Estimation récursive de fonctionnelles

Baba Thiam

THÈSE

présentée devant

L' UNIVERSITE DE VERSAILLES SAINT-QUENTIN

pour obtenir le titre de :

DOCTEUR DE L'UNIVERSITÉ DE VERSAILLES-SAINT-QUENTIN

Mention : Mathématiques et Applications

PAR

Baba THIAM

TITRE DE LA THÈSE :

ESTIMATION RÉCURSIVE DE FONCTIONNELLES

Soutenue publiquement le 05 Décembre 2006 devant le jury composé de :

M.	Gérard	BIAU
M.	Paul	DOUKHAN
M.	Fabrice	GAMBOA
M.	Abdelkader	MOKKADEM
Mme	Mariane	PELLETIER
M.	Alain	ROUAULT

RÉSUMÉ

L'objet de cette thèse est l'étude du comportement asymptotique d'estimateurs à noyau d'une densité de probabilité et de ses dérivées, d'une fonction de régression, ainsi que du mode et de la valeur modale d'une densité de probabilité. Le but est d'établir certaines propriétés des estimateurs à noyau récursifs ou semi-récursifs afin de comparer leur comportement asymptotique à celui des estimateurs classiques. Dans le premier chapitre, nous établissons des principes de grandes déviations (PGD) et des principes de déviations modérées (PDM) pour l'estimateur récursif d'une densité de probabilité et pour ses dérivées. Il s'avère que, dans les principes de déviations vérifiés par les estimateurs des dérivées, la fonction de taux est toujours une fonction quadratique, que les déviations soient grandes ou modérées. Contrairement, pour l'estimateur de la densité, les fonctions de taux qui apparaissent sont de nature différente selon que les déviations sont grandes ou modérées. Les fonctions de taux qui apparaissent tant dans les PGD pour les dérivées que dans les PDM pour la densité et pour les dérivées sont plus grandes dans le cas où l'estimateur récursif est utilisé. Dans le deuxième chapitre, nous établissons des PGD et des PDM pour des estimateurs à noyau d'une fonction de régression. Nous généralisons les résultats déjà obtenus dans le cas unidimensionnel pour l'estimateur de Nadaraya-Watson. Nous étudions ensuite le comportement en déviations de la version semi-récursive de cet estimateur en établissant des PGD et des PDM. Les fonctions de taux qui apparaissent dans les PDM sont plus grandes pour l'estimateur semi-récursif que pour l'estimateur classique. Dans le troisième chapitre, nous nous intéressons à l'estimation jointe du mode et de la valeur modale d'une densité de probabilité basée sur l'estimateur à noyau récursif de la densité. Nous étudions la vitesse de convergence en loi et presque sûre du couple formé par ces deux estimateurs. Pour estimer simultanément les deux paramètres de façon optimale, il faut utiliser des fenêtres différentes pour définir chacun des deux estimateurs. Les estimateurs semi-récursifs conduisent à des variances asymptotiques plus petites que les estimateurs classiques.

ABSTRACT

The aim of this thesis is the study of the asymptotic behaviour of the kernel estimator of a probability density function and its derivatives, of a regression function, as well as of the location and of the size of the mode of a probability density. The goal is to establish several properties of the recursive or semi-recursive kernel estimators in order to compare their asymptotic behaviour with that of the classical estimators. In the first chapter, we establish a large deviations principle (LDP) and a moderate deviations principle (MDP) for the recursive estimator of a probability density and for its derivatives. It turns out that, in the deviations principles for the derivatives estimators, the rate function is always quadratic, the deviations being either large or moderate. On the other hand, for the density estimator, the rate function which appears is of different nature according to whether the deviations are large or moderate. The rate functions which appear in the LDP for the derivatives and in the MDP for the density and its derivatives are larger in the case the recursive estimator is used. In the second chapter, we establish LDP and MDP for kernel estimators of the regression. We generalize the results already obtained in the unidimensional case for the Nadaraya-Watson estimator. We then study the behaviour in deviations of the semi-recursive version of this estimator by establishing a LDP and MDP. The rate function which appears in the MDP are larger for the semi-recursive estimator than for the classical estimator. In the third chapter, we are interested in the joint estimation of the location and of the size of the mode of a probability density based on the recursive kernel density estimator. We study the weak and almost sure convergence rates of the couple formed by these two estimators. To estimate the two parameters simultaneously in an optimal way, it is necessary to use different bandwidths to define each of the two estimators. The semi-recursive estimators lead to asymptotic variances smaller than the classical estimators.

Remerciements

Je suis particulièrement reconnaissant à mes directeurs de thèse, Mariane Pelletier et Abdelkader Mokkadem qui m'ont fait bénéficier de leur expérience ; ils ont dirigé ce travail avec enthousiasme, disponibilité et une grande générosité. Ce travail leur doit beaucoup.

Un grand merci aux rapporteurs Gérard Biau et Paul Doukhan d'avoir eu la gentillesse et la patience de lire en détails mes travaux.

Merci à Fabrice Gamboa, Alain Rouault de me faire l'honneur de faire partie de mon jury.

Je remercie tous les membres du laboratoire LMV de Versailles qui m'ont toujours accueilli avec chaleur et amitié pendant toutes ces années de thèse.

Enfin j'adresse une pensée à ma famille pour sa patience, ses encouragements et son soutien ; qu'elle soit assurée de mon affection.

Table des matières

Chapitre 1

INTRODUCTION

1.1 Introduction

L'objectif de cette thèse est l'étude du comportement asymptotique d'estimateurs à noyau d'une densité de probabilité et de ses dérivées, d'une fonction de régression, ainsi que du mode et de la valeur modale d'une densité de probabilité. Les estimateurs à noyaux classiques ont été introduits par Rosenblatt pour estimer des densités de probabilité, par Parzen pour estimer le mode d'une densité de probabilité et par Nadaraya et Watson pour estimer une fonction de régression. Le comportement asymptotique de ces estimateurs (consistance, vitesse de convergence faible et forte, principes de déviations grandes et modérées) a été étudié par de nombreux auteurs. Des versions récursives ou semi-récursives de ces estimateurs ont été introduites, mais leur comportement asymptotique a été relativement moins étudié que celui des estimateurs classiques. Le but de ce travail est d'établir certaines propriétés des estimateurs à noyau récursifs ou semi-récursifs afin de comparer leur comportement asymptotique à celui des estimateurs classiques.

Avant de présenter nos résultats de façon détaillée, nous en donnons tout d'abord les grandes lignes.

Dans le deuxième chapitre, nous nous intéressons à l'estimateur récursif f_n d'une densité de probabilité f introduit par Wolverton et Wagner. Plus précisément, nous établissons des principes de grandes déviations et des principes de déviations modérées ponctuelles et uniformes pour f_n et pour ses dérivées. Il s'avère que, dans les principes de déviations vérifiés par les estimateurs des dérivées, la fonction de taux est toujours une fonction quadratique, que les déviations soient grandes ou modérées. A contrario, pour l'estimateur de la densité, les fonctions de taux qui apparaissent sont de nature différente selon que les déviations sont grandes ou modérées (la fonction de taux étant quadratique dans ce dernier cas). Cette différence dans le comportement en grandes déviations entre f_n et ses dérivées n'est pas due au fait que cet estimateur est récursif ; elle a été mise en évidence dans [27] pour l'estimateur de Rosenblatt. Les principales différences dans le comportement en déviations entre l'estimateur récursif et l'estimateur de Rosenblatt sont les suivantes. D'une part, dans les principes de grandes déviations pour la densité elle-même, l'expression de la fonction de taux est plus complexe dans le cas récursif que dans le cas non-récursif (et il semble difficile de comparer plus précisément ces deux fonctions). D'autre part, les fonctions de taux qui apparaissent tant dans les principes de grandes déviations pour les dérivées que dans les principes de déviations modérées pour la densité et pour les dérivées sont plus grandes dans le cas où l'estimateur récursif est utilisé. Ceci est bien-sûr totalement expliqué par le fait que la variance de l'estimateur récursif est plus petite que celle de l'estimateur de Rosenblatt.

Dans le troisième chapitre, nous établissons des principes de grandes déviations et des principes de déviations modérées pour des estimateurs à noyau d'une fonction de régression r. Des résultats de grandes déviations pour l'estimateur classique à noyau r_n introduit par Nadaraya et Watson ont été démontrés dans [22] et [15] dans le cas unidimensionnel. Nous généralisons ces résultats au cas multidimensionnel. Nous établissons également des principes de déviations modérées pour r_n. De plus, nous étudions le comportement en déviations de la version semi-récursive $\tilde{r}_n$ de l'estimateur de Nadaraya-Watson en établissant des principes de grandes déviations et des principes de déviations modérées. Les principales différences dans le comportement en déviations entre l'estimateur classique r_n et l'estimateur semi-récursif $\tilde{r}_n$ sont du même type que celles observées dans le cadre de l'estimation d'une densité de probabilité. D'une part, dans les principes de grandes déviations, l'expression de la fonction de taux est plus complexe pour l'estimateur semi-récursif $\tilde{r}_n$ que pour l'estimateur classique r_n. D'autre part, les fonctions de taux qui apparaissent dans les principes de déviations modérées sont plus grandes pour l'estimateur semi-récursif que pour l'estimateur classique.

Dans le quatrième chapitre, nous nous intéressons à l'estimation jointe du mode θ et de la valeur modale μ d'une densité de probabilité. L'estimateur à noyau du mode introduit par Parzen a été largement étudié, mais, curieusement, très peu de résultats existent sur l'estimateur de la valeur modale. Les résultats obtenus aux chapitres 2 et 3 montrant que les estimateurs récursifs ou semi-récursifs sont plus concentrés sur le paramètre à estimer que les estimateurs classiques, nous privilégions, dans ce quatrième chapitre, l'étude des estimateurs à noyau semi-récursifs θ_n et μ_n du mode et de la valeur modale. Nos principaux résultats sont un théorème de la limite centrale et une loi du logarithme itérée compacte pour le couple d'estimateurs (θ_n, μ_n). Nous montrons en particulier que pour estimer simultanément les deux paramètres θ et μ de façon optimale, il faut utiliser des fenêtres différentes pour définir chacun des deux estimateurs. Nous donnons également le théorème de la limite centrale vérifié par le couple d'estimateurs (θ_n^*, μ_n^*) définis à l'aide de l'estimateur (non récursif) de Rosenblatt de la densité, et remarquons à nouveau que les estimateurs semi-récursifs conduisent à des variances asymptotiques plus petites que les estimateurs classiques. Les résultats de simulations comparant les niveaux empiriques des régions de confiance du couple (θ, μ) construits en utilisant d'une part soit les estimateurs classiques (θ_n^*, μ_n^*), soit les estimateurs récursifs (θ_n, μ_n) et, d'autre part, soit la même fenêtre pour estimer θ et μ, soit des fenêtres différentes corroborent nos résultats théoriques.

Nous présentons maintenant de manière détaillée le contenu des trois chapitres de cette thèse.

1.2 Grandes déviations et déviations modérées pour l'estimateur à noyau récursif de la densité et de ses dérivées partielles.

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et de même loi, à valeurs dans $\mathbb{R}^d$ et de densité de probabilité f. L'estimateur à noyau de f introduit par Rosenblatt est défini par

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

où le noyau K est une fonction telle que $\int_{\mathbb{R}^d} K(x)dx = 1$ et où la fenêtre (h_n) est une suite de nombres réels positifs qui tend vers zéro. Cet estimateur a été introduit par Rosenblatt [33] et par Parzen [30]. La consistance faible et forte a été étudié par plusieurs auteurs. On peut en citer quelques uns dont Singh ([39],[40]), Rüschendorf [35], Silverman [38], Giné et Guillou [13], Hall [14], Stute [41] et Arcones [2]. Des principes de grandes déviations ont été établis par Louani [21], Worms [49], Gao [12] et Mokkadem et al. [27], des principes de déviations modérées montrés par Gao [12] et Mokkadem et al. [27]. Des résultats de grandes déviations et de déviations modérées pour les dérivées de f_n^* ont été prouvés par Mokkadem et al. [27]. Mentionnons que des principes de grandes déviations pour le risque L_1 ont été obtenus par Louani [23] et Lei et al. [19] dans le cas i.i.d., par Lei et Wu ([17], [18]) et Lei [20] dans le cas markovien ou mélangeant.

Wolverton et Wagner [47] ont introduit une version récursive de l'estimateur à noyau de Rosenblatt. Celle-ci s'écrit récursivement en posant $f_0 = 0$ et, pour tout $n \geq 1$,

$$f_n(x) = \frac{n-1}{n} f_{n-1}(x) + \frac{1}{nh_n^d} K\left(\frac{x - X_n}{h_n}\right)$$

ou sous la forme compacte

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}K\left(\frac{x-X_i}{h_i}\right).$$

La consistance faible et forte de cet estimateur a été étudié entre autres par Devroye [6], Menon et al. [25] et Wertz [46]. Sa vitesse de convergence presque sûre a été étudiée par Wegman et Davies [45], et par Roussas [34].

L'objectif de ce chapitre est d'établir des principes de grandes déviations (PGD) et de déviations modérées (PDM) ponctuels et uniformes pour l'estimateur f_n et pour ses dérivées partielles.

Rappelons tout d'abord les notions de grandes déviations et de déviations modérées.

Définition 1 *Soit I une fonction définie sur $\mathbb{R}^m$ et à valeurs dans $[0,+\infty]$.*

- *On dit que I est une fonction de taux si ses ensembles de niveau sont fermés ; c'est-à-dire pour tout $\alpha \in \mathbb{R}$, l'ensemble $\{x,\ I(x) \leq \alpha\}$ est fermé.*
- *On dit que I est une bonne fonction de taux si ses ensembles de niveau sont compacts.*

Définition 2 *Une suite de vecteurs $(Z_n)_{n\geq 1}$ de $\mathbb{R}^m$ satisfait un PGD de vitesse (ν_n) et de fonction de taux I si*

(a) (ν_n) est une suite positive telle que $\lim_{n\to\infty}\nu_n = +\infty$;

(b) Pour tout ouvert U de $\mathbb{R}^m$, et pour tout fermé V de $\mathbb{R}^m$,

$$\begin{aligned}\liminf_{n\to\infty}\nu_n^{-1}\log\mathbb{P}\left[Z_n\in U\right] &\geq -\inf_{x\in U} I(x)\\ \limsup_{n\to\infty}\nu_n^{-1}\log\mathbb{P}\left[Z_n\in V\right] &\leq -\inf_{x\in V} I(x).\end{aligned}$$

Définition 3 *Soit (v_n) une suite réelle telle que $\lim_{n\to\infty} v_n = \infty$. On dit qu'une suite de vecteurs $(Z_n)_{n\geq 1}$ de $\mathbb{R}^m$ satisfait un PDM si la suite $(v_nZ_n)_{n\geq 1}$ satisfait un PGD.*

Dans la partie 1.2.1, nous donnons un PGD ponctuel pour l'estimateur récursif de la densité, dans la partie 1.2.2, un PGD pour les estimateurs récursifs des dérivées. La partie 1.2.3 est consacrée au PDM ponctuel (pour les estimateurs de la densité et des dérivées). Enfin, des principes de déviations (grandes et modérées) uniformes (pour les estimateurs de la densité et des dérivées) sont énoncés dans la partie 1.2.4.

1.2.1 PGD ponctuel pour l'estimateur récursif de la densité

Nos hypothèses sur le noyau K et la fenêtre (h_n) sont les suivantes.

(H1) $K : \mathbb{R}^d \to \mathbb{R}$ est une fonction bornée et intégrable, $\int_{\mathbb{R}^d} K(z)dz = 1$ et $\lim_{\|z\|\to\infty} K(z) = 0$.

(H2) $h_n = cn^{-a}$ avec $0 < a < 1/d$ et $c > 0$.

Avant de donner les résultats, nous introduisons la fonction de taux pour le PGD de l'estimateur de la densité. Soit $\psi_a : \mathbb{R} \to \mathbb{R}$ et $I_a : \mathbb{R} \to \mathbb{R}$ les fonctions définies par :

$$\psi_a(u) = \int_{[0,1]\times\mathbb{R}^d} s^{-ad}\left(e^{s^{ad}uK(z)} - 1\right) dsdz \quad \text{et} \quad I_a(t) = \sup_{u\in\mathbb{R}}\{ut - \psi_a(u)\}$$

(où $s \in [0,1]$, $z \in \mathbb{R}^d$). La proposition suivante donne les propriétés des fonctions ψ_a et I_a et précise le comportement de la fonction de taux I_a en fonction de la nature du noyau. Notons λ la mesure de Lebesgue sur $\mathbb{R}^d$ et posons

$$S_+ = \left\{x \in \mathbb{R}^d; K(x) > 0\right\} \quad \text{et} \quad S_- = \left\{x \in \mathbb{R}^d; K(x) < 0\right\}.$$

Proposition 1 *Supposons (H1) vérifiée.*

(i) ψ_a *est strictement convexe, deux fois continuement différentiable sur* $\mathbb{R}$*, et* I_a *est une bonne fonction de taux sur* $\mathbb{R}$*.*

(ii) Si $\lambda(S_-) = 0$*, alors* $I_a(t) = +\infty$ *lorsque* $t < 0$*,* $I_a(0) = \lambda(S_+)$*,* I_a *est strictement convexe sur* $\mathbb{R}$ *et continue sur* $]0, +\infty[$*, et pour tout* $t > 0$*,*

$$I_a(t) = t\left(\psi_a'\right)^{-1}(t) - \psi_a\left(\left(\psi_a'\right)^{-1}(t)\right). \tag{1.1}$$

(iii) Si $\lambda(S_-) > 0$*, alors* I_a *est fini et strictement convexe sur* $\mathbb{R}$ *et* (1.1) *est satisfaite pour tout* $t \in \mathbb{R}$*.*

(iv) Dans tous les cas, le minimum strict de I_a *est atteint pour* $I_a(1) = 0$*.*

On peut maintenant énoncer un PGD pour l'estimateur de la densité.

