x}) - f(\mathbf{x}) \xrightarrow[n\to\infty]{a.s} 0$.

Notre approche consiste à introduire une version de $(2)^\star$ dans le cas continu. Par fois, on remplace les densités conditionnelles par d'autres fonctions aléatoires liées au processus et obtenues par le calcul des probabilités des petites boules.

La variété des phénomènes modélisés par des processus ergodiques motive notre travail et montre son importance. Les processus ergodiques sont largement utilisés dans la modélisation de problématiques appliqués. Sans être exhaustif, nous citons à titre d'exemples

– sciences biomédicales, voir Banks (1975)[5],
– économie, voir Bergstrom (1990)[9], Hale et Verduyn Lunel (1993)[70],
– analyse génétique, voir Lange (2002)[81],
– mécaniques, voir Arnold (1973)[4],
– physiques, voir Papanicolaou (1995)[106],
– mathématiques financières, voir Karatzas et Shreve (1991)[75],
– démographie, voir Cohen (1979)[36].

La description de certains phénomènes par des processus à temps discret peut provoquer une importante perte d'information. Le préjudice subi alors peut remettre en cause toute la modélisation mathématique. De ce fait, un choix de modèle en temps continu peut être déterminant pour la bonne description des phénomènes et les applications à la prévision de leurs évolutions.

Nous sommes maintenant en mesure de donner le plan de la thèse dans le détail.

0.3 Plan de la thèse

Dans le cadre de l'estimation non-paramétrique à noyau, la taille de la fenêtre h joue, comme on peut s'y attendre, un rôle central et influe sur la convergence. Dans tout ce qui suit, nous utilisons le paramètre de lissage $h = h_T$ vérifiant les conditions asymptotiques (H.1-2) données ci-dessous (ces conditions sont souvent appelées conditions de de Csörgö-Révész-Stute (Csörgö et Révész (1981)[37], Stute (1982)[125]) ou [CRS]). Soit $\{h_T\}_{T\geq 0}$ une suite de réels positifs, vérifiant

(H.1) $h_T \downarrow 0$ et $Th_T \uparrow \infty$ lorsque $T \to \infty$,

(H.2) $Th_T/\log T \to \infty$ lorsque $T \to \infty$.

Chapitre 1. Estimation de la densité

Dans le chapitre 1, nous nous intéressons au comportement asymptotique de l'estimateur à noyau de la densité d'un processus stationnaire à temps continu.

Soit $(X_t)_{t\geq 0}$ un processus réel stationnaire à temps continu, admettant une densité f. Pour un T fixé, l'estimateur à noyau de la densité de type Parzen-Rosenblatt est défini, pour tout $x \in \mathbb{R}$, par

$$f_T(x) = \frac{1}{Th_T}\int_0^T K\left(\frac{x - X_t}{h_T}\right) dt,$$

où K est le noyau et h_T est le paramètre de lissage. Sous certains conditions de régularité sur f et K ainsi qu'un choix approprié de $h_T > 0$, nous obtenons la convergence de f_T vers f. Plus précisément, nous présenterons des propriétés asymptotiques de consistance presque sûre, ponctuelle et uniforme, avec des vitesses de convergence. Nous obtenons nos résultats, en faisant usage d'une approche basée sur l'utilisation des différence de martingales en liaison avec des techniques de projection sur des tribus. Nous attacherons une importance primordiale à l'étude de la vitesse de convergence presque sûre, ponctuelle et uniforme du " terme stochastique " obtenu par les estimateurs centrés (par des facteurs n'étant pas nécessairement égaux à l'espérance de l'estimateur). En effet, il peut être montré que, pour les tailles de la fenêtre h les plus appropriées aux applications statistiques (vérifiant les conditions [CRS] avec un ordre de convergence convenable vers 0), et sous des conditions générales de régularité, le " terme de biais " (i.e., le facteur de centrage moins la fonction à estimer) converge avec une vitesse vers zéro. Au préalable, rappelons qu'un noyau K défini sur $\mathbb{R}$ est dit d'ordre l, avec $l \geq 1$ entier, si

1. $\int K(z)dz = 1$,
2. $\int |z|^k K(z)dz = 0, \ k = 0, \cdots, l-1$,
3. $\int |z|^l K(z)dz < \infty$,

Dans cet esprit, nous établirons dans le chapitre 1, les théorèmes suivants qui traitent de la consistance ponctuelle et uniforme presque sûre.

Théorème 0.2. On suppose que le processus $(X_t)_{t\geq 0}$ est strictement stationnaire et ergodique, autrement dit, pour tout $\delta > 0$ suffisamment petit, on a

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(x)dt = f(x), \quad \text{p.s.} \tag{2}$$

où $f_t^{\mathcal{F}_{t-\delta}}$ est la densité conditionnelle de la v.a X_t sachant le tribu $\mathcal{F}_{t-\delta}$. En outre, supposons que les conditions (A.1)(i)-(ii) soient satisfaites, la densité $f_t^{\mathcal{F}_{t-\delta}}$ est une fonction Lipschitzienne et qu'il existe une constante L telle que

$$0 < L < \frac{Th_T^2}{\log T},$$

alors nous avons

$$\lim_{T\to\infty} f_T(x) = f(x), \quad \text{p.s.} \tag{2}$$

La vitesse de convergence est donnée par le théorème suivant,

Théorème 0.3. Supposons que les hypothèses (A.1)(i)-(ii), (A.2) et (A.3) soient satisfaites. Si de plus $Th_T/\log T \to \infty$ lorsque $T \to \infty$, alors, pour tout $x \in \mathbb{R}$, nous avons

$$f_T(x) - f(x) = O(h_T) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right), \quad \text{lorsque} \quad T \to \infty. \tag{2}$$

La convergence uniforme presque sûre de l'estimateur f_T est établie sous des conditions supplémentaires portant sur le noyau K et qui sont

1. K est à support compact.
2. K est lipschitzien.

Le théorème suivant montre la convergence uniforme presque sûre,

Théorème 0.4. Soit $(X_t)_{t\geq 0}$ un processus strictement stationnaire et ergodique telle que, pour tout $\delta > 0$ suffisamment petit,

$$\lim_{T\to\infty} \sup_{x\in\mathbb{R}} \left| \frac{1}{f(x)} \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(x)dt - 1 \right| = 0, \quad \text{a.s.} \tag{3}$$

En outre, on suppose que les conditions (A.1) et (A.6) soient satisfaites, et que la densité conditionnelle $f_t^{\mathcal{F}_{t-\delta}}$ est une fonction de Lipschitzienne et il existe deux constantes L et A de telle sorte que

$$(i) \quad 0 < L \leq \frac{Th_T^2}{\log T}, \quad (ii) \quad \frac{h_T'}{h_T^2} \leq A < \infty.$$

avec h_T' est la dérivée de h_t par rapport à T, alors,

$$\lim_{T\to\infty} \sup_{x\in\mathbb{R}} |f_T(x) - f(x)| = 0, \quad \text{a.s.}$$

Le dernier résultat de ce chapitre se présente sous la forme suivante

Théorème 0.5. Supposons que les hypothèses $(A.1)$-$(A.6)$ soient satisfaites. Si de plus $Th_T/\log T \to \infty$ quand $T \to \infty$, on a

$$\sup_{x\in\mathbb{R}} |f_T(x) - f(x)| = O(h_T) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right), \quad \text{lorsque} \quad T \to \infty. \tag{4}$$

Cette vitesse de convergence peut être améliorée comme le montre la remarque suivante

Remarque 0.1. Sous les hypothèses du théorème 0.5, nous supposons, en outre, que la densité f est s fois dérivable et que sa dérivée d'ordre s est bornée, le noyau K est d'ordre s, alors nous obtenons le résultat suivant

$$\sup_{x\in\mathbb{R}} |f_T(x) - f(x)| = O(h_T^s) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right), \quad \text{lorsque} \quad T \to \infty.$$

Chapitre 2. Estimation de la régression

L'objet du chapitre 2, est d'étudier le problème de régression d'une variable réelle Y sur la co-vecteur $\mathbf{X}$ dans $\mathbb{R}^d$, la fonction de régression est l'espérence de la variable Y conditionnée par $\mathbf{X}$:

$$m(\mathbf{x}) = \mathbb{E}[l(Y) \mid \mathbf{X} = \mathbf{x}],$$

où l est une fonction mesurable réelle. Pour T fixé, le processus observé $(\mathbf{X}_t)_{0\leq t\leq T}$ est supposé stationnaire et ergodique, nous avons choisi une version de l'estimateur à noyau de Nadaraya-Watson en temps continu de la fonction m, défini comme suit

$$\widehat{m}_T(\mathbf{x}) = \frac{\int_0^T l(Y_t)K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt}{\int_0^T K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt}.$$

Nous établirons les convergences presque sûres, ponctuelles et uniformes, et nous utiliserons ces résultats pour faire une application à la prévision dans le cadre des séries temporelles. L'approche utilisée est analogue à celle élaborée pour étudier la convergence de l'estimateur de la densité. La vitesse de convergence est une synthèse entre la vitesse obtenue pour le terme stochastique et la vitesse établie pour le biais conditionnel. Le résultat est exposé dans le Théorème suivant :

Théorème 0.6. Sous les hypothèses $(H.0)$-$(H.3)$, pour tout $\mathbf{x}\in\mathbb{R}^d$ tel que $f(\mathbf{x})\neq 0$, on a

$$\widehat{m}_T(\mathbf{x}) - m(\mathbf{x}) \overset{a.s}{=} O(h_T^\beta) + O\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right), \quad \text{lorsque} \quad T\to\infty. \tag{5}$$

Le Théorème 2.2 du chapitre 2, montre la convergence uniforme sur un compact dilatant $\mathcal{B}_T$ de $\mathbb{R}^d$, tel que $\mathcal{B}_T = \{\mathbf{x}\in\mathbb{R}^d : \|\mathbf{x}\|\leq T^\gamma\}$, où $\frac{1}{2}<\gamma<1$ et $\|\cdot\|$ est une norme définie sur $\mathbb{R}^d$.

Soit (α_T) une suite des réels positifs tels que $\underset{x\in\mathcal{B}_T}{inf}\, f(x)\geq\alpha_T$.

Théorème 0.7. Suppososns les hypothèses $(H.0)$-$(H.4)$ satisfaites. En outre, assumons que

$$\frac{T^{2\gamma-1}h_T^d\alpha_T^2}{\log T} = O(1), \quad \text{lorsque} \quad T\to\infty.$$

Alors, on a

$$\sup_{\mathbf{x}\in\mathcal{B}_T}|\hat{m}_T(\mathbf{x}) - m(\mathbf{x})| \overset{a.s}{=} O(h_T^\beta) + O\left(\left(\frac{\log T}{Th_T^d\alpha_T^2}\right)^{1/2}\right), \quad \text{lorsque} \quad T\to\infty. \tag{6}$$

Lorsque, au lieu de $\mathcal{B}_T$, nous considérons un compact $\mathcal{C}$ sur lequel la densité f est minorée par une constante, nous obtenons le résultat suivant

Remarque 0.2. Sous les même hypothèses que le Théorème 0.7, nous avons

$$\sup_{\mathbf{x}\in\mathcal{C}}|\hat{m}_T(\mathbf{x}) - m(\mathbf{x})| \overset{a.s}{=} O(h_T^\beta) + O\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right), \quad \text{lorsque} \quad T\to\infty.$$

L'application la plus naturelle de ce type de résultat, est celle relative à la prédiction non-paramétrique d'un processus stationnaire et ergodique $(\xi_t, t \in \mathbb{R})$. À partir d'un échantillon observé $(\xi_t, 0 \leq t \leq T)$, nous approximons la variable aléatoire $\zeta_{T+H} = l(\xi_{T+H})$ de carré intégrable, où $0 < H < T$, par $m(\xi_T) = \mathbb{E}[\zeta_{T+H} \mid \xi_s, s \leq T] = \mathbb{E}[\zeta_{T+H} \mid \xi_T]$.

Par conséquent, le prédicteur non paramétrique de ζ_{T+H} est défini par

$$\hat{\zeta}_{T+H} = \frac{\hat{m}_{T-H,2}(\xi_T)}{\hat{m}_{T-H,1}(\xi_T)} = \frac{\int_0^{T-H} l(\xi_{t+H}) K\left(\frac{\xi_T - \xi_t}{h_T}\right) dt}{\int_0^{T-H} K\left(\frac{\xi_T - \xi_t}{h_T}\right) dt} = \hat{m}_{T,H}(\xi_T).$$

Le corollaire suivant, étudie le comportement asymptotique de ce prédicteur.

Corollaire 0.1. En prenant $d = 1$, et si le processus $Z_t = (\xi_t, l(\xi_{t+H}))$ vérifie les hypothèses du Théorème 0.7. Alors, on a

$$|\hat{m}_{T,H}(\xi_T) - m(\xi_T)| \mathbb{1}_{\{\xi_T \in \mathcal{B}_T\}} \overset{a.s}{=} \mathcal{O}(h_T^{\beta}) + \mathcal{O}\left(\left(\frac{\log T}{T h_T \alpha_T^2}\right)^{1/2}\right) \quad \text{lorsque} \quad T \to \infty.$$

Chapitre 3. Estimation du mode conditionnel

Nous considérons l'estimation de la fonction du mode conditionnel lorsque les co-variables sont des vecteurs réels. Le principal objectif de ce chapitre est d'établir la convergence presque sûre, ponctuelle et uniforme avec des vitesses de convergence de l'estimateur à noyau du mode conditionnel lorsque le processus est supposé strictement stationnaire et ergodique.

Un estimateur à noyau du mode conditionnel $\Theta(\mathbf{x})$ est défini comme la variable aléatoire $\Theta_T(\mathbf{x})$ qui maximise l'estimateur à noyau $f_T(\cdot \mid \mathbf{x})$ de $f(\cdot \mid \mathbf{x})$, c'est à dire,

$$f_T(\Theta_T(\mathbf{x}) \mid \mathbf{x}) = \max_{y \in \mathbb{R}} f_T(y \mid \mathbf{x}). \tag{7}$$

Notons que l'estimateur $\Theta_T(\mathbf{x})$ n'est pas forcément unique, et compte tenu de ce fait, nous traiterons toute valeur $\Theta_T(\mathbf{x})$ vérifiant l'équation (7).

Théorème 0.8. Sous les hypothèses (A.1)(i)-(ii2), (A.2)(i), (A.3), (A.4)(i), (A.5), (A.6), (A.7) et (A.8), on a

$$\begin{aligned}\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}) \quad &= \quad O(h_T^{\beta/2}) + O\left(h_T^{\beta/2}\left(\frac{\log T}{T h_T^d}\right)^{1/4}\right) + O\left(\left(\frac{\log T}{T h_T^d}\right)^{1/4}\right) \\ &\quad + O\left(\left(\frac{\log T}{T h_T^{d+1}}\right)^{1/4}\right), \quad \text{p.s.}\end{aligned}$$

Théorème 0.9. Sous les hypothèses (A.1)(i)-(ii)-(iv)-(v), (A.2)(i), (A.3), (A.4), (A.6), (A.7), et (A.8), on a

$$\begin{aligned}\sup_{x\in\mathcal{C}}|\Theta_T(\mathbf{x})-\Theta(\mathbf{x})| &= O(h_T^{\beta/2})+O\left(\left(\frac{\log T}{Th_T^d}\right)^{1/4}\right)+O\left(h_T^{\beta/2}\left(\frac{\log T}{Th_T^d}\right)^{1/4}\right)\\ &\quad +O\left(\left(\frac{\log T}{Th_T^{d+1}}\right)^{1/4}\right), \quad \text{p.s.}\end{aligned}$$

De façon similaire au travail efféctuée pour la régression usuelle, en considérant l'estimateur $\widehat{\Theta}_T(\xi_T)$ de la valeur prédite $\Theta(\xi_T)$ à l'horizon H, on obtient le résultat suivant

Corollaire 0.2. En prenant $d = 1$ et supposant que les hypothèses du Théorème 0.9 soient satisfaites pour le processus $Z_t = (\xi_t, \xi_{t+H})$. Nous avons alors

$$\begin{aligned}|\widehat{\Theta}_{T-H}(\xi_T)-\Theta(\xi_T)|\mathbb{1}_{\{\xi_T\in\mathcal{C}\}} &= O(h_T^{\beta/2})+O\left(h_T^{\beta/2}\left(\frac{\log T}{Th_T}\right)^{1/4}\right)+O\left(\left(\frac{\log T}{Th_T}\right)^{1/4}\right)\\ &\quad +O\left(\left(\frac{\log T}{Th_T^2}\right)^{1/4}\right), \text{ p.s.}\end{aligned}$$

Chapitre 1

Consistency results for the kernel density estimate on continuous time stationary and dependent data

Ce chapitre à fait l'objet d'une publication au *Journal of Statistics and Probability Letters*

Abstract

The aim of this paper is to study the consistency of the kernel density estimator pertaining to a continuous time stationary process $X = (X_t)_{t\geq 0}$, with an underlying density f. More precisely, in a rather general dependency setting, where we use a martingale difference device and a technique based on a sequence of projections on σ-fields, we establish the almost sure pointwise and uniform consistencies with rates of the estimate f_T of f built upon the part $(X_t)_{0\leq t\leq T}$ of the process X.

AMS 1991 subject classification : Primary : 60F10, 62G07, 62F05, 62H15.
Key words and phrases : Consistency, continuous time, density function, kernel estimator, rate of convergence, stationary.

1.1 Introduction

Let $X = (X_t)_{t\geq 0}$ be a real-valued stationary process having a common density f with respect to the Lebesgue measure. On the basis of the part $(X_t)_{0\leq t\leq T}$ observed from the process X, the kernel estimate of the density function f is defined, for any $x \in \mathbb{R}$, by

$$f_T(x) = \frac{1}{Th_T}\int_0^T K\left(\frac{x - X_t}{h_T}\right) dt,$$

where K is a kernel and h_T is the smoothing parameter, i.e., a sequence of positive real numbers tending to zero with some rate that will be given later on.

The nonparametric function estimation on continuous time processes subject has motivated a number of works studying the consistency with rate of convergence for both the density and the regression function estimates while considering various criteria of convergence together with the asymptotic normality. The kernel estimate of the d-variate density function is studied by Delecroix (1980)[44] and Bosq (1996)[14] where rates of convergence for the MSE and the supremum norm criteria are found to be $O(T^{-4/(4+d)})$ and $o(\log_k(T)(\log(T)/T)^{-4/(4+d)})$ respectively. Here, $\log_k(T) = \max\{1, \log_{k-1}(T)\}$ for $T > 0$ and $k \geq 2$. We cite the paper of Castellana and Leadbetter (1986)[26] who studied the consistency together with the asymptotic normality in the real-valued process case. We cite also Bosq and Blanke (2007)[20] (and reference therein)(chapter 6, chapter 7) for more improvements in large dimensional spaces, When the process $(X_t)_t$ satisfies a local irregularity condition, these results may be sharpened as to reach parametric rates given respectively by $O(1/T)$ and $o(\log_k(T)(\log(T)/T)^{1/2})$ as stated in Bosq (1997)[15]. Asymptotic properties of the kernel regression function estimator for continuous time processes have been investigated in several papers. Parametric rates of convergence are also reached in this framework for various criteria. We refer to Bosq ((1996)[14], (1997)[15]) and the references therein for an account of results on the subject. Notice that all these results are stated while considering the mixing setting for the data. It is well known that the mixing conditions may be viewed as an asymptotic independence. Therefore, many processes fail to satisfy these conditions. Examples of such processes are given in papers of Chernick (1981)[31] and Andrews (1984)[2] in the discrete time case. The motivation of this work is to consider a dependence framework that is not covered by the usual mixing cases.

To establish ours results, the proof methodology uses a martingale difference device to treat the conditional variance term and a technique based on a sequence of projections on an increasing sequence of σ-fields to deal with the conditional bias term. Almost sure pointwise and uniform consistencies with rates of the kernel density estimate are established in this paper. Notice that only nonparametric rates of convergence are reached since no irregularity conditions are assumed on the underlying process.

In order to display our results, some notations are needed. From now on, for a positive real number δ such that $n = \frac{T}{\delta} \in \mathbb{N}$, consider the δ-step partition $(T_j)_{0\leq j\leq n}$ of the interval $[0, T]$. Moreover, for $t > 0$ and $1 \leq j \leq n$, consider the σ-fields $\mathcal{F}_t = \sigma((X_s) : 0 \leq s < t)$ and $\mathcal{G}_j = \sigma((X_s), 0 \leq s \leq T_j)$. For any $t > s$ define $f_{X_t}^{\mathcal{F}_s}$ as the conditional density of X_t given the σ-field $\mathcal{F}_s$ and set $f^{\mathcal{F}_{t-\delta}} := f_{X_t}^{\mathcal{F}_{t-\delta}}$. Whenever $s < 0$, $\mathcal{F}_s$ stands as the trivial σ-field. For a real random variable ξ and any $k \in \mathbb{N}$, we define the projection $\mathcal{P}_k$ by $\mathcal{P}_k\xi = \mathbb{E}[\xi|\mathcal{F}_k] - \mathbb{E}[\xi|\mathcal{F}_{k-1}]$, where $\mathbb{E}[\xi|\mathcal{F}_k]$ stands as the conditional expectation of ξ given the σ-field $\mathcal{F}_k$.

1.2 Results

Set now some hypotheses necessary to state our results.

(A.1) (i) K is a positif bounded kernel and $\int K(z)dz = 1$,

(ii) $\int |z|K(z)dz < \infty$,

(iii) The kernel K is a Lipschitz function.

(A.2) (i) The density f is differentiable with a bounded derivative.

(ii) For any $\delta > 0$, the conditional density $f^{\mathcal{F}_{t-\delta}}$ is differentiable with almost surely bounded derivative.

(iii) For any $t \in [0, T]$, any $\delta > 0$ and any $x \in \mathbb{R}$, the function $f^{\mathcal{F}_{t-\delta}}(x)$ is almost surely bounded by a deterministic function $b_{t,\delta}(x)$,

(iv) For any $\delta > 0$, $T^{-1}\int_0^T b_{t,\delta}(x)dt \to D(x) \neq 0$ as $T \to \infty$.

(A.3) For any δ, we have

$$\sup_y \int_{\mathbb{R}^+} \left\|\mathcal{P}_1 f^{\mathcal{F}_{t-\delta}}(y)\right\|^2 dt < \infty.$$

Comments on assumptions.

Conditions (A.1) are very common in nonparametric function estimation literature. Hypotheses (A.2) impose the needed regularity upon the marginal and the conditional densities to reach the rates of convergence given below. Approximating the integral $\int_{\mathbb{R}^+} \left\|\mathcal{P}_1 f^{\mathcal{F}_{t-\delta}}(y)\right\|^2 dt$ by its Riemann's sum, the assumption (A.3) is similar to the condition assumed by Wu (2003)[127] in the discrete time case which is satisfied by various processes including linear as well as many nonlinear ones. For more details and examples, see Wu (2003)[127]. We refer also to the recent paper of Wu *et al.* (2010)[128].

The following theorems establish the main results of this paper.

1.2.1 Pointwise almost sure consistency

The following theorem deals with simple almost sure convergence of f_T,

Theorem 1.1. Assume that the process $(X_t)_{t\geq 0}$ is strictly stationary and ergodic, in other words, for any $\delta > 0$ small enough,

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(x)dt = f(x), \quad \text{a.s.} \tag{1.1}$$

In addition, suppose that the conditions (A.1)(i)-(ii) hold true, the conditional density $f_t^{\mathcal{F}_{t-\delta}}$ is a lipschitz function and there exists a positive constant L such that

$$0 < L \leq \frac{Th_T^2}{\log T}, \tag{1.2}$$

then, we have

$$\lim_{T\to\infty} f_T(x) = f(x), \quad \text{a.s.}$$

Comments.

Here the main assumption is that (X_t) is strictly stationary and ergodic represented by Assumption (1.1), it takes the form of the Birkhoff's ergodic theorem, spelled out for stationary processes in continuous time, see Krengel (1985)[77]. For more details and explanations see the discussion on assumptions in chapter 2, we refer the papers of Delecroix (1987)[45], Delecroix & Negueira & Rosa (1991)[46], Delecroix & Rosa (1996)[47], Laib & Louani ((2010)[88], (2011)[89]), Peskir (1998)[108] and references therein for an account of results on the subject. The Lipschitz condition assumed on $f_t^{\mathcal{F}_{t-\delta}}$ is the only regularity condition upon the conditional density.

The second theorem of this section state the almost sure convergence with rate,

Theorem 1.2. Assume that hypotheses (A.1)(i)-(ii), (A.2) and (A.3) are satisfied. If $Th_T/\log T \to \infty$ as $T \to \infty$, then, for any $x \in \mathbb{R}$, we have

$$f_T(x) - f(x) = O(h_T) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right), \quad \text{as} \quad T \to \infty.$$

1.2.2 Uniform almost sure consistency

From now on, let γ be a constant taking values in the interval $[0,1]$ and consider the set $\mathcal{B}_T = \{x : |x| \leq T^\gamma\}$.

Our uniform result is stated and proved under the following additional conditions,

(A.4) For any $T > 0$, $\sup_{x\in\mathcal{B}_T} D(x) = D_0 < \infty$.

(A.5) There exists a constant $C > 0$ such that for T large enough, $\dfrac{|h'_T| T^{1/2}}{h_T^{3/2}(\log T)^{1/2}} \leq C$, where h'_T is the derivative of h_T.

(A.6) The kernel K is of bounded support.

Now we can state the first uniform result on $\mathbb{R}$,

Theorem 1.3. Let $(X_t)_{t\geq 0}$ be a strictly stationary and ergodic process such that, for any $\delta > 0$ small enough,

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}}\left|\frac{1}{f(x)}\frac{1}{T}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(x)dt - 1\right| = 0, \quad \text{a.s.} \tag{1.3}$$

Furthermore, assume that conditions (A.1) and (A.6) hold true, the conditional density $f_t^{\mathcal{F}_{t-\delta}}$ is a lipschitz function and there exists two constants L and A such that

$$(i) \quad 0 < L \leq \frac{T h_T^2}{\log T}, \quad (ii) \quad \frac{h'_T}{h_T^2} \leq A < \infty. \tag{1.4}$$

where h'_T is the derivative of h_T with respect to T. Then, we have

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}}|f_T(x) - f(x)| = 0, \quad \text{a.s.}$$

Comments.

As in Theorem 1.1, the underlying process (X_t) is assumed to be strictly stationary and ergodic. In the discrte time, to the best of our knowledge, this kind of assumption, given in the statement (1.3), appeared for the first time in the paper of Delecroix (1980)[44]. it was used in several papers among whom we cite Györfi & *al* (1989)[66], Ould Said (1997)[105] and Laib (2005)[87]. The discussion of this condition is performed in Delecroix (1987)[45]. Notice that in the continuous time case, Peskir (1998)[108] obtened the L^2-convergence version of the Burkholder Theorem. In this thesis, we give a complete discussion of the assumption (1.3) in chapter 2.

The last result of this chapter is presented in the following form

Theorem 1.4. Assume that assumptions (A.1)-(A.6) hold true. If $Th_T/\log T \to \infty$ as $T \to \infty$, then we have

$$\sup_{x\in\mathbb{R}} |f_T(x) - f(x)| = O(h_T) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) \quad \text{as} \quad T \to \infty.$$

This convergence rate may be improved as shown in the following remark,

Remark 1.1. If we assume that the density f is twice differentiable with a bounded second derivative and that $\int zK(z)dz = 0$ and $\int |z|^2K(z)dz < \infty$, then we obtain the following result

$$\sup_{x\in\mathbb{R}} |f_T(x) - f(x)| = O(h_T^2) + O_{a.s}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) \quad as \quad T \to \infty.$$

1.3 Proofs

From now on, for any $1 \le j \le n$, any $T \ge 0$ and any $\delta > 0$, set

$$\begin{aligned}
\bar{f}_T(x) &= \frac{1}{T}\int_0^T \mathbb{E}\left[\frac{1}{h_T}K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt, \\
Z_j &:= Z_j(x) = \int_{T_{j-1}}^{T_j} K\left(\frac{x-X_t}{h_T}\right)dt \qquad (1.5)\\
Y_j &:= Y_j(x) = \int_{T_{j-1}}^{T_j}\left(K\left(\frac{x-X_t}{h_T}\right) - E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt. \qquad (1.6)
\end{aligned}$$

Notice, for any $\delta > 0$, that $(Y_j)_{1\le j\le n}$ is a sequence of martingale differences with respect to the sequence of σ-fields $(\mathcal{G}_j)_{1\le j\le n}$. Indeed, since, for any $t \in [T_{j-1}, T_j]$, $\mathcal{G}_{j-2} \subseteq \mathcal{F}_{t-\delta} \subseteq \mathcal{G}_{j-1}$, it is clear that Y_j is $\mathcal{G}_j$-measurable and satisfies

$$E\left[Y_j|\mathcal{G}_{j-2}\right] = E\left[\int_{T_{j-1}}^{T_j}\left(K\left(\frac{x-X_t}{h_T}\right) - E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt|\mathcal{G}_{j-2}\right] = 0.$$

To establish almost sure pointwise and uniform results, we need an exponential inequality for partial sums of unbounded martingale differences. This inequality is given in the following lemma.

Lemma 1.1. Let $(Z_n)_{n\ge1}$ be a sequence of real martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_n = \sigma(Z_1, \cdots, Z_n))_{n\ge1}$, where $\sigma(Z_1, \cdots, Z_n)$ is the σ-field generated by the random variables $Z_1, \cdots, Z_n$. Set $S_n = \sum_{i=1}^n Z_i$. For any

$p \geq 2$ and any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that

$$\mathbb{E}(|Z_n|^p|\mathcal{F}_{n-1}) \leq C^{p-2}p!d_n^2, \quad \text{almost surely.} \tag{1.7}$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}(|S_n| > \epsilon) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

where $D_n = \sum_{i=1}^n d_i^2$.

Proof. The proof follows as a particular case of Theorem 8.2.2 due to de la Peña and Giné (1999). $\square$

The next Lemma give us a similar exponential inequality for partial sums of bounded martingale differences.

Lemma 1.2 (Laïb(1999)[86]). Let $\{Z_n : n \in \mathbb{N}\}$ be a sequence of martingale differences with respect to the sequence of $\sigma-$fields $\{\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n) : i \in \mathbb{N}\}$, where $\sigma(Z_1, \ldots, Z_n)$ is the $\sigma-$field generated by the random variables $Z_1, , \ldots, Z_n$, such that $\|Z_n\| \leq B$ a.s. for $1 \leq i \leq n$. For all $\epsilon > 0$, one has

$$\mathbb{P}\left\{\max_{1\leq i\leq n}\left\|\sum_{j=1}^{i} Z_j\right\| > \epsilon\right\} \leq 2\exp\left\{-\frac{\epsilon^2}{2nB^2}\right\}. \tag{1.8}$$

Observe now that, for any $x \in \mathbb{R}$, one may write

$$\begin{aligned} f_T(x)-f(x) &= \frac{1}{Th_T}\int_0^T \left[K\left(\frac{x-X_t}{h_T}\right) - E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right) dt \\ &\quad + \frac{1}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right] dt - f(x) \\ &:= R_{1,T}(x) + R_{2,T}(x). \end{aligned} \tag{1.9}$$

1.3.1 Proof of Theorem 1.1

Making use of the decomposition (1.9) and statement (1.6), noting that $T = T_n = \delta n$ and $h_{T_n} = h_T$, we may write

$$R_{1,T_n}(x) = \frac{1}{T_n h_{T_n}}\int_0^T \left[K\left(\frac{x-X_t}{h_{T_n}}\right) - E\left[K\left(\frac{x-X_t}{h_{T_n}}\right)|\mathcal{F}_{t-\delta}\right]\right) dt \tag{1.10}$$

$$= \frac{1}{T_n h_{T_n}}\sum_{k=1}^{n} Y_k, \tag{1.11}$$

since the kernel K is bounded, then

$$\begin{aligned} \frac{1}{T_n h_{T_n}} |Y_k| &\leq \frac{1}{T_n h_{T_n}} \int_{T_{j-1}}^{T_j} \left| K\left(\frac{x - X_t}{h_{T_n}}\right) - E\left[K\left(\frac{x - X_t}{h_{T_n}}\right) | \mathcal{F}_{t-\delta}\right] \right| dt \\ &\leq \frac{2\delta \widetilde{K}}{T_n h_{T_n}}, \end{aligned}$$

with $\widetilde{K} = \max_{u \in \mathbb{R}} K(u)$. Observe that the sequence $\{Y_k\}_{1 \leq k \leq n}$ is a martingale differences sequence with respect to the sequence of $\sigma-$fields $\mathcal{G}_{k-1}$. Recalling $T = \delta n = T_n$ and $h_T = h_{T_n}$, Applying Lemma 1.2 for large enough n and any $\epsilon > 0$, we obtain

$$\begin{aligned} \mathbb{P}\left\{ \left| \sum_{k=1}^n Y_k \right| > \epsilon (T_n h_{T_n}) \right\} &\leq 2 \exp\left\{ -\frac{\epsilon^2 (T_n h_{T_n})^2}{8 n \delta^2 \widetilde{K}^2} \right\} \\ &= 2 \exp\left\{ -\frac{\epsilon^2}{8\delta \widetilde{K}^2} \frac{T_n h_{T_n}^2}{\log T_n} \log T_n \right\} \\ &= 2 \exp\left\{ \log T_n^{-C_k \frac{T_n h_{T_n}^2}{\log T_n}} \right\} \\ &= \frac{2}{T_n^{-C_k u(T_n)}}. \end{aligned}$$

where $C_k = \frac{\epsilon^2}{8\delta \widetilde{K}^2}$ is a positive constant, according to condition (1.2) the function $u(T) := \frac{T_n h_{T_n}^2}{\log T_n}$ is bounded from below by a strictly positive constant L. By Borel-Cantelli Lemma, we obtain

$$R_{1,T_n}(x) \overset{a.s}{=} 0, \quad \text{as} \quad T_n \to \infty. \tag{1.12}$$

We turn our attention to the second term of the decomposition (1.9), under condition (1.1) and since $f_t^{\mathcal{F}_{t-\delta}}$ is a Lipschitz function, by a change of variable, we obtain

$$\begin{aligned} R_{2,T}(x) &= \frac{1}{T h_T} \int_0^T E\left[K\left(\frac{x - X_t}{h_T}\right) | \mathcal{F}_{t-\delta}\right] dt - f(x) \\ &= \frac{1}{T h_T} \int_0^T \int_{\mathbb{R}} K\left(\frac{x - z}{h_T}\right) f_t^{\mathcal{F}_{t-\delta}}(z) dz dt - f(x) \\ &= \frac{1}{T} \int_0^T \int_{\mathbb{R}} K(u) \left(f_t^{\mathcal{F}_{t-\delta}}(x - h_T u) - f_t^{\mathcal{F}_{t-\delta}}(x) \right) du dt \\ &\quad + \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(x) dt - f(x) \\ &\leq \frac{1}{T} \int_0^T f_t^{\mathcal{F}_{t-\delta}}(x) dt - f(x) + O(h_T) = o(1). \end{aligned}$$

Therefore,

$$R_{2,T}(x) \overset{a.s}{=} 0, \quad \text{as} \quad T \to \infty. \tag{1.13}$$

Proof of Theorem 1.1 In conjunction with equations (1.12) and (1.13) the proof of theorem 1.1 is achieved.

□

1.3.2 Proof of Theorem 1.2

Hereafter we give some further lemmas that are steps towards stating our results.

Lemma 1.3. Under hypotheses (A.1)-(A.2), for $x \in \mathbb{R}$, we have

$$R_{1,T}(x) = O\left(\left(\frac{\log T}{Th_T}\right)^{\frac{1}{2}}\right) \quad \text{a.s.} \tag{1.14}$$

Proof. The proof needs to use Lemma 1.1. Therefore, we have to check its conditions. In this respect, observe by Minkowski and Jensen inequalities that

$$\begin{aligned}
|\mathbb{E}\left[Y_j^p|\mathcal{G}_{j-2}\right]| &\leq \mathbb{E}\left[\left|\int_{T_{j-1}}^{T_j}\left(K\left(\frac{x-X_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt\right|^p|\mathcal{G}_{j-2}\right] \\
&\leq \int_{T_{j-1}}^{T_j}\mathbb{E}\left[\left|K\left(\frac{x-X_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right|^p|\mathcal{G}_{j-2}\right]dt \\
&\leq \int_{T_{j-1}}^{T_j}\left(\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{G}_{j-2}\right]^{1/p} + \mathbb{E}\left[\mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]^p|\mathcal{G}_{j-2}\right]^{1/p}\right)^p dt \\
&\leq \int_{T_{j-1}}^{T_j}\left(\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{G}_{j-2}\right]^{1/p} + \mathbb{E}\left[\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]|\mathcal{G}_{j-2}\right]^{1/p}\right)^p dt \\
&= \int_{T_{j-1}}^{T_j}\left(2\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{G}_{j-2}\right]^{1/p}\right)^p dt = 2^p\int_{T_{j-1}}^{T_j}\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{G}_{j-2}\right]dt.
\end{aligned}$$

Making use of the change of variable $u = (x-y)/h_T$, it follows from conditions (A.2)(ii)-(iii), with $b_{j-2,\delta} = b_{T_{j-2},\delta}$, that

$$\begin{aligned}
&\mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right)^p|\mathcal{G}_{j-2}\right] \\
&= \int K^p\left(\frac{x-y}{h_T}\right)f_t^{\mathcal{G}_{j-2}}(y)dy \\
&= h_T\int K^p(u)f_t^{\mathcal{G}_{j-2}}(x-h_Tu)du \\
&\leq h_T\|K\|_\infty^{p-1}\int K(z)\left(f_t^{\mathcal{G}_{j-2}}(x) - h_Tz\left(f_t^{\mathcal{G}_{j-2}}\right)'(x_T^*(z))\right)du \\
&= h_T\|K\|_\infty^{p-1}(f_t^{\mathcal{G}_{j-2}}(x)+o(1)) \\
&\leq h_T\|K\|_\infty^{p-1}(b_{j-2,\delta}(x)+o(1)).
\end{aligned} \tag{1.15}$$

Therefore, we have

$$|E\left[Y_j^p|\mathcal{G}_{j-2}\right]| \le 2^p\delta h_{T_n}\|K\|_\infty^{p-1}(b_{j-2,\delta}(x)+o(1)) \le p!C^{p-2}d_j^2,$$

where $C = 2\|K\|_\infty$, $d_j^2 = 2^3\delta h_{T_n}\|K\|_\infty b_{j-2,\delta}(x)$. Notice that the integral $\int_0^T b_{t,\delta}(x)dt$ may be approached by the Riemann sum $\delta\sum_{j=2}^n b_{j-2,\delta}(x)$. Since the condition (A.2)(iv) holds true, it follows that

$$\begin{aligned} D_n &= \sum_{i=2}^n d_j^2 = 8h_{T_n}\|K\|_\infty\left(\delta\frac{1}{n}\sum_{i=2}^n b_{j-2,\delta}(x)\right) \\ &\le 8T_nh_{T_n}\|K\|_\infty\left(\frac{1}{T}\int_0^T b_{t,\delta}(x)\right) = 8T_nh_{T_n}\|K\|_\infty O\Big(D(x)\Big). \end{aligned}$$

It is then clear by Lemma 1.1, for any $\varepsilon > 0$, that

$$\begin{aligned} & P\left(|R_{1,T_n}(x)| > \varepsilon\left(\frac{\log T_n}{T_nh_{T_n}}\right)^{\frac{1}{2}}\right) = P\left(\left|\sum_{j=1}^n Y_j\right| > (T_nh_{T_n})\epsilon\left(\frac{\log T_n}{T_nh_{T_n}}\right)^{\frac{1}{2}}\right) \\ \le &\ 2\exp\left\{-\frac{\epsilon^2(T_nh_{T_n})^2\left(\log T_n/T_nh_{T_n}\right)}{8\|K\|_\infty T_nh_{T_n}O(D(x)) + 4\|K\|_\infty T_nh_{T_n}\epsilon\left(\log T_n/T_nh_{T_n}\right)^{\frac{1}{2}}}\right\} \\ \le &\ 2\exp\left\{-\frac{\epsilon^2(T_nh_{T_n})^2\left(\log T_n/T_nh_{T_n}\right)}{8C_1D(x)\|K\|_\infty T_nh_{T_n}\left(1+\frac{\epsilon}{2D(x)}\left(\log T_n/T_nh_{T_n}\right)^{\frac{1}{2}}\right)}\right\} \\ \le &\ 2\exp\left\{-\frac{\epsilon^2}{8C_1D(x)\|K\|_\infty}\log T_n\right\} \\ = &\ 2\exp\left\{\log\left(T_n^{-\frac{\epsilon^2}{8C_1D(x)\|K\|_\infty}}\right)\right\} \\ = &\ 2T_n^{-\frac{\epsilon^2}{8C_1D(x)\|K\|_\infty}}. \end{aligned} \tag{1.16}$$

It suffices then to consider a suitable choice of ε and to use the Borel-Cantelli Lemma to achieve the proof.

Lemma 1.4. Under hypotheses (A.1)(i)-(ii), (A.2)(i) and (A.3), for $x \in \mathbb{R}$, we have

$$R_{2,T}(x) = O_{a.s.}\left(T^{-1/2}\right) + O(h_T). \tag{1.17}$$

Proof. For every $x \in \mathbb{R}$, consider the following decomposition,

$$R_2(x) = \bar{f}_T(x) - \mathbb{E}[f_T(x)] + \mathbb{E}[f_T(x)] - f(x) = B_T^{(1)}(x) + B_T^{(2)}(x).$$

Let $K_h(\cdot) = 1/hK(\cdot/h)$, observe that

$$\begin{aligned}
|TB_T^{(1)}(x)| &= \left| \int_0^T \left(\mathbb{E}\left[K_h\Big(x - X_t\Big) \Big| \mathcal{F}_{t-\delta} \right] - \mathbb{E}\left[K_h\Big(x - X_t\Big) \right] \right) dt \right| \\
&= \left| \int_{\mathbb{R}} K_h\Big(x - y\Big) \left(\int_0^T f^{\mathcal{F}_{t-\delta}}(y) dt - Tf(y) \right) dy \right| \\
&:= \left| \int_{\mathbb{R}} K_h\Big(x - y\Big) H_T(y) dy \right| \\
&\leq \int_{\mathbb{R}} K_h\Big(x - y\Big) \|H_T(y)\| dy.
\end{aligned}$$

Considering the projection $\mathcal{P}_k$ defined above, following Wu (2003) and making use of Cauchy-Schwarz inequality, we obtain

$$\|H_T(y)\|^2 = \left\| \int_0^T f^{\mathcal{F}_{t-\delta}}(y) dt - Tf(y) \right\|^2 = T \left(\int_0^T \left\| \mathcal{P}_1 f^{\mathcal{F}_{t-\delta}}(y) \right\|^2 dt \right),$$

therefore,

$$\begin{aligned}
\sup_y \|H_T(y)\|^2 &\leq T \left(\sup_y \int_0^T \left\| \mathcal{P}_1 f^{\mathcal{F}_{t-\delta}}(y) \right\|^2 dt \right) \\
&\leq T \left(\sup_y \int_{\mathbb{R}^+} \left\| \mathcal{P}_1 f^{\mathcal{F}_{t-\delta}}(y) \right\|^2 dt \right).
\end{aligned}$$

Subsequently, assuming the condition (A.3), we obtain

$$\sup_y \|H_T(y)\|^2 = O(T). \tag{1.18}$$

Thus,

$$B_T^{(1)}(x) = O_{a.s.}(T^{-\frac{1}{2}}).$$

Moreover, assuming the conditions (A.1)(ii) and (A.2)(i), it is easily seen that

$$\begin{aligned}
B_T^{(2)}(x) &= \mathbb{E}[f_T(x)] - f(x) = \frac{1}{h_T} \int_{\mathbb{R}} K\left(\frac{x - y}{h_T} \right) f(y) dy - f(x) \\
&= \int_{\mathbb{R}} K(z) \left(f(x) - h_T z f'(x^*) \right) dz - f(x) \\
&= O(h_T),
\end{aligned}$$

where $x^* \in [x - h_T z, x]$. Consequently, we have

$$R_2(x) = O_{a.s.}\left(T^{-1/2} \right) + O(h_T).$$

Proof of Theorem 1.2 Making use of the decomposition (1.9), the proof follows as a direct consequence of Lemma 1.3 and Lemma 1.4. $\square$

1.3.3 Proof of Theorem 1.3

The idea of the proof is to consider a covering of

$$\mathcal{B}_T = \{x \in \mathbb{R} : |x| \leq T^{\nu}\}.$$

Let ν_T be a non-decreasing integer function tending to infinity as $T \to \infty$. Consider the partition $\{\mathcal{B}_{T,i}\}_{1\leq i\leq \nu_T}$ of the set $\mathcal{B}_T$ defined by

$$\mathcal{B}_{T,i} = \{x : \|x - x_i\| \leq T^{\gamma}\nu_T^{-1}\}$$

where $(x_i)_{1\leq i\leq \nu_T}$ is a sequence of elements of $\mathcal{B}_T$. Recall that

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |f_T(x) - f(x)| &\leq \sup_{x\in\mathbb{R}} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\quad + \sup_{x\in\mathbb{R}} \left| \frac{1}{Th_T} \int_0^T E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right] dt - f(x) \right| \\
&= F_{T,1} + F_{T,2}. \qquad (1.19)
\end{aligned}$$

Lemma 1.5. Under assumptions (A.1)(i) and (A.1)(iii) commbined with condition (1.4) we have

$$\lim_{T\to\infty} F_{T,1} = 0, \quad \text{a.s.} \qquad (1.20)$$

Proof. Recalling that

$$\begin{aligned}
F_{T,1} &= \sup_{x\in\mathbb{R}} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\leq \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\quad + \sup_{x\in\mathcal{B}_T^c} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&= F_{T,1}^{(1)} + F_{T,1}^{(2)}, \qquad (1.21)
\end{aligned}$$

using the covering of the compact set $\mathcal{B}_T$ we may write

$$\begin{aligned}
F_{T,1}^{(1)} &= \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\leq \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - K\left(\frac{x_k - X_t}{h_T}\right)\right) dt \right| \\
&\quad + \max_{1\leq k\leq \nu_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x_k - X_t}{h_T}\right) - E\left[K\left(\frac{x_k - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\quad + \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} \left| \frac{1}{Th_T} \int_0^T \left(E\left[K\left(\frac{x_k - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right] - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&= F_{T,1}^{(11)} + F_{T,1}^{(12)} + F_{T,1}^{(13)}. \qquad (1.22)
\end{aligned}$$

Now by assumption there exists $\ell > 0$ such that

$$|K(u) - K(u'| \leq \ell|u - u'|, \quad (u, u') \in \mathbb{R}^2.$$

Consequently,

$$\begin{aligned}
F_{T,1}^{(11)} &= \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x - X_t}{h_T}\right) - K\left(\frac{x_k - X_t}{h_T}\right)\right) dt \right| \\
&\leq \frac{1}{h_T^2} \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} |x - x_k| \\
&\leq \frac{T^\gamma}{h_T^2 \nu_T},
\end{aligned}$$

and similarly,

$$\begin{aligned}
F_{T,1}^{(13)} &= \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} \left| \frac{1}{Th_T} \int_0^T \left(E\left[K\left(\frac{x_k - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right] - E\left[K\left(\frac{x - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\leq \frac{1}{h_T^2} \max_{1\leq k\leq \nu_T} \sup_{x\in\mathcal{B}_{T,k}} |x - x_k| \\
&\leq \frac{T^\gamma}{\nu_T h_T^2},
\end{aligned}$$

then choosing $\nu_T = \lfloor T^{\gamma+1}/h_T \rfloor$, it follows that

$$F_{T,1}^{(11)} = O\left(\frac{1}{Th_T}\right), \quad \text{a.s.} \tag{1.23}$$

moreover

$$F_{T,1}^{(13)} = O\left(\frac{1}{Th_T}\right), \quad \text{a.s.} \tag{1.24}$$

using the notation (1.6) we notice

$$\begin{aligned}
F_{T,1}^{(12)} &= \max_{1\leq k\leq \nu_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x_k - X_t}{h_T}\right) - E\left[K\left(\frac{x_k - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&= \max_{1\leq k\leq \nu_T} \left| \frac{1}{Th_T} \sum_{i=1}^n Y_i(x_k) \right|.
\end{aligned}$$

Since assumption (A.6) holds true then we may write

$$\begin{aligned}
|Y_i(x_k)| &= \left| \int_{T_{i-1}}^{T_i} \left(K\left(\frac{x_k - X_t}{h_T}\right) - E\left[K\left(\frac{x_k - X_t}{h_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&\leq 2\delta \sup_{y\in\mathbb{R}} K(y) = 2\delta \widetilde{K}.
\end{aligned}$$

Now, for all $\epsilon > 0$, and by a simple application of Lemma 1.2, we note $T = T_n = \delta n$, $h_T = h_{T_n}$ and $\nu_T = \nu_{T_n}$, then using condition (1.4) we obtain

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq \nu_{T_n}}\left|\sum_{i=1}^{n} Y_i(x_k)\right| > \epsilon(T_n h_{T_n})\right) &\leq \sum_{k=1}^{\nu_{T_n}} \mathbb{P}\left(\left|\sum_{i=1}^{n} Y_i(x_k)\right| > \epsilon(T_n h_{T_n})\right) \quad (1.25)\\
&\leq 2\nu_{T_n}\exp\left\{-\frac{\epsilon^2 T_n^2 h_{T_n}^2}{8n\delta^2\widetilde{K}^2}\right\}\\
&= 2\nu_{T_n}\exp\left\{-\frac{\epsilon^2 T_n h_{T_n}^2}{8\delta\widetilde{K}^2\log T_n}\log T_n\right\}\\
&\leq 2\nu_{T_n}\exp\left\{\log T_n^{-\frac{\epsilon^2 L}{8\delta\widetilde{K}^2}}\right\}\\
&= \frac{2}{T_n^{-\gamma-1+\epsilon^2 L/8\delta\widetilde{K}^2} h_{T_n}}, \quad (1.26)
\end{aligned}$$

taking ϵ large enough to have $T_n^{-\gamma-1+\epsilon^2 L/8\delta\widetilde{K}^2} h_T \to \infty$ as $T_n \to \infty$, therefore the right-side of statement (1.25) converge, it suffices then to use the Borel-Cantelli Lemma to deduce that

$$F_{T,1}^{(12)} = o(1), \quad \text{a.s.} \tag{1.27}$$

Combining Equations (1.23), (1.24) and (1.23) we obtain that

$$F_{T,1}^{(1)} = o(1), \quad \text{a.s.} \tag{1.28}$$

We turn our attention now to the second term of decomposition (1.21). It remains to show that

$$F_{T,1}^{(2)} = o(1), \quad \text{a.s.} \tag{1.29}$$

The fact that $\sup_{0\leq t\leq T}|X_t| \leq T^{2\gamma}/2$ combined with the condition $|x| > T^{2\gamma}$ obviously imply, for $0 \leq t \leq T$, that $\left|\frac{x - X_t}{h_T}\right| > \frac{T^{2\gamma}}{2h_T}$. Furthermore, since by condition (A.6) the kernel K is of bounded support, it follows that there exists $T_0 > 0$ such that, for $T > T_0$, we have $K\left(\frac{x-X_t}{h_T}\right) = 0$. Consequently, we can write

$$\left\{\sup_{0\leq t\leq T}|X_t| \leq \frac{T^{2\gamma}}{2}, |x| > T^{2\gamma}\right\} \subset \left\{\sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{j=1}^{n} Y_j\right| = 0\right\}.$$

Therefore, for T large enough, we have

$$P\left(\sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{1}^{n} Y_j\right| > \eta\right) \leq P\left(\sup_{0\leq t\leq T}|X_t| > \frac{T^{2\gamma}}{2}\right) \leq P\left(\sup_{0\leq t\leq T}|X_t| > T^{\gamma}\right).$$

Now, we need to show that the application $T \mapsto \sup\left|\frac{1}{Th_T}\sum_1^n Y_j\right|$ is uniformly continuous. This property holds provided that there exists a constant $\Lambda > 0$ such that

$$\sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{i=1}^{n} Y_i - \frac{1}{Sh_S}\sum_{j=1}^{n'} Y_j'\right| \leq \Lambda|T - S|, \tag{1.30}$$

when $S=\delta n'$, $S_j=j\delta$ and $Y'_j:=Y'_j(x)=\int_{S_{j-1}}^{S_j}\left(K(\frac{x-X_s}{h_S})-E\left[K(\frac{x-X_s}{h_S})|\mathcal{F}_{s-\delta}\right]\right)ds$. Taking $Z'_j=\int_{S_{j-1}}^{S_j}K\left(\frac{x-X_s}{h_S}\right)ds$, observe that

$$\begin{aligned}
\sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{i=1}^{n}Y_i-\frac{1}{Sh_T}\sum_{j=1}^{n'}Y'_j\right| &\leq \sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{i=1}^{n}Z_i-\frac{1}{Sh_S}\sum_{j=1}^{n'}Z'_j\right| \\
&\quad+\sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\right. \\
&\quad\left.-\frac{1}{Sh_S}\int_0^S E\left[K\left(\frac{x-X_s}{h_S}\right)|\mathcal{F}_{s-\delta}\right]ds\right| \\
&= \sup_{|x|>T^{2\gamma}}|f_T(x)-f_S(x)| \\
&\quad+\sup_{|x|>T^{2\gamma}}\left|\bar{f}_T(x)-\bar{f}_S(x)\right|. \qquad (1.31)
\end{aligned}$$

Note that

$$\log\left(f_T(x)\right) = -\log\left(Th_T\right)+\log\left(\int_0^T K\left(\frac{x-X_t}{h_T}\right)dt\right),$$

differentiating with respect to T, we obtain

$$\begin{aligned}
\frac{d\log\left(f_T(x)\right)}{dT} &= -\frac{h_T-Th'_T}{Th_T}-\frac{h'_T}{h_T^2}\left(\int_0^T K\left(\frac{x-X_t}{h_T}\right)dt\right)^{-1}\left(\int_0^T(x-X_t)K'\left(\frac{x-X_t}{h_T}\right)dt\right) \\
&\quad+\left(\int_0^T K\left(\frac{x-X_t}{h_T}\right)dt\right)^{-1}K\left(\frac{x-X_T}{h_T}\right).
\end{aligned}$$

Recalling the property of the function logarithm log :

$$\frac{d\log g(T)}{dT}=\frac{dg(T)/dT}{g(T)},$$

where $g(T)$ is a function. We obtain then

$$\begin{aligned}
\frac{df_T(x)}{dT} &= f_T(x)\frac{d\log f_T(x)}{dT} \\
&= -\frac{1}{T^2h_T}\left(\int_0^T K\left(\frac{x-X_t}{h_T}\right)dt\right)+\frac{h'_T}{Th_T^2}\left(\int_0^T K\left(\frac{x-X_t}{h_T}\right)dt\right) \\
&\quad-\frac{h'_T}{Th_T^3}\left(\int_0^T(x-X_t)K'\left(\frac{x-X_t}{h_T}\right)dt\right)+\frac{1}{Th_T}K\left(\frac{x-X_T}{h_T}\right).
\end{aligned}$$

Assumption (A.6) allows us to suppose the existing of a positive constant $c_K>0$, $K'(u)=0$ whenever $|u|\geq c_K$,

$$\begin{aligned}
\left|\frac{df_T(x)}{dT}\right| &= \left|f_T(x)\frac{d\log f_T(x)}{dT}\right| \\
&\leq \frac{1}{Th_T}\widetilde{K}+\frac{h'_T}{h_T^2}\widetilde{K}+\frac{h'_T}{h_T^2}c_K\widetilde{K'}+\frac{1}{Th_T}\widetilde{K},
\end{aligned}$$

where $\widetilde{K} = \sup_{y\in\mathbb{R}} K(y)$ and $\widetilde{K}' = \sup_{y\in\mathbb{R}} K(y)$. Therefore

$$\sup_{|x|>T^{2\gamma}} |f_T(x) - f_S(x)| \;\leq\; \left(\frac{2}{Th_T}\widetilde{K} + \frac{h_T'}{h_T^2}\left(\widetilde{K} + c_K\widetilde{K}'\right)\right)|T-S|. \qquad (1.32)$$

Considering the second part of decomposition (1.31), at first, observe that

$$\log\left(\bar{f}_T(x)\right) \;=\; -\log\left(Th_T\right) + \log\left(\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right),$$

therefore,

$$\begin{aligned}\frac{d\log\left(\bar{f}_T(x)\right)}{dT} \;=\;& -\frac{1}{T} - \frac{h_T'}{h_T} - \frac{h_T'}{h_T^2}\left(\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right)^{-1}\\ &\times\left(\int_0^T \mathbb{E}\left[(x-X_t)K'\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right)\\ &+\left(\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right)^{-1}\mathbb{E}\left[K\left(\frac{x-X_T}{h_T}\right) \mid \mathcal{F}_{T-\delta}\right].\end{aligned}$$

Proceeding as above, we obtain

$$\begin{aligned}&\frac{d\bar{f}_T(x)}{dT}\\ &= \bar{f}_T(x)\frac{d\log\bar{f}_T(x)}{dT}\\ &= -\frac{1}{T^2h_T}\left(\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right) + \frac{h_T'}{Th_T^2}\left(\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right)\\ &\quad -\frac{h_T'}{Th_T^3}\left(\int_0^T \mathbb{E}\left[(x-X_t)K'\left(\frac{x-X_t}{h_T}\right) \mid \mathcal{F}_{t-\delta}\right]dt\right) + \frac{1}{Th_T}\mathbb{E}\left[K\left(\frac{x-X_T}{h_T}\right) \mid \mathcal{F}_{T-\delta}\right].\end{aligned}$$

Under Assumption (A.6) we have

$$\left|\frac{d\bar{f}_T(x)}{dT}\right| \;\leq\; \frac{2}{Th_T}\widetilde{K} + \left(\widetilde{K} + c_K\widetilde{K}'\right)\frac{h_T'}{h_T^2}.$$

Hence,

$$\sup_{|x|>T^{2\gamma}} \left|\bar{f}_T(x) - \bar{f}_S(x)\right| \;\leq\; \left(\frac{2}{Th_T}\widetilde{K} + \frac{h_T'}{h_T^2}\left(\widetilde{K} + c_K\widetilde{K}'\right)\right)|T-S|. \qquad (1.33)$$

The statement (1.30) follows from the decomposition (1.31) together with condition (1.4)(ii) and equations (1.32) and (1.33). Taking $h_T = T^{-1/5}$ we notice that

$$\max\left\{\left|\frac{d(f_T(x)}{dT}\right|, \left|\frac{d(\bar{f}_T(x)}{dT}\right|\right\} \leq CT^{-\frac{4}{5}},$$

where C is a positive constant. A straightforward application of Lemma 1.2 combined with condition (1.4) give us

$$\begin{aligned} P\left(\left|\sum_{i=1}^{n} Y_i\right| > Th_T\epsilon_0\right) &\leq 2\exp\left\{-\frac{T^2h_T^2\epsilon_0^2}{8n\delta^2\widetilde{K}^2}\right\} \\ &= 2\exp\left\{\log T^{-\frac{\epsilon_0^2}{8\delta\widetilde{K}^2}\frac{Th_T^2}{\log T}}\right\} \leq 2T^{-\frac{\epsilon_0^2 L}{8\delta\widetilde{K}^2}}. \end{aligned}$$

Hence, the application $T \mapsto \sup_{|x|>T^{2\gamma}}\left|\frac{1}{Th_T}\sum_{j=1}^{n} Y_j\right|$ is uniformly continuous for each ω and the result (1.29) follows from making use of Borel-Cantelli Lemma for continuous time processes (see, for instance, Bosq (1998), Lemma 4.2). we achieve the proof of Lemma 1.5 from statements (1.28) and (1.29). □

Lemma 1.6. Assuming that the conditional density $f_t^{\mathcal{F}_{t-\delta}}$ is a Lipschitz function together with assumption (A.1)(ii) and condition (1.3) we have

$$\lim_{T\to\infty} F_{T,2} = 0, \quad \text{a.s.} \tag{1.34}$$

Proof. Observe that

$$\begin{aligned} F_{T,2} &= \sup_{x\in\mathbb{R}}\left|\frac{1}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt - f(x)\right| \\ &= \sup_{x\in\mathbb{R}}\left|\frac{1}{Th_T}\int_0^T\int_{\mathbb{R}} K\left(\frac{x-u}{h_T}\right) f_t^{\mathcal{F}_{t-\delta}}(u)dudt - f(x)\right| \\ &= \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T\int_{\mathbb{R}} K(v)f_t^{\mathcal{F}_{t-\delta}}(x-h_Tv)dvdt - f(x)\right| \\ &\leq \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T\int_{\mathbb{R}} K(v)\left(f_t^{\mathcal{F}_{t-\delta}}(x-h_Tv) - f_t^{\mathcal{F}_{t-\delta}}(x)\right)dvdt\right| \\ &\quad + \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T\left(f_t^{\mathcal{F}_{t-\delta}}(x) - f(x)\right)dt\right| = F_{T,2}^{(1)} + F_{T,2}^{(2)}, \end{aligned}$$

since $f_t^{\mathcal{F}_{t-\delta}}(x)$ is Lipschitzian and under assumption (A1)(ii) we get

$$\begin{aligned} F_{T,2}^{(1)} &\leq \frac{1}{T}\int_0^T\int_{\mathbb{R}} K(v)\sup_{x\in\mathbb{R}}\left|f_t^{\mathcal{F}_{t-\delta}}(x-h_Tv) - f_t^{\mathcal{F}_{t-\delta}}(x)\right|dvdt \\ &\leq h_T\int_{\mathbb{R}} vK(v)dv = O(h_T). \end{aligned} \tag{1.35}$$

On the other side, condition (1.3) gives us

$$F_{T,2}^{(2)} = \sup_{x\in\mathbb{R}}\left|\frac{1}{Tf(x)}\int_0^T f_t^{\mathcal{F}_{t-\delta}}(x)dt - 1\right| = o(1). \tag{1.36}$$

combining statements (1.35) and (1.36) we obtain

$$F_{T,2} \stackrel{\mathbb{P}}{=} o(1), \quad \text{as} \quad T \to \infty.$$

□

Proof of Theorem 1.3 The proof of Theorem 1.3 is achieved by combining decomposition (1.19) with Lemmas 1.5 and 1.6. □

1.3.4 Proof of Theorem 1.4

We need here the decomposition (1.9) and some intermediate results investigating the behaviour of every term of the decomposition.

Observe first that

$$\begin{aligned}\sup_{x\in\mathcal{B}_T} |f_T(x) - f(x)| &\leq \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x-X_t}{h_T}\right) - E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right) dt \right| \\ &+ \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right] dt - f(x) \right| = A_T + B_T.\end{aligned}$$

where

$$\begin{aligned}A_T &= \sup_{x\in\mathcal{B}_T} |A_T(x)| = \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T \left(K\left(\frac{x-X_t}{h_T}\right) - E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right) dt \right|. \\ B_T &= \sup_{x\in\mathcal{B}_T} |B_T(x)| = \sup_{x\in\mathcal{B}_T} \left| \frac{1}{Th_T} \int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right] dt - f(x) \right|\end{aligned}$$

Let ν_T be a non-decreasing integer function tending to infinity as $T \to \infty$. Consider the partition $\{\mathcal{B}_{T,i}\}_{1\leq i\leq \nu_T}$ of the set $\mathcal{B}_T$ defined by

$$\mathcal{B}_{T,i} = \{x : \|x - x_i\| \leq T^\gamma \nu_T^{-1}\}$$

where $(x_i)_{1\leq i\leq\nu_T}$ is a sequence of elements of $\mathcal{B}_T$. Therefore, we have

$$A_T \leq \max_{1\leq i\leq\nu_T} \sup_{x\in\mathcal{B}_{T,i}} |A_T(x) - A_T(x_i)| + \max_{1\leq i\leq\nu_T} |A_T(x_i)| := I_1 + I_2. \tag{1.37}$$

Making use of assumption (A.1)(iii), it follows that there exists a positive constant C_K such that

$$\begin{aligned}\left|K\left(\frac{x-X_t}{h_T}\right) - K\left(\frac{x_i-X_t}{h_T}\right)\right| &\leq C_K \left|\frac{x-X_t}{h_T} - \frac{x_i-X_t}{h_T}\right| \\ &= \frac{C_K}{h_T}|x-x_i| \leq \frac{C_K T^\gamma}{h_T \nu_T}.\end{aligned}$$

Consequently, we have

$$\begin{aligned} I_1 &\leq \frac{1}{Th_T}\sum_1^n\int_{T_{j-1}}^{T_j}\Bigg\{\max_{1\leq i\leq \nu_T}\sup_{x\in\mathcal{B}_T\bigcap\mathcal{B}_{T,i}}\left|K\left(\frac{x-X_t}{h_T}\right)-K\left(\frac{x_i-X_t}{h_T}\right)\right| \\ &\quad + E\left[\max_{1\leq i\leq \nu_T}\sup_{x\in\mathcal{B}_T\bigcap\mathcal{B}_{T,i}}\left|K\left(\frac{x-X_t}{h_T}\right)-K\left(\frac{x_i-X_t}{h_T}\right)\right||\mathcal{G}_{j-1}\right]\Bigg\}\,dt. \\ &\leq \frac{2\delta n}{Th_T}\times\frac{C_K T^\gamma}{h_T\nu_T}=\frac{2C_K T^\gamma}{h_T^2\nu_T} \end{aligned}$$

Taking $\nu_T = T^{\gamma+2}$ and $\epsilon_T = \left(\frac{\log T}{Th_T}\right)^{1/2}$, we obtain

$$\epsilon_T^{-1}I_1 \leq \left(\frac{Th_T}{\log T}\right)^{1/2}\frac{2C_K T^\gamma}{h_T^2 T^{\gamma+2}} = \frac{2C_K}{\left((Th_T)^3\log T\right)^{1/2}}.$$

Therefore,

$$I_1 = o\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) \quad \text{a.s.} \tag{1.38}$$

Using notation (1.6), we observe that

$$\max_{1\leq i\leq \nu_T}|A_T(x_i)| = \max_{1\leq i\leq \nu_T}|\sum_{j=1}^n Y_j(x_i)|.$$

The order of the term I_2 may be stated in much the same way as in the statement (1.14) with the only difference that the assumption (A.4) holds true. Therefore, noting $T_n = T = \delta n$, taking $\nu_{T_n} = \nu_T = T_n^{\gamma+2}$ and $\epsilon_{T_n} = \epsilon_0\left(\frac{\log T_n}{T_n h_{T_n}}\right)^{1/2}$ with a positive constant ϵ_0, we obtain

$$\begin{aligned} P\left(\max_{1\leq k\leq \nu_{T_n}}\frac{1}{T_n h_{T_n}}\left|\sum_{i=1}^n Y_i(x_k)\right| > \epsilon_{T_n}\right) &\leq \sum_{k=1}^{\nu_{T_n}} P\left(\left|\sum_{j=1}^n Y_i(x_k)\right| > (T_n h_{T_n})\epsilon_0\left(\frac{\log T_n}{T_n h_{T_n}}\right)^{1/2}\right) \\ &\leq 2\nu_{T_n}\exp\left\{-\frac{\epsilon_0^2(T_n h_{T_n})^2\left(\log T_n/T_n h_{T_n}\right)}{O(T_n h_{T_n}) + 2CT_n h_{T_n}\epsilon_0\left(\log T_n/T_n h_{T_n}\right)^{\frac{1}{2}}}\right\} \\ &\leq 2\nu_{T_n}\exp\left\{-\epsilon_0^2 O(T_n h_{T_n})\frac{\log T_n}{T_n h_{T_n}}\right\} \\ &\leq 2\nu_{T_n}\exp\left\{-\epsilon_0^2 C\log T_n\right\} \leq 2T_n^{\gamma+2-\epsilon_0^2 C}, \end{aligned}$$

for some positive constant C. Proceeding as in the statement (1.16) and considering a choice of ϵ_0 that allows to use the Borel-Cantelli Lemma, we obtain

$$I_2 = O\left(\left(\frac{\log T_n}{T_n h_{T_n}}\right)^{1/2}\right) = O\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) \quad \text{a.s.} \tag{1.39}$$

It follows then from the statements (1.37)-(1.39) that

$$A_T = O\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) \quad \text{a.s.} \tag{1.40}$$

Moreover, taking into account the conditions (A.2) and (A.3) together with the statement (1.18), we have

$$\begin{aligned}
B_T &= \sup_{x\in\mathcal{B}_T}\left|\frac{1}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)\middle|\mathcal{F}_{t-\delta}\right]dt - f(x)\right| \\
&\le \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T E\left[K_h\left(x-X_t\right)\middle|\mathcal{F}_{t-\delta}\right]dt - \frac{1}{T}\int_0^T E\left[K_h\left(x-X_t\right)\right]dt\right| \\
&\quad + \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_0^T E\left[K_h\left(x-X_t\right)\right]dt - f(x)\right| \\
&= \sup_{x\in\mathbb{R}}\left|\frac{1}{T}\int_{\mathbb{R}} K_h\left(x-y\right)H_T(y)dy\right| + \sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} K(z)f(x-h_Tz)dz - f(x)\right| \\
&\le \sup_{x\in\mathbb{R}}\frac{1}{T}\int_{\mathbb{R}} K_h\left(x-y\right)\|H_T(y)\|dy + \sup_{x\in\mathbb{R}}\left|\int_{\mathbb{R}} K(z)\left(f(x)-h_Tzf'(x^*)\right)dz - f(x)\right| \\
&= O\Big(\sqrt{T}\Big)\sup_{x\in\mathbb{R}}\frac{1}{Th_T}\int_{\mathbb{R}} K\left(\frac{x-y}{h_T}\right)dy + \sup_{x\in\mathbb{R}}\left|h_Tf'(x^*)\int_{\mathbb{R}} K(z)dz\right|.
\end{aligned}$$

Thus,

$$B_T = O_{a.s.}\Big(T^{-1/2}\Big) + O\Big(h_T\Big). \tag{1.41}$$

Combining the statement (1.40) with (1.41), we obtain the following result

$$\sup_{x\in\mathcal{B}_T}|\hat{f}_T(x) - f(x)| = O_{a.s.}\left(\left(\frac{\log T}{Th_T}\right)^{1/2}\right) + O_{a.s.}\Big(T^{-1/2}\Big) + O(h_T). \tag{1.42}$$

Our task now is to handle the term $\sup_{x\in\mathcal{B}_T^c}|\hat{f}_T(x) - f(x)|$, where $\mathcal{B}_T^c$ is the complementary set of $\mathcal{B}_T$ with respect to the real line. Proceeding as for the statement (1.41), it is clear that

$$\sup_{x\in\mathcal{B}_T^c}|\bar{f}_T(x) - f(x)| = O_{a.s.}\Big(T^{-1/2}\Big) + O\Big(h_T\Big) \quad a.s. \tag{1.43}$$

It remains then to prove hereafter that

$$\lim_{T\to\infty}\sqrt{\frac{Th_T}{\log T}}\sup_{x\in\mathcal{B}_T^c}|\hat{f}_T(x) - \bar{f}(x)| = 0, a.s. \tag{1.44}$$

The fact that $\sup_{0\le t\le T}|X_t| \le T^{2\gamma}/2$ combined with the condition $|x| > T^{2\gamma}$ obviously imply, for $0\le t\le T$, that $\left|\dfrac{x-X_t}{h_T}\right| > \dfrac{T^{2\gamma}}{2h_T}$. Furthermore, since by condition

(A.6) the kernel K is of bounded support, it follows that there exists $T_0 > 0$ such that, for $T > T_0$, we have $K\left(\frac{x-X_t}{h_T}\right) = 0$. Consequently, we can write

$$\left\{\sup_{0\le t\le T} |X_t| \le \frac{T^{2\gamma}}{2}, |x| > T^{2\gamma}\right\} \subset \left\{\sup_{|x|>T^{2\gamma}} \left|\frac{1}{Th_T}\sum_{j=1}^{n} Y_j\right| = 0\right\}.$$

Therefore, for T large enough, we have

$$P\left(\sup_{|x|>T^{2\gamma}} \left|\frac{1}{Th_T}\sum_{1}^{n} Y_j\right| > \eta\right) \le P\left(\sup_{0\le t\le T} |X_t| > \frac{T^{2\gamma}}{2}\right) \le P\left(\sup_{0\le t\le T} |X_t| > T^{\gamma}\right).$$

Now, we need to show that the application $T \mapsto \sup\left|\frac{1}{Th_T}\sum_1^n Y_j\right|$ is uniformly continuous. This property holds provided that there exists a constant $\Lambda > 0$ such that

$$\sup_{|x|>T^{2\gamma}} \left|\frac{\epsilon_T}{Th_T}\sum_{i=1}^{n} Y_i - \frac{\epsilon_S}{Sh_S}\sum_{j=1}^{n'} Y'_j\right| \le \Lambda|T-S|, \tag{1.45}$$

when $S = \delta n'$, $S_j = j\delta$, $Y'_j := Y'_j(x) = \int_{S_{j-1}}^{S_j}\left(K(\frac{x-X_s}{h_S}) - E\left[K(\frac{x-X_s}{h_S})|\mathcal{F}_{s-\delta}\right]\right) ds$ and $\epsilon_T = \left(\frac{Th_T}{\log T}\right)^{1/2}$. Taking $Z'_j = \int_{S_{j-1}}^{S_j} K\left(\frac{x-X_s}{h_S}\right) ds$, observe that

$$\begin{aligned}
\sup_{|x|>T^{2\gamma}} \left|\frac{\epsilon_T}{Th_T}\sum_{i=1}^{n} Y_i - \frac{\epsilon_S}{Sh_T}\sum_{j=1}^{n'} Y'_j\right| &\le \sup_{|x|>T^{2\gamma}} \left|\frac{\epsilon_T}{Th_T}\sum_{i=1}^{n} Z_i - \frac{\epsilon_S}{Sh_S}\sum_{j=1}^{n'} Z'_j\right| \\
&\quad + \sup_{|x|>T^{2\gamma}} \left|\frac{\epsilon_T}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right] dt\right. \\
&\quad \left. - \frac{\epsilon_S}{Sh_S}\int_0^S E\left[K\left(\frac{x-X_s}{h_S}\right)|\mathcal{F}_{s-\delta}\right] ds\right| \\
&= \sup_{|x|>T^{2\gamma}} \left|\epsilon_T \hat{f}_T(x) - \epsilon_S \hat{f}_S(x)\right| \\
&\quad + \sup_{|x|>T^{2\gamma}} \left|\epsilon_T \bar{f}_T(x) - \epsilon_S \bar{f}_S(x)\right|.
\end{aligned} \tag{1.46}$$

Note also that

$$\begin{aligned}
\log\left(\epsilon_T \hat{f}_T(x)\right) &= \log(\epsilon_T) - \log(Th_T) + \log I_T \\
&= \frac{1}{2}\log(Th_T) + \frac{1}{2}\log(\log T) - \log(Th_T) + \log I_T \\
&= \frac{1}{2}\log(\log T) - \frac{1}{2}\log(Th_T) + \log I_T,
\end{aligned}$$

where $I_T = \int_0^T K\left(\frac{x-X_t}{h_T}\right) dt$. Differenciating with respect to T, we obtain

$$\begin{aligned}
\frac{d\log \epsilon_T \hat{f}_T(x)}{dT} &= \frac{1}{2T\log T} - \frac{h_T - Th'_T}{2Th_T} - \frac{h'_T}{h_T^2} I_T^{-1}\int_0^T (x-X_t)K'\left(\frac{x-X_t}{h_T}\right) dt \\
&+ I_T^{-1} K\left(\frac{x-X_T}{h_T}\right).
\end{aligned}$$

Since by condition (A.6), for some $c_K > 0$, $K'(u) = 0$ whenever $|u| \geq c_K$, we have

$$\begin{aligned}\left|\frac{d\log \epsilon_T \hat{f}_T(x)}{dT}\right| &\leq \frac{1}{2T\log T} + \frac{1}{2T} + \frac{|h_T'|}{2h_T} + \left|I_T^{-1}\right| K\left(\frac{x-X_T}{h_T}\right) \\ &\quad +2c_K\frac{|h_T'|}{h_T^2}\left|I_T^{-1}\right| Th_T\|K'\|_\infty.\end{aligned}$$

As $\dfrac{d\epsilon_T f_T(x)}{dT} = \epsilon_T f_T(x)\dfrac{d\log\epsilon_T f_T(x)}{dT}$, it is easily seen that

$$\left|\frac{d\epsilon_T f_T(x)}{dT}\right| \leq \frac{\|K\|_\infty}{(Th_T\log T)^{\frac{1}{2}}}\left(\frac{1}{2\log T}+\frac{3}{2}\right) + \frac{|h_T'|(Th_T)^{\frac{1}{2}}}{h_T^2(\log T)^{\frac{1}{2}}}\left(\frac{\|K\|_\infty}{2} + 2c_K\|K'\|_\infty\right).$$

Therefore, we have

$$\begin{aligned}&\sup_{|x|>T^{2\gamma}}\left|\epsilon_T\hat{f}_T(x) - \epsilon_S\hat{f}_S(x)\right| \leq \qquad (1.47)\\ &\left(\frac{\|K\|_\infty}{(Th_T\log T)^{\frac{1}{2}}}\left(\frac{1}{2\log T}+\frac{3}{2}\right) + \frac{|h_T'|(Th_T)^{\frac{1}{2}}}{h_T^2(\log T)^{\frac{1}{2}}}\left(\frac{\|K\|_\infty}{2} + 2c_K\|K'\|_\infty\right)\right)|T-S|.\end{aligned}$$

Considering now the second part of (1.46), observe first that

$$\begin{aligned}\log\left(\epsilon_T\bar{f}_T(x)\right) &= \log\left(\frac{\epsilon_T}{Th_T}\int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\right)\\ &= -\frac{1}{2}\log(\log T) - \frac{1}{2}\log(Th_T) + \log L_T,\end{aligned}$$

where $L_T = \int_0^T E\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt$. Therefore, we have

$$\begin{aligned}\frac{d}{dT}\log\left(\epsilon_T\bar{f}_T(x)\right) &= \frac{d}{dT}\left(-\frac{1}{2}\log(\log T) - \frac{1}{2}\log(Th_T) + \log L_T\right)\\ &= -\frac{1}{2T\log T} - \frac{1}{2T} - \frac{h_T'}{2h_T}\\ &\quad -\frac{h_T'}{h_T^2}L_T^{-1}\int_0^T E\left[(x-X_t)K'\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\\ &\quad +L_T^{-1}E\left[K\left(\frac{x-X_T}{h_T}\right)|\mathcal{F}_{T-\delta}\right].\end{aligned}$$

Thus, similarly as above, we obtain

$$\begin{aligned}\frac{d}{dT}\left(\epsilon_T\bar{f}_T(x)\right) &= \left(\epsilon_T\bar{f}_T(x)\right)\frac{d}{dT}\log\left(\epsilon_T\bar{f}_T(x)\right)\\ &= \epsilon_T\bar{f}_T(x)\left(-\frac{1}{2T\log T} - \frac{1}{2T} - \frac{h_T'}{2h_T}\right)\\ &\quad -\frac{h_T'}{h_T^2}\frac{\epsilon_T}{Th_T}\int_0^T E\left[(x-X_t)K'\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\\ &\quad +\frac{\epsilon_T}{Th_T}E\left[K\left(\frac{x-X_T}{h_T}\right)|\mathcal{F}_{T-\delta}\right],\end{aligned}$$

and

$$\left|\frac{d\left(\epsilon_T \bar{f}_T(x)\right)}{dT}\right| \leq \frac{\|K\|_\infty}{(Th_T\log T)^{1/2}}\left(\frac{1}{2\log T}+\frac{3}{2}\right) + \frac{|h'_T|}{h_T^2}\frac{(Th_T)^{1/2}}{(\log T)^{1/2}}\left(\frac{1}{2}\|K\|_\infty + 2c_K\|K'\|_\infty\right).$$

Consequently, we have

$$\sup_{|x|>T^{2\gamma}}\left|\epsilon_T\bar{f}_T(x)-\epsilon_S\bar{f}_S(x)\right| \tag{1.48}$$
$$\leq \left(\frac{\|K\|_\infty}{(Th_T\log T)^{1/2}}\left(\frac{1}{2\log T}+\frac{3}{2}\right)+\frac{|h'_T|}{2h_T^2}\frac{(Th_T)^{1/2}}{(\log T)^{1/2}}\left(\frac{1}{2}\|K\|_\infty+\|K'\|_\infty\right)\right)|T-S|,$$

The statement (1.45) follows immediately from the condition (A.5) and the statements (1.46), (1.47) and (1.48), Notice, for the particular choice of the smoothing parameter $h_T = T^{-1/5}$, that we have

$$\max\left\{\left|\frac{d(\epsilon_T f_T(x))}{dT}\right|,\left|\frac{d(\epsilon_T\bar{f}_T(x))}{dT}\right|\right\}\leq CT^{-\frac{2}{5}}\Big(\log T\Big)^{-1/2},$$

where C is a positive constant.

A further use of Lemma 1.1 enables us to write

$$P\left(\left|\sum_{i=1}^n Y_i\right| > Th_T\epsilon_T\epsilon_0\right)\leq 2\exp\{-\epsilon_0^2\log T\} = 2T^{-\epsilon_0^2}.$$

Hence, the application

$$T \mapsto \sup_{|x|>T^{2\gamma}}\left|\frac{\epsilon_T}{Th_T}\sum_{j=1}^n Y_j\right|$$

is uniformly continuous for each ω and the result (1.44) follows by making use of Borel-Cantelli Lemma for continuous time processes (see, for instance, Bosq (1998), Lemma 4.2). □

Proof of Theorem 1.4 We achieve the proof of Theorem 1.4 by combining the statements (1.42), (1.43) and (1.44). □

Chapitre 2

Asymptotic results for the regression function estimate on continuous time stationary and ergodic data

Ce chapitre à fait l'objet d'une publication au *Journal of Statistics and Risk Modelling*

Sultana DIDI and Djamal LOUANI
L.S.T.A., Université de Paris 6.
4, Place Jussieu, 75252 PARIS, FRANCE.

Abstract This paper is devoted to the study of asymptotic properties of the regression function kernel estimate in the setting of continuous time stationary and ergodic data. More precisely, considering the Nadaraya-Watson type estimator, say $\hat{m}_T(x)$, of the l-indexed regression function $m(x) = \mathbb{E}(l(Y)|X = x)$ built upon continuous time stationary and ergodic data $(X_t, Y_t)_{0\leq t\leq T}$, we establish its pointwise and uniform, over a dilative compact set, convergences with rates. Notice that the ergodic setting covers and completes various situations as compared to the mixing case and stands as more convenient to use in practice.

AMS 1991 subject classification : Primary : 60F10, 62G07, 62F05, 62H15.
Key words and phrases : Consistency, continuous time processes, ergodic data, kernel estimator, rate of convergence, regression function.