Théorème 1 ***(PGD ponctuel pour l'estimateur de la densité)***
Supposons (H1)-(H2) vérifiées, et supposons que f *est continue en* x*. Alors, la suite* $(f_n(x) - f(x))$ *satisfait un PGD de vitesse* (nh_n^d) *et de bonne fonction de taux définie de la façon suivante :*

$$\begin{cases} Si \quad f(x) \neq 0, & I_{a,x} : t \mapsto f(x) I_a\left(1 + \frac{t}{f(x)}\right) \\ Si \quad f(x) = 0, & I_{a,x}(0) = 0 \quad et \quad I_{a,x}(t) = +\infty \quad pour \quad t \neq 0. \end{cases}$$

Pour conclure cette partie, soulignons que la différence dans le comportement en grandes déviations entre l'estimateur f_n^* de Rosenblatt et l'estimateur récursif f_n de Wolverton et Wagner apparait dans l'expression de la fonction de taux. Rappelons en effet que, sous l'hypothèse (H1), si f est continue en x et si la fenêtre vérifie les conditions $\lim_{n\to\infty} h_n = 0$ et $\lim_{n\to\infty} nh_n^d = \infty$, alors la suite $(f_n^*(x) - f(x))$ satisfait un PGD de vitesse (nh_n^d) et de bonne fonction de taux définie par

$$\begin{cases} \text{Si} \quad f(x) \neq 0, & I_x^* : t \mapsto f(x) I^*\left(1 + \frac{t}{f(x)}\right) \\ \text{Si} \quad f(x) = 0, & I_x^*(0) = 0 \quad \text{et} \quad I_x^*(t) = +\infty \quad \text{pour} \quad t \neq 0, \end{cases}$$

où $I^*(t) = \sup_{u\in\mathbb{R}}\{ut - \psi^*(t)\}$, la fonction ψ^* étant définie par

$$\psi^*(u) = \int_{\mathbb{R}^d} \left(e^{uK(z)} - 1\right) dz$$

(voir Mokkadem et al. [27]). L'expression de la fonction de taux est ainsi plus compliquée pour l'estimateur récursif que pour l'estimateur classique. En particulier, dans le cas où l'on considère l'estimateur récursif, on a besoin de connaître la vitesse de convergence de la fenêtre vers zéro, l'exposant a qui apparait dans la définition de la fenêtre $(h_n) = (n^{-a})$ étant présent dans l'expression de la fonction de taux I_a (contrairement au cas où l'on considère l'estimateur de Rosenblatt).

1.2.2 PGD ponctuel pour les estimateurs récursifs des dérivées

L'objectif de cette partie est d'énoncer un PGD ponctuel pour les estimateurs récursifs des dérivées de f, et de comparer ce PGD aux cas où l'on considère des estimateurs non récursifs.

Soit $[\alpha] = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}^d$; posons $|\alpha| = \alpha_1 + \cdots + \alpha_d$ et notons $\partial^{[\alpha]} f$ la $[\alpha]$-ième dérivée partielle de f

$$\partial^{[\alpha]} f(x) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(x).$$

Lorsque le noyau K est choisi tel que $\partial^{[\alpha]}K \not\equiv 0$, et lorsque la fenêtre est choisie telle que $\lim_{n\to\infty} nh_n^{d+2|\alpha|} = \infty$, $\partial^{[\alpha]}f$ peut être estimée à l'aide de l'estimateur à noyau non récursif

$$\partial^{[\alpha]}f_n^*(x) = \frac{1}{nh_n^{d+|\alpha|}}\sum_{i=1}^{n}\partial^{[\alpha]}K\left(\frac{x-X_i}{h_n}\right)$$

(f_n^* étant l'estimateur de Rosenblatt) où à l'aide de l'estimateur à noyau récursif

$$\partial^{[\alpha]}f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^{d+|\alpha|}}\partial^{[\alpha]}K\left(\frac{x-X_i}{h_n}\right)$$

(f_n étant l'estimateur de Wolverton et Wagner). Pour énoncer un PGD pour $\partial^{[\alpha]}f_n$ ($|\alpha| \geq 1$), nous avons besoin des hypothèses suivantes.

(H3) $h_n = h(n)$ où la fonction h est localement bornée et à variations régulières d'ordre $(-a)$, $0 < a < 1/(d+2|\alpha|)$.

(H4) i) K est $|\alpha|$-fois différentiable sur $\mathbb{R}^d$ et $\lim_{\|x\|\to\infty}\|D^{(j)}K(x)\| = 0$ pour tout $j \in \{0,\ldots,|\alpha|-1\}$.
ii) $\partial^{[\alpha]}K : \mathbb{R}^d \to \mathbb{R}$ est une fonction intégrable, bornée et $\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(x)\right]^2 dx \neq 0$.

(H5) f est $|\alpha|$-fois différentiable sur $\mathbb{R}^d$ et sa différentielle d'ordre j, $D^{(j)}f$, est bornée sur $\mathbb{R}^d$ pour tout $j \in \{0,\ldots,|\alpha|-1\}$.

Théorème 2 ***(PGD ponctuel pour l'estimateur des dérivées)***
Soit $|\alpha| \geq 1$; supposons (H1), (H3)-(H5) vérifiées et $\partial^{[\alpha]}f$ continue en x. Alors, la suite $\left(\partial^{[\alpha]}f_n(x) - \partial^{[\alpha]}f(x)\right)$ satisfait un PGD de vitesse $(nh_n^{d+2|\alpha|})$ et de bonne fonction de taux $J_{a,[\alpha],x}$ définie par :

$$\begin{cases} Si \quad f(x) \neq 0, \quad J_{a,[\alpha],x} : t \mapsto \frac{t^2(1+a(d+2|\alpha|))}{2f(x)\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(z)\right]^2 dz} \\ Si \quad f(x) = 0, \quad J_{a,[\alpha],x}(0) = 0 \quad et \quad J_{a,[\alpha],x}(t) = \infty \quad pour \quad t \neq 0. \end{cases} \tag{1.2}$$

Sous des hypothèses analogues (excepté l'hypothèse sur la fenêtre qui est simplement $\lim_{n\to\infty} h_n = 0$ et $\lim_{n\to\infty} nh_n^{d+2|\alpha|} = \infty$), il est montré dans Mokkadem et al. [27] que la suite $\left(\partial^{[\alpha]}f_n^*(x) - \partial^{[\alpha]}f(x)\right)$ satisfait un PGD de vitesse $(nh_n^{d+2|\alpha|})$ et de bonne fonction de taux $J^*_{[\alpha],x}$ définie par :

$$\begin{cases} \text{Si} \quad f(x) \neq 0, \quad J^*_{[\alpha],x} : t \mapsto \frac{t^2}{2f(x)\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(z)\right]^2 dz} \\ \text{Si} \quad f(x) = 0, \quad J^*_{[\alpha],x}(0) = 0 \quad \text{et} \quad J^*_{[\alpha],x}(t) = \infty \quad \text{pour} \quad t \neq 0. \end{cases} \tag{1.3}$$

La différence dans le comportement en grandes déviations entre l'estimateur non-récursif $\partial^{[\alpha]}f_n^*$ et l'estimateur récursif $\partial^{[\alpha]}f_n$ est le facteur $1+a(d+2|\alpha|)$ qui apparait dans l'expression de la fonction de taux $J_{a,[\alpha],x}$ mais pas dans celle de $J^*_{[\alpha],x}$. En tout point x tel que $f(x) \neq 0$, on note que, pour tout t, $J_{a,[\alpha],x}(t) > J^*_{[\alpha],x}(t)$. Ceci est bien-sûr à relier au fait que la variance de l'estimateur récursif $\partial^{[\alpha]}f_n$ est plus petite que celle de l'estimateur non-récursif $\partial^{[\alpha]}f_n^*$.

1.2.3 PDM ponctuel pour les estimateurs récursifs de la densité et de ses dérivées

Soit (v_n) une suite positive. Pour donner le comportement en déviations modérées des estimateurs récursifs de la densité et de ses dérivées, nous avons besoin des hypothèses suivantes.

(H6) $\lim_{n\to\infty} v_n = \infty$ et $\displaystyle\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} = 0$.

(H7) i) Il existe un entier $q \geq 2$ tel que $\forall s \in \{1, \ldots, q-1\}, \forall j \in \{1, \ldots, d\}$, $\displaystyle\int_{\mathbb{R}^d} y_j^s K(y) dy_j = 0$, et $\displaystyle\int_{\mathbb{R}^d} \left| y_j^q K(y) \right| dy < \infty$.

ii) $\displaystyle\lim_{n\to\infty} \frac{v_n}{n} \sum_{i=1}^{n} h_i^q = 0$.

iii) $\partial^{[\alpha]} f$ est q-fois différentiable sur $\mathbb{R}^d$ et $M_q = \sup_{x\in\mathbb{R}^d} \|D^q \partial^{[\alpha]} f(x)\| < +\infty$.

Remarque 1 *Lorsque $h_n = O(n^{-a})$, avec $0 < a < 1/(d+2|\alpha|)$, les hypothèses (H6) et (H7)ii) sont satisfaites pour $(v_n) \equiv (n^b)$ pour tout $b \in]0, \min\{aq\ ;\ (1 - a(d+2|\alpha|))/2\}[$.*

Le théorème suivant établit le comportement en déviations modérées des estimateurs récursifs de la densité et de ses dérivées.

Théorème 3 ***(PDM ponctuel)***
Pour $|\alpha| = 0$, supposons les hypothèses (H1), (H3), (H6) et (H7) vérifiées ; pour $|\alpha| \geq 1$, supposons (H1), (H3)-(H7) vérifiées. Si $\partial^{[\alpha]} f$ est q-fois différentiable en x, alors la suite $\left(v_n \left(\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right)\right)$ satisfait un PGD de vitesse $\left(nh_n^{d+2|\alpha|}/v_n^2\right)$ et de bonne fonction de taux $J_{a,[\alpha],x}$ définie en (1.2).

Sous des hypothèses analogues (excepté toujours l'hypothèse sur la fenêtre), les estimateurs non-récursifs $\partial^{[\alpha]} f_n^*$ satisfont également un PDM. Plus précisément, il est montré dans Mokkadem et al. [27] que la suite $\left(v_n \left(\partial^{[\alpha]} f_n^*(x) - \partial^{[\alpha]} f(x)\right)\right)$ satisfait un PGD de vitesse $\left(nh_n^{d+2|\alpha|}/v_n^2\right)$ et de bonne fonction de taux $J^*_{[\alpha],x}$ définie en (1.3). On remarque à nouveau que la fonction de taux est plus grande dans le cas où les estimateurs récursifs $\partial^{[\alpha]} f_n$ sont utilisés.

Soulignons que, dans le comportement en déviations, il existe une grande différence entre l'estimateur de la densité (qu'il soit récursif ou non) et les estimateurs des dérivées (qu'ils soient récursifs ou non). En effet, les fonctions de taux qui apparaissent dans les principes de déviations - qu'elles soient grandes ou modérées - des estimateurs des dérivées sont toutes quadratiques. A contrario, la nature des fonctions de taux qui apparaissent dans les principes de déviations des estimateurs de la densité est totalement différente selon que les déviations sont grandes ou modérées. Ce phénomène s'explique de la façon suivante. (Nous donnons naturellement l'explication dans le cadre de l'estimateur récursif, mais l'explication est la même dans le cas du non-récursif). Remarquons que

$$\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x) = h_n^{-|\alpha|} V_n^{[\alpha]}(x) + B_n^{[\alpha]}(x), \tag{1.4}$$

avec

$$\begin{aligned} V_n^{[\alpha]}(x) &= \frac{h_n^{|\alpha|}}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \left(\partial^{[\alpha]} K\left(\frac{x - X_i}{h_i}\right) - \mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x - X_i}{h_i}\right)\right]\right), \\ B_n^{[\alpha]}(x) &= \mathbb{E}\left[\partial^{[\alpha]} f_n(x)\right] - \partial^{[\alpha]} f(x). \end{aligned}$$

La décomposition (1.4) implique que le comportement en grandes déviations de $\partial^{[\alpha]} f_n$ est donné par celui du terme de variance $h_n^{-|\alpha|} V_n^{[\alpha]}$ dès que $\partial^{[\alpha]} f_n$ est asymptotiquement sans biais, *i.e.* dès que $\lim_{n\to\infty} B_n^{[\alpha]}(x) = 0$. D'une part, en posant $|\alpha| = 0$ dans (1.4), on voit que les grandes déviations pour l'estimateur de la densité sont déduites du PGD pour le terme de variance $V_n^{[0]}(x)$. D'autre part, lorsque $|\alpha| \geq 1$, on voit que le PGD pour l'estimateur de la $[\alpha]$-ième dérivée est obtenu en prouvant un PGD pour le terme de variance $h_n^{-|\alpha|} V_n^{[\alpha]}(x)$; comme $\lim_{n\to\infty} h_n^{-|\alpha|} = +\infty$, ce PGD est en fait un PDM pour la suite $V_n^{[\alpha]}(x)$, ce qui explique que la fonction de taux soit alors quadratique.

1.2.4 PGD et PDM uniformes pour les estimateurs récursifs de la densité et de ses dérivées

Pour établir des principes de déviations uniformes pour les estimateurs récursifs de la densité et de ses dérivées, nous rajoutons les hypothèses suivantes :

(H8) i) Il existe $\xi > 0$ tel que $\int_{\mathbb{R}^d} \|x\|^{\xi} f(x)dx < \infty$.
 ii) f est uniformément continue.

(H9) i) $\partial^{[\alpha]}K$ est une fonction höldérienne.
 ii) Il existe $\gamma > 0$ tel que la fonction $z \mapsto \|z\|^{\gamma}\partial^{[\alpha]}K(z)$ soit bornée.

(H10) $\lim_{n\to\infty} \frac{v_n^2 \log(1/h_n)}{nh_n^{d+2|\alpha|}} = 0$ et $\lim_{n\to\infty} \frac{v_n^2 \log v_n}{nh_n^{d+2|\alpha|}} = 0$.

(H11) i) Il existe $\zeta > 0$ tel que $\int_{\mathbb{R}^d} \|z\|^{\zeta} |K(z)| dz < \infty$.
 ii) Il existe $\eta > 0$ tel que la fonction $z \mapsto \|z\|^{\eta}\partial^{[\alpha]}f(z)$ soit bornée.

Remarque 2 *Lorsque $h_n = O(n^{-a})$ avec $a \in]0, 1/(d + 2|\alpha|)[$, l'hypothèse (H10) est satisfaite par exemple pour $(v_n) \equiv (n^b)$ pour tout $b \in]0, (1 - a(d + 2|\alpha|))/2[$.*

Soit $U \subseteq \mathbb{R}^d$; pour énoncer les principes de déviations grandes et modérées uniformes pour les estimateurs de la densité et de ses dérivées sur l'ensemble U de façon compacte, nous considérons les grandes déviations comme le cas spécial où $(v_n) \equiv 1$ et posons :

$$\begin{aligned} g_U(\delta) &= \begin{cases} \|f\|_{U,\infty} I_a\left(1 + \frac{\delta}{\|f\|_{U,\infty}}\right) & \text{si } |\alpha| = 0 \text{ et } (v_n) \equiv 1 \\ \frac{\delta^2(1+a(d+2|\alpha|))}{2\|f\|_{U,\infty}\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K\right]^2(z)dz} & \text{sinon,} \end{cases} \\ \tilde{g}_U(\delta) &= \min\{g_U(\delta), g_U(-\delta)\}, \end{aligned}$$

où $\|f\|_{U,\infty} = \sup_{x\in U} |f(x)|$.

Remarque 3 *Les fonctions $g_U(\cdot)$ et $\tilde{g}_U(\cdot)$ sont positives ou nulles, continues, croissantes sur $]0, +\infty[$ et décroissantes sur $]-\infty, 0[$, avec un minimum global unique en 0 ($\tilde{g}_U(0) = g_U(0) = 0$). Elles sont donc de bonnes fonctions de taux (et $g_U(\cdot)$ est strictement convexe).*

Nous montrons un PGD et un PDM uniformes sur U pour la suite $\left(\partial^{[\alpha]}f_n - \partial^{[\alpha]}f\right)$. Le théorème 4 correspond au cas où U est borné, le théorème 5 au cas où U est non borné.

Théorème 4 ***(déviations uniformes sur un ensemble borné)***
Dans le cas où $|\alpha| = 0$, supposons (H1), (H2), (H7), (H9)i), et (H10) vérifiées. Dans le cas où $|\alpha| \geq 1$, supposons (H3)-(H5), (H7), (H9)i) et (H10) vérifiées. En outre, supposons soit que $(v_n) \equiv 1$ soit que (v_n) satisfait (H6). Alors, pour tout sous-ensemble borné U de $\mathbb{R}^d$ et pour tout $\delta > 0$,

$$\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]}f_n(x) - \partial^{[\alpha]}f(x)\right| \geq \delta\right] = -\tilde{g}_U(\delta). \tag{1.5}$$

Théorème 5 ***(déviations uniformes sur un ensemble non borné)***
Supposons (H1), (H7)-(H11) vérifiées. De plus,
- *dans le cas où $|\alpha| = 0$ et $(v_n) \equiv 1$, supposons (H2) satisfaite ;*
- *dans le cas où soit $|\alpha| \geq 1$ et $(v_n) \equiv 1$, soit $|\alpha| \geq 0$ et (v_n) satisfait (H6), supposons (H3)-(H5) vérifiées.*

Alors, pour tout sous-ensemble U de $\mathbb{R}^d$ et pour tout $\delta > 0$,

$$\begin{aligned}
-\tilde{g}_U(\delta) &\leq \liminf_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] \\
&\leq \limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] \leq -\frac{\xi}{\xi+d}\tilde{g}_U(\delta).
\end{aligned}$$

Le corollaire suivant est une conséquence immédiate du théorème 5.

Corollaire 1
Sous les hypothèses du théorème 5, si $\int_{\mathbb{R}^d} \|x\|^\xi f(x)dx < \infty \ \forall \xi \in \mathbb{R}$, alors pour tout sous-ensemble U de $\mathbb{R}^d$,

$$\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] = -\tilde{g}_U(\delta). \tag{1.6}$$

Remarque : Les résultats énoncés dans le théorème 4 et le corollaire 1 sont des PGD. En effet, puisque la suite $(\sup_{x\in U} |f_n(x) - f(x)|)$ est positive et puisque $\tilde{g}_U$ est continue sur $[0, +\infty[$, croissante et tend vers l'infini quand $\delta \to \infty$, l'application du lemme 5 dans Worms [49] permet de déduire de (1.5) ou (1.6) que la suite $(\sup_{x\in U} |f_n(x) - f(x)|)$ satisfait un PGD de vitesse $\left(nh_n^d\right)$ et de bonne fonction de taux $\tilde{g}_U$ sur $\mathbb{R}_+$.