2.1 Introduction

Let $(\mathbf{X}_t, Y_t)_{t\geq 0}$ be a stationary ergodic process taking values in $\mathbb{R}^d \times \mathbb{R}$ and distributed as the random vector $(\mathbf{X}, Y)$. Let l be a real measurable function defined upon the real space $\mathbb{R}$. The regression function of the random variable $l(Y)$ given the event $X = x$ is defined, for any $\mathbf{x} \in \mathbb{R}^d$, by $m(\mathbf{x}) = \mathbb{E}[l(Y)|\mathbf{X} = \mathbf{x}]$. On the basis of the part $(\mathbf{X}_t, Y_t)_{0\leq t\leq T}$ observed from the above process, the estimate of the function m is defined, for any $\mathbf{x} \in \mathbb{R}^d$, by

$$\hat{m}_T(\mathbf{x}) = \frac{\int_0^T l(Y_t) K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) dt}{\int_0^T K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) dt},$$

where h_T is the smoothing parameter satisfying the following conditions

$$\lim_{T\to\infty} h_T = 0, \quad \lim_{T\to\infty} Th_T^d = +\infty, \quad \lim_{T\to\infty} \frac{\log T}{Th_T^d} = 0 \tag{2.1}$$

and K is a nonnegative kernel defined on $\mathbb{R}^d$.

The topic of nonparametric function estimation on continuous time processes has motivated a number of works studying the consistency with rate of convergence for both the density and the regression function estimates while considering various criteria of convergence together with the asymptotic normality. The kernel estimate of the d-variate density function is studied by Delecroix (1980) and Bosq (1996) where the rates of convergence for the MSE and the supremum norm criteria are found to be $O(T^{-4/(4+d)})$ and $o(\log_k(T)(\log(T)/T)^{-4/(4+d)})$ respectively. Here, $\log_k(T) = \max\{1, \log_{k-1}(T)\}$ for $T > 0$ and $k \geq 2$. We cite also the paper of Castellana and Leadbetter (1986) where the consistency together with the asymptotic normality are given in the real valued process case. When the process $(X_t)_t$ satisfies a local irregularity condition, these results may be sharpened as to reach the parametric rates given respectively by $O(1/T)$ and $o(\log_k(T)(\log(T)/T)^{1/2})$ and stated in Bosq (1997). Asymptotic properties of the kernel regression function estimator for continuous time processes have been investigated in several papers. There also parametric rates of convergence are reached for various criteria. We refer to Bosq (1996, 1997) and the references therein for an account of results on the subject. Notice that all these results are stated while considering the mixing setting for the data. It is well known that the mixing conditions may be viewed as an asymptotic independence. Therefore, many processes fail to satisfy these conditions. Examples of such processes are given in the papers of Chernick (1981) and Andrews (1984) in the discrete time case.

To be more convenient towards a number of applications in practice, we consider in this paper, the regression function estimation when the data are assumed to be drawn from a stationary and ergodic continuous time process to allow the maximum possible generality in regard to the dependence setting. Notice that continuous time recordings of phenomena are available in various areas of the economical life as the weather, the energy and the health care for example. In the sequel, We establish the pointwise and uniform consistencies with rates of the regression function estimate. Here, only nonparametric rates of convergence are reached since no irregularity conditions upon the process $(\mathbf{X}_t, Y_t)_{t\geq 0}$ are assumed.

The study of the nonparametric regression estimate in the ergodic discrete time setting has motivated a number of papers in the literature. We refer among others to the works of Delecroix and Rosa (1996), Laib (1999,2005), Morvai *et al.* (1996) and Yakowitz *et al.* (1999) where some asymptotic properties are studied. Notice also the consistency with rates together with the asymptotic normality stated by Laïb and Louani (2010,2011) for the regression estimate on discrete time functional data.

In order to display our results, some notations are needed. From now on, for a positive real number δ such that $n = \frac{T}{\delta} \in \mathbb{N}$ and $j \in \mathbb{N} \cap [1, n]$, consider the partition $(T_j)_{1\leq j\leq n}$ of the interval $[0, T]$. Moreover, for $t > 0$, $\delta > 0$ and $1 \leq j \leq n$, consider the σ-fields $\mathcal{F}_{t-\delta} = \sigma((\mathbf{X}_s, Y_s) : 0 \leq s < t - \delta)$, $\mathcal{F}_j = \sigma((\mathbf{X}_s, Y_s), 0 \leq s < T_j)$ and $\mathcal{S}_{t,\delta} = \sigma((\mathbf{X}_s, Y_s); (\mathbf{X}_r) : 0 \leq s < t, \quad t \leq r \leq t + \delta)$. Whenever $s < 0$, $\mathcal{F}_s$ stands as the trivial σ-field. In the sequel, for any function q, q' stands as the generic notation of its derivative. For $i = 1, 2$, set

$$\hat{m}_{T,i}(\mathbf{x}) = \frac{1}{n\mathbb{E}[Z_1(x)]} \int_0^T (l(Y_t))^{i-1} K\left(\frac{x - X_t}{h_T}\right) dt$$

and

$$\tilde{m}_{T,i}(x) = \frac{1}{n\mathbb{E}[Z_1(x)]} \int_0^T \mathbb{E}\left[(l(Y_t))^{i-1} K\left(\frac{x - X_t}{h_T}\right) dt|\mathcal{F}_{t-\delta}\right] dt,$$

where $Z_1(x) = \displaystyle\int_0^\delta K\left(\frac{x - X_t}{h_T}\right) dt.$

For the sake of clarity, introduce some details defining the ergodic property of continuous time processes. Let $X = \{X_t\}_{t\geq 0}$ be a continuous time process taking values in some measurable space $(E, \mathcal{X})$ on which is defined a probability measure μ. For $\delta \geq 0$, let Υ^δ be a $\delta-$shift transformation, i.e. $(\Upsilon^\delta(x))_s = x_{s+\delta}$. A measurable set $\mathcal{A}$ is $\delta-$invariant if it does not change under $\delta-$shift transformation , i.e. $\Upsilon^\delta(\mathcal{A}) = \mathcal{A}$.

Définition 2.1. ($\delta-$ergodicity) A continuous time process $X = \{X_t\}_{t\geq 0}$ is $\delta-$ergodic if every measurable $\delta-$invariant set related to the process X has probability either 1 or 0, in other words, for any $\delta-$invariant set $\mathcal{A}$, $\mu(\mathcal{A}) = (\mu(\mathcal{A}))^2$.

This definition means that if we take the process X and slice it into time blocks of length δ then the new discrete time process $(X_0^\delta, X_\delta^{2\delta}, X_{2\delta}^{3\delta}, X_{3\delta}^{4\delta}, \ldots)$ is ergodic. For discrete time processes, we refer, for instance, to Krengel (1985)[77] or Laïb & Louani (2010) for the definition and details on the ergodic property.

Définition 2.2. (Ergodicity) A continuous time process $X = \{X_t\}_{t\geq 0}$ is ergodic if it is δ-ergodic for every $\delta > 0$.

It is well known from the ergodic theorem that, for a measurable function g and a stationary ergodic process $X = \{X_t\}_{t\geq 0}$, we have

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T g(X_t)dt = \mathbb{E}(g(X_0)). \tag{2.2}$$

Therefore, the ergodic property in our setting is formulated on the basis of the statement (2.2) and the requirements are considered in conditions (H.2) and (H.4)(vi) below. We refer to the book of Krengel (1985) for an account of details and results on the ergodic theory.

2.2 Results

Set now some hypotheses necessary to state our results.

(H.0) (i) K is a spherically symmetric bounded kernel on its compact support, say $[-\lambda, \lambda]^d$, and $\int K(z)dz = 1$; there exists a nonnegative differentiable integrable function k over its support $[0, \lambda c_1]$ such that $K(z) = k(\|z\|)$ and $\int_0^{\lambda c_1} k'(u)du < \infty$, where $\|.\|$ is a norm on $\mathbb{R}^d$ and c_1 is a positive constant.

(ii) The kernel K is a Lipschitz function.

(H.1) (i) For any $0 \leq s < t$, there exists a positive constant c_0 and a nonnegative continuous and bounded random function $f_{t,s}$ defined on $\mathbb{R}^d$ such that, almost surely,

$$\mathbb{P}^{\mathcal{F}_s}\Big(X_t \in S_{r,x}\Big) := \mathbb{P}\Big(\|X_t - x\| \leq r\Big|\mathcal{F}_s\Big) = c_0 r^d f_{t,s}(x) + o(r^d), as\ r \to 0,$$

(ii) For any $0 \leq s < t \leq T$ such that $t - s \leq \alpha$ and any $x \in \mathbb{R}^d$, the function $f_{t,s}(x)$ is almost surely bounded by a deterministic function $b_{s,\alpha}(x)$,

(iii) $T^{-1}\int_0^T b_{t,\alpha}(x)dt \to D_\alpha(x) \neq 0$ as $T \to \infty$.

(H.2) For any $x \in \mathbb{R}^d$ and any $\delta > 0$, there exists a nonnegative deterministic function f such that

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f_{t,t-\delta}(x)dt = f(x), \ a.s.$$

(H.3) For any t and r such that $t \in [0,T]$ and $t \leq r \leq t+\delta$, we have

(i) $\mathbb{E}[l(Y_r)|\mathcal{S}_{t,\delta}] = \mathbb{E}[l(Y_r)|X_r] = m(X_r)$ almost surely.

(ii) There exist constants $\beta > 0$ and $c_3 > 0$ such that, for any $(u,v) \in \mathbb{R}^{2d}$, $|m(u) - m(v)| \leq c_3\|u-v\|^\beta$.

(iii) For any $k \geq 2$ and any $\delta > 0$, $\mathbb{E}[|l(Y_r)|^k|\mathcal{S}_{t,\delta}] = \mathbb{E}[|l(Y_r)|^k|X_r]$ and the function $h_k(x) = \mathbb{E}[|l(Y)|^k|X = x]$ is continuous in the neighborhood of x.

Remark 2.1. For $s \leq 0$, $\mathcal{F}_s$ stands as the trivial σ-field. Therefore, the condition (H.1)(i) takes the form

$$\mathbb{P}\Big(\|X_t - x\| \leq r \Big| \mathcal{F}_s\Big) = \mathbb{P}\Big(\|X_t - x\| \leq r\Big) = c_0 r^d f(x) + o(r^d), as\ r \to 0, \tag{2.3}$$

where f is a positive deterministic function. Notice that the function f matches the density of X_t whenever we assume that this density exists. In such a case, the constant c_0 is given by $c_0 = \pi^{\frac{d}{2}}/\Gamma\left((d+2)/2\right)$, where Γ is the usual gamma function. The statement (2.3) holds also whenever $s = t = 0$.

Comments on hypotheses. Conditions (H.0) are very common in nonparametric function estimation literature. Hypothesis (H.1)(i) assume that the conditional probability of the sphere $S_{r,x}$ given the $\sigma-$field $\mathcal{F}_s$ is asymptotically governed by a local dimension when the radius r tends to zero without assuming the existence of marginal and conditional densities to reach the rates of convergence given below. Note that the condition (H.1)(i) holds true while the conditional distribution $\mathbb{P}_{X_t}^{\mathcal{F}_s}$ has a continuous density function $f_{X_t}^{\mathcal{F}_s}$ on $\mathbb{R}^d$. Condition (H.2) involves the ergodic nature of the data as given, for instance, in Györfi *et al.* (1989). Notice, by (H.1)(ii), that the random functions $f_{t,t-\delta}(x)$ belong to the space $\mathcal{C}^0$ of continuous functions, which is a separable Banach space. Moreover, approaching the integral $\displaystyle\int_0^T f_{t,t-\delta}(x)dt$ by its Riemann's sum, it follows that

$$T^{-1}\int_0^T f_{t,t-\delta}(x)dt \backsimeq n^{-1}\sum_{i=1}^n f_{T_i,T_i-\delta}(x) = n^{-1}\sum_{j=1}^n f_{j\delta,(j-1)\delta}(x).$$

Since the process $(X_{T_j})_{j\geq 1}$ is stationary and ergodic, following Delecroix (1987) (see, Lemma 4 and Corollary 1 together with their proofs), one may prove that the sequence $(f_{j\delta,(j-1)\delta}(x))_{j\geq 1}$ of random functions is stationary and ergodic. Indeed, It suffices to replace in the work of Delecroix the conditional densities by $f_{j\delta,(j-1)\delta}$'s and the density by the function f. Moreover, making use of Beck's Theorem (see, for instance, Györfi *et al.* (1989), Theorem 2.1.1), it follows that

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T f_{t,t-\delta}(x)dt-\mathbb{E}(f_{\delta,0}(x))\right|=\lim_{T\to\infty}\sup_{x\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T f_{t,t-\delta}(x)dt-f(x)\right|=0, a.s.$$

It is then clear that both the conditions (H.2)(i) and (H.4)(vi), given below, are satisfied. Notice that the transform of the stationary ergodic process $(X_t, Y_t)_{t\geq 0}$ into the process $(l^2(Y_t))_{t\geq 0}$ is a measurable function. Therefore, making use of Proposition 4.3 of Krengel (1985) and then the ergodic Theorem, we obtain

$$\lim_{T\to\infty}\frac{1}{T}\int_0^T (l(Y_t))^2 dt = \mathbb{E}((l^2(Y_0)))\ a.s., \tag{2.4}$$

a result needed as a condition in Lemma 2.5 below. Condition (H.3)(iii) is usual in the literature dealing with the study of ergodic processes. The hypothesis (H.3)(ii) is a regularity condition upon the regression function m.

We give now some examples for which the condition (H.3)(i) is satisfied.

Example 1. Consider the regression model $Y_t = m(X_t) + \epsilon_t$, where the random variables ϵ_t's stand as martingale differences with respect to the σ-field $\mathcal{S}_{r,\delta}, r\leq t\leq r+\delta$, generated by $\Big\{(X_s,\epsilon_s),(X_t) : 0\leq s<r, r\leq t\leq r+\delta\Big\}$. Clearly, we have $\mathbb{E}[Y_t|\mathcal{S}_{r,\delta}] = m(X_t)$ almost surely.

Example 2. Consider the regression model $Y_t = m(X_t)+\sigma(X_t)\epsilon_t$, where the random variables ϵ_t are centered and independent of the process $(X_t)_{t\geq 0}$. Taking $\mathcal{S}_{r,\delta}$ as the σ-field generated by $\Big\{(X_s) : 0\leq s\leq r\Big\}$, it follows, for $t\leq r$, that $\mathbb{E}[Y_t|\mathcal{S}_{r,\delta}] = \mathbb{E}[m(X_t)+\sigma(X_t)\epsilon_t|\mathcal{S}_{r,\delta}] = m(X_t)+\sigma(X_t)\mathbb{E}[\epsilon_t] = m(X_t)$ almost surely.

The first result hereafter gives the pointwise convergence with rate result for the regression estimate $\hat{m}_T(x)$.

Theorem 2.1. Under hypotheses $(H.0)$-$(H.3)$, for any $x\in\mathbb{R}^d$ such that $f(x)\neq 0$, we have

$$\hat{m}_T(x) - m(x) = O_{a.s}(h_T^{\beta}) + O_{a.s}\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right), \quad as\ \ T\to\infty.$$

Towards establishing the uniform result, for $\frac{1}{2} < \gamma < 1$, consider the following dilative compact set $\mathcal{B}_T = \{x \in \mathbb{R}^d : \|x\| \leq T^\gamma\}$, where $\|.\|$ is a norm on $\mathbb{R}^d$. Notice that the following additional assumptions are needed.

(H.4)(i) There exists a sequence (α_T) of positive real numbers such that $\inf_{x\in\mathcal{B}_T} f(x) \geq \alpha_T$.

(ii) For any $T > 0$ and any $\alpha > 0$, $\sup_{x\in\mathcal{B}_T} |D_\alpha(x)| \leq D < \infty$,

(iii) For any $k \geq 2$, the function h_k is uniformly continuous and bounded over the set $\mathcal{B}_T$.

(iv) The function f is bounded by a constant $C > 0$.

(v) There exist two constants $\tau > 0$ and $u > 0$ such that, for T large enough, $\frac{1}{\alpha_T^2 h_T^{2+d} T^\tau} \leq \frac{1}{T^{1+u}}$.

(vi) For any $\delta > 0$ small enough, we have

$$\lim_{T\to\infty} \sup_{x\in\mathcal{B}_T} \left| \frac{1}{f(x)T} \int_0^T f_{t,t-\delta}(x)dt - 1 \right| = 0, \quad a.s.$$

Comments on additional hypotheses. Condition $(H.4)(i)$ is satisfied when we consider, for example, a zero-mean Gaussian real random variable with a density f and a finite variance σ^2 and we take $|x| \leq O\Big(\sigma(u \log T)^{1/2}\Big)$ and $\alpha_T = O\Big(T^{-v}/\sigma^2 2\pi\Big)$ with $v \geq u$. Conditions (H.4)(ii) and (H.4)(iii) are uniform versions of conditions (H.1)(iv) and (H.3)(iii) respectively. Hypothesis (H.4)(v) set the needed condition to use the Borel-Cantelli Lemma..

The following theorem gives the uniform version of our results.

Theorem 2.2. Assume that assumptions $(H.0)$-$(H.4)$ hold true. Furthermore, suppose that

$$\frac{T^{2\gamma-1} h_T^d \alpha_T^2}{\log T} = O(1), \quad \text{as} \quad T \to \infty. \tag{2.5}$$

Then, we have

$$\sup_{x\in\mathcal{B}_T} |\hat{m}_T(x) - m(x)| = O_{a.s}(h_T^\beta) + O_{a.s}\left(\left(\frac{\log T}{T h_T^d \alpha_T^2}\right)^{1/2}\right) \quad \text{as} \quad T \to \infty. \tag{2.6}$$

Remark 2.2. Whenever, instead of the set $\mathcal{B}_T$, we consider a compact $\mathcal{C}$ on which the function f is bounded from below by a constant $\alpha > 0$, we obtain the following result

$$\sup_{x\in\mathcal{C}} |\hat{m}_T(x) - m(x)| = O_{a.s}(h_T^\beta) + O_{a.s}\left(\left(\frac{\log T}{T h_T^d}\right)^{1/2}\right), \quad as \quad T \to \infty$$

when the hypothesis (H.4)(v) and the condition (2.5) are replaced by the fact that there exist two constants $\tau > 0$ and $u > 0$ such that $\dfrac{1}{T^{\tau}\left(h_T^5 \log T\right)^{\frac{1}{2}}} \leq \dfrac{1}{T^{1+u}}$.

Nonparametric prediction. Let $(\xi_t, t \in \mathbb{R})$ be a measurable stationary and ergodic process. Given the data $(\xi_t, 0 \leq t \leq T)$, we would like to predict the non-observed square integrable real random variable $\zeta_{T+H} := l(\xi_{T+H})$, where the horizon H is such that $0 < H < T$ and where l is a measurable and bounded function on compact sets. For this purpose, let us construct the predictor. Set

$$Z_t = (X_t, l(Y_t)) = (\xi_t, l(\xi_{t+H})), \quad t \in \mathbb{R},$$

and consider the regression kernel estimator $\hat{m}_T$ based on the data $(Z_t, 0 \leq t \leq T-H)$. Therefore, the nonparametric predictor of ζ_{T+H} is defined by

$$\hat{\zeta}_{T+H} = \frac{\hat{m}_{T-H,2}(\xi_T)}{\hat{m}_{T-H,1}(\xi_T)} = \frac{\int_0^{T-H} l(\xi_{t+H}) K\left(\frac{\xi_T - \xi_t}{h_T}\right) dt}{\int_0^{T-H} K\left(\frac{\xi_T - \xi_t}{h_T}\right) dt} = \hat{m}_{T,H}(\xi_T).$$

We study now the asymptotic behavior of $\hat{\zeta}_{T+H}$ as T tends to infinity, whenever H remains fixed. As usual $\hat{\zeta}_{T+H}$ is an approximation of $m(\xi_T) = \mathbb{E}(\zeta_{T+H}|\xi_s, s \leq T) = \mathbb{E}(\zeta_{T+H}|\xi_T)$. The almost sure convergence with rate of the predictor is given in the following corollary..

Corollary 2.1. Take $d = 1$ and assume that the conditions of Theorem 2.2 are satisfied for the process $Z_t = (\xi_t, l(\xi_{t+H}))$. Then, we have

$$|\hat{m}_{T,H}(\xi_T) - m(\xi_T)| \mathbb{1}_{\{\xi_T \in \mathcal{B}_T\}} = O_{a.s}(h_T^{\beta}) + O_{a.s}\left(\left(\frac{\log T}{T h_T \alpha_T^2}\right)^{1/2}\right) \quad \text{a.s.}$$

2.3 Proofs

For any $x \in \mathbb{R}^d$, consider the following decomposition

$$Q_T(x) := (\hat{m}_{T,2}(x) - \tilde{m}_{T,2}(x)) - m(x)(\hat{m}_{T,1}(x) - \tilde{m}_{T,1}(x)) \tag{2.7}$$

$$R_T(x) := -B_T(x)(\hat{m}_{T,1}(x) - \tilde{m}_{T,1}(x)) \tag{2.8}$$

$$\hat{m}_T(x) - m(x) = \hat{m}_T(x) - C_T(x) + B_T(x) = B_T(x) + \frac{Q_T(x) + R_T(x)}{\hat{m}_{T,1}(x)}, \tag{2.9}$$

where

$$B_T(x) = C_T(x) - m(x) := \frac{\tilde{m}_{T,2}(x)}{\tilde{m}_{T,1}(x)} - m(x)$$

stands as the conditional bias. In this paper we need an exponential inequality for partial sums of unbounded martingale differences that we use to establish our results. This inequality is given in the following lemma.

Lemma 2.1. Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of σ-fields $(\mathcal{F}_n = \sigma(Z_1, \cdots, Z_n))_{n\geq 1}$, where $\sigma(Z_1, \cdots, Z_n)$ is the σ-field generated by the random variables $Z_1, \cdots, Z_n$. Set $S_n = \sum_{i=1}^n Z_i$. For any $p \geq 2$ and any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that

$$\mathbb{E}(Z_n^p|\mathcal{F}_{n-1}) \leq C^{p-2}p!d_n^2, \quad \text{almost surely.} \tag{2.10}$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}(|S_n| > \epsilon) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

where $D_n = \sum_{i=1}^n d_i^2$.

Proof. The proof follows as a particular case of Theorem 8.2.2 due to de la Peña and Giné (1999). □

From now on, for $1 \leq j \leq n$ set

$$Z_j(x) = \int_{T_{j-1}}^{T_j} K\left(\frac{x - X_t}{h_T}\right) dt.$$

2.3.1 Proof of Theorem 2.1

In the sequel we give a sequence of lemmas that are helpful in proving our results.

Lemma 2.2. Under hypotheses (H.0)(i), (H.1)(ii)-(iv) and (H.3)(iii), for any $x \in \mathbb{R}^d$, we have

$$\hat{m}_{T,2}(x) - \tilde{m}_{T,2}(x) = O_{a.s}\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right).$$

Proof. Under hypotheses (H.0)(i) and (H.1)(i), since $P(\|x - X_t\| \leq 0) = 0$, integrating by parts, we obtain

$$\begin{aligned}
\frac{1}{h_T^d}\mathbb{E}[Z_1(x)] &= \frac{1}{h_T^d}\mathbb{E}\left[\int_0^\delta K\left(\frac{x-X_t}{h_T}\right)dt\right]\\
&= \frac{1}{h_T^d}\int_0^\delta \mathbb{E}\left[k\left(\left\|\frac{x-X_t}{h_T}\right\|\right)\right]dt\\
&= \frac{1}{h_T^d}\int_0^\delta\int_0^\lambda k(u)d\mathbb{P}(\|x-X_t\|\leq h_T u)\\
&= \frac{1}{h_T^d}\int_0^\delta\left([k(u)P\left(\|x-X_t\|\leq h_T u\right)]_0^\lambda-\int_0^\lambda k'(u)P\left(\|x-X_t\|\leq h_T u\right)du\right)dt\\
&= c_0\delta f(x)\left(k(\lambda)\lambda^d-\int_0^\lambda u^d k'(u)du\right)+o(1)\\
&= d\ c_0\delta f(x)\left(\int_0^\lambda u^{d-1}k(z)du\right)+o(1). \qquad (2.11)
\end{aligned}$$

Observe that

$$\begin{aligned}
&\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)\\
&\quad= \frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T\left(l(Y_t)K\left(\frac{x-X_t}{h_T}\right)-\mathbb{E}\left[l(Y_t)K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt\\
&\quad= \frac{1}{n\mathbb{E}[Z_1(x)]}\sum_{j=1}^n\int_{T_{j-1}}^{T_j}\left(l(Y_t)K\left(\frac{x-X_t}{h_T}\right)-\mathbb{E}\left[l(Y_t)K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt\\
&\quad:= \frac{1}{n\mathbb{E}[Z_1(x)]}\sum_{j=1}^n L_{T,j}(x),
\end{aligned}$$

where $(L_{T,j}(x))_{j\geq 1}$ is a martingale difference sequence with respect to the sequence of σ-fields $(\mathcal{F}_j)_{j\geq 1}$. We have to check the conditions of Lemma 2.1 to obtain an exponential upper bound allowing to investigate the limiting behaviour of the quantity $\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)$. As a first step, using Jensen and Minkowski inequalities, for any $p\geq 2$, observe that

$$\begin{aligned}
&|\mathbb{E}\left[L_{T,j}^p(x)|\mathcal{F}_{j-2}\right]|\\
&\quad\leq \int_{T_{j-1}}^{T_j}\left(\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{j-2}\right]^{\frac{1}{p}}+\mathbb{E}\left[\left(\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right||\mathcal{F}_{t-\delta}\right]\right)^p|\mathcal{F}_{j-2}\right]^{\frac{1}{p}}\right)^p dt\\
&\quad\leq \int_{T_{j-1}}^{T_j}\left(\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{j-2}\right]^{\frac{1}{p}}+\mathbb{E}\left[\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{t-\delta}\right]|\mathcal{F}_{j-2}\right]^{\frac{1}{p}}\right)^p dt\\
&\quad= 2^p\int_{T_{j-1}}^{T_j}\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{j-2}\right]dt.
\end{aligned}$$

Making use of the condition (H.3)(iii), it follows, for any $T_{j-1}\leq t\leq T_j$ and any

$p \geq 2$, that

$$\begin{aligned}
&\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{j-2}\right] = \mathbb{E}\left[\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{S}_{t,\delta}\right]|\mathcal{F}_{j-2}\right]\\
&= \ \mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)\mathbb{E}\left[|l(Y_t)|^p|\mathcal{S}_{t,\delta}\right]|\mathcal{F}_{j-2}\right] = \mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|h_p(X_t)||\mathcal{F}_{j-2}\right].
\end{aligned}$$

Furthermore, we have

$$\begin{aligned}
&\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|h_p(X_t)||\mathcal{F}_{j-2}\right].\\
&\leq \ \mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|h_p(X_t)-h_p(x)||\mathcal{F}_{j-2}\right] + |h_p(x)|\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{j-2}\right]\\
&\leq \ \mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{j-2}\right]\left(\sup_{\|u-x\|\leq\lambda h_T}|h_p(u)-h_p(x)| + |h_p(x)|\right)\\
&\leq \ C_2(x)\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{j-2}\right],
\end{aligned}$$

where $C_2(x)$ is a positive constant. Integrating by parts and using the condition (H.0)(i) together with the condition (H.1)(ii), we obtain

$$\begin{aligned}
&\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{j-2}\right] = \mathbb{E}\left[k^p\left(\left\|\frac{x-X_t}{h_T}\right\|\right)|\mathcal{F}_{j-2}\right]\\
&= \ \int_0^\lambda k^p(u)d\mathbb{P}(\|x-X_t\| \leq h_T u|\mathcal{F}_{j-2})\\
&= \ \left[k^p(u)P\left(\|x-X_t\| \leq h_T u|\mathcal{F}_{j-2}\right)\right]_0^\lambda - \int_0^\lambda (k^p)'(u)P\left(\|x-X_t\| \leq h_T u|\mathcal{F}_{j-2}\right)du\\
&= \ c_0 h_T^d f_{t,T_{j-2}}(x)\left(k^p(\lambda)\lambda^d - \int_0^\lambda u^d (k^p)'(u)du\right) + o(h_T^d)\\
&= \ d\ c_0 h_T^d f_{t,T_{j-2}}(x)\left(\int_0^\lambda u^{d-1}k^p(u)du\right) + o(h_T^d). \qquad (2.12)
\end{aligned}$$

Assuming that the kernel k is such that $\int_0^\lambda z^{d-1}k(z)dz \leq \infty$ and taking B as to have $\int_0^\lambda z^{d-1}k^p(z)dz \leq B^{p-1}$, for any $T_{j-1} \leq t \leq T_j$, we obtain

$$\mathbb{E}\left[K^p\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{j-2}\right] = d\ c_0 B^{p-1}h_T^d f_{t,T_{j-2}}(x) + o(h_T^d).$$

Consequently, we have

$$\mathbb{E}\left[\left|l(Y_t)K\left(\frac{x-X_t}{h_T}\right)\right|^p|\mathcal{F}_{j-2}\right] = C_2(x)d\ c_0 B^{p-1}h_T^d f_{t,T_{j-2}}(x) + o(h_T^d),$$

which gives

$$\left|\mathbb{E}\left[L_{T,j}^p(x)|\mathcal{F}_{j-2}\right]\right| = 2^p\delta C_2(x)d\ c_0 B^{p-1}h_T^d f_{t,T_{j-2}}(x) + 2^p\delta o(h_T^d).$$

Thus,

$$|\mathbb{E}\left[L_{T,j}^{p}(x)|\mathcal{F}_{j-2}\right]| \;=\; p!C_3^{p-2}\delta h_T^d[dc_0C_3C_2(x)f_{t,T_{j-2}}(x)+o(1)],$$

where $C_3 = 2B$. Considering the condition (H.1)(iii) that imposes to the function $f_{t,T_{j-2}}(x)$ to be bounded by a deterministic quantity $b_{T_{j-2},2\delta}(x)$ since $T_{j-1} \leq t \leq T_j$, we can take $d_{j-2}^2 = \delta h_T^d[dc_0C_3C_2(x)b_{T_{j-2},2\delta}(x) + o_{a.s}(1)]$. Therefore, we have

$$\frac{1}{n}D_n = \frac{1}{n}\sum_{2}^{n} d_{j-2}^2 = \delta h_T^d\left[dc_0C_2(x)C_3\frac{1}{n}\sum_{2}^{n} b_{T_{j-2},2\delta}(x)+o(1)\right].$$

Using the Riemann sum, it is easily seen that

$$\delta\sum_{j=2}^{n} b_{T_{j-2},2\delta}(x) \leq \int_0^T b_{t,2\delta}(x)dt.$$

Taking into account the condition (H.1)(iv), it follows, as $T \to \infty$, that

$$\begin{aligned}\frac{D_n}{n} &= \frac{1}{n}\sum_{2}^{n} d_{j-2}^2 \leq \delta h_T^d\left[dc_0C_2(x)C_3\frac{1}{T}\int_0^T b_{t,2\delta}(x)dt+o(1)\right]\\ &= \delta h_T^d[dc_0C_2(x)C_3D_{2\delta}(x)+o(1)].\end{aligned}$$

Thus, we have

$$D_n = \delta n h_T^d[dc_0C_2(x)C_3D_{2\delta}(x)+o(1)] = O(Th_T^d).$$

Since $\mathbb{E}(Z_1(x)) = dc_0\delta h_T^d f(x)\left(\int_0^\lambda u^{d-1}k(u)du\right) + o(h_T^d)$, we note $T_n = T = \delta n$ and $h_{T_n} = h_T$ using Lemma 2.1, for any $\epsilon_0 > 0$, we obtain

$$\begin{aligned}\mathbb{P}\left(|\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)|>\epsilon_0\left(\frac{\log T}{Th_T^d}\right)^{\frac{1}{2}}\right) &= \mathbb{P}\left(\left|\sum_{j=1}^{n} L_{T,j}(x)\right| > n\mathbb{E}(Z_1(x))\epsilon_0\left(\frac{\log T_n}{T_nh_{T_n}^d}\right)^{\frac{1}{2}}\right)\\ &\leq 2\exp\left\{-\frac{(n\mathbb{E}(Z_1(x))^2\epsilon_0^2(\log T_n/T_nh_{T_n}^d)}{2D_n+2C_3n\mathbb{E}(Z_1(x))\epsilon_0\sqrt{\frac{\log T_n}{T_nh_{T_n}^d}}}\right\}\\ &\leq 2\exp\left\{-\frac{\epsilon_0^2(O(T_nh_{T_n}^d))^2(\log T_n/T_nh_{T_n}^d)}{O(T_nh_{T_n}^d)\left(1+\sqrt{\frac{\log T_n}{T_nh_{T_n}^d}}\right)}\right\}\\ &\leq 2\exp\{-C_4\epsilon_0^2\log T_n\} = \frac{2}{T_n^{C_4\epsilon_0^2}}, \qquad (2.13)\end{aligned}$$

where C_4 is a positive constant. Taking ϵ_0 sufficiently large and using the Borel-Cantelli Lemma, we obtain

$$\sum_{n\geq 1}\mathbb{P}\left(\left|\sum_{j=1}^{n} L_{T,j}(x)\right| > n\mathbb{E}(Z_1(x))\epsilon_0\left(\frac{\log T_n}{T_nh_{T_n}^d}\right)^{1/2}\right) < \infty.$$

Thus, we have

$$\hat{m}_{T,2}(x) - \tilde{m}_{T,2}(x) = O\left(\left(\frac{\log T_n}{T_n h_{T_n}^d}\right)^{1/2}\right) = O\left(\left(\frac{\log T}{T h_T^d}\right)^{1/2}\right), \quad \text{a.s.}$$

□

Lemma 2.3. Under hypotheses of Lemma 2.2, with the function l taken to be identically equal to 1, combined with the conditions (H.1)(ii) and (H.2)(i), we have

$$\hat{m}_{T,1}(x) - 1 = O_{a.s}(h_T) + O_{a.s}\left(\left(\frac{\log T}{T h_T^d}\right)^{1/2}\right).$$

Proof. Observe that

$$\hat{m}_{T,1}(x) - 1 = (\hat{m}_{T,1}(x) - \tilde{m}_{T,1}(x)) + (\tilde{m}_{T,1}(x) - 1) := M_{T,1}(x) + M_{T,2}(x)).$$

Proceeding as in Lemma 2.2 with the function l taken to be identically equal to 1, it follows, under hypotheses of Lemma 2.2, that

$$M_{T,1}(x) = O_{a.s}\left(\left(\frac{\log T}{T h_T^d}\right)^{1/2}\right). \tag{2.14}$$

Moreover, under conditions (H.0)(i), (H.1)(ii) and (H.2)(i), and using statements (2.11) and (2.12) for $p = 1$ we have

$$\begin{aligned}
M_{T,2}(x) + 1 &= \frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \mathbb{E}\left(K\left(\frac{x - X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right)dt \\
&= \frac{1}{ndc_0\delta h_T^d f(x)\left(\int_0^\lambda u^{d-1}k(u)du\right)} \\
&\quad \times \left(d\ c_0 h_T^d\left(\int_0^\lambda u^{d-1}k(u)du\right)\int_0^T f_{t,t-\delta}(x)dt + O(Th_T^d)\right) \\
&= \frac{1}{n\delta f(x)}\int_0^T f_{t,t-\delta}(x)dt + o(1) \\
&= \frac{1}{f(x)}\left(\frac{1}{T}\int_0^T f_{t,t-\delta}(x)dt\right) + o(1) \\
&= 1 + o(1).
\end{aligned}$$

Thus,

$$M_{T,2}(x) = o(1), a.s. \tag{2.15}$$

Combining the statements (2.14) and (2.15), we obtain the result and achieve the proof. □

Lemma 2.4. Under hypotheses $(H.0)(i)$, $(H.1)$-$(H.3)(ii)$, for any $x \in \mathbb{R}$, we have

$$B_T(x) = O_{a.s}(h_T^\beta),$$
$$R_T(x) \;=\; O_{a.s}\left(h_T^\beta\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right)$$

Démonstration. Observe that we can write

$$B_T(x) = \frac{\tilde{m}_{T,2}(x) - m(x)\tilde{m}_{T,1}(x)}{\tilde{m}_{T,1}(x)} = \frac{\tilde{B}_T(x)}{\tilde{m}_{T,1}(x)}. \tag{2.16}$$

Since, from the statement (2.15), we have $\tilde{m}_{T,1}(x) - 1 = o_{a.s}(1)$, it suffices then to establish that $\tilde{B}_T(x) = O_{a.s}(h_T^\beta)$. Under hypotheses (H.0)(i), (H.2)(i) and (H.3)(i)-(ii), for any x such that $f(x) \neq 0$, we have

$$\begin{aligned}
|\tilde{B}_T(x)| &= |\tilde{m}_{T,2}(x) - m(x)\tilde{m}_{T,1}(x)| \\
&= \left|\frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \left(\mathbb{E}\Big[l(Y_t)K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\Big] - m(x)\mathbb{E}\Big[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\Big]\right)dt\right| \\
&= \left|\frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \mathbb{E}\left[\mathbb{E}\left[(l(Y_t)-m(x))K\left(\frac{x-X_t}{h_T}\right)|\mathcal{S}_{t-\delta,\delta}\right]|\mathcal{F}_{t-\delta}\right]dt\right| \\
&= \left|\frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \mathbb{E}\left[(m(X_t)-m(x))K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\right| \\
&\leq \sup_{\|u-x\|\leq\lambda h_T}|m(u)-m(x)| \times \left|\frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \mathbb{E}\left[K\left(\frac{x-X_t}{h_T}\right)|\mathcal{F}_{t-\delta}\right]dt\right| \\
&\leq \sup_{\|u-x\|\leq\lambda h_T}|m(u)-m(x)|\left|\frac{d\; c_0 h_T^d\left(\int_0^\lambda u^{d-1}k(u)du\right)}{ndc_0\delta h_T^d f(x)\left(\int_0^\lambda u^{d-1}k(u)du\right)}\int_0^T f_{t,t-\delta}(x)dt + o(1)\right| \\
&\leq h_T^\beta\lambda^\beta\left|\frac{1}{f(x)}\left(\frac{1}{T}\int_0^T f_{t,t-\delta}(x)dt\right) + o(1)\right| = O_{a.s}(h_T^\beta).
\end{aligned} \tag{2.17}$$

The second part of the lemma follows by combining the decomposition (2.8) together with the statements (2.14) and (2.17). □

Proof of Theorem 2.1. Combining the decomposition (2.7) and the statement (2.14) together with Lemma 2.2, we obtain

$$Q_T(x) = O_{a.s}\left(\left(\frac{\log T}{Th_T^d}\right)^{1/2}\right). \tag{2.18}$$

Therefore, the result follows from the decomposition (2.9) combined with results of Lemma 2.4 and the statement (2.18). □

2.3.2 Proof of Theorem 2.2

The proof of Theorem 3.1 needs some intermediate results that we give hereafter as lemmas.

Lemma 2.5. Under hypotheses (H.0) and (H.4)(i)-(v) together with the condition (2.5), we have

$$\sup_{x\in\mathcal{B}_T}|\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)| = O_{a.s}\left(\left(\frac{\log T}{Th_T^d\alpha_T^2}\right)^{1/2}\right), \quad as \quad T\to\infty. \tag{2.19}$$

Démonstration. Let $x_1,\ldots,x_{\nu_T}$ be a grid of the set $\mathcal{B}_T$ in such a way that we have $\mathcal{B}_T \subseteqq \bigcup_{i=1}^{\nu_T}\mathcal{B}_{T,i}$ with $\mathcal{B}_{T,i}=\{x:\|x-x_i\|\leq T^\gamma\nu_T^{-1}\}$. Observe that

$$\begin{aligned}
\sup_{x\in\mathcal{B}_T}|\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)| &\leq \max_{1\leq k\leq\nu_T}\sup_{x\in\mathcal{B}_{T,k}}|\hat{m}_{T,2}(x)-\tilde{m}_{T,2}(x)| \\
&\leq \max_{1\leq k\leq\nu_T}\sup_{x\in\mathcal{B}_{T,k}}|\hat{m}_{T,2}(x)-\hat{m}_{T,2}(x_k)| \\
&\quad +\max_{1\leq k\leq\nu_T}|\hat{m}_{T,2}(x_k)-\tilde{m}_{T,2}(x_k)| \\
&\quad +\max_{1\leq k\leq\nu_T}\sup_{x\in\mathcal{B}_{T,k}}|\tilde{m}_{T,2}(x)-\tilde{m}_{T,2}(x_k)| \\
&:= H_{T,1}+H_{T,2}+H_{T,3}.
\end{aligned}$$

Moreover, for any $1\leq k\leq\nu_T$ and any $x\in\mathcal{B}_{T,k}$, we have

$$\begin{aligned}
\hat{m}_{T,2}(x)-\hat{m}_{T,2}(x_k) &= \frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T l(Y_t)\left(K\left(\frac{x-X_t}{h_T}\right)-K\left(\frac{x_k-X_t}{h_T}\right)\right)dt \\
&- \frac{1}{n\mathbb{E}[Z_1(x)]\mathbb{E}[Z_1(x_k)]}\int_0^T l(Y_t)K\left(\frac{x_k-X_t}{h_T}\right)(\mathbb{E}[Z_1(x_k)]-\mathbb{E}[Z_1(x)])\,dt \\
&:= I_1(x_k)+I_2(x_k).
\end{aligned}$$

Making use of Cauchy-Schwarz inequality and subsequently considering the conditions (H.0)(ii) and the statement (2.4), it follows that there exists a positive constant C_K such that

$$\begin{aligned}
|I_1(x_k)| &\leq \frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T |l(Y_t)|\left|K\left(\frac{x-X_t}{h_T}\right)-K\left(\frac{x_k-X_t}{h_T}\right)\right|dt \\
&\leq \frac{\sqrt{T}}{n\mathbb{E}[Z_1(x)]}\left(\frac{1}{T}\int_0^T(l(Y_t))^2dt\right)^{1/2}\left(\int_0^T\left|K\left(\frac{x-X_t}{h_T}\right)-K\left(\frac{x_k-X_t}{h_T}\right)\right|^2dt\right)^{1/2} \\
&\leq \frac{\sqrt{T}}{n\mathbb{E}[Z_1(x)]}O_{a.s}\left(\mathbb{E}^{1/2}[(l(Y_0))^2]\right)\left(\int_0^T C_K^2\left\|\frac{x-x_k}{h_T}\right\|^2dt\right)^{1/2} \\
&= \frac{\delta C_K T^\gamma\nu_T^{-1}}{h_T\mathbb{E}[Z_1(x)]}O_{a.s}\left(\mathbb{E}^{1/2}[(l(Y_0))^2]\right).
\end{aligned}$$

Considering the hypothesis (H.4)(i) together with the statement (2.11), we obtain

$$|I_1(x_k)| \;\leq\; \frac{C_K T^{\gamma}\nu_T^{-1}}{h_T^{d+1}\alpha_T} O_{a.s}\left(E^{1/2}[(l(Y_0))^2]\right).$$

Moreover, we have

$$\begin{aligned}
|\mathbb{E}[Z_1(x_k)] - \mathbb{E}[Z_1(x)]| &\leq \mathbb{E}\left[\int_0^{\delta}\left|K\left(\frac{x-X_t}{h_T}\right) - K\left(\frac{x_k-X_t}{h_T}\right)\right| dt\right]\\
&\leq \delta\int\left|K\left(\frac{x-y}{h_T}\right) - K\left(\frac{x_k-y}{h_T}\right)\right| f(y)dy\\
&\leq \delta C_K\left\|\frac{x-x_k}{h_T}\right\|\\
&\leq \frac{\delta C_K T^{\gamma}\nu_T^{-1}}{h_T}.
\end{aligned}$$

Therefore, taking K to be a bounded kernel by a constant a_1, we obtain

$$\begin{aligned}
|I_2(x_k)| &\leq \frac{1}{n\mathbb{E}[Z_1(x)]\mathbb{E}[Z_1(x_k)]}\int_0^T |l(Y_t)|K\left(\frac{x_k-X_t}{h_T}\right)|\mathbb{E}[Z_1(x_k)] - \mathbb{E}[Z_1(x)]|\, dt\\
&\leq \frac{a_1\delta}{\sqrt{T}\mathbb{E}[Z_1(x)]\mathbb{E}[Z_1(x_k)]}\left(\frac{1}{T}\int_0^T (l(Y_t))^2 dt\right)^{1/2}\left(\int_0^T \left(\mathbb{E}[Z_1(x_k)] - \mathbb{E}[Z_1(x)]\right)^2 dt\right)^{1/2}\\
&\leq \frac{a_1\delta}{\sqrt{T}\mathbb{E}[Z_1(x)]\mathbb{E}[Z_1(x_k)]} O_{a.s}\left(\mathbb{E}^{1/2}[(l(Y_0))^2]\right)\left(T\left(\frac{\delta C_K T^{\gamma}\nu_T^{-1}}{h_T}\right)^2\right)^{1/2}\\
&= \frac{a_1 C_K T^{\gamma}\nu_T^{-1}}{h_T^{2d+1}\alpha_T^2} O_{a.s}\left(\mathbb{E}^{1/2}[(l(Y_0))^2]\right).
\end{aligned}$$

Consequently, it follows that

$$I_1(x_k) + I_2(x_k) \;\leq\; \frac{C_K T^{\gamma}\nu_T^{-1}}{h_T^{2d+1}\alpha_T^2}(a_1 + h_T^d\alpha_T) O_{a.s}\left(\mathbb{E}^{1/2}[(l(Y_0))^2]\right). \tag{2.20}$$

Taking now $\nu_T = \lfloor T/h_T^{2d+1}\alpha_T^2\rfloor$ and considering the right hand side of the statement (2.20) together with the fact that $\mathbb{E}[(l(Y_0))^2] < \infty$, we obtain

$$H_{T,1} = \max_{1\leq k\leq \nu_T}\sup_{x\in\mathcal{B}_{T,k}} |\hat{m}_{T,2}(x) - \hat{m}_{T,2}(x)| = O_{a.s.}(T^{\gamma-1}), \quad \text{as} \quad T\to\infty. \tag{2.21}$$

Observe now that

$$\begin{aligned}
\tilde{m}_{T,2}(x_k) - \tilde{m}_{T,2}(x) \;=\; & \frac{1}{n\mathbb{E}[Z_1(x)]}\int_0^T \mathbb{E}\left[l(Y_t)\left(K\left(\frac{x-X_t}{h_T}\right) - K\left(\frac{x_k-X_t}{h_T}\right)\right)|\mathcal{F}_{t-\delta}\right] dt\\
& - \frac{1}{n\mathbb{E}[Z_1(x)]\mathbb{E}[Z_1(x_k)]}\int_0^T \mathbb{E}\left[l(Y_t)K\left(\frac{x_k-X_t}{h_T}\right)\right.\\
& \left.\times\left(\mathbb{E}[Z_1(x_k)] - \mathbb{E}[Z_1(x)]\right) \mid \mathcal{F}_{t-\delta}\right] dt.
\end{aligned}$$

Proceeding similarly as for $H_{T,1}$, we obtain

$$H_{T,3} = \max_{1\leq k\leq \nu_T}\sup_{x\in\mathcal{B}_{T,k}} |\tilde{m}_{T,2}(x_k) - \tilde{m}_{T,2}(x)| = O_{a.s.}(T^{\gamma-1}), \;\; \text{as} \;\; T\to\infty. \tag{2.22}$$

To handle the term $H_{T,2}$, we proceed similarly as in the statement (2.13). Considering hypotheses (H.4)(ii)-(iii), and noting $T = T_n = \delta n$, $h_T = h_{T_n}$, $\nu_T = \nu_{T_n}$ and $\alpha_T = \alpha_{T_n}$, it follows that $D_n = O(T_n h_{T_n}^d)$ uniformly in $x \in \mathcal{B}_T$. Moreover, assuming the conditions (H.4)(i) and (H.4)(iv), for any $T \geq 0$, we have

$$\inf_{x\in\mathcal{B}_T} \mathbb{E}[Z_1(x)] \geq \delta h_{T_n}^d \alpha_{T_n} + O(h_{T_n}^d) \quad \text{and} \quad \sup_{x\in\mathcal{B}_T} \mathbb{E}[Z_1(x)] \leq C\delta h_{T_n}^d + O(h_{T_n}^d).$$

Therefore, taking $\nu_{T_n} = \left[\dfrac{T_n}{\alpha_{T_n}^2 h_{T_n}^{2d+1}}\right]$ and $\epsilon_{T_n} = \epsilon_0 \left(\dfrac{\log T_n}{T_n h_{T_n}^d \alpha_{T_n}^2}\right)^{\frac{1}{2}}$, we obtain

$$\begin{aligned}
\mathbb{P}\left(\max_{1\leq k\leq \nu_{T_n}} |\hat{m}_{T,2}(x_k) - \tilde{m}_{T,2}(x_k)| > \epsilon_T\right) &\leq \sum_{k=1}^{\nu_{T_n}} \mathbb{P}\left(\left|\sum_{j=1}^n L_{T,j}(x_k)\right| > \epsilon_{T_n}\Big(n\mathbb{E}[Z_1(x_k)]\Big)\right) \\
&\leq 2\nu_{T_n} \exp\left\{-\frac{\Big(n\mathbb{E}[Z_1(x_k)]\Big)^2 \epsilon_{T_n}^2}{O(T_n h_{T_n}) + Cn\mathbb{E}[Z_1(x_k)]\epsilon_{T_n}}\right\} \\
&\leq 2\nu_{T_n} \exp\left\{-\frac{\delta^2 n^2 h_{T_n}^{2d} \alpha_{T_n}^2 \epsilon_0^2 \Big(\log T_n / T_n h_{T_n}^d \alpha_{T_n}^2\Big)}{O(T_n h_{T_n}^d) + C\delta n h_{T_n} \Big(\frac{\log T_n}{T_n h_{T_n}^d \alpha_{T_n}^2}\Big)^{1/2}}\right\} \\
&\leq 2\nu_{T_n} \exp\left\{-\frac{\epsilon_0^2 O\Big((T_n h_{T_n}^d)^2\Big) \log T_n / T_n h_{T_n}^d}{O(T_n h_{T_n}^d)\left(1 + C\Big(\frac{\log T_n}{T_n h_{T_n}^d \alpha_{T_n}^2}\Big)^{1/2}\right)}\right\} \\
&\leq 2\nu_{T_n} \exp\left\{-\frac{\epsilon_0^2 \log T_n}{\left(1 + C\Big(\frac{\log T_n}{T_n h_{T_n}^d \alpha_{T_n}^2}\Big)^{1/2}\right)}\right\} \\
&\leq \frac{2\Big(T_n / \alpha_{T_n}^2 h_{T_n}^{2d+1}\Big)}{T_n^{\epsilon_0^2 \theta}} \\
&= \frac{2}{\alpha_{T_n}^2 h_{T_n}^{2d+1} T_n^{\epsilon_0^2 \theta - 1}}
\end{aligned}$$

for some positive constant θ. Making use of the condition (H.4)(v) combined with the Borel-Cantelli Lemma, we obtain

$$\max_{1\leq k\leq \nu_T} |\hat{m}_{T,2}(x_k) - \tilde{m}_{T,2}(x_k)| = O_{a.s}\left(\left(\frac{\log T}{T h_T^d \alpha_T^2}\right)^{1/2}\right). \tag{2.23}$$

Combining the statements (2.20)-(2.23) together with the condition (2.5), the result follows. □

Lemma 2.6. Under hypotheses of Lemma 2.5, with the function l taken to be identically equal to 1, combined with the conditions (H.1)(ii) and (H.2)(i), we have

$$\sup_{x\in\mathcal{B}_T} |\hat{m}_{T,1}(x) - 1| = o_{a.s}(1), \quad \text{as } T \to \infty. \tag{2.24}$$

Proof. Since

$$\hat{m}_{T,1}(x) - 1 = (\hat{m}_{T,1}(x) - \tilde{m}_{T,1}(x)) + (\tilde{m}_{T,1}(x) - 1) := R_{T,1}(x) + R_{T,2}(x)),$$

proceeding as in Lemma 2.5 with the function l taken to be identically equal to 1, it follows, under hypotheses of Lemma 2.5, that

$$\sup_{x\in\mathcal{B}_T} |R_{T,1}(x)| = \sup_{x\in\mathcal{B}_T} |\hat{m}_{T,1}(x) - \tilde{m}_{T,1}(x)| = O_{a.s}\left(\left(\frac{\log T}{Th_T^d\alpha_T^2}\right)^{1/2}\right). \tag{2.25}$$

Moreover, under conditions (H.0)(i), (H.1)(ii) and (H.4)(vi), following similar lines as to establish the statement (2.15), we obtain

$$\sup_{x\in\mathcal{B}_T} |R_{T,2}(x)| \;\; \leq \;\; \sup_{x\in\mathcal{B}_T}\left|\frac{1}{Tf(x)}\int_0^T f_{t,t-\delta}(x)dt - 1\right| + o_{a.s}(1) = o_{a.s}(1). \tag{2.26}$$

In view of the condition (b), we achieve the proof by considering the statements (2.25) and (2.26). □

Lemma 2.7. Under hypotheses $(H.0)(ii)$, $(H.1)$-$(H.3)(ii)$ and the condition (2.5), for any $x \in \mathbb{R}$, we have

$$\sup_{x\in\mathcal{B}_T} |B_T(x)| \;\; = \;\; O_{a.s}\left(h_T^\beta\right), \tag{2.27}$$

$$\sup_{x\in\mathcal{B}_T} |R_T(x)| \;\; = \;\; O_{a.s}\left(h_T^\beta\left(\frac{\log T}{Th_T^d\alpha_T^2}\right)^{1/2}\right) \tag{2.28}$$

Démonstration. From Lemma 2.6, it is easily seen that $\inf_{x\in\mathcal{B}_T} \hat{m}_{T,1}(x) = 1 + o_{a.s}(1)$. Considering the statement (2.16), it suffices to establish that $\tilde{B}_T(x) = O_{a.s}\left(h_T^\beta\right)$. Under hypotheses (H.0)(ii) and (H.3)(i)-(ii), by the same arguments as in the proof of Lemma 4, for any x, we have

$$|\tilde{B}_T(x)| = |\tilde{m}_{T,2}(x) - m(x)\tilde{m}_{T,1}(x)| \;\; \leq \;\; h_T^\beta\lambda^\beta\left|\frac{1}{f(x)T}\int_0^T f_{t,t-\delta}(x)dt + o(1)\right|. \tag{2.29}$$

Therefore, making use of the conditions (H.4)(i), (H.4)(vi), we obtain

$$\begin{aligned}\sup_{x\in\mathcal{B}_T} |\tilde{B}_T(x)| \;\; &\leq \;\; h_T^\beta\left(\sup_{x\in\mathcal{B}_T}\left|\frac{1}{f(x)T}\int_0^T f_{t,t-\delta}(x)dt - 1\right| + 1 + o(1)\right) \\ &= \;\; O_{a.s}\left(h_T^\beta\right).\end{aligned}$$

The second part of the lemma follows by combining the decomposition (2.8) together with the statements (2.25) and (2.27). □

Proof of Theorem 2.2.

Combining the decomposition (2.7) together with Lemma 2.5 and the statement (2.25), it follows that

$$\sup_{x\in\mathcal{B}_T} |Q_T(x)| = O_{a.s}\left(\left(\frac{\log T}{Th_T^d\alpha_T^2}\right)^{1/2}\right). \tag{2.30}$$

Therefore, the result follows from the decomposition (2.9) combined with results of Lemma 2.7 and the statement (2.30). □

2.3.3 Proof of Corollary 2.1

Observe that we have

$$|\hat{m}_{T,H}(\xi_T) - m(\xi_T)|\mathbb{1}_{\{\xi_T\in\mathcal{B}_T\}} \leq \sup_{x\in\mathcal{B}_T}\left|\frac{\hat{m}_{T-H,2}(x)}{\hat{m}_{T-H,1}(x)} - m(x)\right|.$$

Hence, using the same method to prove Theorem 2.2 with $d = 1$, we obtain the result. □

Chapitre 3

Consistency results for the conditional mode estimate on continuous time stationary and ergodic data

Abstract

Let $(\mathbf{X}_t, Y_t)_{t\geq 0}$ be a $\mathbb{R}^d \times \mathbb{R}$-valued strictly stationary and ergodic continuous time process. Define the estimator of the conditional mode of Y given $\mathbf{X} = \mathbf{x}$ as the random variable $\Theta_T(\mathbf{x})$ that maximizes a kernel estimator of the conditional density of Y given $\mathbf{X} = \mathbf{x}$. We establish the almost sure pointwise and uniform consistencies with rates of $\Theta_T(\mathbf{x})$. Furthermore, considering a real stationary and ergodic continuous time process $(Z_t)_{t\geq 0}$, we derive the convergence with a rate of a predictor at the horizon H, built on using the conditional mode estimation device, whenever $0 < H < T$ and only the part $(Z_t)_{0\leq t\leq T-H}$ of the process is observed. It is worth noticing that the ergodic setting covers and completes various situations as compared to the mixing case and stands as more convenient to use in practice.

Keywords : Conditional density, Conditional mode, kernel estimate, Ergodic processes, Martingale difference, Prediction

3.1 Introduction

Let $(\mathbf{X}_t, Y_t)_{t\geq 0}$ be a $\mathbb{R}^d \times \mathbb{R}$-valued strictly stationary and ergodic continuous time process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let h be the density function of the random vector $(\mathbf{X}, Y)$. Our concern in this paper is to build a nonparametric predictor of Y given that one has observed $\mathbf{X} = \mathbf{x}$. This problem has received a lot of attention throughout the time. In the discrete time case, the conditional expectation is used to construct a predictor of the random variable Y given $\mathbf{X} = \mathbf{x}$, We refer to Watson (1964)[126] and Rosenblatt (1969)[118] where results of consistency of the predictor estimate are established for the mixing processes. Considering the same mixing framework, Gannoun (1990)[63] used the conditional median. The conditional mode function estimate of the predictor is used for the first time by Collomb *et al* (1987)[35]. A number of publications followed considering various prediction situations. We refer among others to Ould said (1993, 1997)[104],[105], Louani & Ould said (1999)[84], Ezzahrioui & Ould said (2008, 2010)[56] [57]. Finally, when the process is considered to be i.i.d., the almost sure convergence together with mean convergence of the conditional density were obtained by Youndjé (1993)[130]. Note that the conditional density estimate is used to establish results pertaining to the conditional mode estimate.

Denote by $f(./\mathbf{x})$ the conditional density of Y given that $\mathbf{X} = \mathbf{x}$. Assume that $f(./\mathbf{x})$ has a unique mode and that the conditional mode of Y given $\mathbf{X} = \mathbf{x}$ is denoted by

$$\Theta(\mathbf{x}) = \arg\max_{y\in\mathcal{G}_T} f(y/\mathbf{x}), \tag{3.1}$$

where $\mathcal{G}_T = [-\upsilon_T, \upsilon_T]$ is a compact set on $\mathbb{R}$, with $\upsilon_T = a_T^{-\mu}$, where μ is a positive constant.

The kernel estimator of the conditional mode $\Theta(x)$ is defined as the random variable $\Theta_T(\mathbf{x})$ which maximizes the kernel estimator $f_T(./\mathbf{x})$ of $f(./\mathbf{x})$, that is,

$$\Theta_T(\mathbf{x}) = \arg\max_{y\in\mathcal{G}_T} f_T(y/\mathbf{x}), \tag{3.2}$$

with

$$f_T(y/\mathbf{x}) = \frac{h_T(\mathbf{x}, y)}{g_T(\mathbf{x})}, \tag{3.3}$$

where

$$h_T(\mathbf{x}, y) = \frac{1}{Ta_T^{d+1}} \int_0^T K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) K_2\left(\frac{y - Y_t}{a_T}\right) dt = \int_0^T \varphi_t(\mathbf{x}, y)dt, \tag{3.4}$$

and

$$g_T(\mathbf{x}) = \frac{1}{Ta_T^{d}} \int_0^T K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) dt = \int_0^T \phi_t(\mathbf{x})dt. \tag{3.5}$$

Notice that g_T is the kernel estimate of the marginal density g of $\mathbf{X}$ that we suppose to exist. Here a_T is a sequence of positive real numbers such that

$$(i)\ \lim_{T\to\infty} a_T = 0, \qquad (ii)\ \lim_{T\to\infty} Ta_T^{d+1} = +\infty, \qquad (iii)\ \lim_{T\to\infty} \frac{Ta_T^{d+1}}{\log T} = +\infty,$$
$$(iv)\ \lim_{T\to\infty} \frac{Ta_T^{d}}{\log T} = +\infty. \tag{3.6}$$

This paper is devoted to the conditional mode function estimation when the data are assumed to be sampled from a stationary and ergodic continuous time process to allow the maximum possible generality in regard to the dependence setting. We establish here pointwise and uniform consistencies with rates of the conditional mode kernel estimator $\Theta_T(\mathbf{x})$. An application to the process prediction based on the conditional mode estimation is also given. Note that the ergodic framework avoid the widely used strong mixing condition and its variants to measure the dependency and the very involved probabilistic calculations that it implies (see, for instance, Masry (2005)[98]). Further motivations to consider ergodic data are discussed in Didi & Louani (2013b)[51], and in chapter 2 of this thesis, where details defining the ergodic property of continuous time processes are also given. The paper is organized as follows : In Section 1, we state our assumptions. Section 2 is devoted to the main results to nonparametric estimation in continuous time. Section 3 deals with the proof of the main results.

3.2 Assumptions and notation

To formulate our assumptions, some additional notations are required, for some $\delta > 0$, let $n \in \mathbb{N}$ be such that $T = \delta n$, and $T_j = j\delta$, for $j = 1, \dots, n$. Let $\mathcal{F}_t$ be the $\sigma-$field generated by $\{(\mathbf{X}_s, Y_s) : 0 \le s < t\}$. Set $\mathcal{F}_j$ to be the σ-field generated by $\{(\mathbf{X}_s, Y_s) : 0 \le s \le T_j\}$. The σ-field generated by $\{(\mathbf{X}_s, Y_s), (\mathbf{X}_r) : 0 \le s < t; t \le r \le t + \delta\}$ is denoted by $\mathcal{S}_{t,\delta}$. The $\sigma-$field $\mathcal{G}_t$ is generated by $\{(\mathbf{X}_s) : 0 \le s < t\}$, while that generated by $\{(\mathbf{X}_s) : 0 \le s \le T_j\}$ is denoted by $\mathcal{G}_j$. Introduce now the following assumptions

(A.1) The kernels K_j,$j = 1, 2$ are probability density functions,

- (i) Kernels K_j,$j = 1, 2$ are assumed to be compactly supported.
- (ii) Kernels K_j,$j = 1, 2$ are assumed to be Lipschitz of order γ with a constant $0 < C_K < \infty$, i.e.,
  - (ii1) $|K_1(\mathbf{x}) - K_1(\mathbf{x}')| \le C_K|\mathbf{x} - \mathbf{x}'|^\gamma, \quad (\mathbf{x}, \mathbf{x}') \in \mathbb{R}^{2d}$,

(ii2) $|K_2(y) - K_2(y^{'})| \leq C_K|y - y^{'}|^{\gamma}, \quad (y, y^{'}) \in \mathbb{R}^2\,;$

(iii) $\int_{\mathbb{R}^d} \|\mathbf{x}\| K_1(\mathbf{x})d\mathbf{x} < \infty$;

(iv) $\int_{\mathbb{R}} yK_2(y)dy = 0$ and $\int_{\mathbb{R}} y^2K_2(y)dy < \infty$.

(A.2) There exists a constant $0 < \Gamma_1 < \infty$, such that, for all $(\mathbf{x}, y) \in \mathbb{R}^{d+1}$, $h(\mathbf{x}, y) < \Gamma_1$.

(A.3) For any $\mathbf{x} \in \mathcal{C}$, there exists a constant $\lambda > 0$ such that $\lambda \leq g(\mathbf{x})$.

(A.4) For every $t \in \mathbb{R}_+$, for every couple $(\mathbf{x}, y) \in \mathbb{R}^{d+1}$,

(i) The conditional density $g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}$ of $\mathbf{X}_t$ given the σ-field $\mathcal{G}_{t-\delta}$ exists and is a Lipschitz function.

(ii) The conditional density $h_{\mathbf{X}_t,Y_t}^{\mathcal{F}_{t-\delta}}$ of $(\mathbf{X}_t, Y_t)$ given the σ-field $\mathcal{F}_{t-\delta}$ exists and is continuous.

(iii) $\lim\limits_{T\to\infty} \frac{1}{T}\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt = g(\mathbf{x}), \quad \text{a.s.}\,;$

(iv) The conditional density $g_{\mathbf{X}_t}^{\mathcal{F}_{t-\delta}}$ of $\mathbf{X}_t$ given the σ-field $\mathcal{F}_{t-\delta}$ exists and is continuous.

(A.5) For every $t \in [0, T]$, for every couple $(\mathbf{x}, y) \in \mathbb{R}^{d+1}$,

$$\lim_{T\to\infty} \sup_{\mathbf{x}\in\mathbb{R}^d} \left| \frac{1}{T}\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt - g(\mathbf{x}) \right| = 0, \quad \text{a.s..}$$

(A.6) For any t and any r such that $t \in [0, T]$ and $t \leq r \leq t+\delta$, we have $\mathbb{E}\left[Y_r|\mathcal{S}_{t,\delta}\right] = \mathbb{E}\left[Y_r|\mathbf{X}_r\right]$.

(A.7) For any fixed $\mathbf{x} \in \mathbb{R}^d$,

(i) $f(\cdot/\mathbf{x})$ is differentiable up to order 2 and $\sup_y \left|f^{(2)}(y|\mathbf{x})\right| < \infty$ uniformly on $\mathbf{x}$.

(ii) $\sup\limits_{\mathbf{x}\in\mathcal{C}} |f^{(2)}(\Theta(\mathbf{x}) \mid \mathbf{x})| := \Phi \neq 0.$

(A.8) The conditional density and its two first derivatives satisfy the Hölder condition with respect to each variable, that is, there exist constants $\beta > 0$ and $\nu \in]0, 1]$ such that

$$\begin{aligned} &\forall (y_1, y_2) \in \mathbb{R}^2, \forall (\mathbf{x}_1, \mathbf{x}_2) \in \vartheta(\mathbf{x}) \times \vartheta(\mathbf{x}), \\ &\left|f^{(j)}(y_1|\mathbf{x}_1) - f^{(j)}(y_2|\mathbf{x}_2)\right| \leq C_{\mathbf{x}} \left(|y_1 - y_2|^{\nu} + \|\mathbf{x}_1 - \mathbf{x}_2\|^{\beta}\right) \end{aligned}$$

for $j = 0, 1, 2$, where $\vartheta(\mathbf{x})$ is a neighbourhood of $\mathbf{x}$, and $C_{\mathbf{x}}$ is a constant that depends only on $\mathbf{x}$. Here, we take the convention $f^{(0)}(\cdot|\cdot) = f(\cdot|\cdot)$.

Comments on hypotheses. Conditions (A.1) are very common in nonparametric function estimation literature. They set some kind of regularity upon the kernels used in our estimates. Hypotheses (A.2)-(A.3) and (A.7)-(A.8) introduce the needed regularity conditions on the distribution functions related to the process $(\mathbf{X}_t, Y_t)_{t\geq 0}$. The conditions (A.4)-(A.5) involve the ergodic nature of the data as given, for instance, in Györfi *et al.* (1989). Notice, by (A.4)(i), that the random functions $g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})$ belong to the space $\mathcal{C}^0$ of continuous functions, which is a separable Banach space. Moreover, approaching the integral $\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt$ by its Riemann's sum, it follows that

$$T^{-1}\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt \backsimeq n^{-1}\sum_{i=1}^n g_{\mathbf{X}_{T_i}}^{\mathcal{G}_{T_i-\delta}}(\mathbf{x}).$$

Since the process $(X_{T_j})_{j\geq 1}$ is stationary and ergodic, following Delecroix (1987) (see, Lemma 4 and Corollary 1 together with their proofs), one may prove that the sequence $(g_{\mathbf{X}_{T_i}}^{\mathcal{G}_{T_i-\delta}}(\mathbf{x}))_{i\geq 1}$ of random functions is stationary and ergodic. Indeed, It suffices to replace in the work of Delecroix the conditional densities by $g_{\mathbf{X}_{T_i}}^{\mathcal{G}_{T_i-\delta}}(\mathbf{x})$'s and the density by the function g. Moreover, making use of Beck's Theorem (see, for instance, Györfi *et al.* (1989), Theorem 2.1.1), it follows that

$$\lim_{T\to\infty}\sup_{x\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt - \mathbb{E}(g_{\mathbf{X}_0}^{\mathcal{G}_{-\delta}}(\mathbf{x}))\right| = \lim_{T\to\infty}\sup_{x\in\mathbb{R}^d}\left|\frac{1}{T}\int_0^T g_{\mathbf{X}_t}^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt - g(\mathbf{x})\right| = 0, a.s.$$

It is then clear that the condition (A.5) is satisfied. We can proceed similarly with the function $h_{X_t,Y_t}^{\mathcal{F}_{t-\delta}}(\mathbf{x}, y)$. Condition (A.6) is usual in the literature dealing with the study of ergodic processes.

3.3 Theoretical properties

We denote by $f^{(1)}(\cdot/\mathbf{x})$ the first derivative of $f(\cdot/\mathbf{x})$. Using the definition of the conditional mode function, we have

$$f^{(1)}(\Theta(\mathbf{x})/\mathbf{x}) = 0,$$

Similarly, it follows from the statement (3.2) that

$$f_T^{(1)}(\Theta_T(\mathbf{x})/\mathbf{x}) = 0.$$

Furthermore, assumption (A.7) implies

$$f^{(2)}(\Theta(\mathbf{x})/\mathbf{x}) < 0, \quad \text{and} \quad f_T^{(2)}(\Theta_T(\mathbf{x})/\mathbf{x}) < 0.$$

3.3.1 Consistency

The first result considers the almost sure pointwise convergence with rate of the kernel estimate $\Theta_T(\mathbf{x})$ of the conditional mode function $\Theta(\mathbf{x})$.

Theorem 3.1. Under assumptions (A.1)(i), (A.1)(ii2), (A.1)(iv), (A.2), (A.4), (A.6), (A.7) and (A.8), we have

$$\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}) \overset{a.s}{=} O(a_T^{\beta/2} + a_T^{\nu/2}) + O\left(\left(\frac{\log T}{Ta_T^d}\right)^{1/4}\right) + O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/4}\right).$$

The following theorem gives the almost sure uniform with rate, over a compact set $\mathcal{C}$, consistency result.

Theorem 3.2. Under the hypotheses (A.1), (A.2), (A.3), (A.4)(i)-(ii), (A.4)(iv), (A.5), (A.6), (A.7), and (A.8), we have

$$\sup_{x\in\mathcal{C}} |\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})| \overset{a.s}{=} O(a_T^{\beta/2} + a_T^{\nu/2}) + O\left(\left(\frac{\log T}{Ta_T^d}\right)^{1/4}\right) + O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/4}\right).$$

3.3.2 Nonparametric prediction.

Let $(\xi_t, t \in \mathbb{R}_+)$ be a measurable stationary and ergodic process. Given the data $(\xi_t, 0 \le t \le T)$, we would like to predict the non-observed real random variable ξ_{T+H}, where the horizon H is such that $0 < H < T$. For this purpose, let us construct the predictor. Set

$$Z_t = (X_t, Y_t) = (\xi_t, \xi_{t+H}), \quad t \ge 0,$$

and consider the conditional mode kernel estimator $\Theta_T(\mathbf{x})$ based on the data $(Z_t, 0 \le t \le T - H)$. Therefore, the nonparametric predictor of ξ_T is defined by

$$\hat{\xi}_T = \widehat{\Theta}_{T-H}(\xi_T).$$

We study now the asymptotic behavior of $\hat{\xi}_T$ as T tends to infinity, whenever H remains fixed. As usual $\hat{\xi}_T$ is an approximation of $\Theta(\xi_T))$. The almost sure convergence with rate of the predictor is given in the following corollary.

Corollary 3.1. Take $d = 1$ and assume that the conditions of Theorem 3.2 are satisfied for the process $Z_t = (\xi_t, \xi_{t+H})$. Then, we have

$$|\widehat{\Theta}_{T-H}(\xi_T) - \Theta(\xi_T)| \mathbb{1}_{\{\xi_T\in\mathcal{C}\}} \overset{a.s}{=} O(a_T^{\beta/2} + a_T^{\nu/2}) + O\left(\left(\frac{\log T}{Ta_T}\right)^{1/4}\right) + O\left(\left(\frac{\log T}{Ta_T^2}\right)^{1/4}\right).$$

3.4 Proofs

From the definition of $\Theta_T(\mathbf{x})$ and $\Theta(\mathbf{x})$, we have

$$
\begin{aligned}
&|f(\Theta_T(\mathbf{x})/\mathbf{x}) - f(\Theta(\mathbf{x})/\mathbf{x})| \\
&\leq |f_T(\Theta_T(\mathbf{x})/\mathbf{x}) - f(\Theta_T(\mathbf{x})/\mathbf{x})| + |f_T(\Theta_T(\mathbf{x})/\mathbf{x}) - f(\Theta(\mathbf{x})/\mathbf{x})| \\
&\leq \sup_{y\in\mathcal{G}_T} |f_T(y/\mathbf{x}) - f(y/\mathbf{x})| + |\sup_{y\in\mathcal{G}_T} f_T(y/\mathbf{x}) - \sup_{y\in\mathcal{G}_T} f(y/\mathbf{x})| \\
&\leq 2\sup_{y\in\mathcal{G}_T} |f_T(y/\mathbf{x}) - f(y/\mathbf{x})|. \qquad (3.7)
\end{aligned}
$$

Consider the next decomposition

$$
\begin{aligned}
Q_T(\mathbf{x},y) &:= (h_T(\mathbf{x},y) - \bar{h}_T(\mathbf{x},y)) - \frac{h(\mathbf{x},y)}{g(\mathbf{x})}(g_T(\mathbf{x}) - \bar{g}_{T,1}(\mathbf{x})) \qquad (3.8)\\
R_T(\mathbf{x},y) &:= -B_T(\mathbf{x},y)(g_T(\mathbf{x}) - \bar{g}_{T,1}(\mathbf{x})) \qquad (3.9)\\
B_T(\mathbf{x},y) &:= \frac{\bar{h}_T(\mathbf{x},y) - \Big(h(\mathbf{x},y)/g(\mathbf{x})\Big)\bar{g}_{T,1}(\mathbf{x})}{\bar{g}_{T,1}(\mathbf{x})} \qquad (3.10)\\
f_T(y \mid \mathbf{x}) - f(y \mid \mathbf{x}) &= B_T(\mathbf{x},y) + \frac{Q_T(\mathbf{x},y) + R_T(\mathbf{x},y)}{g_T(\mathbf{x})}, \qquad (3.11)
\end{aligned}
$$

where

$$\bar{g}_{T,1}(\mathbf{x}) = \frac{1}{Ta_T^d}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) \mid \mathcal{F}_{t-\delta}\right] dt,$$

and

$$\bar{h}_T(\mathbf{x},y) = \frac{1}{Ta_T^{d+1}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) K_2\left(\frac{y-Y_t}{a_T}\right) \mid \mathcal{F}_{t-\delta}\right] dt.$$

Lemma 3.1. Let $(Z_n)_{n\geq 1}$ be a sequence of real martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_n = \sigma(Z_1,\ldots,Z_n))_{n\geq 1}$, where is the σ-field generated by the random variables $Z_1,\ldots,Z_n$. Set $S_n = \sum_{i=1}^n Z_i$ For any $p \geq 2$ and any $n \geq 1$, assume that there exist some nonnegative constants C and d_n such that

$$\mathbb{E}\left[Z_n^p|\mathcal{F}_{n-1}\right] \leq C^{p-1}p!\ d_n^2, \quad \text{almost sure.}$$

Then, for any $\epsilon > 0$, we have

$$\mathbb{P}\left(|S_n| > \epsilon\right) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\}.$$

where $D_n = \sum_{i=1}^n d_i^2$

Lemma 3.2. Let $\Lambda \times \Lambda'$ be an index set and for each $(\lambda,\lambda') \in \Lambda \times \Lambda'$, let $\{Z_i(\lambda,\lambda'), i \geq 1\}$, be a sequence of martingale differences such that $|Z_i(\lambda,\lambda')| \leq B$ a.s. Then, for any $\epsilon > 0$ and any n sufficiently large, we have

$$P\left\{\left|\sum_{i=1}^n Z_i(\lambda,\lambda')\right| > \epsilon\right\} \leq 2\exp\left\{-\frac{\epsilon^2}{2nB^2}\right\}$$

3.4.1 Proof of Theorem 3.1

Proposition 3.1. Under assumptions (A.1)(i)-(ii2)-(iii)-(iv), (A.2), (A.4), (A.6) and (A.8), we have

$$\sup_{y\in\mathcal{G}_T}|f_T(y\mid\mathbf{x})-f(y\mid\mathbf{x}) \overset{a.s}{=} O(a_T^{\beta}+a_T^{\nu})+O\left(\left(\frac{\log T}{Ta_T^d}\right)^{1/2}\right)+O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right).$$

Proof of Proposition 3.1. Condition (A.2) implies that

$$\begin{aligned}
\sup_{y\in\mathcal{G}_T}|f_T(y/\mathbf{x})-f(y/\mathbf{x})| &= \sup_{y\in\mathcal{G}_T}\left|B_T(\mathbf{x},y)+\frac{Q_T(\mathbf{x},y)+R_T(\mathbf{x},y)}{g_T(\mathbf{x})}\right| \\
&\leq \sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|+\frac{1}{g_T(\mathbf{x})}\left(\sup_{y\in\mathcal{G}_T}|Q_T(\mathbf{x},y)|+\sup_{y\in\mathcal{G}_T}|R_T(\mathbf{x},y)|\right) \\
&\leq \sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|+\frac{1}{g_T(\mathbf{x})}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y)-\bar{h}_T(\mathbf{x},y)\right| \\
&\quad +\frac{1}{g_T(\mathbf{x})}\left(\lambda^{-1}\Gamma_1+\sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|\right)|g_T(\mathbf{x})-\bar{g}_{T,1}(\mathbf{x})|. \quad (3.12)
\end{aligned}$$

Lemma 3.3. Under conditions (A.1)(i), (A.1)(iii), (A.4)(i) and (A.4)(iii), one has

$$\lim_{T\to\infty} g_T(\mathbf{x}) \overset{a.s}{=} g(\mathbf{x}). \quad (3.13)$$

Proof of Lemma 3.3. Observe that

$$g_T(\mathbf{x})-g(\mathbf{x})=g_T(\mathbf{x})-\bar{g}_T(\mathbf{x})+\bar{g}_T(\mathbf{x})-g(\mathbf{x})=G_{1,T}(\mathbf{x})+G_{2,T}(\mathbf{x}), \quad (3.14)$$

with $\bar{g}_T(\mathbf{x})=\frac{1}{Ta_T^d}\int_0^T\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{G}_{t-\delta}\right]dt$. Two terms have to be investigated. We first take a closer look to the second term $G_{2,T}(\mathbf{x})$ given by

$$\begin{aligned}
G_{2,T}(\mathbf{x}) &= \frac{1}{Ta_T^d}\int_0^T E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{G}_{t-\delta}\right]dt-g(\mathbf{x}) \\
&= \frac{1}{Ta_T^d}\int_0^T\int_{\mathbb{R}^d}K_1\left(\frac{\mathbf{x}-\mathbf{u}}{a_T}\right)g^{\mathcal{G}_{t-\delta}}(\mathbf{u})d\mathbf{u}dt-g(\mathbf{x}) \\
&= \frac{1}{T}\int_0^T\int_{\mathbb{R}^d}K_1(\mathbf{r})g^{\mathcal{G}_{t-\delta}}(\mathbf{x}-a_T\mathbf{r})d\mathbf{r}dt-g(\mathbf{x}).
\end{aligned}$$

It follows from assumption (A.4)(i) that

$$\left|g^{\mathcal{G}_{t-\delta}}(\mathbf{x}-a_T\mathbf{r})-g^{\mathcal{G}_{t-\delta}}(\mathbf{x})\right|\leq a_T\|\mathbf{r}\|.$$

Making use of assumptions (A.1)(iii) and (A.4)(iii), we obtain

$$G_{2,T}(\mathbf{x})=a_T\int_{\mathbb{R}^d}\|\mathbf{r}\|K_1(\mathbf{r})d\mathbf{r}+\frac{1}{T}\int_0^T g^{\mathcal{G}_{t-\delta}}(\mathbf{x})dt-g(\mathbf{x})=o(1), \quad \text{a.s.} \quad (3.15)$$

Let us return our attention on the first term of the decomposition (3.14), $G_{1,T}(\mathbf{x})$. It is clear that

$$G_{1,T}(\mathbf{x}) = \frac{1}{Ta_T^d}\int_0^T \left(K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) - E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{G}_{t-\delta}\right]\right)dt = \sum_{k=1}^{n} Z_{T,k}(\mathbf{x}),$$

with $T = n\delta$, $T_k = k\delta$ and $Z_{T,k}(\mathbf{x}) = \int_{T_{k-1}}^{T_k}\left(K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) - E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{G}_{t-\delta}\right]\right)dt$. Clearly, the sequence $\{Z_{T,k}(\mathbf{x})\}, k = 1,\ldots,n$, is a sequence of martingale differences with respect to the $\sigma-$fields $\mathcal{G}_{k-1} = \sigma(X_s : 0 \leq s < T_{k-1})$. Observe that

$$\begin{aligned}
|Z_{T,k}(\mathbf{x})| &\leq \int_{T_{k-1}}^{T_k}\left|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) - E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{G}_{t-\delta}\right]\right|dt \\
&\leq 2\delta\sup_{\mathbf{x}\in\mathbb{R}^d}|K_1(\mathbf{x})| = 2\delta\widetilde{K}_1,
\end{aligned}$$

with $\widetilde{K}_1 = \sup_{\mathbf{x}\in\mathbb{R}^d}|K_1(\mathbf{x})|$. Now by Lemma 3.2, for any $\epsilon > 0$, we have

$$\mathbb{P}\left\{\left|\sum_{k=1}^{n} Z_{T,k}(\mathbf{x})\right| > \epsilon(Ta_T^d)\right\} \leq 2\exp\left\{-\frac{\epsilon^2(Ta_T^d)^2}{8n\delta^2\widetilde{K}_1^2}\right\} = 2\exp\left\{-\frac{\epsilon^2 Ta_T^{2d}}{8\delta\widetilde{K}_1^2}\right\}.$$

The right-hand side of the last inequality is the general term of a convergent series. Thus, for sufficiently large T, we conclude by Borel-Cantelli lemma that

$$\sum_{n=1}^{\infty}\mathbb{P}\left\{\left|\sum_{k=1}^{n} Z_{T,k}(\mathbf{x})\right| > \epsilon(Ta_T^d)\right\} < \infty.$$

which means that

$$G_{1,T}(\mathbf{x}) = 0, \quad \text{a.s.} \tag{3.16}$$

The proof is achieved considering the statements (3.15) and (3.16). □

Lemma 3.4. Under assumptions (A.1)(i), (A.1)(iii), (A.4)(i) and (A.4)(iv), we have

$$g_T(\mathbf{x}) - \bar{g}_{T,1}(\mathbf{x}) = O\left(\left(\frac{\log T}{Ta_T^d}\right)^{1/2}\right), \quad \text{a.s.} \tag{3.17}$$

Proof of Lemma 3.4. Proceeding as in Lemma 1.3 in chapter 1 of this thesis we obtain the result. □

Now, to end the proof of Proposition 3.1, it suffices to consider the uniformity over $y \in \mathcal{G}$. This is the object of Lemma 3.5 and Lemma 3.6 given hereafter.