Le comportement en déviations uniformes de l'estimateur non-récursif et de ses dérivées a été établi dans Mokkadem et al. [27]. Les deux principales différences dans le comportement en déviations uniformes entre les estimateurs non-récursifs et les estimateurs récursifs sont à nouveau les suivantes : (i) concernant le comportement en grandes déviations, l'expression de la fonction de taux est plus compliquée dans le cadre de l'estimation récursive ; (ii) concernant le comportement en déviations modérées, la fonction de taux est plus grande dans le cadre de l'estimation récursive. Soulignons à nouveau que, que les déviations soient grandes ou modérées, uniformes ou ponctuelles, les hypothèses sur la fenêtre (h_n) sont plus contraignantes dans le cas récursif que dans le cas non-récursif : dans le cas récursif, il est nécessaire de connaitre la vitesse de convergence de (h_n) vers zéro (et on pose $(h_n) = h(n)$ avec $h(x) = cx^{-a}$ pour les PGD ou h à variations régulières d'ordre $-a$ pour les PDM), alors que, dans le cas non-récursif, il suffit de supposer que (h_n) tend vers zéro et que $nh_n^{d+2|\alpha|}$ tend vers l'infini. Cette hypothèse supplémentaire sur la fenêtre dans le cas récursif ne peut pas être supprimée : l'exposant a qui intervient dans la définition de (h_n) apparait dans l'expression des fonctions de taux pour les PGD et PDM ponctuels et uniformes des estimateurs récursifs de la densité et de ses dérivées.

1.3 Principe de grandes déviations et principes de déviations modérées pour l'estimateur multivarié de la régression.

Soit $(X, Y), (X_1, Y_1), \ldots, (X_n, Y_n)$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\mathbb{R}^d \times \mathbb{R}^q$ de densité de probabilité $f(x, y)$ avec $\mathbb{E}|Y| < \infty$. Soient $g(x)$ la densité marginale de X et $r(x) = \mathbb{E}(Y|X = x)$ la régression de Y sur X. L'objectif de ce chapitre est d'établir le comportement en déviations de l'estimateur de la régression de Nadaraya-Watson et de sa version semi-récursive.

1.3.1 PGD et PDM pour l'estimateur de Nadaraya-Watson

L'estimateur de Nadaraya-Watson ([28], [44]) de la régression est défini par

$$r_n(x) = \begin{cases} \dfrac{m_n(x)}{g_n(x)} & si \quad g_n(x) \neq 0 \\ 0 \text{ sinon,} \end{cases}$$

avec

$$m_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n Y_i K\left(\frac{x - X_i}{h_n}\right)$$
$$g_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x - X_i}{h_n}\right),$$

où la fenêtre (h_n) est une suite positive de nombres réels telle que

$$\lim_{n\to\infty} h_n = 0 \quad \text{et} \quad \lim_{n\to\infty} nh_n^d = \infty, \tag{1.7}$$

où le noyau K est une fonction continue telle que $\lim_{\|x\|\to\infty} K(x) = 0$ et $\int_{\mathbb{R}^d} K(x)dx = 1$. La consistance faible et forte de r_n a été étudiée entre autres par Collomb [4], Collomb et Härdle [5], Devroye [7], Mack et Silverman [24] et Senoussi [37]. Pour d'autres travaux sur la consistance, nous renvoyons aux monographies telles que celles de Bosq [3] et de Prakasa Rao [31]. Le comportement en grandes déviations de r_n a été étudié dans un premier temps par Louani [22], puis par Joutard [15] dans le cas unidimensionnel. Un principe de déviations modérées a été obtenu par Worms [49] dans le cas particulier où $Y = r(X) + \varepsilon$ avec ε et X indépendants. Dans cette partie, nous généralisons ces différents résultats.

L'approche utilisée par Louani [22] et Joutard [15] pour étudier les grandes déviations de r_n est de remarquer que si $d = q = 1$ et si le noyau est positif, alors, pour tout $\delta > 0$,

$$\mathbb{P}\left[r_n(x) - r(x) \geq \delta\right] = \mathbb{P}\left[\frac{1}{nh_n}\sum_{j=1}^n \left[Y_j - r(x) - \delta\right] K\left(\frac{x - X_j}{h_n}\right) \geq 0\right].$$

Evidemment, cette approche ne peut pas être étendue au cas multivarié. Pour étudier le comportement en grandes déviations de r_n dans le cas multivarié, nous procédons donc de façon totalement différente : nous établissons tout d'abord un PGD pour le couple $(m_n(x), g_n(x))$, puis montrons comment le comportement en grandes déviations de $r_n(x)$ peut être déduit de celui du couple. Les hypothèses dont nous avons besoin sont les suivantes :

(A1) $K : \mathbb{R}^d \to \mathbb{R}$ est une fonction bornée intégrable, $\int_{\mathbb{R}^d} K(z)dz = 1$ et $\lim_{\|z\|\to\infty} K(z) = 0$.

(A2) Pour tout $u \in \mathbb{R}^q$, $t \mapsto \displaystyle\int_{\mathbb{R}^q} e^{\langle u,y\rangle} f(t,y)dy$ est bornée et continue en x.

Commentaires sur les hypothèses

- Notons que l'hypothèse (A2) implique que la densité g est continue en x et bornée.
- Considérons le modèle $Y = r(X) + \varepsilon$ avec ε et X indépendants, et notons h la densité de probabilité de ε. On a alors

$$\begin{aligned} f(t,y) &= g(t)h\left(y - r(t)\right) \\ \int_{\mathbb{R}^q} \|y\| f(t,y)dy &= g(t) \int_{\mathbb{R}^q} \|y + r(t)\| h(y)dy \\ \int_{\mathbb{R}^q} e^{\langle u,y\rangle} f(t,y)dy &= g(t) e^{\langle u,r(t)\rangle} \int_{\mathbb{R}^q} e^{\langle u,y\rangle} h(y)dy. \end{aligned}$$

Ainsi, l'hypothèse (A2) peut être traduite comme une hypothèse sur g, r et sur les moments de ε.

- Notons aussi que la condition de bornitude dans l'hypothèse (A2) est inutile lorsque K est à support compact.

Introduisons les fonctions de taux du PGD pour l'estimateur de Nadaraya-Watson. Soient $\Psi_x : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ et $I_x, \hat{I}_x : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ les fonctions définies par

$$\begin{aligned} \Psi_x(u,v) &= \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle u,y\rangle + v)K(z)} - 1 \right) f(x,y) dz dy, \\ I_x(t_1,t_2) &= \sup_{(u,v)\in\mathbb{R}^q\times\mathbb{R}} \left\{ \langle u, t_1\rangle + vt_2 - \Psi_x(u,v) \right\}, \\ \hat{I}_x(s,t) &= I_x(st,t). \end{aligned}$$

En outre, pour tout $s \in \mathbb{R}^q$, posons

$$\begin{aligned} J^*(s) &= \inf_{t\in\mathbb{R}^*} I_x(st,t) \\ &= \inf_{t\in\mathbb{R}^*} \hat{I}_x(s,t), \\ J(s) &= J^*(s) \wedge I_x(\vec{0},0) \\ &= \inf_{t\in\mathbb{R}} \hat{I}_x(s,t). \end{aligned}$$

Pour prouver que J est une fonction de taux, nous avons besoin de la condition suivante.

(C) $\inf_{s\in\mathbb{R}^q} I_x(s,0) = I_x(\vec{0},0)$.

Sous les hypothèses (A1) et (A2), la condition (C) est satisfaite dans les cas suivants.

Exemple 1 : Noyau positif ou nul
La condition (C) est satisfaite lorsque K est positif ou nul, puisque, dans ce cas, $I_x(s,0) = +\infty$ pour tout $s \neq \vec{0}$ (voir la proposition 7 du chapitre 3).

Exemple 2 : Modèle avec symétrie
La condition (C) est satisfaite lorsque f est symétrique par rapport à chaque coordonnée de la seconde variable $y \in \mathbb{R}^q$.

Exemple 3 : Un noyau négatif sans hypothèse de symétrie sur f, et pour $d = q = 1$
La condition (C) est satisfaite lorsque le noyau s'écrit sous la forme $K = \mathbb{1}_D - \mathbb{1}_{D'}$ où D et D' sont des sous-ensembles de $\mathbb{R}$ tels que $D \cap D' = \emptyset$ et $\lambda(D) - \lambda(D') = 1$. Un exemple de noyau d'ordre 4 vérifiant ces conditions est $K = \mathbb{1}_{[-a,a]} - \mathbb{1}_{[-b,-a[\cup]a,b]}$, avec

$$\begin{aligned} a &= \frac{1}{6}\sqrt[3]{2} + \frac{1}{12}\left(\sqrt[3]{2}\right)^2 + \frac{1}{3} \\ b &= \frac{1}{3}\sqrt[3]{2} + \frac{1}{6}\left(\sqrt[3]{2}\right)^2 + \frac{1}{6}. \end{aligned}$$

La proposition suivante donne les propriétés de J.

Proposition 2 *Supposons les hypothèses (A1), (A2) et (C) satisfaites. Alors,*

(i) J est une fonction de taux sur $\mathbb{R}^q$. Plus précisément, pour $\alpha \in \mathbb{R}$,
- *si $\alpha < I_x(\vec{0}, 0)$, alors $\{J(s) \leq \alpha\}$ est compact.*
- *si $\alpha \geq I_x(\vec{0}, 0)$, alors $\{J(s) \leq \alpha\} = \mathbb{R}^q$.*

(ii) si $I_x(\vec{0}, 0) = \infty$, alors J est une bonne fonction de taux sur $\mathbb{R}^q$ et $J = J^*$.

(iii) si $J^*(s) < \infty$, alors $J(s) = J^*(s)$.

(iv) si $\alpha < I_x(\vec{0}, 0)$, alors $\{J^*(s) \leq \alpha\} = \{J(s) \leq \alpha\}$.

Remarque 4 *D'après les définitions de J et J^* et la proposition 2 (iii), on a :*

$$J(s) \quad = \quad \begin{cases} J^*(s) & si \quad J^*(s) < \infty \\ I_x(\vec{0}, 0) & si \quad J^*(s) = \infty. \end{cases}$$

Nous pouvons maintenant énoncer un PGD pour l'estimateur de Nadaraya-Watson.

Théorème 6 ***(PGD ponctuel pour l'estimateur de Nadaraya-Watson)***
Supposons que les hypothèses (A1), (A2) et (C) sont vérifiées, et que (h_n) satisfait la condition (1.7)*. Alors, pour tout sous-ensemble ouvert U de $\mathbb{R}^q$,*

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in U\right] \geq - \inf_{s\in U} J^*(s),$$

et pour tout sous-ensemble fermé V de $\mathbb{R}^q$,

$$\limsup_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in V\right] \leq - \inf_{s\in V} J(s).$$

Commentaires.

1) Posons $E = \{J^*(s) < \infty\}$. Pour tout sous-ensemble ouvert U de $\mathbb{R}^q$ tel que $U \cap E \neq \emptyset$, on a

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in U\right] \geq - \inf_{s\in U} J(s).$$

2) Si I_x est fini dans un voisinage de $(\vec{0}, 0)$, alors J^* est fini et, d'après la proposition 2 (iii), $J(s) = J^*(s) < \infty \quad \forall s$; ainsi (r_n) satisfait un PGD de vitesse (nh_n^d) et de fonction de taux J. Bien sûr, ceci n'est pas vrai lorsque le noyau est positif ou nul puisque dans ce cas $I_x(s, 0) = +\infty$ pour tout s (voir la proposition 7 du chapitre 3). Cependant, on peut avoir cette conclusion pour des noyaux pouvant prendre des valeurs négatives.

3) Lorsque $I_x(\vec{0}, 0) = \infty$, la proposition 2 et le théorème 6 nous assurent que (r_n) satisfait un PGD de vitesse (nh_n^d) et de bonne fonction de taux J.

Dans le cas où K est un noyau positif ou nul dont le support est de mesure infinie, $I_x(\vec{0}, 0) = \infty$ (voir la proposition 7 du chapitre 3). On obtient donc le corollaire suivant (où l'on note λ la mesure de Lebesgue).

Corollaire 2 *Supposons les hypothèses du théorème 6 vérifiées. Si le noyau K est positif ou nul et tel que $\lambda\left(\left\{x \in \mathbb{R}^d, K(x) > 0\right\}\right) = \infty$, alors la suite (r_n) satisfait un PGD de vitesse (nh_n^d) et de bonne fonction de taux J.*

Ce corollaire est une extension des résultats de Louani [22] et Joutard [15] au cas multivarié (et au cas où le noyau peut s'annuler). De plus, il montre que la fonction de taux qui apparait dans leurs résultats de grandes déviations est en fait une bonne fonction de taux.

Pour établir un PDM pour l'estimateur de Nadaraya-Watson, nous rajoutons les hypothèses suivantes :

(A3) Pour tout $u \in \mathbb{R}^q$, les fonctions $t \mapsto \int_{\mathbb{R}^q} \langle u, y\rangle^2 f(t,y)dy$ et $t \mapsto \int_{\mathbb{R}^q} \langle u, y\rangle f(t,y)dy$ sont continues en x et $g(x) \neq 0$.

(A4) $\lim_{n\to\infty} v_n = \infty$ et $\lim_{n\to\infty} \frac{nh_n^d}{v_n^2} = \infty$.

(A5) i) Il existe un entier $p \geq 2$ tel que $\forall s \in \{1, \dots, p-1\}, \forall j \in \{1, \dots, d\}$, $\int_{\mathbb{R}^d} y_j^s K(y)dy_j = 0$ et $\int_{\mathbb{R}^d} \left|y_j^p K(y)\right| dy < \infty$.
ii) $\lim_{n\to\infty} v_n h_n^p = 0$.
iii) m et g sont p-fois différentiables sur $\mathbb{R}^d$, et leurs différentielles d'ordre p sont bornées et continues en x.

Nous pouvons maintenant énoncer un PDM pour l'estimateur de Nadaraya-Watson.

Théorème 7 ***(PDM ponctuel pour l'estimateur de Nadaraya-Watson)***
Supposons que les hypothèses (A1)-(A5) sont vérifiées. Alors, la suite $(v_n(r_n(x) - r(x)))$ satisfait un PGD de vitesse $\left(\frac{nh_n^d}{v_n^2}\right)$ et de bonne fonction de taux G_x définie par :

$$G_x(v) \quad = \quad \frac{g(x)}{2\int_{\mathbb{R}^d} K^2(z)dz} v^T \Sigma_x^{-1} v,$$

où Σ_x est la matrice de covariance conditionnelle de Y sachant $X = x$.

1.3.2 PGD et PDM pour l'estimateur semi-récursif de la régression

La version semi-récursive de l'estimateur de Nadaraya-Watson est définie par

$$\tilde{r}_n(x) = \begin{cases} \dfrac{\tilde{m}_n(x)}{\tilde{g}_n(x)} & \text{si } \tilde{g}_n(x) \neq 0 \\ 0 & \text{sinon,} \end{cases}$$

où

$$\tilde{m}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{Y_i}{h_i^d} K\left(\frac{x - X_i}{h_i}\right) \quad \text{et} \quad \tilde{g}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right)$$

($\tilde{m}_n$ et $\tilde{g}_n$ sont les versions récursives de m_n et g_n respectivement). Des conditions faibles sur les différentes formes de consistance de $\tilde{r}_n$ ont été obtenues par Ahmad et Lin [1] et Devroye et Wagner [8]. Roussas [34] a étudié sa vitesse de convergence presque sûre. L'objectif de cette partie est d'établir un PGD et un PDM pour $\tilde{r}_n$.

La fonction de taux apparaissant dans le PGD pour $\tilde{r}_n$ étant plus complexe que celle pour l'estimateur de Nadaraya-Watson, nous considérons les fenêtres de la forme $h_n = cn^{-a}$, avec $c > 0$ et $0 < a < 1/d$ (au lieu de fenêtres satisfaisant (1.7)). Pour $a \in]0, 1/d[$, soient $\tilde{\Psi}_{a,x} : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ et $\tilde{I}_{a,x} : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ les fonctions définies par :

$$\begin{aligned} \tilde{\Psi}_{a,x}(u,v) &= \int_{[0,1]\times\mathbb{R}^d\times\mathbb{R}^q} s^{-ad}\left(e^{s^{ad}(\langle u,y\rangle+v)K(z)} - 1\right) f(x,y)dsdzdy, \\ \tilde{I}_{a,x}(t_1,t_2) &= \sup_{(u,v)\in\mathbb{R}^q\times\mathbb{R}} \left\{\langle u, t_1\rangle + vt_2 - \tilde{\Psi}_x(u,v)\right\}. \end{aligned}$$

De plus, soient $\tilde{J}_a$ et $\tilde{J}_a^*$ les fonctions définies de la façon suivante : pour tout $s \in \mathbb{R}^q$,

$$\begin{aligned} \tilde{J}_a^*(s) &= \inf_{t\in\mathbb{R}^*} \tilde{I}_{a,x}(st,t) \\ \tilde{J}_a(s) &= \tilde{J}_a^*(s) \wedge \tilde{I}_{a,x}(\vec{0},0). \end{aligned}$$

Pour donner le comportement en grandes déviations de l'estimateur semi-récursif $\tilde{r}_n$, nous avons besoin de l'hypothèse suivante.

(A'1) Pour tout $u \in \mathbb{R}^q$, $t \mapsto \int_{\mathbb{R}^q} e^{\alpha\langle u,y\rangle} f(t,y)dy$ est continue en x uniformément par rapport à $\alpha \in [0,1]$.

Nous devons également substituer la condition (C) par la suivante :

(C') $\inf_{s\in\mathbb{R}^q} \tilde{I}_{a,x}(s,0) = \tilde{I}_{a,x}(\vec{0},0)$.

Des exemples pour lesquels la condition (C') est vérifiée sont les exemples 1 et 2 donnés pour (C). La proposition suivante donne les propriétés de la fonction $\tilde{J}_a$.

Proposition 3 *Supposons les hypothèses (A1), (A2), (A'1) et (C') vérifiées. Alors,*

(i) $\tilde{J}_a$ *est une fonction de taux sur* $\mathbb{R}^q$*. Plus précisément, pour* $\alpha \in \mathbb{R}$*,*
- *si* $\alpha < \tilde{I}_{a,x}(\vec{0},0)$*, alors* $\left\{\tilde{J}_a(s) \leq \alpha\right\}$ *est compact.*
- *si* $\alpha \geq \tilde{I}_{a,x}(\vec{0},0)$*, alors* $\left\{\tilde{J}_a(s) \leq \alpha\right\} = \mathbb{R}^q$*.*

(ii) si $\tilde{I}_{a,x}(\vec{0},0) = \infty$*, alors* $\tilde{J}_a$ *est une bonne fonction de taux sur* $\mathbb{R}^q$ *et* $\tilde{J}_a = \tilde{J}_a^*$*.*

(iii) si $\tilde{J}_a^*(s) < \infty$*, alors* $\tilde{J}_a(s) = \tilde{J}_a^*(s)$*.*

(iv) si $\alpha < \tilde{I}_{a,x}(\vec{0},0)$*, alors* $\left\{\tilde{J}_a^*(s) \leq \alpha\right\} = \left\{\tilde{J}_a(s) \leq \alpha\right\}$*.*

Notons que, comme pour J et J^*, on a

$$\tilde{J}_a(s) = \begin{cases} \tilde{J}_a^*(s) & \text{si } \tilde{J}_a^*(s) < \infty \\ \tilde{I}_{a,x}(\vec{0},0) & \text{si } \tilde{J}_a^*(s) = \infty. \end{cases}$$

Nous pouvons maintenant énoncer un PGD pour l'estimateur semi-récursif de la régression.