Lemma 3.5. If hypothesis (A.1)(i)-(ii2)-(iii), (A.4)(ii) and (A.4)(iii) are fulfilled, then we have

$$\frac{1}{g_T(\mathbf{x})}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y) - \bar{h}_T(\mathbf{x},y)\right| = O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right), \quad \text{a.s.} \tag{3.18}$$

Proof of Lemma 3.5. Consider a coverage of the compact set $\mathcal{G}_T = [-\upsilon_T, \upsilon_T]$ by a finite number l_T of intervals $\mathcal{G}_{T,k}$ of the form $\mathcal{G}_{T,k} = [s_k - w_T, s_k + w_T]$, where $w_T = \upsilon_T / l_T \to 0$ as $T \to \infty$. Clearly, we have $\mathcal{G}_T = \cup_{k=1}^{l_n} \mathcal{G}_{T,k}$. Therefore,

$$\begin{aligned}
\frac{1}{g_T(\mathbf{x})} \sup_{y\in\mathcal{G}_T} \left| h_T(\mathbf{x},y) - \bar{h}_T(\mathbf{x},y)\right| &\leq \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \sup_{y\in\mathcal{G}_{T,k}} |h_T(\mathbf{x},y) - h_T(\mathbf{x},s_k)| \\
&\quad + \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \left| h_T(\mathbf{x},s_k) - \bar{h}_T(\mathbf{x},s_k)\right| \\
&\quad + \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \sup_{y\in\mathcal{G}_{T,k}} \left| \bar{h}_T(\mathbf{x},s_k) - \bar{h}_T(\mathbf{x},y)\right| \\
&= F_{1,T}(\mathbf{x},y,s_k) + F_{2,T}(\mathbf{x},s_k) + F_{3,T}(\mathbf{x},y,s_k).
\end{aligned}$$

Regarding $F_{1,T}(\mathbf{x},y,s_k)$, using the last part of Assumption (A.1)(ii2), we obtain

$$\begin{aligned}
&F_{1,T}(\mathbf{x},y,s_k) \\
&\leq \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \sup_{y\in\mathcal{G}_{T,k}} \frac{1}{Ta_T^{d+1}} \int_0^T K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) \left| K_2\left(\frac{y-Y_t}{a_T}\right) - K_2\left(\frac{s_k-Y_t}{a_T}\right)\right| dt \\
&\leq \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \sup_{y\in\mathcal{G}_{T,k}} \frac{1}{Ta_T^{d+1}} \int_0^T K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) \frac{C_K}{a_T} |y-s_k| dt \\
&\leq \frac{C_K w_T}{a_T^2} \frac{1}{g_T(\mathbf{x})} \left(\frac{1}{Ta_T^d} \int_0^T K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) dt\right) \\
&= \frac{C_K \upsilon_T}{a_T^2 l_T} = \frac{C_K}{a_T^{2+\mu} l_T}.
\end{aligned}$$

Choosing now $l_T = \lfloor T^\gamma \rfloor$, where γ is a positive constant, for $\epsilon_T = \epsilon_0 \left(\log T / T a_T^{d+1}\right)^{1/2}$, we obtain

$$\epsilon_T^{-1} F_{1,T}(\mathbf{x},y,s_k) = O\left(\left(\frac{1}{T^{2\gamma-1} a_T^{3+2\mu-d} \log T} \right)^{1/2} \right). \tag{3.19}$$

We deal now with $F_{2,T}(\mathbf{x},y,s_k)$. Observe that

$$\begin{aligned}
F_{2,T}(\mathbf{x},s_k) &= \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \left| \frac{1}{Ta_T^{d+1}} \int_0^T \left(K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) K_2\left(\frac{s_k-Y_t}{a_T}\right) \right.\right. \\
&\quad \left.\left. - \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) K_2\left(\frac{s_k-Y_t}{a_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt \right| \\
&= \frac{1}{Ta_T^{d+1}} \frac{1}{g_T(\mathbf{x})} \max_{1\leq k\leq l_T} \left| \sum_{j=1}^n R_{T,j}(\mathbf{x},s_k)\right|,
\end{aligned}$$

where

$$R_{T,j}(\mathbf{x},s_k) = \int_{T_{j-i}}^{T_j} \left(K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) K_2\left(\frac{s_k-Y_t}{a_T}\right) - \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) K_2\left(\frac{s_k-Y_t}{a_T}\right) |\mathcal{F}_{t-\delta}\right]\right) dt.$$

We observe that the sequence $\left\{R_{T,j}(\mathbf{x},s_k)\right\}_{0\leq j\leq n}$ is a sequence of martingale differences adapted to the filtration $\mathcal{F}_j = \sigma((X_s,Y_s) : 0 \leq s < T_j)$. Making use of Jensen and Minkowski's inequalities with $p \geq 2$, we obtain

$$\begin{aligned}
\left|\mathbb{E}\left[R^p_{T,j}(\mathbf{x}, s_k) \mid \mathcal{F}_{j-2}\right]\right| &= \Bigg|\mathbb{E}\Bigg[\Bigg(\int_{T_{j-i}}^{T_j}\left(K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{s_k-Y_t}{a_T}\right)\right. \\
&\quad \left.\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{s_k-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\right)dt\Bigg)^p \mid \mathcal{F}_{j-2}\Bigg]\Bigg| \\
&\leq \int_{T_{j-i}}^{T_j}\mathbb{E}\Bigg[\Bigg|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{s_k-Y_t}{a_T}\right) \\
&\quad \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{s_k-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\Bigg|^p \mid \mathcal{F}_{j-2}\Bigg]dt \\
&\leq \int_{T_{j-i}}^{T_j}\Bigg(\mathbb{E}\left[K_1^p\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^p\left(\frac{s_k-Y_t}{a_T}\right) \mid \mathcal{F}_{j-2}\right]^{1/p} \\
&\quad -\mathbb{E}\left[\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{s_k-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]^p \mid \mathcal{F}_{j-2}\right]^{1/p}\Bigg)^p dt \\
&\leq \int_{T_{j-i}}^{T_j}\left(2\mathbb{E}\left[K_1^p\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^p\left(\frac{s_k-Y_t}{a_T}\right) \mid \mathcal{F}_{j-2}\right]^{1/p}\right)^p dt \\
&= 2^p\int_{T_{j-i}}^{T_j}\mathbb{E}\left[K_1^p\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^p\left(\frac{s_k-Y_t}{a_T}\right) \mid \mathcal{F}_{j-2}\right]dt. \qquad (3.20)
\end{aligned}$$

Under assumptions (A.1)(i) and (A.4)(ii), we have

$$\begin{aligned}
&\mathbb{E}\left[K_1^p\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^p\left(\frac{s_k-Y_t}{a_T}\right) \mid \mathcal{F}_{j-2}\right] \\
&\quad= \int_{\mathbb{R}^{d+1}}K_1^p\left(\frac{\mathbf{x}-\mathbf{v}}{a_T}\right)K_2^p\left(\frac{s_k-u}{a_T}\right)h_t^{\mathcal{F}_{j-2}}(\mathbf{v},u)d\mathbf{v}du \\
&\quad= a_T^{d+1}\int_{\mathbb{R}^{d+1}}K_1^p(\mathbf{w})K_2^p(z)h_t^{\mathcal{F}_{j-2}}(\mathbf{x}-a_T\mathbf{w}, y-a_Tz)d\mathbf{w}dz \\
&\quad\leq a_T^{d+1}\left\|K_1\right\|^p\left\|K_2\right\|^p. \qquad (3.21)
\end{aligned}$$

Noting that the statement (3.20) may be rewritten using the statement (3.21), we obtain

$$\mathbb{E}\left[R^p_{T,j}(\mathbf{x}, s_k) \mid \mathcal{F}_{j-2}\right] \;\leq\; 2^p\delta a_T^{d+1}\left\|K_1\right\|^p\left\|K_2\right\|^p \leq p!C^{p-2}d_j^2,$$

where $C = 2\left\|K_1\right\|\left\|K_2\right\|$ and $d_j^2 = 2\delta a_T^{d+1}\|K_1\|\|K_2\|$. Denote $T_n = T = n\delta$, $a_{T_n} = a_T$, $l_{T_n} = l_T$, and $\epsilon_{T_n} = \epsilon_T = \epsilon_0\left(\log T_n/T_na_{T_n}^{d+1}\right)^{1/2}$. Therefore,

$$D_n = \sum_{j=1}^n d_j^2 = \sum_{j=1}^n 2\delta a_T^{d+1}\|K_1\|\|K_2\| = O(T_na_{T_n}^{d+1}).$$

Making use of Lemma 3.1, for T_n large enough, we obtain

$$\begin{aligned}
\mathbb{P}\left\{\max_{1\le k\le l_{T_n}}\left|\sum_{j=1}^{n}R_{T,j}(\mathbf{x},s_k)\right|>\epsilon_{T_n}(T_na_{T_n}^{d+1})\right\} &\le \sum_{k=1}^{l_{T_n}}\mathbb{P}\left\{\left|\sum_{j=1}^{n}R_{T,j}(\mathbf{x},s_k)\right|>\epsilon_{T_n}(T_na_{T_n}^{d+1})\right\}\\
&\le 2l_{T_n}\exp\left\{-\frac{\epsilon_{T_n}^2(T_na_{T_n}^{d+1})^2}{82(D_n+C(T_na_{T_n}^{d+1})\epsilon_{T_n})}\right\}\\
&\le 2T_n^{\gamma}\exp\left\{-\frac{\epsilon_0^2(T_na_{T_n}^{d+1})^2\log T_n/T_na_{T_n}^{d+1}}{O(T_na_{T_n}^{d+1})+2C(T_na_{T_n}^{d+1})\epsilon_0\left(\frac{\log T_n}{T_na_{T_n}^{d+1}}\right)^{1/2}}\right\}\\
&:= 2T_n^{\gamma}\exp\left\{\log T_n^{-\epsilon_0^2C_1}\right\}=2T_n^{\gamma-\epsilon_0^2C_1}.
\end{aligned}$$

Taking ϵ_0 large enough so to have $\gamma-\epsilon_0^2C_1<0$, where C_1 is a positive constant, The right-hand side of the previous inequality, is the general term of convergent series. Hence, Borel-Cantelli Lemma, shows that

$$\sum_{n=1}^{\infty}\mathbb{P}\left\{\max_{1\le k\le l_{T_n}}\left|\sum_{j=1}^{n}R_{T,j}(\mathbf{x},s_k)\right|>\epsilon_{T_n}(T_na_{T_n}^{d+1})^{-1}\right\}<\infty.$$

Hence,

$$\frac{1}{T_na_{T_n}^{d+1}}\max_{1\le k\le l_{T_n}}\left|\sum_{j=1}^{n}R_{T,j}(\mathbf{x},s_k)\right|=O\left(\left(\frac{\log T_n}{T_na_{T_n}^{d+1}}\right)^{1/2}\right),\quad \text{a.s.}$$

It follows that

$$F_{2,T}=O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right),\quad \text{a.s.} \tag{3.22}$$

Finally, using the same idea as for $F_{1,T}(\mathbf{x},y,s_k)$ together with Assumption (A.1)(ii2), we obtain

$$\begin{aligned}
&F_{3,T}(\mathbf{x},y,s_k)\\
&\le \frac{1}{g_T(\mathbf{x})}\max_{1\le k\le l_T}\sup_{y\in\mathcal{G}_{T,k}}\frac{1}{Ta_T^{d+1}}\int_0^T\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left|K_2\left(\frac{y-Y_t}{a_T}\right)-K_2\left(\frac{s_k-Y_t}{a_T}\right)\right||\mathcal{F}_{t-\delta}\right]dt\\
&\le \frac{1}{g_T(\mathbf{x})}\max_{1\le k\le l_T}\sup_{y\in\mathcal{G}_k}\frac{1}{Ta_T^{d+1}}\int_0^T\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\frac{C_K}{a_T}|y-s_k||\mathcal{F}_{t-\delta}\right]dt\\
&\le \frac{C_Kw_T}{a_T^2}\frac{1}{g_T(\mathbf{x})}\left(\frac{1}{Ta_T^{d}}\int_0^T\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]dt\right).
\end{aligned}$$

We conclude from Lemmas 3.3 and 3.4 that

$$\epsilon_T^{-1}F_{3,T}(\mathbf{x},y,s_k) = O\left(\left(\frac{1}{T^{2\gamma-1}a_T^{3+2\mu-d}\log T}\right)^{1/2}\right). \tag{3.23}$$

In conjunction with Equation (3.19), (3.22) and (3.23) and lemma 3.3 this concludes that

$$\frac{1}{g_T(\mathbf{x})}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y)-\bar{h}_T(\mathbf{x},y)\right| = O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right), \quad \text{a.s.} \tag{3.24}$$

The lemma is proved. □

Lemma 3.6. Let assumptions (A.1)(i), (A.1)(iv), (A.6) and (A.8) be satisfied, we have

$$B_T(\mathbf{x},y) = O(a_T^{\beta}+a_T^{\nu}). \tag{3.25}$$

Proof of lemma 3.6 Let $K_{1,a}(\cdot) = \frac{1}{a_T^d}K_1\left(\frac{\cdot}{a_T}\right)$ and $K_{2,a}(\cdot) = \frac{1}{a_T}K_2\left(\frac{\cdot}{a_T}\right)$. Under assumption (A.6), we have

$$\begin{aligned}
B_T(\mathbf{x},y) &= \frac{\bar{h}_T(\mathbf{x},y)-\bar{g}_{T,1}(\mathbf{x})h(\mathbf{x},y)/g(\mathbf{x})}{\bar{g}_{T,1}(\mathbf{x})}\\
&= \frac{1}{\bar{g}_{T,1}(\mathbf{x})}\frac{1}{T}\int_0^T \mathbb{E}\left[\left(K_{2,a}(y-Y_t)-\frac{h(\mathbf{x},y)}{g(\mathbf{x})}\right)K_{1,a}(\mathbf{x}-\mathbf{X}_t)\mid \mathcal{F}_{t-\delta}\right]dt\\
&= \frac{1}{\bar{g}_{T,1}(\mathbf{x})}\frac{1}{T}\int_0^T \mathbb{E}\left[K_{1,a}(\mathbf{x}-\mathbf{X}_t)\mathbb{E}\left[\left(K_{2,a}(y-Y_t)-\frac{h(\mathbf{x},y)}{g(\mathbf{x})}\right)\mid \mathcal{S}_{t-\delta,\delta}\right]\mid \mathcal{F}_{t-\delta}\right]dt\\
&= \frac{1}{\bar{g}_{T,1}(\mathbf{x})}\frac{1}{T}\int_0^T \mathbb{E}\left[K_{1,a}(\mathbf{x}-\mathbf{X}_t)\left(\frac{1}{a_T}\mathbb{E}\left[K_2\left(\frac{y-Y_t}{a_T}\right)\mid \mathbf{X}_t\right]-f(y\mid\mathbf{x})\right)\mid \mathcal{F}_{t-\delta}\right]dt.
\end{aligned}$$

Therefore, by assumption (A.8), we obtain

$$\begin{aligned}
&|B_T(\mathbf{x},y)|\\
&= \left|\frac{1}{\bar{g}_{T,1}(\mathbf{x})}\frac{1}{T}\int_0^T \mathbb{E}\left[K_{1,a}(\mathbf{x}-\mathbf{X}_t)\int_{\mathbb{R}}K_2(u)\left(f(y-a_Tu\mid\mathbf{X}_t)-f(y\mid\mathbf{x})\right)du\mid\mathcal{F}_{t-\delta}\right]dt\right|\\
&\le \frac{1}{\left|\bar{g}_{T,1}(\mathbf{x})\right|}\frac{1}{T}\int_0^T \mathbb{E}\left[K_{1,a}(\mathbf{x}-\mathbf{X}_t)\int_{\mathbb{R}}K_2(u)\left|f(y-a_Tu\mid\mathbf{X}_t)-f(y\mid\mathbf{x})\right|du\mid\mathcal{F}_{t-\delta}\right]dt\\
&\le \frac{C_{\mathbf{x}}}{\left|\bar{g}_{T,1}(\mathbf{x})\right|}\frac{1}{T}\int_0^T \mathbb{E}\left[K_{1,a}(\mathbf{x}-\mathbf{X}_t)\int_{\mathbb{R}}K_2(u)\left(|a_Tu|^{\nu}+\|\mathbf{X}_t-\mathbf{x}\|^{\beta}\right)du\mid\mathcal{F}_{t-\delta}\right]dt.
\end{aligned}$$

According to assumption (A.1)(i) the kernel K_1 has a support compact, then $\|\mathbf{X}_t-\mathbf{x}\|\le a_TC_{K_1}$, together with (A.1)(iv), we have

$$\begin{aligned}
|B_T(\mathbf{x},y)| &= C_{\mathbf{x}}(a_T^{\beta}C_{K_1}^{\beta}+a_T^{\nu})\frac{1}{\left|\bar{g}_{T,1}(\mathbf{x})\right|}\frac{1}{Ta_T^d}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mid\mathcal{F}_{t-\delta}\right]dt\\
&= O(a_T^{\beta}+a_T^{\nu}).
\end{aligned}$$

□

Combining Lemmas 3.3, 3.4, 3.5 and 3.6 we achieve the proof of Proposition 3.1. □

Lemma 3.7. Under the same assumptions of Proposition 3.1, and if the function Θ fulfils

$$\begin{aligned} &\forall \epsilon > 0, \quad \exists \xi > 0, \quad \forall \mu(\mathbf{x}), \\ &|\Theta(\mathbf{x}) - \mu(\mathbf{x})| \geq \epsilon \Rightarrow |f(\Theta(\mathbf{x})|\mathbf{x}) - f(\mu(\mathbf{x}) \mid \mathbf{x})| \geq \xi, \end{aligned} \tag{3.26}$$

then

$$\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}) \xrightarrow{a.s} 0, \quad \text{as} \quad T \to \infty. \tag{3.27}$$

Proof of Lemma 3.7. Combining conditions (3.26) and (3.7) one obtain that for any fixed $\mathbf{x} \in \mathcal{C}$ all $\epsilon > 0$, there exists a $\xi > 0$ such that

$$P\left\{|\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})| \geq \epsilon\right\} \leq P\left\{\sup_{y \in \mathcal{G}_T} |f_T(y/\mathbf{x}) - f(y/\mathbf{x})| \geq \xi\right\}. \tag{3.28}$$

The uniqueness hypothesis of the conditional mode function gives us the result provided we prove that the right-hand side of Equation (3.28) converges almost surely to zero. Proposition 3.1 completes the proof. □

Proof of Theorem 3.1

Under assumtion (A.7)(i), using a second order of Taylor expansion of the function $f(\Theta_T(\mathbf{x}) \mid \mathbf{x})$ around $\Theta(\mathbf{x})$, we obtain

$$f(\Theta_T(\mathbf{x}) \mid \mathbf{x}) = f(\Theta(\mathbf{x}) \mid \mathbf{x}) + (\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}))^2 \frac{1}{2} f^{(2)}(\Theta_T^\star(\mathbf{x}) \mid \mathbf{x}), \tag{3.29}$$

where $\Theta_T^\star(\mathbf{x})$ is between $\Theta_T(\mathbf{x})$ and $\Theta(\mathbf{x})$. Taking into account the statements (3.7) and (3.29), it follows that

$$(\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}))^2 \left|f^{(2)}(\Theta_T^\star(\mathbf{x}) \mid \mathbf{x})\right| = O\left(\sup_{y \in \mathcal{G}_T} |f_T(y \mid \mathbf{x}) - f(y \mid \mathbf{x})|\right).$$

Combining Lemma 3.7 and condition (A.7)(ii), we have

$$\lim_{T \to \infty} \left|f^{(2)}(\Theta_T^\star(\mathbf{x}) \mid \mathbf{x})\right| = \left|f^{(2)}(\Theta(\mathbf{x}) \mid \mathbf{x})\right| = \Phi(\mathbf{x}) \neq 0.$$

Consequently,

$$(\Theta_T(\mathbf{x}) - \Theta(\mathbf{x}))^2 = O\left(\sup_{y \in \mathcal{G}_T} |f_T(y \mid \mathbf{x}) - f(y \mid \mathbf{x})|\right),$$

which is enough, while considering Proposition 3.1, to complete the proof. □

3.4.2 Proof of Theorem 3.2

Proposition 3.2. Under hypotheses (A.1)(i)-(ii), (A.2), (A.3), (A.4)(i)-(ii), (A.4)(iv), (A.5), (A.6) and (A.8) together with condition (3.6), we have

$$\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|f_T(y/\mathbf{x})-f(y/\mathbf{x})| \overset{a.s}{=} O(a_T^\beta+a_T^\nu)+O\left(\left(\frac{\log T}{Ta_T^d}\right)^{1/2}\right)+O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right).$$

Proof of Proposition 3.2. Conditions (A.2) and (A.3) imply that

$$\begin{aligned}
&\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|f_T(y/\mathbf{x})-f(y/\mathbf{x})| \\
&= \sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}\left|B_T(\mathbf{x},y)+\frac{Q_T(\mathbf{x},y)+R_T(\mathbf{x},y)}{g_T(\mathbf{x})}\right| \\
&\leq \sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|+\frac{1}{\inf\limits_{\mathbf{x}\in\mathcal{C}}g_T(\mathbf{x})}\left(\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|Q_T(\mathbf{x},y)|+\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|R_T(\mathbf{x},y)|\right) \\
&\leq \sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|+\frac{1}{\inf\limits_{\mathbf{x}\in\mathcal{C}}g_T(\mathbf{x})}\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y)-\bar{h}_T(\mathbf{x},y)\right| \\
&\quad +\frac{1}{\inf\limits_{\mathbf{x}\in\mathcal{C}}g_T(\mathbf{x})}\left(\lambda^{-1}\Gamma_1+\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x},y)|\right)\sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x})-\bar{g}_{T,1}(\mathbf{x})|. \qquad (3.30)
\end{aligned}$$

The following lemma deals with the second term of the decomposition (3.30)

Lemma 3.8. Under assumptions (A.1)(i)-(ii), (A.4)(ii) and the condition (3.6)(ii), we have

$$\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y)-\bar{h}_T(\mathbf{x},y)\right| \overset{a.s.}{=} O\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right), \quad \text{as} \quad T\to\infty. \qquad (3.31)$$

Proof of Lemma 3.8. Our first concern is to give an upper bound for the main term in (3.31). Towards this end, we make use of Lemma 3.2 and gather the needed elements to apply the Borel-Cantelli Lemma. Let

$$\Psi_j(\mathbf{x},y)=\int_{T_{j-1}}^{T_j}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)-E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\right]dt$$

where $j=1,\ldots,n$. Observe, for all pairs $(\mathbf{x},y)$ and all T fixe, that the sequence $(\Psi_j(\mathbf{x},y))_{1\leq j\leq n}$ is a martingale difference with respect to the $\sigma-$fields $(\mathcal{F}_j)_{1\leq j\leq n}$.

For any $p \geq 2$, we have

$$\begin{aligned}
\mathbb{E}\left[|\Psi_j(\mathbf{x},y)|^p \Big| \mathcal{F}_{i-2}\right] &= \int_{T_{j-1}}^{T_j} \mathbb{E}\left[\Big|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)\right.\\
&\quad \left. -E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\Big|^p \Big|\mathcal{F}_{i-2}\right]dt\\
&\leq \int_{T_{j-1}}^{T_j}\left(\mathbb{E}\left[K_1^P\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^P\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{j-2}\right]^{1/p}\right.\\
&\quad \left.+\mathbb{E}\left[\mathbb{E}\left[K_1^P\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^P\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]|\mathcal{F}_{j-2}\right]^{1/p}\right)^p dt\\
&\leq \int_{T_{j-1}}^{T_j}\left(2\mathbb{E}\left[K_1^P\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^P\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{j-2}\right]^{1/p}\right)^p dt\\
&= 2^p\int_{T_{j-1}}^{T_j}\mathbb{E}\left[K_1^P\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^P\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{j-2}\right]dt. \qquad (3.32)
\end{aligned}$$

Consider now the change of variable $\mathbf{w} = (\mathbf{x}-\mathbf{u})/a_T$ and $z = (y-v)/a_T$. Under assumptions (A.1)(i) and (A.4)(ii), we may write

$$\begin{aligned}
&\mathbb{E}\left[K_1^P\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^P\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{j-2}\right]\\
&= \int_{\mathbb{R}^{d+1}} K_1^P\left(\frac{\mathbf{x}-\mathbf{u}}{a_T}\right)K_2^P\left(\frac{y-v}{a_T}\right)h_t^{\mathcal{F}_{j-2}}(\mathbf{u},v)d\mathbf{u}dv\\
&= a_T^{d+1}\int_{\mathbb{R}^{d+1}} K_1^P(\mathbf{w})K_2^P(z)h_t^{\mathcal{F}_{j-2}}(\mathbf{x}-a_T\mathbf{w}, y-a_Tz)d\mathbf{w}dz\\
&\leq a_T^{d+1}\|K_1\|_\infty^p\|K_2\|_\infty^p\int_{\mathbb{R}^{d+1}} h_t^{\mathcal{F}_{j-2}}(\mathbf{x}-a_T\mathbf{w}, y-a_Tz)d\mathbf{w}dz\\
&\leq a_T^{d+1}\|K_1\|_\infty^p\|K_2\|_\infty^p. \qquad (3.33)
\end{aligned}$$

Therefore, using the statement (3.33), the statement (3.32) may be rewritten so that

$$\mathbb{E}\left[|\Psi_j(\mathbf{x},y)|^p\Big|\mathcal{F}_{i-2}\right] \leq 2^p\delta a_T^{d+1}\|K_1\|_\infty^p\|K_2\|_\infty^p \leq p!C^{p-1}d_j^2,$$

where $C = 2\|K_1\|_\infty^{p-1}\|K_2\|_\infty^{p-1}$ and $d_j^2 = \delta a_T^{d+1}\|K_1\|_\infty\|K_2\|_\infty$. Define $T_n = T = n\delta$ and $a_{T_n} = a_T$. Hence

$$D_n = \sum_{j=1}^n d_j^2 = n\delta a_T^{d+1}\|K_1\|_\infty\|K_2\|_\infty = O(T_n a_{T_n}^{d+1}).$$

We are in position to apply lemma 3.1 to the martingale differences sum. For n

sufficiently large and any $\epsilon_{T_n} = \epsilon_T = \epsilon_0 \left(\log T_n / T_n a_{T_n}^{d+1}\right)$, where $\epsilon_0 > 0$, we have

$$\begin{aligned}
\mathbb{P}\left(|\sum_{j=1}^{n} \Psi_j(\mathbf{x}, y)| > (T_n a_{T_n}^{d+1})\epsilon_{T_n} \right) &\leq 2\exp\left\{ -\frac{(T_n a_{T_n}^{d+1})^2 \epsilon_{T_n}^2}{2(D_n + C(T_n a_{T_n}^{d+1})\epsilon_{T_n})} \right\} \\
&= 2\exp\left\{ -\frac{(T_n a_{T_n}^{d+1})^2 \epsilon_0^2 \log T_n / T_n a_{T_n}^{d+1}}{O(T_n a_{T_n}^{d+1})(1 + C\epsilon_0 \sqrt{\frac{\log T_n}{T_n a_{T_n}^{d+1}}})} \right\} \\
&= 2\exp\left\{ -\epsilon_0^2 C_1 \log T_n \right\} = \frac{2}{T_n^{\epsilon_0^2 C_1}}, \qquad (3.34)
\end{aligned}$$

where C_1 is a positive constant. Observe, for ϵ_0 sufficiently large, that the right-hand side of the last inequality is the general term of a convergent series. A simple use of Borel-Cantelli Lemma shows that

$$\left(\sum_{j=1}^{n} \Psi_j(\mathbf{x}, y) \right) \overset{a.s.}{\to} 0 \quad \text{as} \quad T \to \infty.$$

Therefore,

$$\left| h_T(\mathbf{x}, y) - \bar{h}_T(\mathbf{x}, y) \right| = O\left(\left(\frac{\log T}{T a_T^{d+1}} \right)^{1/2} \right), \quad \text{a.s.} \qquad (3.35)$$

In what is coming next, we will investigate the uniform convergence of $\left((T a_T^{d+1})^{-1} \int_0^T \Psi_t(\mathbf{x}, y) dt \right)$, where $\Psi_t(\mathbf{x}, y) = \varphi_t(\mathbf{x}, y) - E\left[\varphi_t(\mathbf{x}, y) \mid \mathcal{F}_{t-\delta}\right]$. Recalling that $\mathcal{G}_T = [-\upsilon_T, \upsilon_T] = [-a_T^{-\mu}, a_T^{-\mu}]$, where μ is a positive constant, we may write

$$\begin{aligned}
\sup_{\mathbf{x}\in\mathcal{C}} \sup_{y\in\mathcal{G}_T} \left| (T a_T^{d+1})^{-1} \int_0^T \Psi_t(\mathbf{x}, y) dt \right| &= \sup_{\mathbf{x}\in\mathcal{C}} \sup_{|y|\leq a_T^{-\mu}} \left| (T a_T^{d+1})^{-1} \int_0^T \Psi_t(\mathbf{x}, y) dt \right| \\
&= U_T. \qquad (3.36)
\end{aligned}$$

where $a_T^{-\mu} \to \infty$, as $T \to \infty$. We shall show then that $U_T \underset{T\to\infty}{\overset{a.s.}{\to}} 0$. Let η be a positive constant and consider a cover of the compact set $\mathcal{C} \times \mathcal{G}_T$ of $\mathbb{R}^{d+1}$ by a finite number of spheres $B_j = B((\mathbf{x}_j, y_j), \frac{a_T^{-\mu}}{\nu_T})$ of radius a_T^{η}, where $j = 1, \ldots, \nu_T$. Whenever $(\mathbf{x}, y) \in \mathcal{C} \times \mathcal{G}_T$, there exists j and a point $(\mathbf{x}_j, y_j)$ such that $(x, y) \in B_j$. Observe that

$$|U_T(\mathbf{x}, y)| \leq |U_T(\mathbf{x}, y) - U_T(\mathbf{x}_j, y_j)| + |U_T(\mathbf{x}_j, y_j)| = |\tilde{U}_T(\mathbf{x}, y)| + |U_T(\mathbf{x}_j, y_j)|.$$

Subsequently, we have

$$\sup_{\mathbf{x}\in\mathcal{C}} \sup_{(\mathbf{x},y)\in B_j} |U_T(\mathbf{x}, y)| \leq \max_{1\leq j\leq\nu_T} \sup_{(\mathbf{x},y)\in B_j} |\tilde{U}_T(\mathbf{x}, y)| + \max_{1\leq j\leq\nu_T} |U_T(x_j, y_j)|, \qquad (3.37)$$

where

$$
\begin{aligned}
&|\tilde{U}_T(\mathbf{x},y)| \\
&\leq\ (Ta_T^{d+1})^{-1}\int_0^T \left|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)-E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\right. \\
&-\ \left.K_1\left(\frac{\mathbf{x}_j-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y_j-Y_t}{a_T}\right)-E\left[K_1\left(\frac{\mathbf{x}_j-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y_j-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right]\right|dt.
\end{aligned}
$$

Adding and then subtracting the term

$$
K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y_j-Y_t}{a_T}\right)-E\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2\left(\frac{y_j-Y_t}{a_T}\right)|\mathcal{F}_{t-\delta}\right],
$$

and under assumption (A.1)(ii), we obtain

$$
\begin{aligned}
|\tilde{U}_T(\mathbf{x},y)| &\leq (Ta_T^{d+1})^{-1}2\tilde{K}C_K a_T^{-\gamma}(|\mathbf{x}-\mathbf{x}_j|^\gamma+|y-y_j|^\gamma)\int_0^T dt \\
&= 2C_K a_T^{-\gamma-d-1}\tilde{K}(|\mathbf{x}-\mathbf{x}_j|^\gamma+|y-y_j|^\gamma),
\end{aligned}
$$

where C_K is a Lipschitz constant and $\widetilde{K}=\max\{\sup_{\mathbf{x}\in\mathcal{C}}|K_1(\mathbf{x})|,\sup_{y\in\mathcal{G}_T}|K_2(y)|\}$. For any $(\mathbf{x},y)\in\mathcal{C}\times\mathcal{G}_T$, we have

$$
|\mathbf{x}-\mathbf{x}_j|^\gamma\leq a_T^{-\gamma\mu}\nu_T^{-\gamma}\quad\text{and}\quad |y-y_j|^\gamma\leq a_T^{-\gamma\mu}\nu_T^{-\gamma}.
$$

Therefore, for $\epsilon_0>0$, take $\epsilon_T=\epsilon_0\left(\log T/Ta_T^{d+1}\right)^{1/2}$, we have

$$
\begin{aligned}
\epsilon_T^{-1}\max_{1\leq j\leq\nu_T}\sup_{(\mathbf{x},y)\in B_j}|\tilde{U}_T(\mathbf{x},y)| &\leq C\left(\frac{Ta_T^{d+1}}{\log T}\right)^{1/2}a_T^{-(d+1)-\gamma(\mu+1)}\nu_T^{-\gamma} \\
&\leq C\left(\frac{Ta_T^{d+1}}{\log T}\right)^{1/2}\frac{1}{a_T^{d+1+\gamma(\mu+1)}\nu_T^{\gamma}}
\end{aligned}
$$

Taking $\nu_T=T/a_T^{1+\mu}$, we obtain

$$
\epsilon_T^{-1}\max_{1\leq j\leq\nu_T}\sup_{(\mathbf{x},y)\in B_j}|\tilde{U}_T(\mathbf{x},y)|\leq\frac{1}{(\log T)^{1/2}T^{\gamma-1}(Ta_T^{d+1})^{1/2}}
$$

Taking $\gamma>1$, we obtain

$$
\max_{1\leq j\leq\nu_T}\sup_{(\mathbf{x},y)\in B_j}|\tilde{U}_T(\mathbf{x},y)|=o\left(\left(\frac{\log T}{Ta_T^{d+1}}\right)^{1/2}\right). \tag{3.38}
$$

Recall that

$$
\max_{1\leq j\leq\nu_T}|U_T(\mathbf{x}_j,y_j)|=\max_{1\leq j\leq\nu_T}\left|h_T(\mathbf{x}_j,y_j)-\bar{h}_T(\mathbf{x}_j,y_j)\right|.
$$

Denote now $T_n=T=\delta n$, $a_{T_n}=a_T$, $\nu_{T_n}=\nu_T$ and $\epsilon_{T_n}=\epsilon_T$. The proof is similar to the one given above to set the statement (3.34). Details of the calculation are then omitted. We have

$$\begin{aligned}
P\left\{\max_{1\le j\le \nu_{T_n}} |U_T(\mathbf{x}_j,y_j)| > \epsilon_{T_n}\right\} &\le \sum_{j=1}^{\nu_{T_n}} P\left\{\left|h_T(\mathbf{x}_j,y_j)-\bar{h}_T(\mathbf{x}_j,y_j)\right| > \epsilon_{T_n}\right\}\\
&\le 2\nu_{T_n} {T_n}^{-C_1\epsilon_0^2}\\
&\le 2\frac{T_n^{1-C_1\epsilon_0^2}}{a_{T_n}^{\mu+1}}\\
&\le \frac{2}{(T_n a_{T_n}^{d+1})^{\frac{\mu+1}{d+1}} T_n^{\epsilon_0^2 C_1-1-(\mu+1)/(d+1)}}.
\end{aligned}$$

where C_1 is a positive constant. Taking ϵ_0 sufficiently large to have $C_1\epsilon_0^2-\frac{\mu+1}{d+1}-1>0$ and assuming the condition (3.6)(ii), we conclude that, the right-hand side in the last inequality converges to 0. Borel-Cantelli Lemma shows then that

$$\max_{1\le j\le \nu_T} |U_T(\mathbf{x}_j,y_j)| = O\left(\left(\frac{\log T}{T a_T^{d+1}}\right)^{1/2}\right), \quad \text{a.s.} \tag{3.39}$$

Combining statements (3.38) with (3.39), we obtain

$$\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}\left|h_T(\mathbf{x},y)-\bar{h}_T(\mathbf{x},y)\right| = O\left(\left(\frac{\log T}{T a_T^{d+1}}\right)^{1/2}\right), \quad \text{a.s.} \tag{3.40}$$

The proof is then achieved. □

Now, we are in position to establish the following lemma dealing with the limiting behaviour of the last term in the decomposition 3.30).

Lemma 3.9. Let assumptions (A.1)(i)-(ii1), (A.4)(i) and (A.4)(iv) be satisfied. Then, we have

$$\sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x})-\bar{g}_{T,1}(\mathbf{x})| = O\left(\left(\frac{\log T}{T a_T^{d}}\right)^{1/2}\right). \tag{3.41}$$

Proof Lemma 3.9. The proof is analogous to that of Lemma 3.4. For the uniform convergence case, we refer to Theorem 1.2 in chapter 1 of this manuscript. We refer also to Theorem 2 of Didi & Louani (2013a)[50].

□

Lemma 3.10. Under assumptions (A.1)(i), (A.1)(ii1), (A.1)(iii), (A.3), (A.4)(i) and (A.5), we have

$$\inf_{x\in\mathcal{C}}|g_T(\mathbf{x})| \ge \lambda.$$

Proof Lemma 3.10 Observe that

$$\inf_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x})| \geq \inf_{\mathbf{x}\in\mathcal{C}}|g(\mathbf{x})| - \sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x}) - g(\mathbf{x})|. \tag{3.42}$$

We aim to show that,

$$\sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x}) - g(\mathbf{x})| \to 0, \quad \text{as} \quad T \to \infty. \tag{3.43}$$

Recall that

$$\sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x}) - g(\mathbf{x})| \leq \sup_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x}) - \bar{g}_T(\mathbf{x})| + \sup_{\mathbf{x}\in\mathcal{C}}|\bar{g}_T(\mathbf{x}) - g(\mathbf{x})| = G_1 + G_2, \tag{3.44}$$

wehere $\bar{g}_T(\mathbf{x}) = \frac{1}{Ta_T^d}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right) \mid \mathcal{G}_{t-\delta}\right] dt$. Under hypotheses (A.1)(i), (A.1)(ii1) and (A.4)(i) and following Lemma 3.9 step by step, the first term is vanishing

$$G_1 \xrightarrow{a.s} 0, \quad \text{as} \quad T \to \infty. \tag{3.45}$$

Finally, under assumptions (A.1)(iii), (A.4)(i) and (A.5) and by the same arguments as for $G_{T.2}$ in Lemma 3.3, we have

$$G_2 \xrightarrow{a.s} 0, \quad \text{as} \quad T \to \infty. \tag{3.46}$$

We conclude the proof of (3.43) from statements (3.45) and (3.46). Noting that, by combing statements (3.42) and (3.43) with assumption (A.3) we obtain

$$\inf_{\mathbf{x}\in\mathcal{C}}|g_T(\mathbf{x})| \geq \lambda,$$

□

The next lemma deals with the uniform almost sure convergence of the bias term $B_T(\mathbf{x}, y)$ defined by statement (3.10)

Lemma 3.11. Under assumptions (A.1)(i), (A.1)(iv), (A.6) and (A.8), we have

$$\sup_{\mathbf{x}\in\mathcal{C}}\sup_{y\in\mathcal{G}_T}|B_T(\mathbf{x}, y)| = O\Big(a_T^{\beta} + a_T^{\nu}\Big). \tag{3.47}$$

Proof of Lemma 3.11 Since $\mathcal{C}$ is a compact set, the condition (A.8) holds with a constant C that does not depend on $\mathbf{x}$. Then, proceeding as in the proof of Lemma 3.6, the result follows. □

It suffices to make use of the decomposition (3.30) and lemmas 3.10, 3.8, 3.9 and 3.11 to achieve the proof of Proposition 3.2. □

Lemma 3.12. Under the hypotheses of Proposition 3.2. In addition, if the function Θ is such that

$$\begin{aligned} &\forall \epsilon > 0, \quad \exists \beta > 0, \quad \forall \xi : \mathcal{C} \to \mathbb{R} : \\ &\sup_{\mathbf{x}\in\mathcal{C}} |\Theta(\mathbf{x}) - \xi(\mathbf{x})| \geq \epsilon \Rightarrow \sup_{x\in\mathcal{C}} |f_T(\Theta(\mathbf{x})/\mathbf{x}) - f(\xi(\mathbf{x})/\mathbf{x})| \geq \beta, \end{aligned} \tag{3.48}$$

then

$$\sup_{\mathbf{x}\in\mathcal{C}} |\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})| \overset{a.s.}{\to} 0, \quad T \to \infty. \tag{3.49}$$

Proof of Lemma 3.12. The statements (3.7) and (3.48) imply that, for all $\epsilon > 0$, there exists a constant $\beta > 0$ such that

$$P\left\{\sup_{x\in\mathcal{C}} |\Theta_T(x) - \Theta(x)| > \epsilon\right\} \leq P\left\{\sup_{x\in\mathcal{C}} \sup_{y\in\mathcal{G}_T} |f_T(y/x) - f(y/x)| > \beta\right\}. \tag{3.50}$$

According to proposition 3.2 the right side of equation (3.50) converges almost surely to zero, the uniqueness hypothesis of the conditional mode gives us the desired result. □

Proof of Theorem 3.2

By a second order of Taylor expansion of $f(\Theta_T(\mathbf{x}) \mid \mathbf{x})$ around $\Theta(\mathbf{x})$, given by the statement (3.29), combining with Lemma 3.12 and condition (A.7)(ii), we obtain

$$\sup_{\mathbf{x}\in\mathcal{C}} |\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})|^2 = O\left(\sup_{\mathbf{x}\in\mathcal{C}} \sup_{y\in\mathcal{G}_T} |f_T(y \mid \mathbf{x}) - f(y \mid \mathbf{x})|\right), \tag{3.51}$$

the result follows form considering Proposition 3.2. □

Proof of Corollary 3.1. Observe that

$$|\Theta_T(\hat{\xi}_T) - \Theta(\hat{\xi}_T)| \mathbb{1}_{\{\hat{\xi}_T \in \mathcal{C}\}} \leq \sup_{\mathbf{x}\in\mathcal{C}} |\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})|.$$

Hence, using Theorem 3.2 with $d = 1$, we obtain the desired result. □

Annexe : Quelques outils de probabilités

Notation et défintion

Soit $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$ un espace probabilisé filtré.