Théorème 8 ***(PGD pour l'estimateur semi-récursif de la régression)***
Soit $(h_n) = (cn^{-a})$ *avec* $c > 0$ *et* $0 < a < 1/d$*. Supposons (A1), (A2), (A'1) et (C') satisfaites. Alors, pour tout sous-ensemble ouvert* U *de* $\mathbb{R}^q$*,*

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[\tilde{r}_n(x) \in U\right] \geq -\inf_{s\in U} \tilde{J}_a^*(s),$$

et pour tout sous-ensemble fermé V *de* $\mathbb{R}^q$*,*

$$\limsup_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[\tilde{r}_n(x) \in V\right] \leq -\inf_{s\in V} \tilde{J}_a(s).$$

Tous les commentaires faits à la suite du théorème 6 restent vrais pour le théorème 8. En particulier, nous avons le corollaire suivant :

Corollaire 3 *Supposons les hypothèses du théorème 8 satisfaites. Si* K *est un noyau positif ou nul tel que* $\lambda\left(\left\{x \in \mathbb{R}^d, K(x) > 0\right\}\right) = \infty$*, alors la suite* $(\tilde{r}_n)$ *satisfait un PGD de vitesse* (nh_n^d) *et de bonne fonction de taux* $\tilde{J}_a$*.*

Avant d'établir un PDM pour l'estimateur semi-récursif de la régression, rappelons qu'une suite (u_n) est dite à variations régulières d'ordre α s'il existe une fonction u à variations régulières d'ordre α telle que $u_n = u(n)$ pour tout n (voir par exemple Feller [11] page 275). Nous supposons aussi que la condition suivante est réalisée.

$$\sup_{i\leq n} \frac{h_n}{h_i} < \infty. \tag{1.8}$$

(Cette condition est satisfaite par exemple lorsque (h_n) est une suite décroissante).

Théorème 9 ***(PDM ponctuel pour l'estimateur semi-récursif de la régression)***
Supposons que (h_n) est une suite à variations régulières d'ordre $(-a)$ avec $a \in]0, 1/d[$ et qu'elle satisfait (1.8). Supposons (A1)-(A5) vérifiées. Alors, la suite $(v_n(\tilde{r}_n(x) - r(x)))$ satisfait un PGD de vitesse $\left(\frac{nh_n^d}{v_n^2}\right)$ et de bonne fonction de taux $\tilde{G}_{a,x}$ définie par :

$$\tilde{G}_{a,x}(v) = \frac{(1+ad)g(x)}{2\int_{\mathbb{R}^d} K^2(z)dz} v^T \Sigma_x^{-1} v.$$

Notons que, comme c'était déjà le cas dans le cadre de l'estimation d'une densité, la fonction de taux qui apparait dans les PDM est plus grande pour l'estimateur semi-récursif que pour l'estimateur de Nadaraya-Watson.

1.4 Comportement joint des estimateurs semi-récursifs du mode et de la valeur modale d'une densité de probabilité.

Soit $X_1, \ldots, X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées à valeurs dans $\mathbb{R}^d$ de densité de probabilité inconnue f. On suppose que f a un unique mode θ, c'est-à-dire on suppose qu'il existe $\theta \in \mathbb{R}^d$ tel que $f(x) < f(\theta)$ pour tout $x \neq \theta$. En outre, on suppose que θ est non dégénéré, *i.e.* que $D^2 f(\theta)$, la différentielle d'ordre 2 de f au point θ, est non singulière (dans toute la suite, $D^m g$ est la différentielle d'ordre m de la fonction multivariée g). L'objectif principal de ce chapitre est l'étude de la vitesse de convergence en loi d'estimateurs à noyau du mode θ et de la valeur modale $\mu = f(\theta)$ et son application à la construction de régions de confiance pour le couple (θ, μ).

L'estimateur à noyau du mode a été introduit par Parzen [30]. Il est définit comme une variable aléatoire θ_n^* satisfaisant

$$f_n^*(\theta_n^*) = \sup_{y\in\mathbb{R}^d} f_n^*(y), \tag{1.9}$$

où f_n^* est l'estimateur de Rosenblatt de la densité f. Lorsque le noyau K est continu et tel que $\lim_{\|x\|\to\infty} K(x) = 0$, le choix de la variable aléatoire θ_n^* vérifiant (1.9) peut être fait à l'aide d'un ordre sur $\mathbb{R}^d$. Par exemple, on peut considérer l'ordre lexicographique suivant : $x \leq y$ si la première coordonnée non nulle de $x - y$ est négative. La définition

$$\theta_n^* = \inf\left\{ y \in \mathbb{R}^d \text{ tel que } f_n^*(y) = \sup_{x\in\mathbb{R}^d} f_n^*(x) \right\},$$

où l'infimum est pris suivant l'ordre lexicographique assure la mesurabilité de l'estimateur à noyau du mode. Il existe de nombreux travaux sur le comportement asymptotique de θ_n^*. Citons, parmi beaucoup d'autres, Parzen [30], Nadaraya [28], Van Ryzin [42], Rüschendorf [35], Konakov [16], Samanta [36], Eddy ([9], [10]), Romano [32], Vieu [43], Mokkadem et Pelletier [26]. Pour estimer la valeur modale $f(\theta)$, il est naturel de prendre $\mu_n^* = f_n^*(\theta_n^*)$, mais, à notre connaissance, les propriétés asymptotiques de cet estimateur n'ont pas été beaucoup étudiées.

Dans les deux chapitres 2 et 3, nous avons vu que les estimateurs récursifs ou semi-récursifs sont plus concentrés sur le paramètre à estimer que les estimateurs classiques. C'est pourquoi, dans ce quatrième chapitre, nous préférons introduire des estimateurs semi-récursifs du mode et de la valeur modale plutôt que d'étudier le couple (θ_n^*, μ_n^*).

La version semi-récursive de l'estimateur à noyau classique du mode est définie par $\theta_n = \inf\left\{y \in \mathbb{R}^d \text{ tel que } f_n(y) = \sup_{x\in\mathbb{R}^d} f_n(x)\right\}$, où f_n est l'estimateur récursif de la densité introduit par Wolverton et Wagner. L'estimateur semi-récursif auquel on pense naturellement pour estimer la valeur modale est $\mu_n = f_n(\theta_n)$. Cependant, nous allons voir que cet estimateur a deux inconvénients majeurs. D'une part, si le noyau K utilisé dans la définition de f_n est d'ordre 2, alors la loi limite de la suite $(\mu_n - \mu)$ convenablement renormalisée est nécessairement dégénérée ; pour construire une région de confiance pour (θ, μ) (ou un intervalle de confiance pour μ), il faut obligatoirement avoir recours à un noyau d'ordre supérieur. D'autre part, quel que soit l'ordre du noyau K utilisé dans la définition de f_n, le choix de la fenêtre qui assure que θ_n converge à la vitesse optimale n'est pas le même que celui qui assure que μ_n converge à la vitesse optimale. C'est la raison pour laquelle nous utilisons deux estimateurs distincts de la densité (définis à l'aide de fenêtres différentes) pour estimer θ et μ.

Plus précisément, soient h et $\tilde{h}$ deux fonctions positives, soient (h_n) et $(\tilde{h}_n)$ les suites définies par $h_n = h(n)$ et $\tilde{h}_n = \tilde{h}(n)$ pour tout $n \geq 1$, et soient f_n et $\tilde{f}_n$ les deux estimateurs récursifs de la densité définis par

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h_i^d}K\left(\frac{x - X_i}{h_i}\right) \quad \text{et} \quad \tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\tilde{h}_i^d}K\left(\frac{x - X_i}{\tilde{h}_i}\right).$$

Pour estimer le mode, nous utilisons le premier estimateur de la densité et posons donc

$$\theta_n = \inf\left\{y \in \mathbb{R}^d \text{ tel que } f_n(y) = \sup_{x\in\mathbb{R}^d} f_n(x)\right\},$$

tandis que, pour estimer la valeur modale, nous utilisons le second estimateur de la densité et posons ainsi

$$\tilde{\mu}_n = \tilde{f}_n(\theta_n).$$

La partie 1.4.1 est consacrée à la consistance des estimateurs θ_n et $\tilde{\mu}_n$, la partie 1.4.2 à la vitesse de convergence en loi du couple $(\theta_n, \tilde{\mu}_n)$. Dans la partie 1.4.3, nous comparons l'utilisation des estimateurs semi-récursifs à celle des estimateurs classiques. Enfin, nous donnons la vitesse de convergence presque sûre du couple $(\theta_n, \tilde{\mu}_n)$ dans la partie 1.4.4.

1.4.1 Consistance des estimateurs θ_n et $\tilde{\mu}_n$

Pour montrer la consistance de θ_n et de $\tilde{\mu}_n$, nous avons besoin des hypothèses suivantes.

(A1) i) K est une fonction différentiable, paire, intégrable et $\int_{\mathbb{R}^d} K(z)dz = 1$.

ii) K est un noyau d'ordre $q \geq 2$ i.e. $\forall s \in \{1, \dots, q-1\}$, $\forall j \in \{1, \dots, d\}$, $\int_{\mathbb{R}^d} y_j^s K(y)dy_j = 0$

et $\int_{\mathbb{R}^d} \left|y_j^q K(y)\right| dy < \infty$.
iii) K est une fonction höldérienne.
iv) Il existe $\gamma > 0$ tel que la fonction $z \mapsto \|z\|^\gamma |K(z)|$ soit bornée.

(A2) i) f est uniformément continue sur $\mathbb{R}^d$, q-fois différentiable et $\sup_{x\in\mathbb{R}^d} \|D^q f(x)\| < \infty$, où q est l'ordre du noyau K.
ii) Il existe $\xi > 0$ tel que $\int_{\mathbb{R}^d} \|x\|^\xi f(x)dx < \infty$.
iii) Il existe $\eta > 0$ tel que la fonction $z \mapsto \|z\|^\eta f(z)$ soit bornée.
iv) Il existe $\theta \in \mathbb{R}^d$ tel que $f(x) < f(\theta)$ pour tout $x \neq \theta$.

(A3) Les fonctions h et $\tilde{h}$ sont locallement bornées et à variations régulières d'ordre $(-a)$ et $(-\tilde{a})$ respectivement, où $a \in]0, 1/(d+4)[$, $\tilde{a} \in]0, 1/(d+2)[$.

Remarque 5 *Notons que l'hypothèse (A1)iv) implique que K est borné.*

Proposition 4 *Sous les hypothèses (A1)-(A3), on a*

$$\lim_{n\to\infty} \theta_n = \theta \quad p.s. \quad et \quad \lim_{n\to\infty} \tilde{\mu}_n = \mu \quad p.s.$$

1.4.2 Vitesse de convergence en loi du couple $(\theta_n, \tilde{\mu}_n)$

Posons

$$\mathcal{L}_\theta(n) = n^a h_n \quad \text{et} \quad \mathcal{L}_\mu(n) = n^{\tilde{a}} \tilde{h}_n. \tag{1.10}$$

Au vu de l'hypothèse (A3), $\mathcal{L}_\theta$ et $\mathcal{L}_\mu$ sont des fonctions positives à variations lentes. En outre, nous introduisons les notations suivantes :

$$B_q(\theta) = \begin{pmatrix} \frac{(-1)^q}{q!(1-aq)} \nabla\left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\right) \\ \frac{(-1)^q}{q!(1-\tilde{a}q)} \sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix} \quad \text{avec} \quad \beta_j^q = \int_{\mathbb{R}^d} y_j^q K(y)dy, \quad aq \neq 1 \ \text{ et } \ \tilde{a}q \neq 1,$$

$$A = \begin{pmatrix} -\left[D^2 f(\theta)\right]^{-1} & 0 \\ 0 & 1 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{f(\theta)G}{1+a(d+2)} & 0 \\ 0 & \frac{f(\theta)\int_{\mathbb{R}^d} K^2(z)dz}{1+\tilde{a}d} \end{pmatrix},$$

où G est la matrice $d \times d$ définie par $G^{(i,j)} = \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_i}(x) \frac{\partial K}{\partial x_j}(x)dx$.

Pour tout $c, \tilde{c} \geq 0$, on pose $D(c, \tilde{c}) = \begin{pmatrix} \sqrt{c} I_d & 0 \\ 0 & \sqrt{\tilde{c}} \end{pmatrix}$ où I_d est la matrice identité $d \times d$.

Pour établir la vitesse de convergence faible de $(\theta_n, \tilde{\mu}_n)$, nous introduisons les conditions suivantes.

(C1) L'une des trois conditions suivantes est satisfaite.
i) $\left(\frac{1}{d+4} < \tilde{a} < \frac{q}{d+2q+2} \text{ et } \frac{\tilde{a}}{q} < a < \frac{1-2\tilde{a}}{d+2}\right)$
ii) $\left(\frac{1}{d+2q} < \tilde{a} \leq \frac{1}{d+4} \text{ et } \frac{1}{d+2q+2} < a < \frac{1+\tilde{a}d}{2(d+2)}\right)$
iii) $\tilde{a} = \frac{1}{d+2q}$, $a = \frac{1}{d+2q+2}$, $\lim_{n\to\infty} \mathcal{L}_\mu(n) = \tilde{\gamma} < \infty$ et $\lim_{n\to\infty} \mathcal{L}_\theta(n) = \gamma < \infty$.

(C2) L'une des deux conditions suivantes est vérifiée.
i) $\left(0 < \tilde{a} < \frac{1}{d+2q} \text{ et } \frac{\tilde{a}}{2} < a < \frac{1}{d+2q+2}\right)$

ii) $\tilde{a} = \dfrac{1}{d+2q}$, $\lim_{n\to\infty} \mathcal{L}_\mu(n) = \infty$ et $\dfrac{1}{2(d+2q)} < a < \dfrac{1}{d+2q+2}$.

(A4) i) K est deux fois différentiable sur $\mathbb{R}^d$.
ii) La fonction $z \mapsto z\nabla K(z)$ est intégrable.
iii) Pour tout $(i,j) \in \{1,\dots,d\}^2$, $\partial^2 K/\partial x_i \partial x_j$ est bornée, intégrable et höldérienne.

(A5) i) $D^2 f(\theta)$ est non singulière.
ii) $D^2 f$ est q-fois différentiable et ∇f est bornée.
iii) Pour tout $(i,j) \in \{1,\dots,d\}^2$, $\sup_{x\in\mathbb{R}^d} \|D^q\left(\partial^2 f/\partial x_i \partial x_j\right)\| < \infty$, et pour tout $k \in \{1,\dots,d\}$, $\sup_{x\in\mathbb{R}^d} \|D^q\left(\partial f/\partial x_k\right)\| < \infty$.

Remarque 6 *Notons que les hypothèses (A4)ii) et (A4)iii) impliquent que* ∇K *est une fonction lipschitzienne et intégrable ; ainsi, on peut voir que* $\lim_{\|x\|\to\infty} \|\nabla K(x)\| = 0$ *(et en particulier* ∇K *est bornée).*

Théorème 10 *Supposons les hypothèses (A1)-(A5) vérifiées.*

i) Si (C1)i) ou (C1)ii) est satisfaite, alors

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, A\Sigma A\right).$$

ii) Si (C1)iii) est satisfaite, on pose $c = \gamma^{d+2q+2}$ *et* $\tilde{c} = \tilde{\gamma}^{d+2q}$ *; alors*

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(D(c,\tilde{c})AB_q(\theta), A\Sigma A\right).$$

iii) Si (C2) est satisfaite, alors

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n - \theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathbb{P}} AB_q(\theta).$$

Remarque. La partie 1 (respectivement la partie 3) du théorème 10 correspond au cas où les biais (respectivement les variances) des estimateurs θ_n et $\tilde{\mu}_n$ sont négligeables devant leurs variances (respectivement leurs biais). Lorsque $\gamma, \tilde{\gamma} > 0$, la partie 2 du théorème 10 correspond au cas où le biais et la variance des estimateurs θ_n et $\tilde{\mu}_n$ ont la même vitesse de convergence. D'autres conditions possibles conduisant à différentes combinaisons sont omises pour des raisons de simplicité.

L'estimateur du mode θ_n converge à la vitesse optimale lorsque la fenêtre (h_n) est choisie telle que $\lim_{n\to\infty} nh_n^{d+2q+2} = \gamma \in]0,\infty[$, tandis que l'estimateur de la valeur modale $\tilde{\mu}_n$ converge à la vitesse optimale lorsque la fenêtre $(\tilde{h}_n)$ est choisie telle que $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = \tilde{\gamma} \in]0,\infty[$. Nous voyons ainsi clairement l'intérêt d'utiliser deux estimateurs différents de la densité (définis à l'aide de fenêtres distinctes) pour estimer le mode et la valeur modale.

Notons également que chacune des conditions (C1)i) et (C1)iii) excluent le choix $a = \tilde{a}$ (et donc $h_n = \tilde{h}_n$) et que la condition (C1) est vide pour $q = 2$. Ainsi, lorsque la même fenêtre est utilisée pour estimer à la fois θ et μ, il faut obligatoirement avoir recours à un noyau d'ordre supérieur pour que la loi asymptotique du couple $(\theta_n - \theta, \tilde{\mu}_n - \mu)$ convenablement normalisé ne soit pas dégénérée (et donc pour construire des régions de confiance de (θ, μ)).