Définition 1 (Processus en temps continu)**.** Soit $X = (X_t, t \geq 0)$ un processus défini sur $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$. Le processus X est dit mesurable si l'application

$$\begin{array}{rcl} X : [0, \infty[\times \Omega & \to & (E, \mathcal{E}) \\ (t, \omega) & \mapsto & X_t(\omega) \end{array}$$

est mesurable par rapport à $\mathcal{B}([0, \infty[) \otimes \mathcal{F}_t$. Le processus X est dit adapté si $\forall t \geq 0,\ X_t$ est $\mathcal{F}_t-$mesurable.

Définition 2 (Processus stationnaire)**.** Un processus $(X_t)_{t\in\mathbb{U}}$, $\mathbb{U} = \{\mathbb{Z}, \mathbb{R}^+\}$ est dit *strictement stationnaire* ou *stationnaire au sens strict* si les lois jointes de $(X_{t_1}, \cdots, X_{t_k})$ et de $(X_{t_1+h}, \cdots, X_{t_k+h})$ sont identiques pour tout entier positif k et pour tous $t_1, \ldots, t_k, h \in \mathbb{Z}$.

Définition 3 (Ensemble invariant)**.** Soit $(X_t)_{t\in\mathbb{U}}$, $\mathbb{U} = \{\mathbb{Z}, \mathbb{R}^+\}$, défini sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, L'ensemble $\mathcal{B} \in \mathcal{F}$ est appelé *invariant* si il existe un ensemble $\mathcal{A}$, tel que $\mathcal{B} = \{(X_n, X_{n+1}, \ldots) \in \mathcal{A}\}$ est vrai pour tout $n \geq 1$.

Définition 4 (Processus ergodique)**.** Un processus $(X_t)_{t\in\mathbb{U}}$, $\mathbb{U} = \{\mathbb{Z}, \mathbb{R}^+\}$ est dit *ergodique* si tout ensemble invariant $\mathcal{B}$ on a $\mathbb{P}(\mathcal{B}) = 1$ ou 0.

Définition 5 (Filtration)**.** Une filtration est une suite croissante de tribus $(\mathcal{F}_t)_{t\geq 0}$, c'est-à-dire

$$\mathcal{F}_t \subset \mathcal{F}_{t+s} \quad \text{pour tout} \quad t, s \geq 0.$$

Définition 6 (Martingales)**.** Soit M un processus adapté avec $\forall t \geq 0, \quad M_t \in L^1$. On dit que M est une $(\mathcal{F}_t)_{t\geq 0}-$martingale si $M_t = E[M_{t+s}|F_t]$ pour tout $t, s \geq 0$

Définition 7 (Différence de martingale en temps continu)**.** La variable aléatoire $M = (M_t; t \geq 0)$ est une différence de martingale par rapport à la filtration $\mathcal{F} = (\mathcal{F}_t; t \geq 0)$ si

1. M_t est $\mathcal{F}_t-$mesurable,
2. $\mathbb{E}[M_t \mid \mathcal{F}_{t-s}] = 0, \quad t \geq 0, \quad s \geq 0.$

Construction d'une différence de martingale

Toutes les suites de différences de martingale présentes dans cette thèse ont été construites de la manière suivante :

Pour un réel $T \geq 0$, soit $\delta > 0$ un réel tel que $n = \frac{T}{\delta} \in \mathbb{N}$. Considérons la partition $(T_j = j\delta; 1 \leq j \leq n)$ de l'intervalle $[0, T]$. En fixant $s = 2\delta$, on observe que

1. M_t est $\mathcal{F}_t-$mesurable,
2. $\mathbb{E}[M_t \mid \mathcal{F}_{t-2\delta}] = 0, \quad t \geq 0.$

Il ressort alors que M_t vérifie la propreté de différence de martingale par rapport à la tribu $\mathcal{F}_t$. En prenant, $t \in [T_{j-1}, T_j]$, on a

1. M_t est $\mathcal{F}_{T_j}-$mesurable,
2. $\mathbb{E}[M_t \mid \mathcal{F}_{t-2\delta}] = \mathbb{E}[M_t \mid \mathcal{F}_{T_j-2\delta}] = \mathbb{E}[M_t \mid \mathcal{F}_{T_{j-2}}] = 0.$

Il s'en suit alors, pour $t \in [T_{j-1}, T_j]$, que M_t est une différence de martingale par rapport à la tribu $\mathcal{F}_{T_j}$. En utilisant à la mesurabilité de l'intégrale, on obtient

1. $A_j = \int_{T_{j-1}}^{T_j} M_t dt$ est $\mathcal{F}_{T_j}-$mesurable,
2. $\mathbb{E}[A_j \mid \mathcal{F}_{T_j-2\delta}] = \mathbb{E}[A_j \mid \mathcal{F}_{T_{j-2}}] = 0.$

Par conséquent, A_j est une différence de martingale par rapport à la tribu $\mathcal{F}_{T_j}$. Ainsi, $(A_j)_{0\leq j\leq n}$ est une suite de différences de martingale par rapport à la filtration $(\mathcal{F}_{T_j})_{0\leq j\leq n}$.

Théorème ergodique

Théorème 1. Si $X = (X_t)_{t\in\mathbb{Z}}$ est un processus ergodique stationnaire et si X_1 est intégrable, on a

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} X_i = \mathbb{E}[X_1].$$

Convergence

Définition 8 (Convergence en probabilité)**.** La suite $X_n \to X$ en probabilité (P.), si

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

Définition 9 (Convergence presque sûre)**.** La suite (X_n) converge presque sûrement (p.s.) vers X, si

$$\mathbb{P}\left\{\omega : \lim_{n\to\infty} X_n(\omega) = X(\omega)\right\} = 1.$$

Remarque : La convergence presque sûre entraîne la convergence en probabilité.

Définition 10. Si $(A_n)_{n\in\mathbb{N}}$ est une suite d'événements on note

$$\limsup A_n := \cap_{n=0}^{\infty}\left(\cup_{k=n}^{\infty} A_k\right),$$

et

$$\liminf A_n := \cup_{n=0}^{\infty}\left(\cap_{k=n}^{\infty} A_k\right).$$

Lemme 1 (Lemme de Borel-Cantelli)**.** Soit $(A_n)_{n\in\mathbb{N}}$ une suite d'événements.

(i) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) < \infty$, alors

$$\mathbb{P}\left(\limsup A_n\right) = 0,$$

ou de manière équivalente,

$$\text{presque sûrement}, \quad \{n \in \mathbb{N}; \omega \in A_n\} \quad \text{est fini.}$$

(ii) Si $\sum_{n\in\mathbb{N}} \mathbb{P}(A_n) = \infty$ et si les événements A_n sont indépendants, alors

$$\mathbb{P}\left(\limsup A_n\right) = 1,$$

ou de manière équivalente,

$$\text{presque sûrement}, \quad \{n \in \mathbb{N}; \omega \in A_n\} \quad \text{est infini.}$$

Le lemme suivant est un type lemme de Borel-Cantelli pour les processus en temps continu.

Lemme 2. Soit $(U_t, t \geq 0)$ un processus réel à temps continu de telle sorte que nous avons

a) Pour tout $\eta > 0$, il existe une fonction réelle décroissante ψ_η intéegrable sur $[0, +\infty)$ et satisfaisante

$$\mathbb{P}\left(|U_t| \geq \eta\right) \leq \psi_\eta(t), \quad t \geq 0,$$

b) Les trajectoires de U_t sont uniformément continue avec une probabilité 1, alors

$$\lim_{T\to\infty} U_t = 0, \quad \text{p.s.}$$

Inégalités exponentielles

Lemme 3. Soit $(Z_n)_{n\geq 1}$ une suite de différences de martingales réelles par rapport à la séquence de σ-algèbre $(\mathcal{F}_n = \sigma(Z_1, \cdots, Z_n))_{n\geq 1}$, où $\sigma(Z_1, \cdots, Z_n)$ est la tribu engendré par les variables aléatoires $Z_1, \cdots, Z_n$. On pose $S_n = \sum_{i=1}^n Z_i$. Pour tout $p \geq 2$ et toute $n \geq 1$, supposons qu'il existe certaines constantes non négatif C et d_n tels que

$$\mathbb{E}(Z_n^p|\mathcal{F}_{n-1}) \leq C^{p-2}p!d_n^2, \quad \text{presque sûrement.} \tag{3.52}$$

Alors, pour tout $\epsilon > 0$, on a

$$\mathbb{P}(|S_n| > \epsilon) \leq 2\exp\left\{-\frac{\epsilon^2}{2(D_n + C\epsilon)}\right\},$$

où $D_n = \sum_{i=1}^n d_i^2$.

Preuve. La preuve suit comme un cas particulier du théorème 8.2.2 due à de la Peña and Giné (1999). □

Le prochain lemme nous donne une inégalité exponentielle similaire pour des sommes partielles de différences de martingales bornées.

Lemme 4 (Laïb(1999)[86])**.** Soit $\{Z_n : n \in \mathbb{N}\}$ une suite des différences de martingale par rapport à la filtration $\{\mathcal{F}_n = \sigma(Z_1, \ldots, Z_n) : i \in \mathbb{N}\}$, où $\sigma(Z_1, \ldots, Z_n)$ est la tribu engendrée par les les variables aléatoires $Z_1, , \ldots, Z_n$, telles que $\|Z_n\| \leq B$ p.s. pour $1 \leq i \leq n$. Pour tout $\epsilon > 0$, on a

$$\mathbb{P}\left\{\max_{1\leq i\leq n}\left\|\sum_{j=1}^{i} Z_j\right\| > \epsilon\right\} \leq 2\exp\left\{-\frac{\epsilon^2}{2nB^2}\right\}. \tag{3.53}$$

Perspectives de recherche

Dans cette thèse, nous proposons des réponses à certaines questions, et contribuons à donner des alternatives aux travaux établis pour l'estimation non paramétrique à noyau des fonctions de densité, de régression (voir Bosq (1998)[17] et Bosq et Blanke ((2007)[20],(2008)[21]), Laïb (2005)[87] et Laïb et Louani ((2010)[88], (2011)[89])) ainsi que pour la fonction mode conditionnel (voir Ould said ((1993)[104], (1997)[105])). Nous avons établi la convergence presque sûre, ponctuelle et uniforme, sur un compact en donnant leurs vitesses de convergence. Un travail sur la normalité asymptotique est en voie d'achèvement puisqu'il ne reste plus qu'à montrer que les variances asymptotiques sont non nulles ce qui confirmera que les lois limites sont non dégénérées.

Les énoncés des théorèmes relatifs à la normalité asymptotique se présentent sous les formes suivantes. Les hypothèses et les démonstrations sont données dans l'annexe aux perspectives.

Théorème 2. Supposons que les conditions (H.0)(i) et (H.2)-(H.3) soient vérifiées, et

$$\lim_{T\to\infty} h_T^2 (Th_T^d)^{\frac{1}{2}} = 0, \tag{3.54}$$

alors, pour tout $\mathbf{x} \in \mathbb{R}^d$, on a

$$(Th_T^d)^{1/2} \left(f_T(\mathbf{x}) - f(\mathbf{x})\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\mathbf{x})\right) \quad \text{lorsque} \quad T \to \infty,$$

où $\xrightarrow{\mathcal{D}}$ est la convergence en distribution et $\mathcal{N}$ est la distribution gaussienne.

Théorème 3. Sous les assomptions (H.0), (H.1) et (H.4), lorsque

$$\lim_{T\to\infty} h_T^\beta (Th_T^d)^{\frac{1}{2}} = 0,$$

pour tout $\mathbf{x} \in \mathbb{R}^d$ tel que $f(\mathbf{x}) \neq 0$, on a

$$(Th_T^d)^{1/2} (\hat{r}_T(\mathbf{x}) - r(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma^2(\mathbf{x})\right), \quad \text{lorsque} \quad T \to \infty.$$

Les hypothèses et les notations mentionnés au prochain théorème sont donnés dans le chapitre 3.

Théorème 4. Considérons les hypothèses (A.1)(i), (A.1)(ii2), (A.1)(v), (A.2), (A.3), (A.4)(i), (A.4)(iii), (A.6)(ii), (A.7)(i) et (A.8). Si, de plus, pour tout $\eta \geq 2$, nous avons

$$(Th_T^{d+1})h_T^{\eta} \longrightarrow 0, \quad \text{lorsque} \quad T \to \infty,$$

alors, on a

$$\left(Th_T^{d+3}\right)^{1/2} (\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2(\mathbf{x}, \Theta(\mathbf{x}))) \quad \text{lorsque } T \to \infty,$$

De façon similaire au travail efféctuée pour la régression usuelle, en considérant l'estimateur $\widehat{\Theta}_T(\xi_T)$ de la valeur prédite $\Theta(\xi_T)$ à l'horizon H, on obtient le résultat suivant

Corollaire 1. Sous les mêmes hypothèses que le Théorème 4, on a

$$\left(Ta_T^4\right)^{1/2} (\Theta_{T-H}(\xi_T) - \Theta(\xi_T)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2(\xi_T, \Theta_{T-H}(\xi_T))) \quad \text{lorsque } T \to \infty.$$

Naturellement, il reste de nombreuses questions qui méritent des réponses. Nous présentons brièvement dans ce qui suit quelques perspectives de recherches qui retiennent notre attention pour de futurs travaux.

Estimation des composantes additives dans le modèle additif de la régression pour les processus en temps continu.

Les modèles additifs (voir Stone (1985)[124]) ont été étudiés dans le cadre de processus vectoriels à temps continu dans plusieurs contributions (voir, Debbarh (2008)[39] et Debbarh et Maillot (2008)[39],[40]). Comme il a été déjà signalé tout le long de cette thèse, divers processus ne vérifient pas les conditions de dépendance faible (α-mélange) utilisées dans ces travaux. Il est de ce fait alors tout à fait approprié d'envisager l'étude de ces modèles dans le cas de données ergodiques pour étendre le champs d'application en pratique de ces modèles qui sont introduits dans ce qui suit.

Soit $Z_t = (\mathbf{X}_t, Y_t), t \geq 0$ un processus mesurable stationnaire et ergodique défini sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$ et à valeurs dans $\mathbb{R}^d \times \mathbb{R}$. Soit ψ une fonction réelle mesurable. On considère le modèle de régression additif, défini par

$$\begin{aligned} m_\psi(\mathbf{x}) &= \mathbb{E}[\psi(Y) \mid \mathbf{X} = \mathbf{x}], \quad \forall \mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d, \\ &= \mu + \sum_{\ell=1}^{d} m_\ell(\mathbf{x}_\ell). \end{aligned}$$

Soit $q_1, \ldots, q_d$ des densités sur $\mathbb{R}$, on pose $q(\mathbf{x}) := \prod_{\ell=1}^{d} q_\ell(x_\ell)$ et $q_{-\ell}(\mathbf{x}) := \prod_{j\neq\ell} q_j(x_j)$. Pour estimer les composantes additives de la fonction de régression la méthode de l'intégration marginal (voir, Linton et Nielsen (1995)[83], Newey (1994)[101]) semble être la plus adaptée du point de vue théorique. En utilisant cette technique la composante additive est alors de la forme

$$\begin{aligned} \eta_\ell(x_\ell) &= \int_{\mathbb{R}^{d-1}} m_\psi(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} m_\psi(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad \ell = 1, \ldots, d, && (3.55) \\ &= m_\ell(x_\ell) - \int_{\mathbb{R}} m_\ell(z) q_\ell(z) dz, \quad \ell = 1, \ldots, d. && (3.56) \end{aligned}$$

De ce fait, on obtient la fonction de régression de départ

$$m_\psi(\mathbf{x}) = \sum_{\ell=1}^{d} \eta_\ell(\mathbf{x}_\ell) + \int_{\mathbb{R}^d} m_\psi(\mathbf{z}) q(\mathbf{z}) d\mathbf{z}. \tag{3.57}$$

Compte tenu de l'équation (3.56), les composantes η_ℓ et m_ℓ sont égales à une constante additionnelle près. Par conséquent, η_ℓ est également une composante additive, vérifiant une condition d'identifiabilité différente. Au vu de l'équation (3.55), l'estimateur naturel de la $\ell^{ème}$ composante est donné par

$$\widetilde{\eta}_\ell(x_\ell) = \int_{\mathbb{R}^{d-1}} \widetilde{m}_{\psi,T,\ell}(\mathbf{x}) q_{-\ell}(\mathbf{x}_{-\ell}) d\mathbf{x}_{-\ell} - \int_{\mathbb{R}^d} \widetilde{m}_{\psi,T,\ell}(\mathbf{x}) q(\mathbf{x}) d\mathbf{x}, \quad \ell = 1, \ldots, d,$$

où

$$\widetilde{m}_{\psi,T,\ell}(\mathbf{x}) = \frac{1}{T h_T^d \widehat{f}_T(\mathbf{x})} \int_0^T \psi(Y_t) K_1\left(\frac{x_\ell - X_{t,\ell}}{h_T}\right) K_2\left(\frac{\mathbf{x}_{-\ell} - \mathbf{X}_{t,-\ell}}{h_T}\right) dt,$$

et

$$\widehat{f}_T(\mathbf{x}) = \frac{1}{T h_T^d} \int_0^T K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt.$$

Ici, K_1, K_2 et K sont des noyaux définis respectivement sur $\mathbb{R}$, $\mathbb{R}^{d-1}$ et $\mathbb{R}^d$ et h_T est une suite des réels positifs tendant vers zero lorsque T tends vers l'infini. On en déduit alors que l'estimateur $\hat{m}_{\psi,T,add}$ de la fonction de régression additive est

$$\widehat{m}_{\psi,T,add}(\mathbf{x}) = \sum_{\ell=1}^{d} \widetilde{\eta}_\ell(\mathbf{x}_\ell) + \int_{\mathbb{R}^d} \widetilde{m}_{\psi,T}(\mathbf{z}) q(\mathbf{z}) d\mathbf{z},$$

où

$$\widetilde{m}_{\psi,T}(\mathbf{x}) = \frac{1}{T h_T^d \widehat{f}_T(\mathbf{x})} \int_0^T \psi(Y_t) K_3\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right) dt,$$

avec K_3 un noyau défini sur $\mathbb{R}^d$.

Les questions posées alors sont relatives au comportement asymptotique des estimateurs $\widehat{m}_{\psi,T,add}$ et $\widetilde{\eta}_\ell$.

1. Convergence presque sûre ponctuelle et uniforme avec vitesses.
2. Normalité asymptotique des composantes additives $\widetilde{\eta}_\ell$.

Estimation de la régression sur données fonctionnelles stationnaires et ergodiques en temps continu.

Il serait intéressant d'étendre les résultats obtenus dans le deuxième chapitre de cette thèse aux données fonctionnelles strictement stationnaire et ergodique. Le cadre de travail est défini comme suit.

Soit $(X_t, Y_t)_{0\le t\le T}$ une réalisation du couple d'éléments aléatoires (X, Y), où Y_t est une variable aléatoire à valeurs réelles et X_t est à valeurs dans un espace fonctionnel $\mathcal{E}$ muni de la semi-métrique d. Pour tout $x \in \mathcal{E}$, considérons la fonction de régression $r(x) = \mathbb{E}[Y \mid X = x]$.
Une extension de l'estimateur de Nadaraya-Watson de la fonction r en temps continu au cas fonctionnel est donné par l'expression (voir, Ferraty et Vieu (2000)[59])

$$\widehat{r}_T(x) = \frac{\int_0^T Y_t K\left(\frac{d(x,X_t)}{h_T}\right) dt}{\int_0^T K\left(\frac{d(x,X_t)}{h_T}\right) dt},$$

où K est une fonction réelle et h_T un paramètre de lissage tend vers zero lorsque T tend vers l'infini.
L'objet de l'étude est détablir quelques propriétés asymptotiques :

1. Convergence presque sûre : ponctuelle et uniforme, avec vitesse.
2. Normalité asymptotique.

Estimation de la densité et de la fonction de régression par la méthode des ondelettes pour des processus stationnaires et ergodiques

Dans les chapitres 2 et 3 nous avons étudié le comportement asymptotique des estimateurs à noyau des fonctions de densité et de régression. D'autre part, en considérant un processus $(\mathbf{X}_t, Y_t)_{t\in\mathcal{T}}$ stationnaire et ergodique, où $\mathcal{T} = \mathbb{N}$ ou $\mathcal{T} = \mathbb{R}_+$, il est possible de traiter le cas de l'estimation de ces fonctionnelles par la méthode des ondelettes. Il est alors intéressant d'établir des propriétés fortes de consistance ponctuelles et uniformes avec des vitesses de convergence ainsi que la normalité asymptotique de ces estimateurs dans des conditions générales d'ergodicité.

Annexe aux perspectives

Some notations are needed. For a positive real number δ such that $n = \frac{T}{\delta} \in \mathbb{N}$, consider the δ-partition $(T_j)_{1\leq j\leq n}$ of the interval $[0, T]$. Moreover, for $t > 0$ and $\delta > 0$, consider the σ-fields $\mathcal{G}_t = \sigma((\mathbf{X}_s) : 0 \leq s \leq t)$, $\mathcal{F}_t = \sigma((\mathbf{X}_s, Y_s) : 0 \leq s < t)$, and $\mathcal{S}_{t,\delta} = \sigma((\mathbf{X}_s, Y_s); (\mathbf{X}_r) : 0 \leq s < t, \quad t \leq r \leq t + \delta)$. Whenever $s < 0$, $\mathcal{G}_s$ and $\mathcal{F}_s$ stand as trivial σ-fields. For $j \in \mathbb{N}$, denote $\mathcal{G}_j = \mathcal{G}_{T_j}$ and $\mathcal{F}_j = \mathcal{F}_{T_j}$. For a real random variable ξ and any $k \in \mathbb{N}$, we define the projection $\mathcal{P}_k$ by $\mathcal{P}_k\xi = \mathbb{E}[\xi|\mathcal{G}_k] - \mathbb{E}[\xi|\mathcal{G}_{k-1}]$, where $\mathbb{E}[\xi|\mathcal{G}_k]$ stands as the conditional expectation of ξ given the σ-field $\mathcal{G}_k$.

The following assumptions will be needed throughout the paper.

(H.0) (i) K is a spherically symetric bounded kernel of order 2 on its compact support, such that $\int K(\mathbf{z})d\mathbf{z} = 1$, and satisfying $\int z_1^{\alpha_1} \cdots z_d^{\alpha_d} K(\mathbf{z})dz = 0$, where $\alpha_1, \ldots, \alpha_d \in \mathbb{N}$; $\alpha_1 + \cdots + \alpha_d = j$, $j = 1, 2.$ and $\int \|\mathbf{z}\|^2 K(\mathbf{z})d\mathbf{z} < \infty$.

(ii) There exists a nonnegative differentiable integrable function k over its support $[0, c_1\lambda]$ such that $K(u) = k(\|u\|)$, and $\int_{\mathbb{R}} u^{d-1}k^2(u)du < \infty$

(H.1) For any $0 \leq s < t \leq T$, there exists a positive constant c_0 and a nonnegative continuous and bounded random function $f_{t,s}$ defined on $\mathbb{R}^d$ such that, almost surely,

$$\mathbb{P}^{\mathcal{F}_s}\Big(\mathbf{X}_t \in S_{r,\mathbf{x}}\Big) := \mathbb{P}\Big(\|\mathbf{X}_t - \mathbf{x}\| \leq r \Big| \mathcal{F}_s\Big) = c_0 r^d f_{t,s}(\mathbf{x}) + o(r^d), as\ r \to 0;$$

Notice for the density estimation case that the function $f_{t,t-\delta}$ stands as the conditional density of $\mathbf{X}_t$ given the σ-field $\mathcal{G}_{t-\delta}$;

(H.2) (i) The function f is twice differentiable with a bounded second derivative.

(ii) The function $f_{t,s}$ is twice differentiable with almost surely bounded second derivative.

(iii) For any r and any $(s,t) \in [0,T]^2$ such that $r < s$ and $r < t$, the conditional density of the couple (X_s, X_t) given the σ field $\mathcal{G}_r$ exists.

(H.3) For any δ, we have

$$\sup_y \int_{\mathbb{R}^+} \|\mathcal{P}_1 f_{t,t-\delta}(\mathbf{y})\|^2 \, dt < \infty.$$

(H.4) For any small $\delta > 0$ and any s, t such that $t \in [0, T]$ and $t \leq s \leq t + \delta$, we have

(i) $\mathbb{E}[Y_r|\mathcal{S}_{t,\delta}] = \mathbb{E}[Y_r|X_r] = r(X_r)$ almost surely.

(ii) There exist constants $\beta > 0$ and $c_3 > 0$ such that, for any $(u, v) \in \mathbb{R}^{2d}$, $|r(\mathbf{u}) - r(\mathbf{v})| \leq c_3|\mathbf{u} - \mathbf{v}|^\beta$

(iii) $\mathbb{E}[(Y_s - r(\mathbf{x}))^2|\mathcal{S}_t] = \mathbb{E}[(Y_s - r(\mathbf{x}))^2|\mathbf{X}_s] = W_2(\mathbf{X}_s)$, moreover the function $W_2(\mathbf{x})$ is uniformly continuous in neighborhood of x, i.e.,

$$\sup_{|\mathbf{u}-\mathbf{x}|\leq h_T} |W_2(\mathbf{u}) - W_2(\mathbf{x})| = o(1) \quad \text{as} \quad h_T \to 0.$$

(iv) For any $\alpha > 0$ small enough, $\mathbb{E}[|Y_s|^{2+\alpha}] < \infty$ and the function $\bar{W}_{2+\alpha}(\mathbf{u}) = \mathbb{E}[|Y_s - r(\mathbf{x})|^{2+\alpha}|\mathbf{X}_s = u]$, is continuous in neighborhood of $\mathbf{x}$.

Proof of Theorem 2.

In order to prove our result, introduce first the following notation

$$\tilde{f}_{T,1}(\mathbf{x}) = \frac{1}{Th_T^d} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[K\left(\frac{\mathbf{x} - \mathbf{X}_t}{h_T}\right)|\mathcal{G}_{i-2}\right] dt.$$

Consider now the following decomposition

$$\begin{aligned}
\sqrt{Th_T^d}\Big(\hat{f}_T(\mathbf{x}) - f(\mathbf{x})\Big) &= \sqrt{Th_T^d}\Big(\hat{f}_T(\mathbf{x}) - \tilde{f}_{T,1}(\mathbf{x})\Big) + \sqrt{Th_T^d}\Big(\tilde{f}_{T,1}(\mathbf{x}) - f(\mathbf{x})\Big) \\
&= (Th_T^d)^{1/2}\left(V_T(\mathbf{x}) + B_T(\mathbf{x})\right).
\end{aligned}$$

Our first concern is to investigate the behaviour of the normalized conditional bias $(Th_T^d)^{1/2}B_T(\mathbf{x})$. For all $\mathbf{x} \in \mathbb{R}^d$, we have

$$\begin{aligned}
(Th_T^d)^{1/2}B_T(\mathbf{x}) &= (Th_T^d)^{1/2}\left(\tilde{f}_{T,1}(\mathbf{x}) - \mathbb{E}\left[\hat{f}_T(\mathbf{x})\right]\right. \\
&\quad \left. + \mathbb{E}\left[\hat{f}_T(\mathbf{x})\right] - f(\mathbf{x})\right) \\
&= (Th_T^d)^{1/2}\Big(B_T^{(1)}(\mathbf{x}) + B_T^{(2)}(\mathbf{x})\Big).
\end{aligned}$$

Hereafter, we use the notation $K_h(\cdot) := \frac{1}{h}K(\frac{\cdot}{h})$. Observe that

$$\begin{aligned}
\left|TB_T^{(1)}(\mathbf{x})\right| &= \left|\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\left(\mathbb{E}\left[K_h\Big(\mathbf{x} - \mathbf{X}_t\Big)\Big|\mathcal{G}_{i-2}\right] - \mathbb{E}\left[K_h\Big(\mathbf{x} - \mathbf{X}_t\Big)\right]\right)dt\right| \\
&= \left|\int_{\mathbb{R}^d} K_h\Big(\mathbf{x} - \mathbf{y}\Big)\left(\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{y})dt - Tf(\mathbf{y})\right)dy\right| \\
&:= \left|\int_{\mathbb{R}^d} K_h\Big(\mathbf{x} - \mathbf{y}\Big)H_T^{(1)}(\mathbf{y})d\mathbf{y}\right| \leq \int_{\mathbb{R}^d} K_h\Big(\mathbf{x} - \mathbf{y}\Big)\|H_T^{(1)}(\mathbf{y})\|d\mathbf{y}.
\end{aligned}$$

Using the lower Riemann's sum, for $\delta > 0$ small enough, it follows that

$$\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{y})dt \simeq \sum_{i=1}^{n}\delta f_{T_{i-1},T_{i-2}} \leq \int_0^T f_{t,t-\delta}(\mathbf{y})dt.$$

Considering the projection $\mathcal{P}_k$ defined above, following Wu (2003) and making use of Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\|H_T^{(1)}(\mathbf{y})\|^2 &= \left\|\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{y})dt - Tf(\mathbf{y})\right\| \\ &\leq \left\|\int_0^T f_{t,t-\delta}(\mathbf{y})dt - Tf(\mathbf{y})\right\|^2 = T\left(\int_0^T \left\|\mathcal{P}_1 f_{t,t-\delta}(\mathbf{y})\right\|^2 dt\right).\end{aligned}$$

Thus,

$$\sup_{\mathbf{y}}\|H_T^{(1)}(\mathbf{y})\|^2 \leq T\left(\sup_{\mathbf{y}}\int_0^T \left\|\mathcal{P}_1 f_{t,t-\delta}(\mathbf{y})\right\|^2 dt\right) \leq T\left(\sup_{\mathbf{y}}\int_{\mathbb{R}^+} \left\|\mathcal{P}_1 f_{t,t-\delta}(\mathbf{y})\right\|^2 dt\right).$$

Subsequently, assuming the condition (H.2), we obtain

$$\sup_{\mathbf{y}}\|H_T^{(1)}(\mathbf{y})\|^2 = O(T), \tag{3.58}$$

which implies that

$$B_T^{(1)}(\mathbf{x}) = O(T^{-1/2}). \tag{3.59}$$

Moreover, assuming the conditions (H.0)(i) and (H.2)(i), it is easily seen that

$$\begin{aligned}B_T^{(2)}(\mathbf{x}) &= \mathbb{E}[\hat{f}_T(\mathbf{x})] - f(\mathbf{x}) \\ &= \int_{\mathbb{R}^d} K_h\Big(\mathbf{x}-\mathbf{y}\Big)f(\mathbf{y})d\mathbf{y} - f(\mathbf{x}) \\ &= \int_{\mathbb{R}^d} K(\mathbf{z})\left(f(\mathbf{x}) - h_T\sum_{i=1}^{d} z_i\frac{\partial f}{\partial x_i}(\mathbf{x}^*) + \frac{h_T^2}{2}\sum_{j=1}^{d}\sum_{k=1}^{d} z_j z_k\frac{\partial^2 f}{\partial x_j \partial x_k}(\mathbf{x}^*)\right)d\mathbf{z} \\ &\quad -f(\mathbf{x}) \\ &= \sup_{x_j,x_k}\left|\frac{\partial^2 f}{\partial x_j\partial x_k}(\mathbf{x}^*)\right|\frac{h_T^2}{2}\int_{\mathbb{R}^d}\|\mathbf{z}\|^2 K(\mathbf{z})\mathbf{z} \\ &= O(h_T^2),\end{aligned}$$

where $\mathbf{x}^*$ ranges between $\mathbf{x}$ and $\mathbf{x} - \mathbf{z}h_T$. Thus,

$$B_T(\mathbf{x}) = O(h_T^2) + O\Big(T^{-1/2}\Big).$$

Obviously, we have

$$B_T(\mathbf{x}) = O(h_T^2).$$

Considering the condition (3.54), we conclude that

$$(Th_T^d)^{1/2}B_T(\mathbf{x}) = O\left(h_T^2(Th_T^d)^{1/2}\right).$$

Therefore, to state Theorem 2, it suffices then to prove the following assertion

$$(Th_T^d)^{1/2} V_T(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \sigma^2(\mathbf{x})\Big).$$

In this respect, observe that

$$\begin{aligned}
(Th_T^d)^{1/2} V_T(\mathbf{x}) &= \sqrt{Th_T^d}\Big(\hat{f}_T(\mathbf{x}) - \tilde{f}_{T,1}(\mathbf{x})\Big) \\
&= (Th_T^d)^{-\frac{1}{2}} \sum_{i=1}^{n} \left(\int_{T_{i-1}}^{T_i} K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) dt \right.\\
&\quad \left. -\mathbb{E}\left[\int_{T_{i-1}}^{T_i} K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) dt \Big| \mathcal{G}_{i-2}\right]\right) \\
&= \sum_{i=1}^{n} \Big(\eta_{T,i}(\mathbf{x}) - \mathbb{E}\left[\eta_{T,i}(\mathbf{x}) \Big| \mathcal{G}_{i-2}\right]\Big) \\
&= \sum_{i=1}^{n} \xi_{T,i}(\mathbf{x}),
\end{aligned}$$

where $\eta_{T,i}(\mathbf{x}) = (Th_T^d)^{-\frac{1}{2}} \int_{T_{i-1}}^{T_i} K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) dt$ and $\xi_{T,i}(\mathbf{x}) = \eta_{T,i}(\mathbf{x}) - \mathbb{E}\left[\eta_{T,i}(\mathbf{x}) \Big| \mathcal{G}_{i-2}\right]$. Furthermore, it is easily seen that $(\xi_{T,i}(\mathbf{x}))_{1\leq i\leq n}$ is a sequence of martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{G}_{i-2})_{1\leq i\leq n}$. Therefore, we have to check the following two conditions

(a) $\sum_{i=1}^{n} \mathbb{E}[\xi_{T,i}^2(\mathbf{x})|\mathcal{G}_{i-2}] \xrightarrow{\mathbb{P}} \sigma^2(\mathbf{x})$

(b) $n\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(\mathbf{x})|>\epsilon\}}] = o(1)$ holds, for any $\epsilon > 0$

that are necessary to establish the asymptotic normality related to discrete time martingale difference sequences (see, for instance, Hall and Heyde (1980)).

First, by a change of variable and a Taylor expansion of the order one of $f_{t,T_{i-2}}(\mathbf{x}-h_T\mathbf{z})$ around $\mathbf{x}$ when assumption (H.2)(ii) is considered, observe that

$$\begin{aligned}
&\left|\sum_{i=1}^{n} \mathbb{E}\left[\eta_{T,i}^2(\mathbf{x}) \Big| \mathcal{G}_{i-2}\right] - \sum_{i=1}^{n} \mathbb{E}\left[\xi_{T,i}^2(\mathbf{x}) \Big| \mathcal{G}_{i-2}\right]\right| = \left|\sum_{i=1}^{n} \Big(\mathbb{E}\left[\eta_{T,i}(\mathbf{x}) \Big| \mathcal{G}_{i-2}\right]\Big)^2\right| \\
&= \left|\frac{1}{Th_T^d} \sum_{i=1}^{n} \left(\int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K\left(\frac{\mathbf{x}-\mathbf{y}}{h_T}\right) f_{t,T_{i-2}}(\mathbf{y}) d\mathbf{y} dt\right)^2\right| \\
&= h_T^d \left|\frac{1}{T} \sum_{i=1}^{n} \left(\int_{T_{i-1}}^{T_i} \int_{\mathbb{R}^d} K(\mathbf{z}) f_{t,T_{i-2}}(\mathbf{x}-h_T\mathbf{z}) d\mathbf{z} dt\right)^2\right| \\
&= h_T^d \left|\frac{1}{T} \sum_{i=1}^{n} \left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x}) dt\right)^2 + O(h_T)\right| \\
&= h_T^d \left|\frac{1}{\delta n} \sum_{i=1}^{n} g_i^{\mathcal{G}_{i-2}}(\mathbf{x}) + O(h_T)\right| \\
&= O(h_T^d).
\end{aligned}$$

where $g_i^{\mathcal{G}_{i-2}}(\mathbf{x}) = \left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt\right)^2$. Making use of the Riemann sum, the quantity $g_i^{\mathcal{G}_{i-2}}(\mathbf{x})$ may be approached, whenever δ is small enough, by $\delta f_{T_{i-1},T_{i-2}}$. It is then clear from the discussion above that the process $(f_{T_{i-1},T_{i-2}})_{i\geq 1}$ is stationary and ergodic. So the sum $\frac{1}{n}\sum_{i=1}^n g_i^{\mathcal{G}_{i-2}}(\mathbf{x})$ has a finite limit, (see Krengel (1985), Theorem 4.4), which is

$$\mathbb{E}\left[g_1^{\mathcal{G}_{-\delta}}(\mathbf{x})\right] = \left(\int_0^\delta f(\mathbf{x})dt\right)^2 = \delta^2 f^2(\mathbf{x}). \tag{3.60}$$

Using Jensen inequality we obtain

$$\begin{aligned}
\mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{G}_{i-2}\right] &= \frac{1}{Th_T^d}\mathbb{E}\left[\left(\int_{T_{i-1}}^{T_i} K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt\right)^2\middle|\mathcal{G}_{i-2}\right] \\
&\leq \frac{1}{Th_T^d}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\middle|\mathcal{G}_{i-2}\right]dt
\end{aligned}$$

By a first odrder of Taylor expansion fonction $f_{t,Ti-2}$, for $\mathbf{x}^*$ in $[\mathbf{x}-h_T\mathbf{v},\mathbf{x}]$, we obtain

$$\begin{aligned}
\sum_{i=1}^n \mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{G}_{i-2}\right] &\leq \frac{1}{Th_T^d}\sum_{i=1}^n\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\middle|\mathcal{G}_{i-2}\right]dt \\
&= \frac{1}{Th_T^d}\sum_{i=1}^n\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d} K^2\left(\frac{\mathbf{x}-\mathbf{u}}{h_T}\right)f_{t,T_{i-2}}(\mathbf{u})d\mathbf{u}dt \\
&= \frac{1}{T}\sum_{i=1}^n\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}^d} K^2(\mathbf{v})f_{t,T_{i-2}}(\mathbf{x}-h_T\mathbf{v})d\mathbf{v}dt \\
&= \frac{1}{\delta}\left(\frac{1}{n}\sum_{i=1}^n\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x}) + O(h_T)\right)\int_{\mathbb{R}^d} K^2(\mathbf{v})d\mathbf{v}dt
\end{aligned}$$

Proceeding similarly as for the statement (3.60), it is clear, whenever δ is small enough, that the quantities $\left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt\right)_{i\in\mathbb{N}}$ may be approached by $\left(\delta f_{T_{i-1},T_{i-2}}(\mathbf{x})\right)_{i\in\mathbb{N}}$. Consequently, using the ergodic and stationarity properties of the process $(\mathbf{X}_t)_{t\geq 0}$, it follows that

$$\begin{aligned}
\frac{1}{n}\sum_{j=1}^n\left(\int_{T_{j-1}}^{T_j} f_{t,T_{i-2}}(\mathbf{x})dt\right) &= \mathbb{E}\left(\int_{T_0}^{T_1} f_t(\mathbf{x})dt\right) + o(1) = \int_0^\delta \mathbb{E}\left(f_t(\mathbf{x})\right)dt + o(1) \\
&= \delta f(\mathbf{x}) + o(1).
\end{aligned} \tag{3.61}$$

It follows that

$$\sum_{i=1}^n \mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{G}_{i-2}\right] \leq f(\mathbf{x})\int_{\mathbb{R}^d} K^2(\mathbf{y})d\mathbf{y} + O(h_T)$$

which yields

$$\sum_{i=1}^n \mathbb{E}\left[\xi_{T,i}^2\mathbf{x})\middle|\mathcal{G}_{i-2}\right] \xrightarrow{\mathcal{D}} \sigma^2(\mathbf{x}) \leq f(\mathbf{x})\int_{\mathbb{R}^d} K^2(\mathbf{y})d\mathbf{y} \quad \text{as} \quad T\to\infty. \tag{3.62}$$

Proof of part (b). Using successively inequalities of Holder, Markov, Jensen and then Minkowski, we obtain for any $\epsilon > 0$ and any p and q such that $1/p + 1/q = 1$

$$\begin{aligned}
\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(x)|>\epsilon\}}] &\leq (\mathbb{E}[\xi_{T,i}^{2q}(\mathbf{x})])^{1/q}(P\{|\xi_{T,i}(\mathbf{x})| > \epsilon\})^{1/p} \leq \epsilon^{-2q/p}\mathbb{E}[|\xi_{T,i}(\mathbf{x})|^{2q}] \\
&\leq \frac{\epsilon^{-2q/p}}{(Th_T^d)^q}\mathbb{E}\left[\int_{T_{i-1}}^{T_i}\left|K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{G}_{t-\delta}\right]\right|^{2q} dt\right] \\
&\leq \frac{\epsilon^{-2q/p}}{(Th_T^d)^q}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K^{2q}\left(\frac{\mathbf{x}-X_t}{h_T}\right)\right]dt \\
&= \frac{\epsilon^{-2q/p}}{(Th_T^d)^q}\int_{T_{i-1}}^{T_i}\int_{\mathbb{R}}K^{2q}\left(\frac{\mathbf{x}-\mathbf{y}}{h_T}\right)f(\mathbf{y})d\mathbf{y}dt \\
&= \frac{\epsilon^{-2q/p}\delta}{T^q h_T^{d(q-1)}}\int_{\mathbb{R}}K^{2q}(\mathbf{z})f(\mathbf{x}-h_T\mathbf{z})d\mathbf{z}.
\end{aligned}$$

Consequently, since f is a density, it follows then that

$$n\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(\mathbf{x})|>\epsilon\}}] = \frac{\epsilon^{-2q/p}\|K\|_\infty^{2q}}{(Th_T^d)^{q-1}} = o(1). \tag{3.63}$$

Combining the statements (3.62) and (3.63), we obtain

$$(Th_T^d)^{1/2}V_T(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \sigma^2(\mathbf{x})\right), \tag{3.64}$$

which completes the proof. □

Proof of Theorem 3.