1.4.3 Comparaison avec les estimateurs classiques

Rappelons que l'estimateur du mode introduit par Parzen [30] est défini par $\theta_n^* = \inf\left\{y \in \mathbb{R}^d \text{ tel que } f_n^*(y) = \sup_{x\in\mathbb{R}^d} f_n^*(x)\right\}$ où

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

est l'estimateur de la densité de Rosenblatt. Pour estimer la valeur modale, on peut bien-sûr poser $\mu_n^* = f_n^*(\theta_n^*)$, mais on peut également, comme dans le cadre de l'estimation semi-récursive, utiliser une fenêtre différente et poser $\tilde{\mu}_n^* = \tilde{f}_n^*(\theta_n^*)$, où

$$\tilde{f}_n^*(x) = \frac{1}{n\tilde{h}_n^d}\sum_{i=1}^n K\left(\frac{x - X_i}{\tilde{h}_n}\right).$$

En suivant les lignes de la démonstration du théorème 10, on peut établir la vitesse de convergence en loi du couple $(\theta_n^* - \theta, \tilde{\mu}_n^* - \mu)$. Plus précisément, on peut montrer que les conclusions du théorème 10 sont valables lorsque l'on remplace θ_n par θ_n^*, $\tilde{\mu}_n$ par $\tilde{\mu}_n^*$, Σ par la matrice

$$\Sigma^* = \begin{pmatrix} f(\theta)G & 0 \\ 0 & f(\theta)\int_{\mathbb{R}^d} K^2(z)dz \end{pmatrix},$$

et $B_q(\theta)$ par le vecteur

$$B_q^*(\theta) = \begin{pmatrix} \frac{(-1)^q}{q!}\nabla\left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\right) \\ \frac{(-1)^q}{q!}\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix}.$$

Notons que Σ, la variance asymptotique des estimateurs semi-récursifs θ_n, $\tilde{\mu}_n$, est plus petite que Σ^*, celle des estimateurs classiques θ_n^*, $\tilde{\mu}_n^*$, puisque

$$\Sigma = \begin{pmatrix} [1 + a(d+2)]^{-1} I_d & 0 \\ 0 & [1 + \tilde{a}d]^{-1} \end{pmatrix}\Sigma^*.$$

Dans le chapitre 4, nous présentons des simulations qui montrent que pour construire des régions de confiance pour (θ, μ), il est - comme nous l'indiquent nos résultats théoriques - préférable : (i) d'utiliser les estimateurs semi-récursifs plutôt que les estimateurs classiques ; (ii) d'utiliser des fenêtres différentes pour les estimations de θ et de μ.

1.4.4 Vitesse de convergence presque sûre du couple $(\theta_n, \tilde{\mu}_n)$

Pour établir la vitesse de convergence presque sûre du couple $(\theta_n, \tilde{\mu}_n)$, nous avons besoin des hypothèses suivantes :

(A6) i) h et $\tilde{h}$ sont différentiables, leurs dérivées sont à variations régulières d'ordre $(-a-1)$ et $(-\tilde{a}-1)$ respectivement.

ii) Il existe $n_0 \in \mathbb{N}$ tel que

$$n \geq m \geq n_0 \quad \Rightarrow \quad \max\left\{\frac{mh_m^{-(d+2)}}{nh_n^{-(d+2)}}; \frac{m\tilde{h}_m^{-d}}{n\tilde{h}_n^{-d}}\right\} = \frac{\min\left\{mh_m^{-(d+2)}; m\tilde{h}_m^{-d}\right\}}{\min\left\{nh_n^{-(d+2)}; n\tilde{h}_n^{-d}\right\}}.$$

Remarque 7 *L'hypothèse (A6)ii) est satisfaite lorsque $a \neq \tilde{a}$ et, dans le cas où $a = \tilde{a}$, elle est satisfaite lorsque $\mathcal{L}_\theta(n) = (\mathcal{L}_\mu(n))^{\frac{d}{d+2}}$ pour n assez grand ($\mathcal{L}_\theta$ et $\mathcal{L}_\mu$ étant définies dans (1.10)).*

En outre, les conditions (C1) et (C2) sont remplacées par les suivantes :

(C'1) L'une des trois conditions suivantes est satisfaite.

i) $\left(\frac{1}{d+4} < \tilde{a} < \frac{q}{d+2q+2} \text{ et } \frac{\tilde{a}}{q} < a < \frac{1-2\tilde{a}}{d+2}\right)$

ii) $\left(\frac{1}{d+2q} < \tilde{a} \leq \frac{1}{d+4} \text{ et } \frac{1}{d+2q+2} < a < \frac{1+\tilde{a}d}{2\,(d+2)}\right)$

iii) $\tilde{a} = \frac{1}{d+2q}$, $a = \frac{1}{d+2q+2}$, $\lim_{n\to\infty} \frac{(\mathcal{L}_\mu(n))^{d+2q}}{2\log\log n} = \tilde{\alpha} < \infty$ et $\lim_{n\to\infty} \frac{(\mathcal{L}_\theta(n))^{d+2q+2}}{2\log\log n} = \alpha < \infty$.

(C'2) L'une des deux conditions suivantes est vérifiée.

i) $\left(0 < \tilde{a} < \frac{1}{d+2q} \text{ et } \frac{\tilde{a}}{2} < a < \frac{1}{d+2q+2}\right)$

ii) $\tilde{a} = \frac{1}{d+2q}$, $\lim_{n\to\infty} \frac{(\mathcal{L}_\mu(n))^{d+2q}}{2\log\log n} = \infty$ et $\frac{1}{2(d+2q)} < a < \frac{1}{d+2q+2}$.

Avant de donner une vitesse de convergence presque sûre de $(\theta_n, \tilde{\mu}_n)$, précisons que la proposition 2.3 dans Mokkadem et Pelletier [26] assure que la matrice G est non singulière, ce qui garantit l'inversibilité de Σ.

Théorème 11 *Supposons les hypothèses (A1)-(A6) vérifiées.*

i) Si (C'1)i) ou (C'1)ii) est satisfaite, alors, avec probabilité un, la suite

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix}$$

est relativement compacte et l'ensemble de ses points d'adhérence est l'ellipsoïde

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ tel que } \nu^T A^{-1}\Sigma^{-1}A^{-1}\nu \leq 1\right\}.$$

ii) Si (C'1)iii) est satisfaite, on pose $c = \alpha$ et $\tilde{c} = \tilde{\alpha}$; alors, avec probabilité un, la suite

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix}$$

est relativement compacte et l'ensemble de ses points d'adhérence est l'ellipsoïde

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ tel que } \left(A^{-1}\nu - D\,(c,\tilde{c})\,B_q(\theta)\right)^T \Sigma^{-1} \left(A^{-1}\nu - D\,(c,\tilde{c})\,B_q(\theta)\right) \leq 1\right\}.$$

iii) Si (C'2) est satisfaite, alors

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n - \theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{p.s.} AB_q(\theta).$$

Remarque. Lorsque la condition (C'1) est satisfaite, on a

$$\lim_{n\to\infty} \frac{nh_n^{d+2q+2}}{2\log\log n} = \alpha \quad \text{et} \quad \lim_{n\to\infty} \frac{n\tilde{h}_n^{d+2q}}{2\log\log n} = \tilde{\alpha},$$

avec $\alpha = \tilde{\alpha} = 0$ dans le cadre de la première partie du théorème 11, et $\alpha, \tilde{\alpha} \in [0, \infty[$ dans le cadre de la deuxième partie du théorème 11. Lorsque la condition (C'2) est satisfaite, on a

$$\lim_{n\to\infty} \frac{nh_n^{d+2q+2}}{2\log\log n} = \infty \quad \text{et} \quad \lim_{n\to\infty} \frac{n\tilde{h}_n^{d+2q}}{2\log\log n} = \infty.$$

Notons qu'une loi du logarithme itérée compacte pour l'estimateur θ_n^* de Parzen a été établie dans [26]. Cependant, les techniques utilisées dans [26] sont complètement différentes de celles que nous utilisons pour démontrer le théorème 11. Soulignons que pour établir la vitesse de convergence du couple d'estimateurs classiques $(\theta_n^*, \tilde{\mu}_n^*)$, ni les techniques utilisées dans [26], ni celles utilisées dans le chapitre 4 ne s'appliquent ; établir une loi du logarithme itérée pour ce couple d'estimateur reste donc une question ouverte.

Bibliographie

[1] Ahmad, I.A. and Lin, P. (1976), Nonparametric sequential estimation of a multiple regression function, *Bull. Math. Statist.*, **17**, pp. 63-75.

[2] Arcones, M.A. (1997), The law of the iterated logarithm for a triangular array of empirical processes, *Electronic Journal of Probab.* **2**, pp. 1-39.

[3] Bosq, D. (1985), Nonparametric statistics for stochastic processes, *Lectures notes in Control and Inform. Sc. Springer.*

[4] Collomb, G. (1984), Propriétés de convergence presque-complète du prédicateur à noyau, *Z. Wahrsch. verw. Gebiete*, **66**, pp. 441-460.

[5] Collomb, G. and Härdle W. (1986), Strong uniform convergence rates in robust nonparametric time series analysis and prediction : kernel regression estimation from dependent observations, *Stoch. Proc. and their Appl.* **23**, pp. 77-89.

[6] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities, *Utilitas Math.* **15**, pp. 113-128.

[7] Devroye, L. (1979), The uniform convergence of the Nadaraya-Watson regression function estimate, *Can. J. Stat.* **6**, pp. 179-191.

[8] Devroye, L. and Wagner, T.J. (1980), On the L_1 convergence of kernel estimators of regression function with applications in discrimination, *Wahrsch. verw. Gebiete*, **51**, pp. 15-25.

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

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

[11] Feller, W. (1970), An introduction to probability theory and its applications, Second edition Volume II, Wiley.

[12] Gao, F. (2003), Moderate deviations and large deviations for kernel density estimators, *Journal of Theoritical Probab.* **16**, pp. 401-418.

[13] Giné, E. and Guillou, A. (2002), Rates of strong uniform consistency for multivariate kernel density estimator, *Ann. Inst. Henri Poincaré Probab. Statist* **6**, pp. 907-921.

[14] Hall, P. (1981), Laws of the iterated logarithm for nonparametric density estimator, *Z. Warsch. Verw. Gebiete* **56**, pp. 47-61.

[15] Joutard, C. (2003), Sharp large deviations in nonparametric estimation, *J. of Nonparam. Statist.*, **18**, pp. 293-306.

[16] Konakov, V.D. (1973), On asymptotic normality of the sample mode of multivariate distributions, *Theory Probab. Appl.* **18**, pp. 836-842.

[17] Lei, L. and Wu, L. (2004), The exponential convergence of kernel density estimator in L_1 for phi-mixing processes, *Ann. I.S.U.P.*, **48**, pp. 59-68.

[18] Lei, L. and Wu, L. (2005), Large deviations of kernel density estimator in $L_1(\mathbb{R}^d)$ for uniformly ergodic Markov processes, *Stoch. Process. Appl.* **115**, pp. 275-298.

[19] Lei, L., Wu, L. and Xie, B. (2003), Large deviations and deviation inequality for kernel density estimator in $L_1(\mathbb{R}^d)$-distance. *Ser. Biostat.*, **1**, pp. 89-97.

[20] Lei, L. (2006), Large deviations of the kernel density estimator in $L_1(\mathbb{R}^d)$ for reversible Markov processes, *Bernouilli*, **12**, pp. 65-83.

[21] Louani, D. (1998), Large deviations limit theorems for the kernel density estimator, *Scand. J. Statist.* **25**, pp. 243-253.

[22] Louani, D. (1999), Some large deviations limit theorems in conditionnal nonparametric statistics, *Statistics*, **33**, pp. 171-196.

[23] Louani, D. (2000), Large deviations for the L_1-distance in kernel density estimation, *J. Statist. Plann. Inference*, **90**, pp. 177-182.

[24] Mack, Y. P. and Silverman, B.W. (1982), Weak and strong uniform consistency of kernel regression estimates, *Z. Wahrsch. verw. Gebiete*, **61**, pp. 405-415.

[25] Menon, V.V., Prasad, B. and Singh R.S. (1984), Non-parametric recursive estimates of a probability density function and its derivatives, *Journal of Statistical Planning and inference* **9**, pp. 73-82.

[26] Mokkadem, A. and Pelletier, M. (2003), The law of the iterated logarithm for the multivariate kernel mode estimator, *ESAIM : Probab. Statist.* **7**, pp. 1-21.

[27] Mokkadem, A., Pelletier, M. and Worms, J. (2005), Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives, *Austral. J. Statist.*, **4**, pp. 489-502.

[28] Nadaraya, E. A. (1964), On estimating regression. *Theory Probab. Appl.* **10**, pp. 186-190.

[29] Nadaraya, E.A. (1965), On non-parametric estimates of density functions and regression curves, *Theory Probab. Appl.* **10**, pp. 186-190.

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

[31] Prakasa Rao, B. L. S. (1983), Nonparametric functionnal estimation, *Academic presse New York*.

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

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

[34] Roussas, G. (1992), Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimate, *J. of Nonparam. Statist.* **3**, pp. 171-195.

[35] Rüschendorf, L. (1977), Consistency of estimators for multivariate density functions and for the mode, *Sankhya Ser. A*, **39**, pp. 243-250.

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

[37] Senoussi, R. (1991), Loi du log itéré et identification, Thèse, *Université Paris-Sud*.

[38] Silverman, B.W. (1978), Weak and strong uniform consistency of the kernel estimate of a density and its derivatives. *Ann. Statist.*, **6**, pp. 177-184.

[39] Singh, R.S. (1976), Nonparametric estimation of mixed partial derivatives of a multivariate density, *J. Multivariate Analysis*, **6**, pp. 111-122.

[40] Singh, R.S. (1981), Speed of convergence in nonparametric estimation of a multivariate μ-density and its mixed partial derivatives. *J. Statist. Planning and Inference*, **5**, pp. 287-298.

[41] Stute, W. (1982), A law of the iterated logarithm for kernel density estimators. *Ann. Probability*, **10**, pp. 414-422.

[42] Van Ryzin, J. (1969), On strong consistency of density estimates, *Ann. Math. Statist.* **40**, pp 1765-1772.

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

[44] Watson, G. S. (1964), Smooth regression analysis, *Sankhya Ser. A.* **26**, pp. 359-372.

[45] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density, *Annals of Statistics* **7**, pp. 316-327.

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

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

[48] Worms, J. (2000), Principes de déviations modérées pour des martingales et applications statistiques, Thèse, *Université Marne-la-vallée*

[49] Worms, J. (2001), Moderate and large deviations of some dependent variables, Part II : some kernel estimators, *Math. Methods Statist.* **10**, pp. 161-193.

Chapitre 2

Large and moderate deviations principles for recursive kernel estimator of a multivariate density and its partial derivatives

2.1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed $\mathbb{R}^d$-valued random vectors with bounded probability density f. Let (h_n) be a positive sequence such that $\lim_{n\to\infty} h_n = 0$ and $\lim_{n\to\infty} nh_n^d = \infty$; the recursive kernel estimator of f is defined as

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right) \tag{2.1}$$

where the kernel K is a continuous function such that $\lim_{\|x\|\to+\infty} K(x) = 0$ and $\int_{\mathbb{R}^d} K(x)dx = 1$. The estimate (2.1) is a recursive version of the well-known Rosenblatt kernel estimate (see Rosenblatt [18] and Parzen [16]) ; it was first discussed by Wolverton and Wagner [23], Yamato [25], and Davies [3]. The estimator (2.1) is easily updated each time an additionnal observation becomes available without resorting to past data, through the recursive relationship

$$f_n(x) = \frac{n-1}{n} f_{n-1}(x) + \frac{1}{nh_n^d} K\left(\frac{x - X_n}{h_n}\right).$$

The weak and strong consistency of the recursive estimator of the density was studied by many authors ; let us cite, among many others, Devroye [5], Menon et al. [14] and Wertz [22]. The law of the iterated logarithm of the recursive density estimator was established by Wegman and Davies [21] and Roussas [19]. For other works on recursive density estimation, the reader is referred to the papers of Wegman [20], Ahmad and Lin [1], and Carroll [2].

Recently, pointwise and uniform large and moderate deviations results have been proved for the Rosenblatt density estimator and its derivatives. The large deviations principle has been studied by Louani [12] and Worms [24]. Gao [7] and Mokkadem et al. [15] extend these results and provide moderate deviations principles. The large and moderate deviations of the derivatives of the Rosenblatt density estimator are given in Mokkadem et al. [15]. Let us mention that large deviations principles for the L_1- distance are investigated by Louani [13] and Lei and al. [10] in the i.i.d. case, by Lei and Wu ([8], [9]) and Lei [11] in the markovian or mixing framework. The purpose of this paper is to establish pointwise and uniform large and moderate deviations principles for the recursive density estimator f_n and its derivatives.

Let us recall that a $\mathbb{R}^m$-valued sequence $(Z_n)_{n\geq 1}$ satisfies a large deviations principle (LDP) with speed (ν_n) and good rate function I if :

(a) (ν_n) is a positive sequence such that $\lim_{n\to\infty} \nu_n = \infty$;

(b) $I : \mathbb{R}^m \to [0, \infty]$ has compact level sets ;

(c) for every borel set $B \subset \mathbb{R}^m$,

$$\begin{aligned} -\inf_{x\in \overset{\circ}{B}} I(x) &\leq \liminf_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in B\right] \\ &\leq \limsup_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in B\right] \leq -\inf_{x\in \overline{B}} I(x), \end{aligned}$$

where $\overset{\circ}{B}$ and $\overline{B}$ denote the interior and the closure of B respectively. Moreover, let (v_n) be a non-random sequence that goes to infinity ; if $(v_n Z_n)$ satisfies a LDP, then (Z_n) is said to satisfy a moderate deviations principle (MDP).

For any d-uplet $[\alpha] = \left(\alpha_1, \ldots, \alpha_d\right) \in \mathbb{N}^d$, set $|\alpha| = \alpha_1 + \cdots + \alpha_d$, let

$$\partial^{[\alpha]} f(x) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}(x)$$

denote the $[\alpha]$-th partial derivative of f (if $|\alpha| = 0$, then $\partial^{[\alpha]} f \equiv f$) and, for any $j \in \mathbb{N}$, let $D^{(j)} f$ denote the j-th differential of f. The recursive kernel estimator of the $[\alpha]$-th partial derivative of f is defined as

$$\partial^{[\alpha]} f_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \partial^{[\alpha]} K\left(\frac{x - X_i}{h_i}\right),$$

where the kernel K is chosen such that $\partial^{[\alpha]} K \not\equiv 0$ and the bandwidth such that $\lim_{n\to\infty} n h_n^{d+2|\alpha|} = \infty$.

Our first aim is to establish pointwise LDP for the recursive kernel density estimator f_n. It turns out that expliciting the rate function in this case is more complex than either for the Rosenblatt kernel estimator, or for the derivatives estimators. That is the reason why, in this particular framework, we only consider bandwidths defined as $(h_n) \equiv (cn^{-a})$ with $c > 0$ and $a \in]0, 1/d[$. We then prove that the sequence $(f_n(x) - f(x))$ satisfies a LDP with speed $\left(nh_n^d\right)$ and rate function

$$I_{a,x} : t \mapsto f(x) I_a\left(\frac{t}{f(x)} + 1\right)$$

where $I_a(t)$ is the Fenchel-Legendre transform of the function ψ_a defined as follows :

$$\psi_a(u) = \int_{[0,1]\times\mathbb{R}^d} s^{-ad}\left(e^{s^{ad} u K(z)} - 1\right) ds dz.$$

Our second aim is to provide pointwise LDP for the derivative estimators $\partial^{[\alpha]} f_n$ (with $|\alpha| \geq 1$). In this case, we consider more general bandwidths defined as $h_n = h(n)$ for all n, where h is a regularly varying function with exponent $(-a)$, $a \in]0, 1/\left(d + 2|\alpha|\right)[$. We prove that the sequence

$$\left(\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right)$$

satisfies a LDP of speed $\left(nh_n^{d+2|\alpha|}\right)$ and quadratic rate function $J_{a,[\alpha],x} : \mathbb{R} \to \mathbb{R}$ defined by

$$\begin{cases} \text{if} \quad f(x) \neq 0, \quad J_{a,[\alpha],x} : t \mapsto \frac{t^2(1+a(d+2|\alpha|))}{2f(x)\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(z)\right]^2 dz} \\ \text{if} \quad f(x) = 0, \quad J_{a,[\alpha],x}(0) = 0 \quad \text{and} \quad J_{a,[\alpha],x}(t) = \infty \quad \text{for} \quad t \neq 0. \end{cases} \tag{2.2}$$

Our third aim is to prove pointwise MDP for the density estimator and for its derivatives. For any d-uplet $[\alpha]$ such that $|\alpha| \geq 0$, any positive sequence (v_n) satisfying

$$\lim_{n\to\infty} v_n = \infty \quad \text{and} \quad \lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} = 0,$$

and general bandwidths (h_n), we prove that the random sequence

$$v_n\left(\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right)$$

satisfies a LDP of speed $\left(nh_n^{d+2|\alpha|}/v_n^2\right)$ and rate function $J_{a,[\alpha],x}(\cdot)$ defined by Equation (2.2).

Let us point out that the rate function $J_{a,[\alpha],x}$ is larger (by a factor $1 + a(d + 2|\alpha|)$) than the rate function obtained for the Rosenblatt estimator in Mokkadem et al. [15] ; this means that the recursive estimators $\partial^{[\alpha]} f_n(x)$, $|\alpha| \geq 0$, are more concentrated around $\partial^{[\alpha]} f(x)$ than the Rosenblatt estimators.

Finally, we give a uniform version of the previous results. More precisely, let U be a subset of $\mathbb{R}^d$; we establish large and moderate deviations principles for the sequence $\left(\sup_{x\in U} \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right|\right)$ in the case either U is bounded or all the moments of f are finite.

2.2 Assumptions and Results

2.2.1 Pointwise LDP for the density estimator

The assumptions required on the kernel K and the bandwidth (h_n) are the following.

(H1) $K : \mathbb{R}^d \to \mathbb{R}$ is a bounded and integrable function , $\int_{\mathbb{R}^d} K(z)dz = 1$ and $\lim_{\|z\|\to\infty} K(z) = 0$.

(H2) $h_n = cn^{-a}$ with $0 < a < 1/d$ and $c > 0$.

Before stating our results, we need to introduce the rate function for the LDP of the density estimator. Let $\psi_a : \mathbb{R} \to \mathbb{R}$ and $I_a : \mathbb{R} \to \mathbb{R}$ be the functions defined as :

$$\psi_a(u) = \int_{[0,1]\times\mathbb{R}^d} s^{-ad}\left(e^{s^{ad}uK(z)} - 1\right) dsdz \quad \text{and} \quad I_a(t) = \sup_{u\in\mathbb{R}} \{ut - \psi(u)\}$$

(where $s \in [0, 1]$, $z \in \mathbb{R}^d$) and set

$$S_+ = \left\{x \in \mathbb{R}^d; K(x) > 0\right\} \quad \text{and} \quad S_- = \left\{x \in \mathbb{R}^d; K(x) < 0\right\}.$$

The following proposition gives the properties of the functions ψ_a and I_a ; in particular, the behaviour of the rate function I_a, which differs depending on whether K is non-negative or not, is explicited.

Proposition 1 *Let λ be the Lebesgue measure on $\mathbb{R}^d$ and let Assumption (H1) holds.*

(i) ψ_a is strictly convex, twice continuously differentiable on $\mathbb{R}$, and I_a is a good rate function on $\mathbb{R}$.

(ii) If $\lambda(S_-) = 0$, then $I_a(t) = +\infty$ when $t < 0$, $I_a(0) = \lambda(S_+)$, I_a is strictly convex on $\mathbb{R}$ and continuous on $]0, +\infty[$, and for any $t > 0$

$$I_a(t) = t\left(\psi_a'\right)^{-1}(t) - \psi_a\left(\left(\psi_a'\right)^{-1}(t)\right). \tag{2.3}$$

(iii) If $\lambda(S_-) > 0$, then I_a is finite and strictly convex on $\mathbb{R}$ and (2.3) holds for any $t \in \mathbb{R}$.

(iv) In both cases, the strict minimum of I_a is achieved by $I_a(1) = 0$.

Remark 1 *The following relations are straightforward, and will be used in the sequel :*

$$I_a(t) = \begin{cases} \sup_{u>0}\{ut - \psi_a(u)\} & \text{if } t > 1 \\ \sup_{u<0}\{ut - \psi_a(u)\} & \text{if } t < 1. \end{cases} \tag{2.4}$$

We can now state the LDP for the density estimator.

Theorem 1 ***(Pointwise LDP for the density estimator)***
Let Hypotheses (H1)-(H2) hold and assume that f is continuous at x. Then, the sequence $(f_n(x) - f(x))$ satisfies a LDP with speed (nh_n^d) and rate function defined as follows :

$$\begin{cases} \text{if } \; f(x) \neq 0, & I_{a,x} : t \mapsto f(x) I_a \left(1 + \frac{t}{f(x)}\right) \\ \text{if } \; f(x) = 0, & I_{a,x}(0) = 0 \;\text{ and }\; I_{a,x}(t) = +\infty \;\text{ for }\; t \neq 0. \end{cases}$$

2.2.2 Pointwise LDP for the derivatives estimators

Let $[\alpha]$ be a d-uplet such that $|\alpha| \geq 1$. To establish pointwise LDP for $\partial^{[\alpha]} f_n$, we need the following assumptions.

(H3) $h_n = h(n)$ where the function h is locally bounded and varies regularly with exponent $(-a)$, $0 < a < 1/(d + 2|\alpha|)$.

(H4) i) K is $|\alpha|$-times differentiable on $\mathbb{R}^d$ and $\lim_{\|x\| \to \infty} \|D^{(j)} K(x)\| = 0$ for any $j \in \{0, \dots, |\alpha| - 1\}$.
ii) $\partial^{[\alpha]} K : \mathbb{R}^d \to \mathbb{R}$ is a bounded and integrable function and $\int_{\mathbb{R}^d} \left[\partial^{[\alpha]} K(x)\right]^2 dx \neq 0$.

(H5) f is $|\alpha|$-times differentiable on $\mathbb{R}^d$ and its j-th differentials $D^{(j)} f$ are bounded on $\mathbb{R}^d$ for any $j \in \{0, \dots, |\alpha| - 1\}$.

Remark 2 *A positive (not necessarily monotone) function L defined on $]0, \infty[$ is slowly varying if $\lim_{t\to\infty} L(tx)/L(t) = 1$; a function G is said to vary regularly with exponent ρ, $\rho \in \mathbb{R}$, if and only if there exists a slowly varying function L such that $G(x) = x^\rho L(x)$ (see, for example, Feller [6] page 275). Typical examples of regularly varying functions with exponent ρ are x^ρ, $x^\rho \log x$, $x^\rho \log\log x$, $x^\rho \log x / \log\log x$, and so on. An important consequence of (H3) which will be used in the sequel is :*

$$\text{if} \quad \beta a < 1, \quad \text{then} \quad \lim_{n\to\infty} \frac{1}{n h_n^\beta} \sum_{i=1}^n h_i^\beta = \frac{1}{1 - a\beta}. \tag{2.5}$$

Theorem 2 ***(Pointwise LDP for the derivatives estimators)***
Let $|\alpha| \geq 1$ and assume that (H1), (H3)-(H5) hold and that $\partial^{[\alpha]} f$ is continuous at x. Then, the sequence $\left(\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right)$ satisfies a LDP with speed $(nh_n^{d+2|\alpha|})$ and rate function $J_{a,[\alpha],x}$ defined by (2.2).

2.2.3 Pointwise MDP for the density estimator and its derivatives

Let (v_n) be a positive sequence; we assume that

(H6) $\lim_{n\to\infty} v_n = \infty$ and $\lim_{n\to\infty} \dfrac{v_n^2}{n h_n^{d+2|\alpha|}} = 0$.

(H7) i) There exists an integer $q \geq 2$ such that $\forall s \in \{1, \dots, q-1\}, \forall j \in \{1, \dots, d\}$, $\int_{\mathbb{R}^d} y_j^s K(y) dy_j = 0$, and $\int_{\mathbb{R}^d} \left|y_j^q K(y)\right| dy < \infty$.
ii) $\lim_{n\to\infty} \dfrac{v_n}{n} \sum_{i=1}^n h_i^q = 0$.
iii) $\partial^{[\alpha]} f$ is q-times differentiable on $\mathbb{R}^d$ and $M_q = \sup_{x\in\mathbb{R}^d} \|D^q \partial^{[\alpha]} f(x)\| < +\infty$.

Remark 3 *When $h_n = O(n^{-a})$, with $0 < a < 1/\left(d + 2|\alpha|\right)$, (H6) and (H7)ii) hold for instance for $(v_n) \equiv (n^b)$ for any $b \in]0, \min\{aq; (1 - a(d + 2|\alpha|))/2\}[$.*

The following theorem gives the MDP for the density estimator and its derivatives.

Theorem 3 ***(Pointwise MDP)***
When $|\alpha| = 0$, let Assumptions (H1), (H3), (H6) and (H7) hold ; when $|\alpha| \geq 1$, let (H1), (H3)-(H7) hold. If $\partial^{[\alpha]} f$ is q-times differentiable at x, then the sequence $\left(v_n\left(\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right)\right)$ satisfies a LDP with speed $\left(nh_n^{d+2|\alpha|}/v_n^2\right)$ and good rate function $J_{a,[\alpha],x}$ defined in (2.2).

2.2.4 Uniform LDP and MDP for the density estimator and its derivatives

To establish uniform large deviations principles for the density estimator and its derivatives, we need the following additionnal assumptions :

(H8) i) There exists $\xi > 0$ such that $\int_{\mathbb{R}^d} \|x\|^{\xi} f(x)dx < \infty$.
ii) f is uniformly continuous.

(H9) i) $\partial^{[\alpha]} K$ is Hölder continuous.
ii) There exists $\gamma > 0$ such that $z \mapsto \|z\|^{\gamma} \partial^{[\alpha]} K(z)$ is a bounded function.

(H10) $\lim_{n\to\infty} \dfrac{v_n^2 \log(1/h_n)}{nh_n^{d+2|\alpha|}} = 0$ and $\lim_{n\to\infty} \dfrac{v_n^2 \log v_n}{nh_n^{d+2|\alpha|}} = 0$.

(H11) i) There exists $\zeta > 0$ such that $\int_{\mathbb{R}^d} \|z\|^{\zeta} |K(z)|\, dz < \infty$.
ii) There exists $\eta > 0$ such that $z \mapsto \|z\|^{\eta} \partial^{[\alpha]} f(z)$ is a bounded function.

Remark 4 *When $h_n = O\left(n^{-a}\right)$ with $a \in]0, 1/(d+2|\alpha|)[$, (H10) holds for instance with $(v_n) \equiv (n^b)$ for any $b \in]0, (1 - a(d + 2|\alpha|))/2[$.*

Set $U \subseteq \mathbb{R}^d$; in order to state in a compact form the uniform large and moderate deviations principles for the density estimator and its derivatives on U, we consider the large deviations case as the special case $(v_n) \equiv 1$ and we set :

$$
\begin{aligned}
g_U(\delta) &= \begin{cases} \|f\|_{U,\infty} I_a\left(1 + \frac{\delta}{\|f\|_{U,\infty}}\right) & \text{if } |\alpha| = 0 \text{ and } (v_n) \equiv 1 \\ \frac{\delta^2 (1+a(d+2|\alpha|))}{2\|f\|_{U,\infty} \int_{\mathbb{R}^d} \left[\partial^{[\alpha]} K\right]^2 (z)dz} & \text{otherwise,} \end{cases} \\
\tilde{g}_U(\delta) &= \min\{g_U(\delta), g_U(-\delta)\},
\end{aligned}
$$

where $\|f\|_{U,\infty} = \sup_{x\in U} |f(x)|$.

Remark 5 *The functions $g_U(\cdot)$ and $\tilde{g}_U(\cdot)$ are non-negative, continuous, increasing on $]0, +\infty[$ and decreasing on $]-\infty, 0[$, with a unique global minimum in 0 ($\tilde{g}_U(0) = g_U(0) = 0$). They are thus good rate functions (and $g_U(\cdot)$ is strictly convex).*

Theorem 4 below states uniform LDP and MDP for $\left(\partial^{[\alpha]} f_n - \partial^{[\alpha]} f\right)$ on U in the case U is bounded, and Theorem 5 in the case U is unbounded.

Theorem 4 ***(Uniform deviations on a bounded set)***
In the case $|\alpha| = 0$, let (H1), (H2), (H7), (H9)i), and (H10) hold. In the case $|\alpha| \geq 1$, let (H3)-(H5), (H7), (H9)i) and (H10) hold. Moreover, assume either that $(v_n) \equiv 1$ or that (v_n) satisfies (H6). Then for any bounded subset U of $\mathbb{R}^d$ and for all $\delta > 0$,

$$\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] = -\tilde{g}_U(\delta). \tag{2.6}$$

Theorem 5 ***(Uniform deviations on an unbounded set)***
Let Assumptions (H1), (H7)-(H11) hold. Moreover,

- *in the case $|\alpha| = 0$ and $(v_n) \equiv 1$, let (H2) holds ;*
- *in the case $|\alpha| \geq 1$ and $(v_n) \equiv 1$, or $|\alpha| \geq 0$ and (v_n) satisfies (H6), let (H3)-(H5) hold.*

Then for any subset U of $\mathbb{R}^d$ and for all $\delta > 0$,

$$\begin{aligned} -\tilde{g}_U(\delta) &\leq \liminf_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] \\ &\leq \limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] \leq -\frac{\xi}{\xi+d}\tilde{g}_U(\delta). \end{aligned}$$

The following corollary is a straightforward consequence of Theorem 5.

Corollary 1
Under the assumptions of Theorem 5, if $\int_{\mathbb{R}^d} \|x\|^\xi f(x)dx < \infty$ $\forall \xi \in \mathbb{R}$, then for any subset U of $\mathbb{R}^d$,

$$\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| \geq \delta\right] = -\tilde{g}_U(\delta). \tag{2.7}$$

Comment Theorem 4 and Corollary 1 are LDP for the sequence $(\sup_{x\in U} |f_n(x) - f(x)|)$. As a matter of fact, since the sequence $(\sup_{x\in U} |f_n(x) - f(x)|)$ is positive and since $\tilde{g}_U$ is continuous on $[0, +\infty[$, increasing and goes to infinity as $\delta \to \infty$, the application of Lemma 5 in Worms [24] allows to deduce from (2.6) or (2.7) that $(\sup_{x\in U} |f_n(x) - f(x)|)$ satisfies a LDP with speed $\left(nh_n^d\right)$ and good rate function $\tilde{g}_U$ on $\mathbb{R}_+$.

2.3 Proofs

Let $(\Psi_n^{[\alpha]})$ and $(B_n^{[\alpha]})$ be the sequences defined as

$$\begin{aligned} \Psi_n^{[\alpha]}(x) &= \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}}\left(\partial^{[\alpha]}K\left(\frac{x-X_i}{h_i}\right) - \mathbb{E}\left[\partial^{[\alpha]}K\left(\frac{x-X_i}{h_i}\right)\right]\right), \\ B_n^{[\alpha]}(x) &= \mathbb{E}\left[\partial^{[\alpha]} f_n(x)\right] - \partial^{[\alpha]} f(x). \end{aligned}$$

We have :

$$\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x) = \Psi_n^{[\alpha]}(x) + B_n^{[\alpha]}(x). \tag{2.8}$$

Theorems 1, 2 and 3 are consequences of (2.8) and the following propositions.

Proposition 2 ***(Pointwise LDP and MDP for $(\Psi_n^{[\alpha]})$)***

1. *Under the assumptions (H1) and (H2), the sequence $(f_n(x) - \mathbb{E}(f_n(x)))$ satisfies a LDP with speed $\left(nh_n^d\right)$ and rate function $I_{a,x}$.*
2. *Let $|\alpha| \geq 1$ and assume that (H3), (H4) hold, then the sequence $\left(\Psi_n^{[\alpha]}(x)\right)$ satisfies a LDP with speed $\left(nh_n^{d+2|\alpha|}\right)$ and rate function $J_{a,[\alpha],x}$.*
3. *When $|\alpha| = 0$, let Assumptions (H1), (H2) and (H6) hold and when $|\alpha| \geq 1$, let (H3), (H4) and (H6) hold, then the sequence $\left(v_n\Psi_n^{[\alpha]}(x)\right)$ satisfies a LDP with speed $\left(nh_n^{d+2|\alpha|}/v_n^2\right)$ and rate function $J_{a,[\alpha],x}$.*

Proposition 3 ***(Uniform LDP and MDP for $(\Psi_n^{[\alpha]})$)***

1. *In the case $|\alpha| = 0$, let (H1), (H2), (H9)i) and (H10) hold. In the case $|\alpha| \geq 1$, let (H3)-(H5), (H9)i) and (H10) hold. Moreover, assume either that $(v_n) \equiv 1$ or that (v_n) satisfies (H6); then for any bounded subset U of $\mathbb{R}^d$ and for all $\delta > 0$,*

$$\lim_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]}\Psi_n(x)\right| \geq \delta\right] = -\tilde{g}_U(\delta).$$

2. *Let Assumptions (H1), (H8)-(H11) hold . Moreover,*
   - *in the case $|\alpha| = 0$ and $(v_n) \equiv 1$, let (H2) holds,*
   - *in the case either $|\alpha| \geq 1$ and $(v_n) \equiv 1$, or $|\alpha| \geq 0$ and (v_n) satisfies (H6), let (H3)-(H5) hold.*

   Then for any subset U of $\mathbb{R}^d$ and for all $\delta > 0$,

$$\begin{aligned} -\tilde{g}_U(\delta) &\leq \liminf_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]}\Psi_n(x)\right| \geq \delta\right] \\ &\leq \limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]}\Psi_n(x)\right| \geq \delta\right] \leq -\frac{\xi}{\xi+d}\tilde{g}_U(\delta). \end{aligned}$$

Proposition 4 ***(Pointwise and uniform convergence rate of $B_n^{[\alpha]}$)***
Let Assumptions (H1), (H3)-(H5) and (H7)i) hold.