Define

$$Z_1(\mathbf{x}) = \int_0^\delta K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt.$$

For $j = 1, 2$, set

$$\tilde{r}_{T,j}(\mathbf{x}) = \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[Y_t^{j-1}K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]dt.$$

Since $P(\|\mathbf{x}-\mathbf{X}_t\| \leq 0) = 0$, integrating by parts and considering hypotheses (H.0)(i) and (H.1)(i), we obtain

$$\begin{aligned}
\frac{1}{h_T^d}\mathbb{E}[Z_1(\mathbf{x})] &= \frac{1}{h_T^d}\mathbb{E}\left[\int_0^\delta K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt\right] \\
&= \frac{1}{h_T^d}\int_0^\delta \mathbb{E}\left[k\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right)\right]dt \\
&= \frac{1}{h_T^d}\int_0^\delta\int_0^\lambda k(u)d\mathbb{P}(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T u) \\
&= \frac{1}{h_T^d}\int_0^\delta\left([k(u)P\left(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T u\right)]_0^\lambda \int_0^\lambda k'(u)P\left(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T u\right)du\right)dt \\
&= c_0\delta f(\mathbf{x})\left(k(\lambda)\lambda^d - \int_0^\lambda u^d k'(u)du\right) + o(1) \\
&= d\ c_0\delta f(\mathbf{x})\left(\int_0^\lambda u^{d-1}k(u)du\right) + o(1) \\
&= d\ c_0\delta f(\mathbf{x})\tau_0 + o(1) \qquad (3.65)
\end{aligned}$$

where $\tau_0 = \left(\int_0^\lambda u^{d-1}k(u)du\right)$. Consider now the following decomposition

$$\begin{aligned}
Q_T(\mathbf{x}) &:= (\hat{r}_{T,2}(\mathbf{x}) - \tilde{r}_{T,2}(\mathbf{x})) - r(\mathbf{x})(\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x})) \qquad (3.66)\\
R_T(\mathbf{x}) &:= -B_T(\mathbf{x})(\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x})) \qquad (3.67)\\
\hat{r}_T(\mathbf{x}) - r(\mathbf{x}) &= \hat{r}_T(\mathbf{x}) - C_T(\mathbf{x}) + B_T(\mathbf{x}) = B_T(\mathbf{x}) + \frac{Q_T(\mathbf{x}) + R_T(\mathbf{x})}{\hat{r}_{T,1}(\mathbf{x})}, \qquad (3.68)
\end{aligned}$$

where

$$B_T(\mathbf{x}) = C_T(\mathbf{x}) - r(\mathbf{x}) := \frac{\tilde{r}_{T,2}(\mathbf{x})}{\tilde{r}_{T,1}(\mathbf{x})} - r(\mathbf{x})$$

stands as the conditional bias. Observe that

$$B_T(\mathbf{x}) = \frac{\tilde{r}_{T,2}(\mathbf{x}) - r(\mathbf{x})\tilde{r}_{T,1}(\mathbf{x})}{\tilde{r}_{T,1}(\mathbf{x})} = \frac{\tilde{B}_T(\mathbf{x})}{\tilde{r}_{T,1}(\mathbf{x})},$$

The assumptions (H.4)(i)-(ii) and the statement (3.61) allow us to write

$$\begin{aligned}
\tilde{B}_T(\mathbf{x}) &= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[(Y_t - r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]dt\\
&= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\mathbb{E}\left[(Y_t - r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{S}_{t-\delta,\delta}\right]|\mathcal{F}_{i-2}\right]dt\\
&= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[(r(\mathbf{X}_t) - r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]dt\\
&\leq \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sup_{\|\mathbf{u}-\mathbf{x}\|\leq h_T}|r(\mathbf{u})-r(\mathbf{x})|\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[k\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right)|\mathcal{F}_{i-2}\right]dt\\
&= \frac{1}{nd\delta c_0\tau_0 h_T^d f(\mathbf{x})}\sup_{\|\mathbf{u}-\mathbf{x}\|\leq h_T}|r(\mathbf{u})-r(\mathbf{x})|\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} dc_0\tau_0 h_T^d f_{t,T_{i-2}}(\mathbf{x})dt\\
&= \frac{h_T^{\beta}}{\delta f(\mathbf{x})}\left(\frac{1}{n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt\right)\\
&= O\Big(h_T^{\beta}\Big). \qquad (3.69)
\end{aligned}$$

Now we attend to show that

$$\tilde{r}_{T,1}(\mathbf{x}) \overset{\mathbb{P}}{=} 1, \text{ as } T\to\infty.$$

Remark, by the statement (3.61), that

$$\begin{aligned}
\tilde{r}_{T,1}(\mathbf{x}) &= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]dt\\
&= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[k\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right)|\mathcal{F}_{i-2}\right]dt\\
&= \frac{1}{nd\delta c_0\tau_0 h_T^d f(\mathbf{x})}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\Big(dc_0\tau_0 h_T^d f_{t,T_{i-2}}(\mathbf{x}) + O\Big(h_T^d\Big)\Big)dt\\
&= \frac{1}{\delta f(\mathbf{x})}\left(\frac{1}{n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt\right) + O\Big(h_T^d\Big) = 1 + o(1). \qquad (3.70)
\end{aligned}$$

Combining the statements (3.69) and (3.70), we have thus proved that

$$Th_T^d)^{1/2}B_T(\mathbf{x}) = O\left(h_T^{\beta}(Th_T^d)^{1/2}\right). \qquad (3.71)$$

We are left with the task of determining the limit of $R_T(\mathbf{x})$, according to the decomposition (3.67), the only point remaining concerns the limiting behaviour of the term $(\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x}))$.

Consider the next decomposition

$$\begin{aligned}
\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x}) &= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\left(K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]\right)dt\\
&= \frac{1}{n\mathbb{E}(Z_1(\mathbf{x}))}\sum_{i=1}^{n}\chi_{T,i}(\mathbf{x}).
\end{aligned}$$

where $\chi_{T,i}(\mathbf{x}) = \int_{T_{i-1}}^{T_i} \left(K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) - \mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) |\mathcal{F}_{i-2}\right] \right) dt$.
It is easy to check that the sequence of random variables $\{\chi_{T,i}(\mathbf{x})\}$ form a triangular array of martingale differences. By Markov, Burkholder and Jensen inequalities together with the fact that $(a+b)^2 \leq 2a^2 + b^2$, for any $\epsilon > 0$, there exists a constant $C_0 > 0$ such that

$$
\begin{aligned}
\mathbb{P}\left(|\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x})| > \epsilon\right) &= \mathbb{P}\left(\left| \sum_{i=1}^{n} \chi_{T,i}(\mathbf{x}) \right| > \epsilon\left(n\mathbb{E}(Z_1(\mathbf{x}))\right) \right) \\
&\leq \frac{1}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \mathbb{E}\left[\left(\sum_{i=1}^{n} \chi_{T,i}(\mathbf{x}) \right)^2 \right] \\
&\leq \frac{C_0}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \sum_{i=1}^{n} \mathbb{E}\left[\chi_{T,i}^2(\mathbf{x})\right] \\
&\leq \frac{C_0}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\left(K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \right.\right. \\
&\quad \left.\left. - \mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) |\mathcal{F}_{i-2}\right] \right)^2 \right] dt \\
&\leq \frac{2C_0}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \left(\mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\right) \right. \\
&\quad \left. + \mathbb{E}\left[\left(\mathbb{E}\left[K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) |\mathcal{F}_{i-2}\right] \right)^2 \right] \right) dt \\
&\leq \frac{4C_0}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\right] dt \\
&\leq \frac{4C_0}{\epsilon^2 \left(n\mathbb{E}(Z_1(\mathbf{x}))\right)^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[k^2\left(\left\| \frac{\mathbf{x}-\mathbf{X}_t}{h_T} \right\| \right)\right] dt.
\end{aligned}
$$

Let assumptions (H.0)(ii)-(H.1) hold. Then, by the same method as in the statement (3.65), we obtain

$$
\begin{aligned}
&\mathbb{P}\left(|\hat{r}_{T,1}(\mathbf{x}) - \tilde{r}_{T,1}(\mathbf{x})| > \epsilon\right) \\
&\leq \frac{4C_0}{\epsilon^2 \left(ndc_0\delta\tau_0 h_T^d f(\mathbf{x})\right)^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} dc_0 h_T^d f(\mathbf{x}) dt \left(\int_0^{\lambda} v^{d-1} k^2(v) dv \right) \\
&= \frac{4C_0}{\epsilon^2 T h_T^d dc_0 \tau_0^2 f(\mathbf{x})} \left(\int_0^{\lambda} v^{d-1} k^2(v) dv \right) = O\left(\frac{1}{T h_T^d} \right). \qquad (3.72)
\end{aligned}
$$

Combining now the decomposition (3.67) the statements (3.71) and (3.72), it follows that

$$
(Th_T^d)^{1/2} R_T(\mathbf{x}) = O\left(\frac{h_T^{\beta}}{(Th_T^d)^{1/2}} \right).
$$

Using the decomposition (3.68) and the conditions (??)(iii) and (iv), it remains then to establish the limiting behaviour of the term $Q_T(\mathbf{x})$. In this respect, note that

$$\begin{aligned}(Th_T^d)^{1/2}Q_T(\mathbf{x}) &= (Th_T^d)^{1/2}(\hat{r}_{T,2}(\mathbf{x})-\tilde{r}_{T,2}(\mathbf{x}))+r(\mathbf{x})(\hat{r}_{T,1}(\mathbf{x})-\tilde{r}_{T,1}(\mathbf{x}))\\ &= \sum_{i=1}^{n}\frac{(Th_T^d)^{1/2}}{nE[Z_1(\mathbf{x})]}\int_{T_{i-1}}^{T_i}\left[(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\right.\\ &\quad\left.-\mathbb{E}\left[(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\mid\mathcal{F}_{i-2}\right]\right]dt\\ &:= \sum_{i=1}^{n}\xi_{T,i}(\mathbf{x}).\end{aligned}$$

Observe that $\xi_{T,i}(\mathbf{x})=\eta_{T,i}(\mathbf{x})-\mathbb{E}\left[\eta_{T,i}(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]$ where $\eta_{T,i}(\mathbf{x})=\frac{(Th_T^d)^{1/2}}{nE[Z_1(\mathbf{x})]}\int_{T_{i-1}}^{T_i}(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt$. As above, $(\xi_{T,i}(\mathbf{x}))_{1\le i\le n}$ is a sequence of martingale differences with respect to the sequence of $\sigma-$fields $(\mathcal{F}_{i-1})_{1\le i\le n}$. Therefore, proceeding similarly as for the density estimate $\hat{f}_T$, we have to check the following two assertions

(a) $\sum_{i=1}^{n}\mathbb{E}[\xi_{T,i}^2(\mathbf{x})|\mathcal{F}_{i-2}]\overset{\mathbb{P}}{\to}\Sigma^2(\mathbf{x})$

(b) $n\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(\mathbf{x})|>\epsilon\}}]=o(1)$ holds, for any $\epsilon>0$.

First of all, observe that

$$\left|\sum_{i=1}^{n}\mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]-\sum_{i=1}^{n}\mathbb{E}\left[\xi_{T,i}^2(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\right|=\left|\sum_{i=1}^{n}\left(\mathbb{E}\left[\eta_{T,i}(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\right)^2\right|.$$

Note that, for $T_{i-1}\le t\le T_i$, we have $\mathcal{F}_{i-2}\subset\mathcal{S}_{t-\delta,\delta}$. Therefore, making use of conditions (H.4)(i) and the statement (3.65), we have

$$\begin{aligned}&\left|\mathbb{E}\left[\eta_{T,i}(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\right|\\ &= \left|\frac{(Th_T^d)^{1/2}}{nE[Z_1(\mathbf{x})]}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\middle|\mathcal{F}_{i-2}\right]dt\right|\\ &= \left|\frac{(Th_T^d)^{1/2}}{nE[Z_1(\mathbf{x})]}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\mathbb{E}\left[(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\middle|\mathcal{S}_{t-\delta,\delta}\right]\middle|\mathcal{F}_{i-2}\right]dt\right|\\ &= \left|\frac{(Th_T^d)^{1/2}}{ndc_0\tau_0\delta f(\mathbf{x})h_T^d}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[\mathbb{E}\left[(r(\mathbf{X}_t)-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\middle|\mathcal{S}_{t-\delta,\delta}\right]\middle|\mathcal{F}_{i-2}\right]dt\right|\\ &\le \left|\frac{(Th_T^d)^{1/2}}{ndc_0\tau_0\delta f(\mathbf{x})h_T^d}\int_{T_{i-1}}^{T_i}\sup_{\|\mathbf{x}-\mathbf{v}\|\le h_T}|r(\mathbf{v})-r(\mathbf{x})|\mathbb{E}\left[k\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right)\middle|\mathcal{F}_{i-2}\right]dt\right|,\end{aligned}$$

by conditions (H.4)(ii) we obtain

$$\begin{aligned}
\left|\mathbb{E}\left[\eta_{T,i}(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\right| &\leq \left(\frac{1}{Th_T^d}\right)^{1/2}\frac{1}{dc_0\delta\tau_0 f(\mathbf{x})}\sup_{\|\mathbf{x}-\mathbf{v}\|\leq h_T}|r(\mathbf{v})-r(\mathbf{x})| \\
&\times\int_{T_{i-1}}^{T_i}\left(\left[k(u)\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\middle|\mathcal{F}_{i-2}\right)\right]_0^\lambda\right. \\
&\left.-\int_0^\lambda k'(u)\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\middle|\mathcal{F}_{i-2}\right)du\right)dt \\
&= \left(\frac{h_T^d}{T}\right)^{1/2}\frac{h_T^\beta}{\delta\tau_0 f(\mathbf{x})}\left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt+o(1)\right)\int_0^\lambda u^{d-1}k(u)dudt \\
&= \left(\frac{h_T^d}{T}\right)^{1/2}\frac{h_T^\beta}{\delta f(\mathbf{x})}\left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt+o(1)\right).
\end{aligned}$$

Thus, by the statement (3.60), we have

$$\begin{aligned}
\left|\sum_{i=1}^n\left(\mathbb{E}\left[\eta_{T,i}(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\right)^2\right| &= \frac{h_T^{2\beta+d}}{\delta^3 f^2(\mathbf{x})}\left(\frac{1}{n}\sum_{i=1}^n\left(\int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x})dt\right)^2\right)+o\left(h_T^{d+2\beta}\right) \\
&:= \frac{h_T^{2\beta+d}}{\delta(\delta f(\mathbf{x}))^2}\frac{1}{\delta n}\sum_{i=1}^n\left(g_i^{\mathcal{F}_{i-2}}(\mathbf{x})\right)^2+o\left(h_T^{d+2\beta}\right) \\
&= \frac{h_T^{2\beta+d}}{\delta(\delta f(\mathbf{x}))^2}\mathbb{E}[g_0^{\mathcal{F}_{-\delta}}(\mathbf{x})]+o\left(h_T^{d+2\beta}\right) \\
&= o\left(h_T^{d+2\beta}\right).
\end{aligned}$$

The statement (a) follows then provided that

$$\sum_{i=1}^n\mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{F}_{i-2}\right]\to\Sigma^2(\mathbf{x}).$$

By Jensen inequality, we obtain

$$\begin{aligned}
\mathbb{E}\left[\eta_{T,i}^2(\mathbf{x})\middle|\mathcal{F}_{i-2}\right] &= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2}\mathbb{E}\left[\left(\int_{T_{i-1}}^{T_i}(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt\right)^2\middle|\mathcal{F}_{i-2}\right] \\
&\leq \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2}\mathbb{E}\left[\int_{T_{i-1}}^{T_i}(Y_t-r(\mathbf{x}))^2K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)dt\middle|\mathcal{F}_{i-2}\right],
\end{aligned}$$

Using the condition (H.4)(iii) and observing that, for any $1\leq i\leq n$ and any

$t \in [T_{i-1}, T_i]$, we have $\mathcal{F}_{i-2} \subseteq \mathcal{S}_{t-\delta}$, it follows that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}\left[\eta_{T,i}^2(\mathbf{x}) \middle| \mathcal{F}_{i-2}\right] \\
&\leq \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[(Y_t - r(\mathbf{x}))^2 K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[\mathbb{E}\left[(Y_t - r(\mathbf{x}))^2 \middle| \mathcal{S}_{t-\delta}\right] K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[W_2(\mathbf{X}_t) K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[(W_2(\mathbf{X}_t) - W_2(\mathbf{x})) K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&\quad + \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[W_2(\mathbf{x}) K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&= I_{T,1,i}(\mathbf{x}) + I_{T,2,i}(\mathbf{x}).
\end{aligned} \tag{3.73}$$

Considering the condition (H.0)(i), it follows that

$$\begin{aligned}
\|I_{T,1,i}(\mathbf{x})\| &= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \mathbb{E}\left[(W_2(\mathbf{X}_t) - W_2(\mathbf{x})) k^2\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right) \middle| \mathcal{F}_{i-2}\right] dt \\
&= \frac{Th_T^d}{(nE[Z_1(\mathbf{x})])^2} \sup_{\|\mathbf{v}-\mathbf{x}\| \leq h_T} |W_2(\mathbf{v}) - W_2(\mathbf{x})| \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} \\
&\quad \times \int_{\mathbb{R}^d} k^2(\mathbf{y}) d\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T \mathbf{y} \middle| \mathcal{F}_{i-2}\right) dt \\
&= \frac{Th_T^d}{(ndc_0\tau_0\delta h_T^d f(\mathbf{x}))^2} \sup_{\|\mathbf{v}-\mathbf{x}\| \leq h_T} |W_2(\mathbf{v}) - W_2(\mathbf{x})| \sum_{i=1}^{n} \\
&\quad \int_{T_{i-1}}^{T_i} \left(\left[k^2(u)\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T u \middle| \mathcal{F}_{i-2}\right)\right]_0^\lambda \right. \\
&\quad \left. - \int_0^\lambda \left(k^2\right)'(u) \mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\| \leq h_T u \middle| \mathcal{F}_{i-2}\right) du \right) dt \\
&= \frac{o(1)}{\delta dc_0(\tau_0 f(\mathbf{x}))^2} \int_0^\lambda u^{d-1} k^2(u) du \left(\frac{1}{n} \sum_{i=1}^{n} \int_{T_{i-1}}^{T_i} f_{t,T_{i-2}}(\mathbf{x}) dt\right).
\end{aligned}$$

Therefore, proceeding as for the statement (3.61), we obtain

$$I_{T,1,i}(\mathbf{x}) = o(1). \tag{3.74}$$

Moreover, we have

$$\begin{aligned}
I_{T,2,i}(\mathbf{x}) &= \frac{W_2(\mathbf{x})}{\delta(dc_0\tau_0 f(\mathbf{x}))^2}\frac{1}{nh_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K^2\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\Big|\mathcal{F}_{i-2}\right]dt \\
&= \frac{W_2(\mathbf{x})}{\delta(dc_0\tau_0 f(\mathbf{x}))^2}\frac{1}{nh_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\int_0^{\lambda} k^2(u)d\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T\Big|\mathcal{F}_{i-2}\right)dt \\
&= \frac{W_2(\mathbf{x})}{\delta(dc_0\tau_0 f(\mathbf{x}))^2}\frac{1}{nh_T^d}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}\left(\left[k^2(u)\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\Big|\mathcal{F}_{i-2}\right)\right]_0^{\lambda}\right. \\
&\quad \left.-\int_0^{\lambda}\left(k^2\right)'(u)\mathbb{P}\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\Big|\mathcal{F}_{i-2}\right)du\right) \\
&= \frac{W_2(\mathbf{x})}{\delta dc_0(\tau_0 f(\mathbf{x}))^2}\int_0^{\lambda}u^{d-1}k^2(u)du\left(\frac{1}{n}\sum_{i=1}^{n}\int_{T_{i-1}}^{T_i}f_{t,T_{i-2}}(\mathbf{x})dt\right).
\end{aligned}$$

Thus, again proceeding as in the statement (3.61) together with a passage to the limit, it follows that

$$I_{T,2,i}(\mathbf{x}) = \frac{W_2(\mathbf{x})}{dc_0\tau_0^2 f(\mathbf{x})}\int_0^{\lambda}u^{d-1}k^2(u)du, \quad \text{as} \quad T\to\infty. \tag{3.75}$$

Hence, statements (3.74) and (3.75) give

$$\lim_{T\to\infty}\sum_{i=1}^{n}\mathbb{E}[\eta_{T,i}^2(\mathbf{x})|\mathcal{F}_{i-2}] \leq \frac{W_2(\mathbf{x})}{dc_0\tau_0^2 f(\mathbf{x})}\int_0^{\lambda}u^{d-1}k^2(u)du, \quad \text{a.s.} \tag{3.76}$$

Proof (b). Using Hölder, Markov, Jensen and Minkowski's inequalities, for any $\epsilon>0$ and any p and q such that $1/p+1/q=1$, we obtain

$$\begin{aligned}
&\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(\mathbf{x})|>\epsilon\}}] \leq (\mathbb{E}[\xi_{T,i}^{2q}(\mathbf{x})])^{1/q}(P\{|\xi_{T,i}(\mathbf{x})|>\epsilon\})^{1/p} \leq \epsilon^{-2q/p}\mathbb{E}[|\xi_{T,i}(\mathbf{x})|^{2q}] \\
&\leq \frac{(Th_T^d)^q\epsilon^{-2q/p}}{(n\mathbb{E}[Z_1(\mathbf{x})])^{2q}}\mathbb{E}\Bigg[\int_{T_{i-1}}^{T_i}\Bigg|(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right) \\
&\quad -\mathbb{E}\left[(Y_t-r(\mathbf{x}))K\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)|\mathcal{F}_{i-2}\right]\Bigg|^{2q}dt\Bigg] \\
&\leq \frac{2^{2q}(Th_T^d)^q\epsilon^{-2q/p}}{(n\mathbb{E}[Z_1(\mathbf{x})])^{2q}}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[|Y_t-r(\mathbf{x})|^{2q}K^{2q}\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\right]dt \\
&= \frac{2^{2q}(Th_T^d)^q\epsilon^{-2q/p}}{(n\mathbb{E}[Z_1(\mathbf{x})])^{2q}}\int_{T_{i-1}}^{T_i}\mathbb{E}\left[K^{2q}\left(\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right)\mathbb{E}\left[|Y_t-r(\mathbf{x})|^{2q}|\mathcal{S}_{t-\delta}\right]\right]dt \\
&= \frac{2^{2q}(Th_T^d)^q\epsilon^{-2q/p}}{(n\mathbb{E}[Z_1(\mathbf{x})])^{2q}}\int_{T_{i-1}}^{T_i}\Big(\sup_{|\mathbf{u}-\mathbf{x}|\leq h_T}|\bar{W}_{2q}(\mathbf{u})-\bar{W}_{2q}(\mathbf{x})|+\bar{W}_{2q}(\mathbf{x})\Big)\mathbb{E}\left[k^{2q}\left(\left\|\frac{\mathbf{x}-\mathbf{X}_t}{h_T}\right\|\right)\right]dt \\
&= \frac{2^{2q}(Th_T^d)^q\epsilon^{-2q/p}}{(dc_0 n\delta\tau_0 h_T^d f(\mathbf{x}))^{2q}}\int_{T_{i-1}}^{T_i}(\bar{W}_{2q}(\mathbf{x})+o(1))\left(\left[k^{2k}(u)P\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\right)\right]_0^{\lambda}\right. \\
&\quad \left.-\int_0^{\lambda}(k^{2q})'(u)P\left(\|\mathbf{x}-\mathbf{X}_t\|\leq h_T u\right)du\right)dt \\
&\leq \frac{2^{2q}\delta\epsilon^{-2q/p}}{(dc_0)^{2q-1}\tau_0^{2q}T^q h_T^{d(q-1)}f^{2q-1}(\mathbf{x})}(\bar{W}_{2q}(\mathbf{x})+o(1))\int_0^{\lambda}u^{d-1}k^{2q}(u)du.
\end{aligned}$$

Therefore, assumptions (H.1)(i) and (H.4)(iii) imply that

$$\begin{aligned} n\mathbb{E}[\xi_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|\xi_{T,i}(\mathbf{x})|>\epsilon\}}] &= \frac{2^{2q}\epsilon^{-2q/p}}{(dc_0)^{2q-1}\tau_0^{2q}(Th_T^d)^{q-1}f(\mathbf{x})^{2q}}(\bar{W}_{2q}(\mathbf{x})+o(1))\int_0^\lambda u^{d-1}k^{2q}(u)du \\ &= O\left((Th_T)^{-(q-1)}\right). \end{aligned}$$

This completes the proof of the part (b). Consequently, we have

$$(Th_T)^{1/2}Q_T(\mathbf{x}) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma^2(\mathbf{x})\right) \tag{3.77}$$

which completes the proof of Theorem 3. □

3.4.3 Proof of Theorem 4.

Note first that we have

$$\begin{aligned} \sqrt{Ta_T^{d+3}}\left(\Theta_T(\mathbf{x})-\Theta(\mathbf{x})\right) &= -\frac{\sqrt{Ta_T^{d+3}}\left(h_T^{(0,1)}(\mathbf{x},\Theta_T(\mathbf{x}))-\bar{h}_T^{(0,1)}(\mathbf{x},\Theta_T(\mathbf{x}))\right)}{h_T^{(0,2)}(\mathbf{x},\Theta_T^*(\mathbf{x}))} \\ &\quad -\frac{\sqrt{Ta_T^{d+3}}\bar{h}_T^{(0,1)}(\mathbf{x},\Theta_T(\mathbf{x}))}{h_T^{(0,2)}(\mathbf{x},\Theta_T^*(\mathbf{x}))} \end{aligned} \tag{3.78}$$

The following lemma deals with the asymptotic behaviour of the bias term of the first derivative when suitably normalized.

Lemma 1. If hypotheses (A.1)(v), (A.3), (A.4)(i)-(iii), (A.6) and (A.8) hold true and if for any $\alpha \geq 2$

$$h_T^\alpha (Ta_T^{d+1})^{1/2} \longrightarrow 0, \quad \text{as} \quad T\to\infty, \tag{3.79}$$

then we have

$$\sqrt{Ta_T^{d+3}}\bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) \longrightarrow 0, \quad \text{as} \quad T\to\infty. \tag{3.80}$$

Proof Lemma 1. We have by assumption (A.6)

$$\begin{aligned} \bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) &= \frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^{(1)}\left(\frac{\Theta(\mathbf{x})-Y_t}{a_T}\right)\Big|\mathcal{F}_{t-\delta}\right]dt \\ &= \frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_0}{a_T}\right)K_2^{(1)}\left(\frac{\Theta(\mathbf{x})-Y_0}{a_T}\right)\Big|\mathcal{S}_{t-\delta,\delta}\right]\Big|\mathcal{F}_{t-\delta}\right]dt \\ &= \frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathbb{E}\left[K_2^{(1)}\left(\frac{\Theta(\mathbf{x})-Y_t}{a_T}\right)|\mathbf{X}_t\right]\Big|\mathcal{F}_{t-\delta}\right]dt. \end{aligned}$$

An integration by parts, together with a change of variables, leads to

$$\begin{aligned}
\mathbb{E}\left[K_2^{(1)}\left(\frac{\Theta(\mathbf{x})-Y_t}{a_T}\right)|\mathbf{X}_t\right] &= \int_{\mathbb{R}} K_2^{(1)}\left(\frac{\Theta(\mathbf{x})-z}{a_T}\right) f(z|\mathbf{X}_t)dz \\
&= a_T \int_{\mathbb{R}} K_2^{(1)}(v) f(\Theta(\mathbf{x})-va_T|\mathbf{X}_t)dv \\
&= a_T \left([K_2(v)f(\Theta(\mathbf{x})-va_T|\mathbf{X}_t)]_{-\infty}^{\infty} + a_T\int_{\mathbb{R}} K_2(v) f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{X}_t)dv\right) \\
&= a_T^2 \int_{\mathbb{R}} K_2(v) f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{X}_t)dv \\
&= a_T^2 \int_{\mathbb{R}} K_2(v)\left(f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{X}_t) - f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{x})\right)dv \\
&\quad + a_T^2 \int_{\mathbb{R}} K_2(v) f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{x})dv =: \mathcal{I}_1+\mathcal{I}_2.
\end{aligned}$$

By a two-order Taylor expansion of $f^{(1)}(\Theta(\mathbf{x})-va_T|\mathbf{x})$ around $\Theta(\mathbf{x})$, the fact that $\Theta(\mathbf{x})$ is the conditional mode function, we get

$$\begin{aligned}
\bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) &= \frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathcal{I}_1\Big|\mathcal{F}_{t-\delta}\right]dt \\
&\quad + \frac{1}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\int_{\mathbb{R}} K_2(v)\left(-a_T v f^{(2)}(\Theta(\mathbf{x})|\mathbf{x})\right.\right. \\
&\quad \left.\left. + \frac{a_T^2v^2}{2} f^{(3)}(\Theta^*(\mathbf{x})|\mathbf{x})\right)dv\Big|\mathcal{F}_{t-\delta}\right]dt,
\end{aligned}$$

where $\Theta^*(\mathbf{x})$ is between $\Theta(\mathbf{x})$ and $\Theta(\mathbf{x})-va_T$. Then, by assumption (A.1)(v), we have

$$\begin{aligned}
\sqrt{Ta_T^{d+3}}\bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) &= \frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathcal{I}_1\Big|\mathcal{F}_{t-\delta}\right]dt \\
&\quad + \frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left(\frac{a_T^2}{2}f^{(3)}(\Theta^*(\mathbf{x})|\mathbf{x})\int_{\mathbb{R}} t^2K_2(t)dt\right)\Big|\mathcal{F}_{t-\delta}\right]dt \\
&= \frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathcal{I}_1\Big|\mathcal{F}_{t-\delta}\right]dt \\
&\quad + \frac{a_T^2\sqrt{Ta_T^{d+3}}}{2} f^{(3)}(\Theta^*(\mathbf{x})|\mathbf{x})\left(\int_{\mathbb{R}} v^2K_2(v)dv\right)\frac{1}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\Big|\mathcal{F}_{t-\delta}\right]dt.
\end{aligned}$$

Under assumptions (A.1)(iv), (A.4)(i) and (A.4)(iii) , we have

$$\begin{aligned}
\frac{1}{Ta_T^d}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\Big|\mathcal{F}_{t-\delta}\right]dt &= \frac{1}{Ta_T^d}\int_0^T\int_{\mathbb{R}^d} K_1\left(\frac{\mathbf{x}-\mathbf{z}}{a_T}\right) g^{\mathcal{F}_{t-\delta}}(\mathbf{z})d\mathbf{z}dt \\
&\leq \frac{1}{T}\int_0^T\int_{\mathbb{R}^d} K_1(\mathbf{u})\left|g^{\mathcal{F}_{t-\delta}}(\mathbf{x}-a_T\mathbf{u})-g^{\mathcal{F}_{t-\delta}}(\mathbf{x})\right|d\mathbf{u}dt \\
&\quad +\frac{1}{T}\int_0^T g^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt \\
&= \frac{1}{T}\int_0^T g^{\mathcal{F}_{t-\delta}}(\mathbf{x})dt + O(a_T) \\
&= g(\mathbf{x}) + o(1) + O(a_T). \qquad (3.81)
\end{aligned}$$

By the statement (3.81) combined with conditions (A.3), one obtains

$$\begin{aligned}
\sqrt{Ta_T^{d+3}}\bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) &= \frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathcal{I}_1\Big|\mathcal{F}_{t-\delta}\right]dt \\
&\quad +O\left(a_T^2\sqrt{Ta_T^{d+1}}\right)+O\left(a_T^3\sqrt{Ta_T^{d+1}}\right).
\end{aligned}$$

On the other hand, by assumptions (A.3) and (A.8) together with the equation (3.81), we have

$$\begin{aligned}
&\left|\frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_s}{a_T}\right)\mathcal{I}_1\Big|\mathcal{F}_{s-\delta}\right]ds\right| \\
&\leq \frac{1}{Ta_T^{d+2}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_s}{a_T}\right)|\mathcal{I}_1|\Big|\mathcal{F}_{s-\delta}\right]ds \\
&\leq \frac{1}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_s}{a_T}\right)\int_{\mathbb{R}} K_2(t)\left|f^{(1)}(\Theta(\mathbf{x})-Ta_T|\mathbf{X}_s)\right.\right. \\
&\quad \left.\left.-f^{(1)}(\Theta(\mathbf{x})-Ta_T|\mathbf{x}))\right|dt\Big|\mathcal{F}_{s-\delta}\right]ds \\
&\leq \frac{C_{\mathbf{x}}}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_s}{a_T}\right)\int_{\mathbb{R}} K_2(t)\left\|\mathbf{X}_s-\mathbf{x}\right\|^\beta dt\Big|\mathcal{F}_{s-\delta}\right]ds \\
&\leq C_{\mathbf{x}}h_T^\beta\left(\frac{1}{Ta_T^{d}}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_s}{a_T}\right)\Big|\mathcal{F}_{s-\delta}\right]ds\right) \\
&= C_{\mathbf{x}}a_T^\beta\left(g(\mathbf{x})+o(1)+O(a_T)\right).
\end{aligned}$$

Hence,

$$\sqrt{Ta_T^{d+3}}\bar{h}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) = O\left(a_T^{\beta+1}\sqrt{Ta_T^{d+1}}\right)+O\left(a_T^3\sqrt{Ta_T^{d+1}}\right).$$

By the condition (3.79), it follows that

$$\sqrt{Ta_T^{d+3}}\bar{f}_T^{(0,1)}(\mathbf{x},\Theta(\mathbf{x})) = o(1),\ \ a.s. \qquad (3.82)$$

□

Lemma 2. Under assumptions (A.1)(ii2), (A.1)(iv)-(v), (A.2), (A.3), (A.4)(i)-(iii), (A.6) and (A.8), for any $\mathbf{x} \in \mathbb{R}^d$, we have

$$h_T^{(0,2)}(\mathbf{x}, \Theta_T^*(\mathbf{x})) \xrightarrow{\mathbb{P}} h^{(0,2)}(\mathbf{x}, \Theta_T(\mathbf{x})), \text{ as } \quad T \to \infty. \tag{3.83}$$

Proof Lemma 2. Set

$$\bar{h}_T^{(0,2)}(\mathbf{x}, y) = \frac{1}{Ta_T^{d+3}} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_0}{a_T}\right) K_2^{(2)}\left(\frac{y - Y_0}{a_T}\right) \Big| \mathcal{F}_{t-\delta}\right] dt$$

and write

$$h_T^{(0,2)}(\mathbf{x}, y) = h_T^{(0,2)}(\mathbf{x}, y) - \bar{h}_T^{(0,2)}(\mathbf{x}, y) + \bar{h}_T^{(0,2)}(\mathbf{x}, y).$$

Proceeding as in Lemma 1, under the assumption (A.6)(ii), we have

$$\begin{aligned}
\bar{h}_T^{(0,2)}(\mathbf{x}, y) &= \frac{1}{Ta_T^{d+3}} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) \mathbb{E}\left[K_2^{(2)}\left(\frac{y - Y_t}{a_T}\right) \Big| \mathcal{S}_{t-\delta,\delta}\right] \Big| \mathcal{F}_{t-\delta}\right] dt \\
&= \frac{1}{Ta_T^{d+3}} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) \mathbb{E}\left[K_2^{(2)}\left(\frac{y - Y_t}{a_T}\right) |\mathbf{X}_t\right] \Big| \mathcal{F}_{t-\delta}\right] dt.
\end{aligned}$$

Recall that

$$h^{(0,2)}(\mathbf{x}, y) = g(\mathbf{x}) f^{(2)}(y|\mathbf{x}).$$

An integration by parts two times together with a change of variables lead to

$$\begin{aligned}
\mathbb{E}\left[K_2^{(2)}\left(\frac{y - Y_t}{a_T}\right) |\mathbf{X}_t\right] &= \left|a_T^3 \int_{\mathbb{R}} K_2(v) f^{(2)}(y - a_T v|\mathbf{X}_t) dv\right| \\
&= a_T^3 \int_{\mathbb{R}} K_2(v) \left(f^{(2)}(y - a_T v|\mathbf{X}_t) - f^{(2)}(y - a_T v|\mathbf{x})\right) dv \\
&\quad + a_T^3 \int_{\mathbb{R}} K_2(v) f^{(2)}(y - a_T v|\mathbf{x}) dv,
\end{aligned}$$

Under assumptions (A.1)(iv) (A.4)(i) and (A.4)(iii), we get

$$\begin{aligned}
\frac{1}{Ta_T^d} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) \Big| \mathcal{F}_{t-\delta}\right] dt &= \frac{1}{Ta_T^d} \int_0^T \int_{\mathbb{R}^d} K_1\left(\frac{\mathbf{x} - \mathbf{z}}{a_T}\right) g_t^{\mathcal{F}_{t-\delta}}(\mathbf{z}) d\mathbf{z} dt \\
&= \frac{1}{T} \int_0^T \int_{\mathbb{R}^d} K_1(\mathbf{u}) \left|g_t^{\mathcal{F}_{t-\delta}}(\mathbf{x} - a_T \mathbf{u}) - g_t^{\mathcal{F}_{t-\delta}}(\mathbf{x})\right| d\mathbf{u} dt \\
&\quad + \frac{1}{T} \int_0^T g_t^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt \\
&= \frac{1}{T} \int_0^T g_t^{\mathcal{F}_{t-\delta}}(\mathbf{x}) dt + O(a_T) \\
&= g(\mathbf{x}) + o(1) + O(a_T).
\end{aligned} \tag{3.84}$$

Here and after , we suppose that $\mathbf{X}_t \in B(\mathbf{x}, a_T)$, by assumptions (A.4)(i) and (A.8) and equation (3.84) we have

$$\begin{aligned}
&\left|\frac{1}{Ta_T^d} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_0}{a_T}\right) \left(f^{(2)}(y - a_T v|\mathbf{X}_0) - f^{(2)}(y - a_T v|\mathbf{x})\right) \Big| \mathcal{F}_{t-\delta}\right] dv\right| \\
&\leq a_T^\beta \frac{1}{Ta_T^d} \int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x} - \mathbf{X}_t}{a_T}\right) \Big| \mathcal{F}_{t-\delta}\right] dv = a_T^\beta \left(g(\mathbf{x}) + o(1)\right).
\end{aligned} \tag{3.85}$$

On the other hand, by a first-order Taylor expansion of $f^{(2)}(y-a_Tv|\mathbf{x})$ around y, we obtain by equation (3.84) with conditions (A.1)(v)

$$\begin{aligned}
&\frac{1}{Ta_T^d}\int_0^T \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\Big|\mathcal{F}_{t-\delta}\right]dt\int_{\mathbb{R}}K_2(v)f^{(2)}(y-a_Tv|\mathbf{x})dv\\
&= (g(\mathbf{x})+o(1))\int_{\mathbb{R}}K_2(v)\left(f^{(2)}(y|\mathbf{x})-a_Tvf^{(3)}(y^*|\mathbf{x})\right)dv\\
&= g(\mathbf{x})f^{(2)}(y|\mathbf{x})+o(1)\\
&= h^{(0,2)}(\mathbf{x},y)+o(1). \qquad (3.86)
\end{aligned}$$

where y^* is between y and $y-a_Tv$. Gathering statements (3.85) and (3.86) with assumption (A.3) we obtain

$$\bar{h}_T^{(0,2)}(\mathbf{x},y)=h^{(0,2)}(\mathbf{x},y)+o(1)+O(a_T^{\beta+1})+O(a_T^{\beta+1}).$$