1) If $\partial^{[\alpha]} f$ is q-times differentiable at x, then

$$\mathbb{E}\left(\partial^{[\alpha]} f_n(x)\right) - \partial^{[\alpha]} f(x) = O\left(\frac{\sum_{i=1}^n h_i^q}{n}\right).$$

2) If (H7)iii) holds, then :

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^n h_i^q} \sup_{x\in\mathbb{R}^d} \left|\mathbb{E}\left(\partial^{[\alpha]} f_n(x)\right) - \partial^{[\alpha]} f(x)\right| \leq \frac{M_q}{q!}\int_{\mathbb{R}^d} \|z\|^q \left|K(z)\right| dz.$$

Set $x \in \mathbb{R}^d$; since the assumptions of Theorems 1 and 2 guarantee that $\lim_{n\to\infty} B_n^{[\alpha]}(x) = 0$, Theorem 1 (respectively Theorem 2) is a straightforward consequence of the application of Part 1 (respectively of Part 2) of Proposition 2. Moreover, under the assumptions of Theorem 3, we have, by application of Part 1 of Proposition 4, $\lim_{n\to\infty} v_n B_n^{[\alpha]}(x) = 0$; Theorem 3 thus straightfully follows from the application of Part 3 of Proposition 2. Finally, Theorems 4 and 5 are obtained by applying Proposition 3 together with the second part of Proposition 4.

We now state a preliminary lemma, which will be used in the proof of Propositions 2 and 3. For any $u \in \mathbb{R}$, set

$$\begin{aligned} \Lambda_{n,x}(u) &= \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{E}\left[\exp\left(u\frac{nh_n^{d+2|\alpha|}}{v_n}\Psi_n^{[\alpha]}(x)\right)\right], \\ \Lambda_x^L(u) &= f(x)\left(\psi_a(u) - u\right), \\ \Lambda_x^M(u) &= \frac{u^2}{2(1+a\,(d+2|\alpha|))} f(x)\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(z)\right]^2 dz. \end{aligned}$$

Lemma 1 ***(Convergence of $\Lambda_{n,x}$)***

- *In the case $|\alpha| = 0$ and $(v_n) \equiv 1$, let (H1) and (H2) hold;*

- *In the case either $|\alpha| \geq 1$ and $(v_n) \equiv 1$, or $|\alpha| \geq 0$ and (v_n) satisfies (H6), let (H1), (H3) and (H4) hold.*

1. *(Pointwise convergence)*
   If f is continuous at x, then for all $u \in \mathbb{R}$,

$$\lim_{n\to\infty} \Lambda_{n,x}(u) = \Lambda_x(u) \tag{2.9}$$

   where

$$\Lambda_x(u) = \begin{cases} \Lambda_x^L(u) & \text{when } v_n \equiv 1 \text{ and } |\alpha| = 0 \\ \Lambda_x^M(u) & \text{when either } v_n \to \infty \text{ and } |\alpha| \geq 0 \text{ or } v_n \equiv 1 \text{ and } |\alpha| \geq 1. \end{cases}$$

2. *(Uniform convergence)*
   If f is uniformly continuous, then the convergence (2.9) holds uniformly in $x \in U$.

Our proofs are now organized as follows : Lemma 1 is proved in Section 2.3.1, Proposition 2 in Section 2.3.2, Proposition 3 in Section 2.3.3 and Proposition 4 in Section 2.3.4. Section 2.3.5 is devoted to the proof of Proposition 1 on the rate function I_a.

2.3.1 Proof of Lemma 1

Set $u \in \mathbb{R}$, $u_n = u/v_n$, $Y_i = \partial^{[\alpha]} K\left(\frac{x-X_i}{h_i}\right)$ and $a_n = nh_n^{d+2|\alpha|}$. We have :

$$\begin{aligned} \Lambda_{n,x}(u) &= \frac{v_n^2}{a_n} \log \mathbb{E}\left[\exp\left(u_n a_n \Psi_n^{[\alpha]}(x)\right)\right] \\ &= \frac{v_n^2}{a_n} \log \mathbb{E}\left[\exp\left(\frac{u_n a_n}{n} \sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \left(Y_i - \mathbb{E}(Y_i)\right)\right)\right] \\ &= \frac{v_n^2}{a_n} \sum_{i=1}^n \log \mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right)\right] - \frac{u v_n}{n} \sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}(Y_i). \end{aligned}$$

By Taylor expansion, there exists $c_{i,n}$ between 1 and $\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right)\right]$ such that

$$\begin{aligned} &\log\left(\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right)\right]\right) \\ &= \mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right) - 1\right] - \frac{1}{2c_{i,n}^2}\left(\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right) - 1\right]\right)^2 \end{aligned}$$

and $\Lambda_{n,x}$ can be rewritten as

$$\begin{aligned} &\Lambda_{n,x}(u) \\ &= \frac{v_n^2}{a_n} \sum_{i=1}^n \mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right) - 1\right] - \frac{v_n^2}{2a_n} \sum_{i=1}^n \frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right)\right] - 1\right)^2 \\ &\quad - \frac{u v_n}{n} \sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}(Y_i). \end{aligned} \tag{2.10}$$

For proving Lemma 1, we consider two cases :

First case : either $v_n \to \infty$ or $|\alpha| \geq 1$.

A Taylor's expansion implies the existence of $c'_{i,n}$ between 0 and $\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i$ such that

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right) - 1\right] \\
&= \frac{u_n a_n}{n h_i^{d+|\alpha|}} \mathbb{E}(Y_i) + \frac{1}{2}\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}}\right)^2 \mathbb{E}(Y_i^2) + \frac{1}{6}\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}}\right)^3 \mathbb{E}\left(Y_i^3 e^{c'_{i,n}}\right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\Lambda_{n,x}(u) &= \frac{u^2 a_n}{2n^2} \sum_{i=1}^{n} \frac{1}{h_i^{2d+2|\alpha|}} \mathbb{E}\left[Y_i^2\right] + R_{n,x}^{(1)}(u) \\
&= f(x) \frac{u^2 a_n}{2n^2} \sum_{i=1}^{n} \frac{1}{h_i^{d+2|\alpha|}} \int_{\mathbb{R}^d} \left(\partial^{[\alpha]} K(z)\right)^2 dz + R_{n,x}^{(1)}(u) + R_{n,x}^{(2)}(u)
\end{aligned}$$

with

$$\begin{aligned}
R_{n,x}^{(1)}(u) &= \frac{1}{6} \frac{u^3 a_n^2}{n^3 v_n} \sum_{i=1}^{n} h_i^{-3d-3|\alpha|} \mathbb{E}\left(Y_i^3 e^{c'_{i,n}}\right) - \frac{v_n^2}{2a_n} \sum_{i=1}^{n} \frac{1}{c_{i,n}^2} \left(\mathbb{E}\left[\exp\left(\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right) - 1\right]\right)^2 \\
R_{n,x}^{(2)}(u) &= \frac{u^2 a_n}{2n^2} \sum_{i=1}^{n} \frac{1}{h_i^{d+2|\alpha|}} \int_{\mathbb{R}^d} \left(\partial^{[\alpha]} K(z)\right)^2 \left[f(x - h_i z) - f(x)\right] dz.
\end{aligned}$$

Since $|Y_i| \leq \|\partial^{[\alpha]} K\|_\infty$, we have

$$\left|\frac{u_n a_n}{n h_i^{d+|\alpha|}} Y_i\right| \leq |u_n| \|\partial^{[\alpha]} K\|_\infty,$$

so that

$$c'_{i,n} \leq |u_n| \|\partial^{[\alpha]} K\|_\infty, \tag{2.11}$$

and

$$\frac{1}{c_{i,n}^2} \leq \exp\left(2|u_n| \|\partial^{[\alpha]} K\|_\infty\right). \tag{2.12}$$

Noting that $\mathbb{E}|Y_i|^3 \leq h_i^d \|f\|_\infty \int_{\mathbb{R}^d} \left|\partial^{[\alpha]} K(z)\right|^3 dz$. Hence, using (2.5) and (2.11), there exists a positive constant c_1 such that, for n large enough,

$$\begin{aligned}
&\left|\frac{u^3 a_n^2}{n^3 v_n} \sum_{i=1}^{n} h_i^{-3d-3|\alpha|} \mathbb{E}\left(Y_i^3 e^{c'_{i,n}}\right)\right| \\
&\leq c_1 \frac{\|f\|_\infty |u|^3 e^{|u_n| \|\partial^{[\alpha]} K\|_\infty} h_n^{|\alpha|}}{\left(1 + a(2d + 3|\alpha|)\right) v_n} \int_{\mathbb{R}^d} \left|\partial^{[\alpha]} K(z)\right|^3 dz
\end{aligned} \tag{2.13}$$

which goes to 0 as $n \to \infty$ since either $v_n \to \infty$ or $|\alpha| \geq 1$.
In the same way, in view of (2.5) and (2.12), there exists a positive constant c_2 such that, for n

large enough,

$$\begin{aligned}
&\left|\frac{v_n^2}{2a_n}\sum_{i=1}^n \frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[\exp\left(\frac{u_n a_n}{nh_i^{d+|\alpha|}}Y_i\right)-1\right]\right)^2\right| \\
&\quad\leq\ c_2 u^2\|f\|_\infty^2 e^{4|u_n|\|\partial^{[\alpha]}K\|_\infty}\left(\int_{\mathbb{R}^d}\left|\partial^{[\alpha]}K(z)\right|dz\right)^2 h_n^d \qquad (2.14)
\end{aligned}$$

which goes to 0 as $n\to\infty$. The combination of (2.13) and (2.14) ensures that

$$\lim_{n\to\infty}\sup_{x\in\mathbb{R}^d}|R_{n,x}^{(1)}(u)| = 0.$$

Now, since f is continuous, we have $\lim_{i\to\infty}|f(x-h_iz)-f(x)| = 0$, and thus, by the dominated convergence theorem, (H4)ii) implies that

$$\lim_{i\to\infty}\int_{\mathbb{R}^d}\left(\partial^{[\alpha]}K(z)\right)^2|f(x-h_iz)-f(x)|\,dz = 0.$$

Since, in view of (2.5), the sequence $\left(\frac{u^2a_n}{2n^2}\sum_{i=1}^n\frac{1}{h_i^{d+2|\alpha|}}\right)$ is bounded, it follows that $\lim_{n\to\infty}|R_{n,x}^{(2)}(u)| = 0$. The pointwise convergence (2.9) then follows.
In the case f is uniformly continuous, we have $\lim_{i\to\infty}\sup_{x\in\mathbb{R}^d}|f(x-h_iz)-f(x)| = 0$ and thus, using the same arguments as previously, we obtain $\lim_{n\to\infty}\sup_{x\in\mathbb{R}^d}|R_{n,x}^{(2)}(u)| = 0$.
We then deduce that $\lim_{n\to\infty}\sup_{x\in U}|\Lambda_{n,x}(u)-\Lambda_x(u)| = 0$ which concludes the proof of Lemma 1 in this case.

Second case : $|\alpha| = 0$ and $(v_n)\equiv 1$.

Using assumption (H2) and in view of (2.10), we have

$$\begin{aligned}
\Lambda_{n,x}(u) &= \frac{1}{a_n}\sum_{i=1}^n\mathbb{E}\left[e^{\frac{ua_n}{nh_i^d}Y_i}-1\right]-\frac{1}{2a_n}\sum_{i=1}^n\frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[e^{\frac{ua_n}{nh_i^d}Y_i}-1\right]\right)^2-\frac{u}{n}\sum_{i=1}^n h_i^{-d}\mathbb{E}(Y_i) \\
&= \frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\mathbb{R}^d}\left[e^{u\left(\frac{i}{n}\right)^{ad}K(z)}-1\right]f(x)dz-u\int_{\mathbb{R}^d}K(z)f(x)dz-R_{n,x}^{(3)}(u)+R_{n,x}^{(4)}(u)
\end{aligned}$$

with

$$\begin{aligned}
R_{n,x}^{(3)}(u) &= \frac{1}{2nh_n^d}\sum_{i=1}^n\frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[e^{\frac{uh_n^d}{h_i^d}Y_i}-1\right]\right)^2 \\
R_{n,x}^{(4)}(u) &= \frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\mathbb{R}^d}\left(e^{\frac{uh_n^d}{h_i^d}K(z)}-1\right)[f(x-h_iz)-f(x)]\,dz \\
&\quad -\frac{u}{n}\sum_{i=1}^n\int_{\mathbb{R}^d}K(z)\,[f(x-h_iz)-f(x)]\,dz.
\end{aligned}$$

Using the bound (2.14), we have $\lim_{n\to\infty}\sup_{x\in\mathbb{R}^d}|R_{n,x}^{(3)}(u)|=0$.
Since $|e^t-1|\leq |t|e^{|t|}$, we have

$$\begin{aligned}\left|R_{n,x}^{(4)}(u)\right| &\leq \frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\mathbb{R}^d}\left|\left(e^{\frac{uh_n^d}{h_i^d}K(z)}-1\right)[f(x-h_iz)-f(x)]\right|dz\\ &\quad+\frac{|u|}{n}\sum_{i=1}^n\int_{\mathbb{R}^d}|K(z)|\,|f(x-h_iz)-f(x)|\\ &\leq |u|e^{|u|\|K\|_\infty}\frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\mathbb{R}^d}|K(z)|\,|f(x-h_iz)-f(x)|\,dz\\ &\quad+\frac{|u|}{n}\sum_{i=1}^n\int_{\mathbb{R}^d}|K(z)|\,|f(x-h_iz)-f(x)|\,.\end{aligned}$$

In the case f is continuous, since in view of (2.5) the sequence $\left(\frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\right)$ is bounded, the dominated convergence theorem ensures that $\lim_{n\to\infty}R_{n,x}^{(4)}(u)=0$.
In the case f is uniformly continuous, set $\varepsilon>0$ and let $M>0$ such that $2\|f\|_\infty\int_{\|z\|\geq M}|K(z)|\,dz<\varepsilon/2$. We need to prove that for n sufficiently large

$$\sup_{x\in\mathbb{R}^d}\int_{\|z\|\leq M}|K(z)|\,|f(x-h_iz)-f(x)|\,dz \quad\leq\quad \varepsilon/2$$

which is a straightforward consequence of the uniform continuity of f. It follows from analysis considerations that

$$\lim_{n\to\infty}\Lambda_{n,x}(u) \;=\; f(x)\int_{\mathbb{R}^d}\left[\int_0^1 s^{-ad}\left(e^{s^{ad}uK(z)}-1-us^{ad}K(z)\right)ds\right]dz,$$

and thus Lemma 1 is proved. ■

2.3.2 Proof of Proposition 2

To prove Proposition 2, we apply Proposition 1, Lemma 1 and the following result (see Puhalskii [17]).

Lemma 2 *Let (Z_n) be a sequence of real random variables, (ν_n) a positive sequence satisfying $\lim_{n\to\infty}\nu_n=+\infty$, and suppose that there exists some convex non-negative function Γ defined (*i.e. *finite) on $\mathbb{R}$ such that*

$$\forall u\in\mathbb{R},\quad \lim_{n\to\infty}\frac{1}{\nu_n}\log\mathbb{E}\left[\exp(u\nu_nZ_n)\right]=\Gamma(u).$$

If the Legendre transform $\tilde{\Gamma}$ of Γ is a strictly convex function, then the sequence (Z_n) satisfies a LDP of speed (ν_n) and good rate function $\tilde{\Gamma}$.

In our framework, when $|\alpha|=0$ and $v_n\equiv 1$, we take $Z_n=f_n(x)-\mathbb{E}(f_n(x))$, $\nu_n=nh_n^d$ with $h_n=cn^{-a}$ where $0<a<1/d$ and $\Gamma=\Lambda_x^L$. In this case, the Legendre transform of $\Gamma=\Lambda_x^L$ is the rate function $I_{a,x}:t\mapsto f(x)I_a\left(\frac{t}{f(x)}+1\right)$ which is strictly convex by Proposition 1. Otherwise, we take $Z_n=v_n\left(\partial^{[\alpha]}f_n(x)-\mathbb{E}\left(\partial^{[\alpha]}f_n(x)\right)\right)$, $\nu_n=nh_n^{d+2|\alpha|}/v_n^2$ and $\Gamma=\Lambda_x^M$; $\tilde{\Gamma}$ is then the quadratic rate function $J_{a,[\alpha],x}$ defined in (2.2) and thus Proposition 2 follows. ■

2.3.3 Proof of Proposition 3

In order to prove Proposition 3, we first establish some lemmas.

Lemma 3 *Let $\phi_a : \mathbb{R}+ \to \mathbb{R}$ be the function defined for $\delta > 0$ as*

$$\phi_a(\delta) \quad = \quad \begin{cases} (\psi_a')^{-1}\left(\frac{\delta}{\|f\|_{U,\infty}} + 1\right) & \text{if} \quad (v_n) \equiv 1 \quad \text{and} \quad |\alpha| = 0, \\ \frac{\delta(1+a(d+2|\alpha|))}{\|f\|_{U,\infty}\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K\right]^2(z)dz} & \text{otherwise.} \end{cases}$$

1. $\sup_{u\in\mathbb{R}}\{u\delta - \sup_{x\in U}\Lambda_x(u)\}$ *equals* $g_U(\delta)$ *and is achieved for* $u = \phi_a(\delta) > 0$.
2. $\sup_{u\in\mathbb{R}}\{-u\delta - \sup_{x\in U}\Lambda_x(u)\}$ *equals* $g_U(-\delta)$ *and is achieved for* $u = \phi_a(-\delta) < 0$.