Hence

$$\sup_{y\in\mathcal{G}}\left|\bar{h}_T^{(0,2)}(\mathbf{x},y)-h^{(0,2)}(\mathbf{x},y)\right|=o(1),\quad \text{a.s.} \qquad (3.87)$$

Now we deal with the variance term expressed in the case of $f_T^{(0,2)}(\mathbf{x},y)$ (the proof is similar for $f_T(\mathbf{x},y)$, Lemma 3.5). Consider a cover of a compact set $\mathcal{C}$ of $\mathbb{R}$ by a finite number of intervals $\mathcal{C}_k=[y_k-h_T^\eta,y_k+h_T^\eta]$, $k=1,\cdots,l_T$ where $\mu>4$, Since is a compact there exists $M>0$ such that $l_T\le Ma_T^{-\mu}$.
Clearly, we have

$$\begin{aligned}
\sup_{y\in\mathcal{G}}\left|h_T^{(0,2)}(\mathbf{x},y)-\bar{h}_T^{(0,2)}(\mathbf{x},y)\right| &\le \max_{1\le k\le l_T}\sup_{y\in\mathcal{C}_k}\left|h_T^{(0,2)}(\mathbf{x},y)-h_T^{(0,2)}(\mathbf{x},y_k)\right|\\
&\quad+\max_{1\le k\le l_T}\left|h_T^{(0,2)}(\mathbf{x},y_k)-\bar{h}_T^{(0,2)}(\mathbf{x},y_k)\right|\\
&\quad+\max_{1\le k\le l_T}\sup_{y\in\mathcal{C}_k}\left|\bar{h}_T^{(0,2)}(\mathbf{x},y_k)-\bar{h}_T^{(0,2)}(\mathbf{x},y)\right|\\
&= \mathcal{J}_1+\mathcal{J}_2+\mathcal{J}_3. \qquad (3.88)
\end{aligned}$$

Notice that $\mathcal{J}_1$ and $\mathcal{J}_3$ may be treated using the same arguments. Undertaking the first term and making use of the assumption (A.2)(ii), it follows that

$$\begin{aligned}
\mathcal{J}_1 &\le \frac{1}{Ta_T^{d+3}}\int_0^T\max_{1\le k\le l_T}\sup_{y\in\mathcal{C}_k}K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left|K_2^{(2)}\left(\frac{y-Y_t}{a_T}\right)-K_2^{(2)}\left(\frac{y_k-Y_t}{a_T}\right)\right|dt\\
&\le \frac{C_K}{Ta_T^{d+3}}\int_0^T\max_{1\le k\le l_T}\sup_{y\in\mathcal{C}_k}K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left|\frac{y-y_k}{a_T}\right|dt\\
&\le \frac{C_Ka_T^\mu}{a_T^4}\left(\frac{1}{Ta_T^d}\int_0^T K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)dt\right)=C_Ka_T^{\mu-4}g_T(\mathbf{x}).
\end{aligned}$$

By Lemma 3.3 and conditions (A.3), we have

$$\mathcal{J}_1=C_Ka_T^{\mu-4}(g(\mathbf{x})+o(1))=o(1),\quad \text{a.s.} \qquad (3.89)$$

Considering the term $\mathcal{J}_2$, observe first that

$$\begin{aligned}
\mathcal{J}_2 &\leq \frac{1}{Ta_T^{d+3}}\int_0^T \max_{1\leq k\leq l_T}\left|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^{(2)}\left(\frac{y-Y_t}{a_T}\right)\right.\\
&\quad \left.-\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^{(2)}\left(\frac{y_k-Y_t}{a_T}\right)\left|\mathcal{F}_{t-\delta}\right.\right]\right|dt\\
&= \frac{1}{Ta_T^{d+3}}\max_{1\leq k\leq l_T}\sum_{j=1}^{n}\Psi_{T,j}(\mathbf{x},y_k),
\end{aligned}$$

with

$$\Psi_{T,j}(\mathbf{x},y_k)=\int_{T_{j-1}}^{T_j}\left|K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^{(2)}\left(\frac{y-Y_t}{a_T}\right)-\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)K_2^{(2)}\left(\frac{y_k-Y_t}{a_T}\right)\left|\mathcal{F}_{t-\delta}\right.\right]\right|dt.$$

By simple calculations, we have

$$|\Psi_{T,j}(\mathbf{x},y_k)|\leq 2\delta\widetilde{K}^2,$$

where $\widetilde{K}=\max\{\sup_{\mathbf{r}\in\mathbb{R}^d}K_1(\mathbf{r}),\sup_{s\in\mathbb{R}}K_2(s)\}$. It follows then, by Lemma 3.2, that for $\epsilon>0$

$$\begin{aligned}
P\left\{\max_{1\leq k\leq l_T}\left|\sum_{i=1}^{n}\Psi_{T,j}(\mathbf{x},y_k)\right|>\epsilon(Ta_T^{d+3})\right\} &\leq \sum_{k=1}^{l_T}P\left\{\left|\sum_{i=1}^{n}\Psi_{T,j}(\mathbf{x},y_k)\right|>\epsilon(Ta_T^{d+3})\right\}\\
&\leq 2l_T\exp\left\{-\frac{T^2a_T^{2(d+3)}\epsilon^2}{8n\delta^2\widetilde{K}^4}\right\}\\
&\leq 2Ma_T^{-\mu}\exp\left\{-\frac{Ta_T^{2(d+3)}\epsilon^2}{8\delta\widetilde{K}^4}\right\}.
\end{aligned}$$

Consequently, under the condition (3.6)(iii), we have

$$P\{\mathcal{J}_2>\epsilon\}\leq 2Ma_T^{-\mu}T^{-\frac{\epsilon^2Ta_T^{2(d+3)}}{8\delta\widetilde{K}^4\log T}}=2Ma_T^{-\mu}\left(T^{-\frac{Ta_T^{d+1}}{\log T}}\right)^{\frac{\epsilon^2a_T^{d+5}}{8\delta\widetilde{K}^4}}. \tag{3.90}$$

The right-hand side of equation (3.90) is the general term of a convergent series. Therefore by Borel-Cantelli's Lemma we obtain

$$\mathcal{J}_2=o(1),\quad \text{a.s.} \tag{3.91}$$

Finally, observe that

$$\begin{aligned}
\mathcal{J}_3 &\leq \frac{1}{Ta_T^{d+3}}\int_0^T\max_{1\leq k\leq l_T}\sup_{y\in\mathcal{C}_k}\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left|K_2^{(2)}\left(\frac{y-Y_t}{a_T}\right)-K_2^{(2)}\left(\frac{y_k-Y_t}{a_T}\right)\right|\left|\mathcal{F}_{t-\delta}\right.\right]dt\\
&\leq \frac{C_K}{a_T^3}\max_{1\leq k\leq l_T}\sup_{y\in\mathcal{C}_k}\left|\frac{y-y_k}{a_T}\right|\left(\frac{1}{Ta_T^d}\int_0^T\mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left|\mathcal{F}_{t-\delta}\right.\right]dt\right)\\
&\leq C_Ka_T^{\mu-4}\bar{g}_T(\mathbf{x}),
\end{aligned}$$

Using the statement (3.15), it follows that

$$\mathcal{J}_3 = C_K a_T^{\mu-4}(g(\mathbf{x}) + o(1)) = o(1), \quad \text{a.s.} \tag{3.92}$$

Combining now the statements (3.87), (3.89), (3.91) and (3.92), we conclude the proof of the lemma. □

We will now prove the asymptotic normality of

$$h_T^{(0,1)}(\mathbf{x}, \Theta(\mathbf{x})) - \bar{h}_T^{(0,1)}(\mathbf{x}, \Theta(\mathbf{x})) = \frac{1}{Ta_T^{d+2}} \int_0^T Z_t(\mathbf{x}, \Theta(\mathbf{x}))dt,$$

after a suitable normalization, where

$$\begin{cases} Z_t(\mathbf{x}, y) = K_{1,t}K_{2,t}^{(1)} - \mathbb{E}[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}] \\ K_{1,t} = K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right), \quad \text{and} \quad K_{2,t}^{(1)} = K_2^{(1)}\left(\frac{y-Y_t}{a_T}\right). \end{cases}$$

Set

$$R_T(\mathbf{x}, y) = \frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d+2}} \int_0^T Z_t(\mathbf{x}, y)dt = \sum_{k=1}^n Z_{T,k}$$

with $\{Z_{T,k}\} = \left\{\frac{\sqrt{Ta_T^{d+3}}}{Ta_T^{d+2}} \int_{T_{k-1}}^{T_k} Z_t(\mathbf{x}, y)dt\right\}$ is a triangular array of martingale differences with respect to the $\sigma-$field $\mathcal{F}_{k-2}$. This allows us to apply the central limit theorem for discrete-time arrays of real-valued martingales, to establish the asymptotic normality of $R_T(\mathbf{x}, y)$. This can be done by establishing the following assertions

(i) $\sum_{k=1}^n \mathbb{E}\left[Z_{T,k}^2|\mathcal{F}_{k-2}\right] \xrightarrow{\mathbb{P}} V(\mathbf{x}, y)$.

(ii) $n\mathbb{E}\left[Z_{T,k}^2 \mathbb{I}_{\{|Z_{T,k}|>\epsilon\}}\right] = o(1)$ holds for any $\epsilon > 0$ (Lindberg condition).

Under the condition (A.6), we have

$$\begin{aligned} \mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}\right] &= \mathbb{E}\left[\mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}\middle|\mathcal{S}_{t-\delta,\delta}\right]\middle|\mathcal{F}_{t-\delta}\right] \\ &= \mathbb{E}\left[K_1\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\mathbb{E}\left[K_2^{(1)}\left(\frac{y-Y_t}{a_T}\right)\middle|\mathbf{X}_t\right]\middle|\mathcal{F}_{t-\delta}\right], \end{aligned}$$

An integration by parts, together with a change of variables, leads to

$$\begin{aligned} \mathbb{E}\left[K_2^{(1)}\left(\frac{y-Y_t}{a_T}\right)\middle|\mathbf{X}_t\right] &= \int_{\mathbb{R}} K_2^{(1)}\left(\frac{y-u}{a_T}\right) f(u|\mathbf{X}_t)du \\ &= a_T^2 \int_{\mathbb{R}} K_2(v) f^{(1)}(y - a_T v|\mathbf{X}_t)dv \\ &= a_T^2 \int_{\mathbb{R}} K_2(v)[f^{(1)}(y - a_T v|\mathbf{X}_t) - f^{(1)}(y - a_T v|\mathbf{x})]dv \\ &\quad + a_T^2 \int_{\mathbb{R}} K_2(v) f^{(1)}(y - a_T v|\mathbf{x})dv. \end{aligned} \tag{3.93}$$

Using the same idea as for equations (3.84), (3.85) and (3.86), we obtain

$$\mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}\right] = O(a_T^{d+2}). \tag{3.94}$$

Therefore,

$$\begin{aligned}\sum_{k=1}^{n}\mathbb{E}\left[Z_{T,k}^2|\mathcal{F}_{k-2}\right] &= \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\mathbb{E}\left[\left(\int_{T_{k-1}}^{T_k} Z_t dt\right)^2\Big|\mathcal{F}_{k-2}\right]\\ &\le \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[Z_t^2\Big|\mathcal{F}_{k-2}\right]dt\\ &= W_T(\mathbf{x},y). \end{aligned}\tag{3.95}$$

Now, we are in position to establish the following lemma.

Lemma 3. If assumptions (A.1)(v), (A.3), (A.4)(i)-(ii) and (A.8) are fulfilled, then

(i) $\lim_{T\to\infty} W_T(\mathbf{x},y) = V(\mathbf{x},y)$,

(ii) $n\mathbb{E}\left[Z_{T,k}^2\mathbb{I}_{\{|Z_{T,k}|>\epsilon\}}\right] = o(1)$ holds for any $\epsilon > 0$ (Lindberg condition).

where $V(\mathbf{x},y) = h(\mathbf{x},y)\int_{\mathbb{R}^{d+1}}[K_1(\mathbf{r})K_2^{(1)}(s)]^2 d\mathbf{r}ds$.

Proof Lemma 3 Part (i) We have

$$\begin{aligned} W_T(\mathbf{x},y) &= \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[Z_t^2\Big|\mathcal{F}_{k-2}\right]dt\\ &= \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[\left(K_{1,t}K_{2,t}^{(1)} - \mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}\right]\right)^2\Big|\mathcal{F}_{k-2}\right]dt\\ &= \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[\left(K_{1,t}K_{2,t}^{(1)}\right)^2 - 2K_{1,t}K_{2,t}^{(1)}\mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}\right]\right.\\ &\quad\left. + \left(\mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}\right]\right)^2\Big|\mathcal{F}_{k-2}\right]dt. \end{aligned}$$

Using the same idea as in the statement (3.94), we obtain

$$\mathbb{E}\left[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{k-2}\right] = O(a_T^{d+2}).\tag{3.96}$$

Combining equations (3.94) and (3.96), we have

$$W_T(\mathbf{x},y) = \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[\left(K_{1,t}K_{2,t}^{(1)}\right)^2\Big|\mathcal{F}_{k-2}\right]dt + O(a_T^{d+3}).$$

By a one-order Taylor expansions of $h(\mathbf{x}-a_T\mathbf{r}, y-a_Ts)$ around $(\mathbf{x},y)$, together with assumption (A.4)(i), we obtain

$$\begin{aligned} &\frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[\left(K_{1,t}K_{2,t}^{(1)}\right)^2\Big|\mathcal{F}_{k-2}\right]dt\\ &= \frac{1}{Ta_T^{d+1}}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\int_{\mathbb{R}^{d+1}}K_1^2\left(\frac{\mathbf{x}-\mathbf{u}}{a_T}\right)\left(K_2^{(1)}\left(\frac{y-v}{a_T}\right)\right)^2 h_t^{\mathcal{F}_{k-2}}(\mathbf{u},v)d\mathbf{u}dvdt\\ &= \frac{1}{T}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\int_{\mathbb{R}^{d+1}}[K_1(\mathbf{r})K_2^{(1)}(s)]^2 h_t^{\mathcal{F}_{k-2}}(\mathbf{x}-a_T\mathbf{r}, y-a_Ts)d\mathbf{r}dsdt\\ &= \frac{1}{T}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k}\left(h_t^{\mathcal{F}_{k-2}}(\mathbf{x},y) + O(a_T)\right)dt\int_{\mathbb{R}^{d+1}}[K_1(\mathbf{r})K_2^{(1)}(s)]^2 d\mathbf{r}ds. \end{aligned}\tag{3.97}$$

The stationarity ane ergodicity of the process $(\mathbf{X}, Y)$ together with the linearity of the integral implies that the process $\left\{\int_{T_{k-1}}^{T_k} h_t^{\mathcal{F}_{k-2}}(\mathbf{x}, y)dt\right\}_{1\leq k\leq n}$ is stationary and ergodic (see Delecroix (1987)[45]). Then we have by the ergodic theorem for stationary processes (see Krengel (1985)[77], Theorem 4.4, p. 26).

$$\lim_{n\to\infty}\sum_{k=1}^{n}\int_{T_{k-1}}^{T_k} h_t^{\mathcal{F}_{k-2}}(\mathbf{x}, y)dt = \mathbb{E}\left[\int_{T_0}^{T_1} h_t^{\mathcal{F}_{-1}}(\mathbf{x}, y)dt\right] = \delta h(\mathbf{x}, y). \tag{3.98}$$

We conclude from equations (3.97) and (3.98) we obtain

$$\lim_{T\to\infty} W_T(\mathbf{x}, y) = h(\mathbf{x}, y)\int_{\mathbb{R}^{d+1}}[K_1(\mathbf{r})K_2^{(1)}(s)]^2 d\mathbf{r}ds. \tag{3.99}$$

Part (iii). Using successively inequalities of Holder, Markov, Jensen and then Minkowski, for any $\epsilon > 0$, any p and q such that $1/p + 1/q = 1$, we obtain

$$\begin{aligned}
\mathbb{E}[Z_{T,k}^2\mathbb{I}_{\{|Z_{T,k}|>\epsilon\}}] &\leq (\mathbb{E}[Z_{T,i}^{2q}(\mathbf{x})])^{1/q}(P\{|Z_{T,i}(\mathbf{x})| > \epsilon\})^{1/p} \leq \epsilon^{-2q/p}\mathbb{E}[|Z_{T,i}(\mathbf{x})|^{2q}]\\
&\leq \frac{\epsilon^{-2q/p}}{(Ta_T^d)^q}\mathbb{E}\left[\int_{T_{k-1}}^{T_k}\left|K_{1,t}K_{2,t}^{(1)} - \mathbb{E}[K_{1,t}K_{2,t}^{(1)}|\mathcal{F}_{t-\delta}]\right|^{2q} dt\right]\\
&\leq \frac{(2\epsilon)^{-2q/p}}{(Ta_T^{d+1})^q}\int_{T_{k-1}}^{T_k}\mathbb{E}\left[K_1^{2q}\left(\frac{\mathbf{x}-\mathbf{X}_t}{a_T}\right)\left(K_{2,t}^{(1)}\left(\frac{y-Y_t}{a_T}\right)\right)^{2q}\right]dt
\end{aligned}$$

$$\begin{aligned}
&= \frac{(2\epsilon)^{-2q/p}}{(Ta_T^{d+1})^q}\int_{T_{k-1}}^{T_k}\int_{\mathbb{R}^{d+1}} K_1^{2q}\left(\frac{\mathbf{x}-\mathbf{v}}{a_T}\right)\left(K_{2,t}^{(1)}\left(\frac{y-u}{a_T}\right)\right)^{2q} f(\mathbf{u}, v)d\mathbf{u}dvdt\\
&= \frac{(2\epsilon)^{-2q/p}\delta}{T^q a_T^{(d+1)(q-1)}}\int_{\mathbb{R}^{d+1}} K_1^{2q}(\mathbf{z})\left(K_{2,t}^{(1)}(w)\right)^{2q} f(\mathbf{x}-a_T\mathbf{z}, y-a_T w)d\mathbf{z}dw.
\end{aligned}$$

Consequently, since the integral $\int_{\mathbb{R}^{d+1}} f(\mathbf{x}-a_T\mathbf{z}, y-a_T w)d\mathbf{z}dw$ can not exceed 1, it follows from assumption (A.1)(i) that

$$n\mathbb{E}[Z_{T,i}^2(\mathbf{x})\mathbb{I}_{\{|Z_{T,i}(\mathbf{x})|>\epsilon\}}] = \frac{\epsilon^{-2q/p}\|K_1\|_\infty^{2q}\|K_2^{(1)}\|_\infty^{2q}}{(Ta_T^d)^{q-1}} = o(1). \tag{3.100}$$

the proof of the lemma is completed.

We conclude from (3.95) and (3.99) that

$$\sum_{k=1}^{n}\mathbb{E}\left[Z_{T,k}^2 \mid \mathcal{F}_{k-2}\right] = \Sigma^2(\mathbf{x}, \Theta(\mathbf{x})) \leq h(\mathbf{x}, y)\int_{\mathbb{R}^{d+1}}[k_1(\mathbf{r})K_2^{(1)}(s)]^2 d\mathbf{r}ds. \tag{3.101}$$

□

Proof Theorem 4.
By Lemma 2, we have $h_T^{(0,2)}(\mathbf{x}, \Theta_T^*(\mathbf{x})) \xrightarrow{\mathbb{P}} h_T^{(0,2)}(\mathbf{x}, \Theta(\mathbf{x}))$. Consequently, by Lemma 1, it follows that the last term of the right-hand side of Equation (3.78) goes to zero as T goes to infinity. It suffices then to combine the statements (3.101)and (3.100) to conclude the proof

$$\sqrt{Ta_T^{d+3}}\,(\Theta_T(\mathbf{x}) - \Theta(\mathbf{x})) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma^2(\mathbf{x}, \Theta(\mathbf{x})). \tag{3.102}$$

$\square$

Bibliographie

[1] Absava, R. M. (1999). *About choice of the window width in the kernel nonparametric estimate of probability density. Appl. Math. Inform.* **4**, p. 13–28, 93.

[2] Andrews, D.W.K. (1984). *Non-strong mixing autoregressive processes. J. Appl. Probab.*, **21**, 930-934.

[3] Ash, R.B. and Gardner, M.F. (1975). *Topics in stochastic processes*, *Academic Press.*

[4] Arnold, L. (1973) *Stochastic Differential Equations : Theory and Practice. New York : Wiley.*

[5] Banks, H.T. (1975) *Modeling and Control in the Biological Sciences. Lect. Notes Biomath*, **6**, Berlin : Springer-Verlag.

[6] Banon, G. (1978). *Nonparametric identification for diffusion processes. SIAM J. Control Optim.* **16**(3), p. 380–395.

[7] Banon, G. Nguyen, H.T. (1978). *Sur l'estimation récurrente de la densité et de sa dérivée pour un processus de Markov. C. R. Acad. Sci. Paris Sér. A Math.* **286(16)**, p. 691–4.

[8] Banon, G. Nguyen, H.T. (1981). *Recursive estimation in diffusion model.SIAM J. Control Optim.* **19(5)**, p. 676–85.

[9] Bergstrom, A.R. (1990) *Continuous Time Eonometrie Modelling. Oxford : Oxford University Press.*

[10] Berlinet, A., Gannoun, A. et Matzner-Løber, E. (1998). *Normalité asymptotique d'estimateurs convergent du mode conditionnel. Canad. J.Statist.* **26**, p. 365–380.

[11] Beran, J. (1994).*Statistics for long memory processes. Chapman and Hall, New York.*

[12] Bosq, D. (1971). *Contribution à la Théorie de l'estimation fonctionnelle.* PhD thesis, Paris VI.

[13] Bosq, D. (1973). *Sur l'estimation de la densité d'un processus stationnaire et mélangeant. C. R. Acad. Sci. Paris SÃ©r. A-B*, **277**. A535–A538

[14] Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes.Lecture Notes in Statistics*. **110**. Springer, New York.

[15] Bosq, D. (1997). *Parametric rates of nonparametric estimators and predictors for continuous time processes. Ann. Statist.*, **25**, p. 982–1000.

[16] Bosq, D and Blanke, D. (1997). *Accurate rates of density estimators for continuous-time processes*, *Statist. Probab. Lett.*, **33** p. 185–191.

[17] Bosq D. (1998). *Nonparametric statistics for stochastic processes : Estimation and prediction*. Springer New york.

[18] Bosq, D. and Merlevède, F. and Peligrad, M. (1999).*Asymptotic normality for density kernel estimators in discrete and continuous time*, *J. Multivariate Anal.*, **68**, p. 78–95.

[19] Bosq, D. (2000). *Linear process in function spaces. Theory and application. Lectures notes in statistics.***129** : *Springer Verlag. Berlin.*

[20] Bosq, D and Blanke, D. (2007). *Inference and prediction in large dimensions*, *Wiley Series in Probability and Statistics*, p. 316.

[21] Bosq, D and Blanke, D. (2008). *Regression estimation and prediction in continuous time*, *Journal of the Japan Statistical Society (Nihon Tôkei Gakkai Kaihô)*, **38**. p. 15–26.

[22] Bradley, R. C. (1981). *Central limit theorems under weak dependence. Journal of Multivariate Analysis*, **11**, p. 11–16.

[23] Bradley, R.C. (1983). *Asymptotic normality of some kernel-type estimators of probability density. Statistics & Probability Letters*, **1**, p. 295–300.

[24] Bradley, R. C. (2007). *Introduction to strong mixing conditions. Vol. 1*. Kendrick Press, Heber City, UT.

[25] Cacoullos, Theophilos, (1966). *Estimation of a multivariate density*, *Ann. Inst. Statist. Math.* 18, p. 179–189.

[26] Castellana, J. V. and Leadbetter, M. R. (1986). *On smoothed probability density estimation for stationary processes. Stochastic Process. Appl.*, **21**, p. 179–193.

[27] Collomb, G. (1983). *Méthodes non paramétriques en régression, analyse des séries temporelles, prédiction et discrimination. Doctorat d'Etat, Toulouse 3.*

[28] Chacón, J.E. and Tenreiro, C. (2012). *Exact and asymptotically optimal bandwidths for kernel estimation of density functionals. Methodol. Comput. Appl. Probab.* **3**,p. 523–548.

[29] Chahboun, M. (1984). *Estimation des densités de la loi et les probabilités de transition par les méthodes du noyau, des δ-suites et des fonctions orthogonales pour un processus de Markov stationnaire et mélangeant.* PhD thesis, Université de Rouen.

[30] Chanda, K.C., (1974). *Strong mixing of linear stochastic processes. J. Appl. Probab.*, **11**,p. 401–408.

[31] Chernick, M.R. (1981). *A limit theorem for the maximum of autoregressive processes with uniform marginal distributions. Ann. Probab.*, **9**, 145-149.

[32] Chèze-Payaud, N. (1994). *Nonparametric regression and prediction for continuous time process. Publ. Inst. Statistics Univ. Paris*, **38**, p. 37–58.

[33] Collomb, G. (1984). *Propriétés de convergence presque complète du prédicteur à noyau. Z. Wahrsch. Verw. Gebiete.* **66**, p. 441–460.

[34] Collomb,G. (1985). *Nonparametric regression : an up-to-date bibliography. Statistics. A Journal of Theoretical and Applied Statistics.* **16**, p. 309–324.

[35] Collomb, G., Härdle, W. et Hassani, S. (1987). *A note on prediction via estimation of the conditional mode function. J. Statist. Plann. and Inf.* **15**, p. 227–236.

[36] Cohen, J. E. (1979). *Ergodic theorems in demography. Bull. Amer. Math. Soc. (N.S.)* **1**. p. 275–295.

[37] Csörgö-Révész P. (1981). *Strong Approximation in Probability and Statistics.* Acedemic Press, New York.

[38] Debbarh, M. (2008). *Some uniform limit results in additive regression model, Comm. Statist. Theory Methods*, **37**, p. 3090–3114.

[39] Debbarh, M. and Maillot, B. (2008a). *Additive regression model for continuous time processes*, ıComm. Statist. Theory Methods, **37**. p. 2416–2432.

[40] Debbarh, M. and Maillot, B. (2008b). *Asymptotic normality of the additive regression components for continuous time processes, C. R. Math. Acad. Sci. Paris*, **346**, p. 901–906.

[41] de la Peña, V.H. and Giné, E. (1999). *Decoupling, from dependence to independence. Probability and its applications.* Springer-Verlag, New York.

[42] Delecroix, M. (1975). *Sur l'estimation des densités marginales et de transition d'un processus stationnaire et mélangeant.* PhD thesis, Université de Lille.

[43] Delecroix, M. (1979).*Sur l'estimation des densités d'un processus stationnaire et mélangeant . Publ. U.E.R. Math. Pures Appl. IRMA*, **1**(4), exp. no. I, 24. Seminar on Mathematical Statistics.

[44] Delecroix, M. (1980). *Sur l'estimation des densités d'un processus stationaire à temps continu. Publ. Inst. Statist. Univ. Paris*, **25**, p. 17–39.

[45] Delecroix, M. (1987). *Sur l'estimation et la prévision nonparamétrique des processus ergodiques.* Doctorat d'état, Université des sciences et techniques de Lille Flandre-Artois.

[46] Delecroix, M. Nogueira, M.E. Rosa, A. C. (1991). *Sur l'estimation de la densité d'observations ergodiques. Statist. Anal. Données.*, **16**, p. 25–38.

[47] Delecroix, M. and Rosa, A. C. (1996). *Nonparametric estimation of a regression function and its derivatives under an ergodic hypothesis. J. Nonparametr. Statist.*, **6**, p. 367–382.

[48] Devroye, L.P. (1978). *The uniform convergence of the Nadaraya-Watson regression function estimate. Canad. J. Statist.*, **6**, p. 179–191.

[49] Devroye, L., Györfi, L. et Lugossi, G. (1986). *A probability Theory of Pattern Recognition. Springer New-York.*

[50] Didi, S. Louani, D. (2013a) *Consistency results for the kernel density estimate on continuous time stationary and dependent data. Journal of Statistics & Probability Letters*, **83**, **4**, p. 1262-1270.

[51] Didi, S. Louani, D. (2013b). *Asymptotic results for regression function estimate on continuous time stationary and ergodic data. Statistics & Risk Modelling, Statistics & Risk Modeling.* **31**, Issue 2, p. 129–150.

[52] Didi, S. Louani, D. (2013b). *Asymptotic normality of kernel density and regression function estimators from continuous time stationary and ergodic processes.* soumis.

[53] Doukhan, P. Louhichi, S. (2001). *Functional estimation of a density under a new weak dependence condition. Scand. J. Statist.* **28(2)**, p. 325–41.

[54] Dvoretzky, A. (1972). *Asymptotic normality for sums of dependent random variables*, in : Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Univ. California, Berkeley, Calif., 1970/1971, in :

Probability theory, **II**. Univ. California Press, Berkeley, Calif., 1972, p. 513–535.

[55] Efromovich, S. (2007). *Optimal nonparametric estimation of the density of regression errors with finite support. Ann. Inst. Statist. Math.* **59(4)**. p. 617–654.

[56] Ezzahrioui, M. and Ould Saïd, E. (2008). *Asymptotic normality of a nonparametric estimator of the conditional mode function for functional data*, *Journal of Nonparametric Statistics.* 20, p. 3–18.

[57] Ezzahrioui, M. and Ould Saïd, E. (2010). *Some asymptotic results of a nonparametric conditional mode estimator for functional time-series data*, *Statistica Neerlandica. Journal of the Netherlands Society for Statistics and Operations Research*, **64**, p. 171–201.

[58] Fan, J. and Yim, T. H. (2004). *A crossvalidation method for estimating conditional densities. Biometrika.* **91(4)**. p. 819–834.

[59] Ferraty, F. et Vieu, P. (2000). *Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. C.R. Acad. Sci., Paris Sér. I,***330** , p. 139–142.

[60] Ferraty, F. Goia, A. and Vieu, P. (2002). *Régression non paramétrique pour des variables aléatoires fonctionnelles mélangeantes. C. R. Math. Acad. Sci. Paris.* **334**, p. 217–220.

[61] Ferraty, F. et Vieu, P. (2006). *Nonparametric Functional Data Analysis, Theory and Practice. Springer-Verlag. Berlin.*

[62] Ferraty, F., Laksaci, A., and Vieu, P. (2006). *Estimating some characteristics of the conditional distribution in nonparametric functional models. Stat. Inference Stoch. Process.* **9(1)**. p. 47–76.

[63] Gannoun, A. (1990) *Estimation non paramétrique de la médiane conditionnelle. Application à la prévision*, Comptes Rendus de l'Académie des Sciences. Série I. Mathématiqu , 310.

[64] Gasser, T. and Müller, H. (1984). *Estimating regression functions and their derivatives by the kernel method. Scand. J. Statist.* **11**, p. 171–185.

[65] Guégan, D., Ladoucette, S., (2001). *Non-mixing properties of long memory processes. C.R. Acad. Sci. Paris.* **333**, p. 373–376.

[66] Györfi,L., Härdle, W., Sarda, P., and Vieu, P. (1989). *Nonparametric curve estimation from time series.* Lecture Notes in Statistics, **60**. Springer-Verlag, Berlin.

[67] Hall, P. and Heyde, C. C. (1980). *Martingale limit theory and its application.* Probability and Mathematical Statistics, Academic Press Inc. New York.

[68] Hall, P. Lahiri, S.N. Truong, Y.K. (1995). *On bandwidth choice for density estimation with dependent data. Ann. Statist.* **23(6)**, p. 2241–63.

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

[70] Hale, J.K., Verduyn Lunel, S.M. (1993) *lntroduction to Functional-differential Equations. New York : Springer-Verlag.*

[71] Härdle, W. and Marron, J. S. (1985). *Optimal bandwidth selection in nonparametric regression function estimation. Ann. Statist.* **13(4)**. p. 1465–1481.

[72] Hart, J.D. Vieu, P. (1990). *Data-driven bandwidth choice for density estimation based on dependent data. Ann. Statist.* **18(2)**, p. 873–90.

[73] Hyndman, R. J., Bashtannyk, D. M., and Grunwald, G. K. (1996). *Estimating and visualizing conditional densities. J. Comput. Graph. Statist.* **5(4)**. p. 315–336.

[74] Ibragimov, I.A., Linnik, Yu.V., (1971). *Independent and Stationary Sequences of Random Variables . Wolters-Nordhoff, Groningen.*

[75] Karatzas, I., Shreve, S.E. (1991) *Brownian Motion and Stochastic Calculus. New York : Springer-Verlag.*

[76] Kim, T.Y. (1997). *Asymptotically optimal bandwidth selection rules for the kernel density estimator with dependent observations. J. Statist. Plann. Inference.* **59(2)**, p. 321–36.

[77] Krengel, U. (1985). *Ergodic Theorems, Walter de Gruyter & Co. Berlin.*

[78] Kutoyants, Y.A. (1995), *On density estimation by the observations of ergodic diffusion process.* Preprint, Universit6 du Maine.

[79] Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes.Springer Series in Statistics. Springer-Verlag London Ltd., London.*

[80] Laksaci, A. (2007). *Convergence en moyenne quadratique de l'estimateur à noyau de la densité conditionnelle avec variable explicative fonctionnelle. Ann. I.S.U.P.* **51(3)**. p. 69–80 (2008).

[81] Lange, K. (2002) *Mathematieal and Statistical Methods for Genetie Analysis. New York : Springer-Verlag.*

[82] Liebscher, E. (2001). *Estimation of the density and the regression function under mixing conditions.Statist. Decisions.* **19(1)**, p. 9–26.

[83] Linton, O., Nielsen, J. (1995). *A kernel method of estimating structured nonparametric regression based on marginal integration. Biometrika*, **82**, p. 93–100.

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

[85] Laïb, N. (1999). *Uniform consistency of the partitioning estimate under ergodic conditions. J. Austral. Math. Soc. Ser. A*, **67**, p. 1–14.

[86] Laib, N. (1999). Exponential-type inequalities for martingale difference sequences. Application to nonparametric regression estimation. *Comm. Statist. Theory Methods*, **28**(7), 1565–1576.

[87] Laïb, N. (2005). *Kernel estimates of the mean and the volatility functions in a nonlinear autoregressive model with ARCH errors. J. Statist. Plann. Inference*, **134**, p. 116–139.

[88] Laïb, N. and Louani D. (2010).*Nonparametric kernel regression estimation for functional stationary ergodic data : Asymptotic properties. Journal of Multivariate Analysis*, **101**, p. 2266–2281.

[89] Laïb, N. and Louani D. (2011). *Rates of strong consistencies of the regression function estimator for functional stationary ergodic data. J. Statist. Plann. Inference*, **141**, p. 359–372.

[90] Laksaci, A. et Yousfate, A. (2002). *Estimation fonctionnelle de la densité de l'opérateur de transition d'un processus de Markov à temps discret. C.R. Math., Acad. Sci. Paris, Sér. I,* **334**, p. 1035–1038.

[91] Leblanc, F. (1995), *Wavelet density estimation of a continuous-time process and application to diffusion process. C.R. Acad. Sci. Paris,* série **I 321**, p. 345–350.

[92] Lejeune, F. X. (2007). *Histogramme, Régressogramme et Polygone de Fréquences en Temps Continu* , PhDthesis, Université Paris 6.

[93] LU, Z. (2009). *Analyse des Processus Longue MÃ©moire Stationnaires et Non-stationnaires : Estimations, Applications et PrÃ©visions.* PhD Thesis.

[94] Mack, Y. P. and Silverman, B. W. (1982). *Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. Verw. Gebiete.* **61**,p. 405–415.

[95] Marron, J. S. (1987). A comparison of cross-validation techniques in density estimation. *Ann. Statist.* **15(1)**. p. 152–162

[96] Maslowski, B,and Pospíšil, J (2008). Ergodicity and parameter estimates for infinite-dimensional fractional Ornstein-Uhlenbeck process. *Appl. Math. Optim.* **3**, p. 401–429.

[97] Morvai, G., Yakowitz, S. and Györfi, L. (1996). *Nonparametric inference for ergodic, stationary time series. Ann. Statist.*, **24**, p. 370–379.

[98] Masry, E. (2005). *Nonparametric regression estimation for dependent functional data : asymptotic normality. Stochastic Process. Appl.*, **115**, 155–177.

[99] Nadaraya, E.N. (1964). *On a regression estimate*, *Teor. Verojatnost. i Primenen.* **9**, p 157–159.

[100] Nadaraya, E.N. (1965). *On nonparametric estimates of density functions and regression curves. Teor. Verojatnost. i Primenen.*, **10**, p. 199–203.

[101] Newey, W. K. (1994). *Kernel estimation of partial means and a general variance estimator. Econometric Theor,* 10(2), p. 233–253.

[102] Nguyen, H.T. (1979). *Density estimation in a continuous-time stationary Markov process.Ann. Statist.* **7(2)**, p. 341–348.

[103] Nguyen, H.T. Pham, T.D. (1980). *Sur l'utilisation du temps local en statistique des processus. C. R. Acad. Sci. Paris SeÂ´r. A Math.* **290(3)**, p. 165–168.

[104] Ould-Said, E. (1993) *Estimation non paramétrique du mode conditionnel. Application à la prévision*, C. R. Acad. Sci. Paris Sér. I Math., 316, p. 943–947.

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

[106] Papanicolaou, G. (1995) *Diffusion in random media.In Surveys in Applied Mathematics*, J.B. Keller, D. McLaughlin and G. Papanicolaou (Eds), Plenum Press, p. 205–255.

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

[108] Peskir, G. (1998). *The uniform mean-square ergodic theorem for wide sense stationary processes*, Stochastic Anal. Appl. , **16**, p. 697–720.

[109] Peskir, G. (2000). *From uniform laws of large numbers to uniform ergodic theorems*, Lecture Notes Series (Aarhus), **66**.

[110] Quintela del Rio, A., Vieu, P. (1997). *A nonparametric conditionnal mode estimate. Nonparametric Statistics.* **8**, p. 253–266.

[111] Ramsay, J.O. et Silverman, B.W. (2002). *Applied function data analysis : Methods and Cases Studies.Springer, New-York.*

[112] Ramsay, J.O. et Silverman, B.W. (2005). *Function data analysis.* 2nd Edition. *Springer, New-York.*

[113] Robinson, P. M. (1983). *Nonparametric estimators for time series. J. Time Ser. Anal.* **4**, p. 185–207.

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

[115] Rossi, V. (2004). *Filtrage non linéaire par noyaux de convolution Application à un procédé de dépollution biologique. Thèse du Doctorat en Science, École Nationale Supérieure Agronomique de Montpellier.*

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

[117] Roussas, G. (1968). *On some properties of nonparametric estimates of probability density functions.Bull. Soc. Math. Grèce (N.S.),* **9**, p. 29–43.

[118] Roussas, G. (1969) *Nonparametric estimation of the transition distribution function of a Markov process*, Annals of Mathematical Statistics, 40, p. 1386–1400.

[119] Roussas, G.G. (1990). *Asymptotic normality of the kernel estimate under dependence conditions : application to hazard rate. Journal of Statistical Planning and Inference*, **25**, p. 81–104.

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

[121] Sarda, P. (1991). *Estimating smooth distribution functions. In Nonparametric functional estimation and related topics (Spetses, 1990)*, volume **335** of NATO *Adv.Sci. Inst. Ser. C Math. Phys. Sci.* p. 261–270. Kluwer Acad. Publ., Dordrecht.

[122] Sarda, P. and Vieu, P. (1991). *Smoothing parameter selection in hazard estimation. Statist. Probab. Lett.* **11(5)**. p. 429–434.

[123] Silverman, B.W. and Jones, MC. 1989 E. Fix and J. L. Hodges (1951). an important contribution to nonparametric discriminant analysis and density estimation. *Int. Stat. Rev.* 57(3) 233–247.

[124] Stone, C. J. (1985). *Additive regression and other nonparametric models. Ann. Statist.*, **13**(2), 689–705.

[125] Stute, W. (1982). *The oscillation behavior of empirical processes. Ann. Prob.*, **10**, p. 86–107.

[126] Watson, G. S., (1964) *Smooth regression analysis* 26 of *Sankhyā Ser. A*, Sankhyā (Statistics). The Indian Journal of Statistics. Series A, p. 359–372.

[127] Wu, W.B., (2003) *Nonparametric Estimation For stationary Processes.* Technical Report 536 *University of Chicago.*

[128] Wu, W. B., Huang, Y. and Huang, Y. (2010). *Kernel estimation for time series : An asymptotic theory. Stochastic Processes and their Applications*, **120**, p. 2412–2431

[129] Yakowitz, S., Györfi, L., Kieffer, J. and Morvai, G. (1999). *Strongly consistent nonparametric forecasting and regression for stationary ergodic sequences. J. Multivariate Anal.*, **71**, p. 24–41.

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

[131] Youndjé, É. Sarda, P. and Vieu, P. (1993a) *Estimateur à noyau d'une densité conditionnelle : choix de la fenêtre pour des observations dépendantes. C. R. Acad. Sci. Paris SÃ©r. I Math.* **316(9)**. p. 935–938.

[132] Youndjé, É. Sarda, P. and Vieu, P. (1994). *Validation croisée pour l'estimation non-paramétrique de la densité conditionnelle. Publ. Inst. Statist. Univ. Paris.* **38(1)**. p. 57–80,

[133] Youndjé, É. Wells, M.T. (2008). *Optimal bandwidth selection for multivariate kernel deconvolution density estimation.TEST*,**17(1)**. p. 138–162.

[134] Yoshihara, K.I. (1994). *Weakly dependent stochastic sequences and their application.* **IV** : *Curve estimation based on on weakly dependent data. Sanseido, Tokyo.*