Proof of Lemma 3

We just prove the first part, the proof of the second one being similar. First, let us consider the case $(v_n) \equiv 1$ and $|\alpha| = 0$. Since $e^t \geq 1 + t$ $(\forall t)$, we have $\psi(u) \geq u$ and therefore,

$$\begin{aligned} u\delta - \sup_{x\in U}\Lambda_x(u) &= u\delta - \|f\|_{U,\infty}(\psi_a(u) - u) \\ &= \|f\|_{U,\infty}\left[u\left(\frac{\delta}{\|f\|_{U,\infty}} + 1\right) - \psi_a(u)\right] \end{aligned}$$

The function $u \mapsto u\delta - \sup_{x\in U}\Lambda_x(u)$ has second derivative $-\|f\|_{U,\infty}\psi_a''(u) < 0$ and thus it has a unique maximum achieved for

$$u_0 = (\psi_a')^{-1}\left(\frac{\delta}{\|f\|_{U,\infty}} + 1\right).$$

Now, since ψ_a' is increasing and since $\psi_a'(0) = 1$, we deduce that $u_0 > 0$.
In the case $\lim_{n\to\infty} v_n = \infty$, Lemma 3 is established in the same way by noting that

$$\begin{aligned} u\delta - \sup_{x\in U}\Lambda_x(u) &= u\delta - \sup_{x\in U}\Lambda_x^M(u) \\ &= u\delta - \frac{u^2}{2\left(1 + a\left(d + 2|\alpha|\right)\right)}\|f\|_{U,\infty}\int_{\mathbb{R}^d}\left[\partial^{[\alpha]}K(z)\right]^2 dz. \quad \blacksquare \end{aligned}$$

Lemma 4
- *In the case $|\alpha| = 0$ and $(v_n) \equiv 1$, let (H1) and (H2) hold ;*
- *In the case either $|\alpha| \geq 1$ and $(v_n) \equiv 1$, or $|\alpha| \geq 0$ and (v_n) satisfies (H6), let (H1), (H3) and (H4) hold.*

Then for any $\delta > 0$,

$$\begin{aligned} \lim_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\sup_{x\in U}\mathbb{P}\left[v_n\left(\partial^{[\alpha]}f_n(x) - \mathbb{E}\left(\partial^{[\alpha]}f_n(x)\right)\right) \geq \delta\right] &= -g_U(\delta) \\ \lim_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\sup_{x\in U}\mathbb{P}\left[v_n\left(\partial^{[\alpha]}f_n(x) - \mathbb{E}\left(\partial^{[\alpha]}f_n(x)\right)\right) \leq -\delta\right] &= -g_U(-\delta) \\ \lim_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\sup_{x\in U}\mathbb{P}\left[v_n\left|\partial^{[\alpha]}f_n(x) - \mathbb{E}\left(\partial^{[\alpha]}f_n(x)\right)\right| \geq \delta\right] &= -\tilde{g}_U(\delta). \end{aligned}$$

Proof of Lemma 4

Set $b_n = nh_n^{d+2|\alpha|}/v_n^2$, $S_n(x) = v_n \Psi_n^{[\alpha]}(x)$, and $\delta > 0$. In the sequel, $\Lambda_x(u)$ is defined as in (2.9). We first note that, for any $u > 0$,

$$\begin{aligned}
\mathbb{P}\left[S_n(x) \geq \delta\right] &= \mathbb{P}\left[e^{b_n u S_n(x)} \geq e^{b_n u \delta}\right] \\
&\leq e^{-b_n u\delta}\mathbb{E}\left[e^{b_n u S_n(x)}\right] \\
&\leq e^{-b_n u\delta} e^{b_n \Lambda_{n,x}(u)} \\
&\leq e^{-b_n(u\delta - \Lambda_x(u))} e^{b_n(\Lambda_{n,x}(u) - \Lambda_x(u))}.
\end{aligned}$$

Therefore, for every $u > 0$,

$$\sup_{x\in U} \mathbb{P}\left[S_n(x) \geq \delta\right] \leq e^{-b_n\left(u\delta - \sup_{x\in U}\Lambda_x(u)\right)} e^{b_n \sup_{x\in U}|\Lambda_{n,x}(u) - \Lambda_x(u)|}. \tag{2.15}$$

Similarly, we prove that, for every $u < 0$,

$$\sup_{x\in U} \mathbb{P}\left[S_n(x) \leq -\delta\right] \leq e^{-b_n\left(-u\delta - \sup_{x\in U}\Lambda_x(u)\right)} e^{b_n \sup_{x\in U}|\Lambda_{n,x}(u) - \Lambda_x(u)|}. \tag{2.16}$$

The application of Lemma 3 to (2.15) and (2.16) yields

$$\begin{aligned}
\sup_{x\in U} \mathbb{P}\left[S_n(x) \geq \delta\right] &\leq e^{-b_n g_U(\delta)} e^{b_n \sup_{x\in U}|\Lambda_{n,x}(\phi_a(\delta)) - \Lambda_x(\phi_a(\delta))|} \\
\sup_{x\in U} \mathbb{P}\left[S_n(x) \leq -\delta\right] &\leq e^{-b_n g_U(-\delta)} e^{b_n \sup_{x\in U}|\Lambda_{n,x}(\phi_a(-\delta)) - \Lambda_x(\phi_a(-\delta))|}
\end{aligned}$$

and the second part of Lemma 1 provides

$$\begin{aligned}
\lim_{n\to\infty} \sup_{x\in U} |\Lambda_{n,x}(\phi_a(\delta)) - \Lambda_x(\phi_a(\delta))| &= 0 \\
\lim_{n\to\infty} \sup_{x\in U} |\Lambda_{n,x}(\phi_a(-\delta)) - \Lambda_x(\phi_a(-\delta))| &= 0.
\end{aligned}$$

Consequently, it follows that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{b_n} \log \sup_{x\in U} \mathbb{P}\left[S_n(x) \geq \delta\right] &\leq -g_U(\delta) \\
\limsup_{n\to\infty} \frac{1}{b_n} \log \sup_{x\in U} \mathbb{P}\left[S_n(x) \leq -\delta\right] &\leq -g_U(-\delta)
\end{aligned}$$

and thus, setting $\tilde{g}_U(\delta) = \min\{g_U(\delta), g_U(-\delta)\}$,

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \sup_{x\in U} \mathbb{P}\left[|S_n(x)| \geq \delta\right] \leq -\tilde{g}_U(\delta).$$

In order to conclude the proof of Lemma 4, let us note that there exists $x_0 \in \overline{U}$ such that $f(x_0) = \|f\|_{U,\infty}$. The application of Proposition 2 at the point x_0 thus yields

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left[S_n(x_0) \geq \delta\right] &= -g_U(\delta) \\
\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left[S_n(x_0) \leq -\delta\right] &= -g_U(-\delta) \\
\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left[|S_n(x_0)| \geq \delta\right] &= -\tilde{g}_U(\delta).
\end{aligned}$$

The latter relation being due to the straightforward bounds

$$\begin{aligned}\max\{\mathbb{P}\left[S_n(x_0)\geq\delta\right],\mathbb{P}\left[S_n(x_0)\leq-\delta\right]\} &\leq \mathbb{P}\left[|S_n(x_0)|\geq\delta\right]\\ &\leq 2\max\{\mathbb{P}\left[S_n(x_0)\geq\delta\right],\mathbb{P}\left[S_n(x_0)\leq-\delta\right]\}. \blacksquare\end{aligned}$$

Lemma 5 *Let Assumptions (H1), (H3), (H4)i), (H9)i) and (H10) hold and assume that either* $(v_n)\equiv 1$ *or (H6) holds.*

1. *If* U *is a bounded set, then, for any* $\delta>0$, *we have*

$$\lim_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\sup_{x\in U}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right] \quad\leq\quad -\tilde{g}_U(\delta).$$

2. *If* U *is an unbounded set, then, for any* $b>0$ *and* $\delta>0$,

$$\limsup_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\sup_{x\in U,\|x\|\leq w_n}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right] \quad\leq\quad db-\tilde{g}_U(\delta)$$

where $w_n=\exp\left(bnh_n^{d+2|\alpha|}/v_n^2\right)$.

Proof of Lemma 5

Set $\rho\in]0,\delta[$, let β denote the Hölder order of $\partial^{[\alpha]}K$, and $\|\partial^{[\alpha]}K\|_H$ its corresponding Hölder norm. Set $w_n=\exp\left(bnh_n^{d+2|\alpha|}/v_n^2\right)$ and

$$R_n \quad=\quad \left(\frac{\rho n}{2\|\partial^{[\alpha]}K\|_H v_n\sum_{j=1}^n h_j^{-(d+\beta+|\alpha|)}}\right)^{\frac{1}{\beta}}.$$

We begin with the proof of the second part of Lemma 5. There exist $N'(n)$ points of $\mathbb{R}^d$, $y_1^{(n)},y_2^{(n)},\ldots,y_{N'(n)}^{(n)}$ such that the ball $\{x\in\mathbb{R}^d;\|x\|\leq w_n\}$ can be covered by the $N'(n)$ balls $B_i^{(n)}=\{x\in\mathbb{R}^d;\|x-y_i^{(n)}\|\leq R_n\}$ and such that $N'(n)\leq 2\left(\frac{2w_n}{R_n}\right)^d$. Considering only the $N(n)$ balls that intersect $\{x\in U;\|x\|\leq w_n\}$, we can write

$$\{x\in U;\|x\|\leq w_n\}\subset\cup_{i=1}^{N(n)}B_i^{(n)}.$$

For each $i\in\{1,\ldots,N(n)\}$, set $x_i^{(n)}\in B_i^{(n)}\cap U$. We then have :

$$\begin{aligned}\mathbb{P}\left[\sup_{x\in U,\|x\|\leq w_n}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right] &\leq \sum_{i=1}^{N(n)}\mathbb{P}\left[\sup_{x\in B_i^{(n)}}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right]\\ &\leq N(n)\max_{1\leq i\leq N(n)}\mathbb{P}\left[\sup_{x\in B_i^{(n)}}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right].\end{aligned}$$

Now, for any $i \in \{1, \ldots, N(n)\}$ and any $x \in B_i^{(n)}$,

$$\begin{aligned}
v_n \left| \Psi_n^{[\alpha]}(x) \right| &\leq v_n \left| \Psi_n^{[\alpha]}(x_i^{(n)}) \right| + \frac{v_n}{n} \sum_{j=1}^{n} \frac{1}{h_j^{d+|\alpha|}} \left| \partial^{[\alpha]} K \left(\frac{x - X_j}{h_j} \right) - \partial^{[\alpha]} K \left(\frac{x_i^{(n)} - X_j}{h_j} \right) \right| \\
&\quad + \frac{v_n}{n} \sum_{j=1}^{n} \frac{1}{h_j^{d+|\alpha|}} \mathbb{E} \left| \partial^{[\alpha]} K \left(\frac{x - X_j}{h_j} \right) - \partial^{[\alpha]} K \left(\frac{x_i^{(n)} - X_j}{h_j} \right) \right| \\
&\leq v_n \left| \Psi_n^{[\alpha]}(x_i^{(n)}) \right| + \frac{2v_n}{n} \|\partial^{[\alpha]} K\|_H \sum_{j=1}^{n} \frac{1}{h_j^{d+|\alpha|}} \left(\frac{\|x - x_i^{(n)}\|}{h_j} \right)^{\beta} \\
&\leq v_n \left| \Psi_n^{[\alpha]}(x_i^{(n)}) \right| + \frac{2v_n}{n} \|\partial^{[\alpha]} K\|_H \sum_{j=1}^{n} h_j^{-(d+\beta+|\alpha|)} R_n^{\beta} \\
&\leq v_n \left| \Psi_n^{[\alpha]}(x_i^{(n)}) \right| + \rho.
\end{aligned}$$

Hence, we deduce that

$$\begin{aligned}
\mathbb{P} \left[\sup_{x \in U, \|x\| \leq w_n} v_n \left| \Psi_n^{[\alpha]}(x) \right| \geq \delta \right] &\leq N(n) \max_{1 \leq i \leq N(n)} \mathbb{P} \left[v_n \left| \Psi_n^{[\alpha]}(x_i^{(n)}) \right| \geq \delta - \rho \right] \\
&\leq N(n) \sup_{x \in U} \mathbb{P} \left[v_n \left| \Psi_n^{[\alpha]}(x) \right| \geq \delta - \rho \right].
\end{aligned}$$

Let us at first assume that

$$\limsup_{n \to \infty} \frac{v_n^2}{n h_n^{d+2|\alpha|}} \log N(n) \leq db. \tag{2.17}$$

The application of Lemma 4 then yields

$$\begin{aligned}
&\limsup_{n \to \infty} \frac{v_n^2}{n h_n^{d+2|\alpha|}} \log \mathbb{P} \left[\sup_{x \in U, \|x\| \leq w_n} v_n \left| \Psi_n^{[\alpha]}(x) \right| \geq \delta \right] \\
&\leq \limsup_{n \to \infty} \frac{v_n^2}{n h_n^{d+2|\alpha|}} \log N(n) - \tilde{g}_U(\delta - \rho) \\
&\leq db - \tilde{g}_U(\delta - \rho).
\end{aligned}$$

Since this inequality holds for any $\rho \in]0, \delta[$, Part 2 of Lemma 5 thus follows from the continuity of $\tilde{g}_U$.

Let us now establish Relation (2.17). By definition of $N(n)$ and w_n, we have $\log N(n) \leq \log N'(n) \leq dbnh_n^{d+2|\alpha|}/v_n^2 + (d+1)\log 2 - d \log R_n$, with

$$\begin{aligned}
&\frac{v_n^2}{n h_n^{d+2|\alpha|}} \log R_n \\
&= \frac{v_n^2}{\beta n h_n^{d+2|\alpha|}} \left[\log \rho + \log n - \log \left(2 \|\partial^{[\alpha]} K\|_H \right) - \log v_n - \log \left(\sum_{j=1}^{n} h_j^{-(d+\beta+|\alpha|)} \right) \right],
\end{aligned}$$

which goes to zero in view of (H10) and (2.5). Thus, (2.17) is proved, and the proof of part 2 of Lemma 5 is completed.

Let us now consider part 1 of Lemma 5. This part is proved by following the same steps as for part 2, except that the number $N(n)$ of balls covering U is at most the integer part of $(\Delta/R_n)^d$, where Δ denotes the diameter of $\overline{U}$. Relation (2.17) then becomes

$$\limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log R_n \leq 0$$

and Lemma 5 is proved. ■

Lemma 6 *Let Assumptions (H1), (H3), (H4)i) and (H11)ii) hold. Assume that either* $(v_n) \equiv 1$ *or (H6), (H10) and (H11)i) hold. Moreover assume that* $\partial^{[\alpha]} f$ *is continuous. For any* $b > 0$ *if we set* $w_n = \exp\left(bnh_n^{d+2|\alpha|}/v_n^2\right)$ *then, for any* $\rho > 0$*, we have, for* n *large enough,*

$$\sup_{x\in U, \|x\|\geq w_n} \frac{v_n}{n} \sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \left| \mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x-X_i}{h_i}\right)\right]\right| \leq \rho.$$

Proof of Lemma 6

We have

$$\frac{v_n}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x-X_i}{h_i}\right)\right] = \frac{v_n}{n}\sum_{i=1}^{n} \int_{\mathbb{R}^d} K(z)\partial^{[\alpha]} f(x-h_i z)dz. \tag{2.18}$$

Set $\rho > 0$. In the case $(v_n) \equiv 1$, set M such that $\|\partial^{[\alpha]} f\|_\infty \int_{\|z\|>M} |K(z)|\, dz \leq \rho/2$; we have

$$\begin{aligned} &\frac{v_n}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \left|\mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x-X_i}{h_i}\right)\right]\right| \\ &\leq \frac{\rho}{2} + \partial^{[\alpha]} f(x) \int_{\|z\|\leq M} |K(z)|\, dz + \frac{1}{n}\sum_{i=1}^{n} \int_{\|z\|\leq M} |K(z)| \left|\partial^{[\alpha]} f(x-h_i z) - \partial^{[\alpha]} f(x)\right| dz. \end{aligned}$$

Lemma 6 then follows from the fact that $\partial^{[\alpha]} f$ fulfills (H11)ii). As a matter of fact, this condition implies that $\lim_{\|x\|\to\infty, x\in\overline{U}} \partial^{[\alpha]} f(x) = 0$ and that the third term in the right-hand-side of the previous inequality goes to 0 as $n \to \infty$ (by the dominated convergence).
Let us now assume that $\lim_{n\to\infty} v_n = \infty$; relation (2.18) can be rewritten as

$$\begin{aligned} \frac{v_n}{n}\sum_{i=1}^{n} \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x-X_i}{h_i}\right)\right] &= \frac{v_n}{n}\sum_{i=1}^{n} \int_{\|z\|\leq w_n/2} K(z)\partial^{[\alpha]} f(x-h_i z)dz \\ &+ \frac{v_n}{n}\sum_{i=1}^{n} \int_{\|z\|> w_n/2} K(z)\partial^{[\alpha]} f(x-h_i z)dz. \end{aligned}$$

Set $\rho > 0$; on the one hand, we have

$$\begin{aligned} \|x\| \geq w_n \quad \text{and} \quad \|z\| \leq w_n/2 \quad &\Rightarrow \quad \|x-h_i z\| \geq w_n (1-h_i/2) \\ &\Rightarrow \quad \|x-h_i z\| \geq w_n/2 \quad \text{for } n \text{ large enough.} \end{aligned}$$

Set $M_f = \sup_{x\in\mathbb{R}^d} \|x\|^\eta \partial^{[\alpha]} f(x)$. Assumption (H11)ii) implies that, for n sufficiently large,

$$\begin{aligned}
&\sup_{\|x\|\geq w_n} \frac{v_n}{n}\sum_{i=1}^{n}\int_{\|z\|\leq w_n/2}\left|K(z)\partial^{[\alpha]}f(x-h_iz)\right|dz \\
&\leq \sup_{\|x\|\geq w_n} \frac{v_n}{n}\sum_{i=1}^{n}\int_{\|z\|\leq w_n/2}|K(z)|\,M_f\|x-h_iz\|^{-\eta}dz \\
&\leq 2^\eta\frac{v_n}{w_n^\eta}M_f\int_{\mathbb{R}^d}|K(z)|\,dz \\
&\leq \frac{\rho}{2}.
\end{aligned}$$

On the other hand, we note that, in view of assumptions (H10) and (H11)i),

$$\begin{aligned}
\sup_{\|x\|\geq w_n} \frac{v_n}{n}\sum_{i=1}^{n}\int_{\|z\|>w_n/2}\left|K(z)\partial^{[\alpha]}f(x-h_iz)\right|dz &\leq 2^\zeta\frac{v_n}{w_n^\zeta}\|\partial^{[\alpha]}f\|_\infty\int_{\|z\|>w_n/2}\|z\|^\zeta\,|K(z)|\,dz \\
&\leq \frac{\rho}{2}
\end{aligned}$$

(for n large enough). As a matter of fact, we have by assumptions (H6) and (H10), $\forall\xi>0$

$$\frac{v_n}{w_n^\xi} = \exp\left\{-\xi\log w_n\left(1-\frac{\log v_n}{\xi\log w_n}\right)\right\}\stackrel{n\to\infty}{\longrightarrow}0.$$

This concludes the proof of Lemma 6. ■

Since $\partial^{[\alpha]}K$ is a bounded function that vanishes at infinity, we have $\lim_{\|x\|\to\infty}|\Psi_n^{[\alpha]}(x)|=0$ for every given $n\geq 1$. Moreover, since $\partial^{[\alpha]}K$ is assumed to be continuous, $\Psi_n^{[\alpha]}$ is continuous, and this ensures the existence of a random variable s_n such that

$$\left|\Psi_n^{[\alpha]}(s_n)\right| = \sup_{x\in U}\left|\Psi_n^{[\alpha]}(x)\right|.$$

Lemma 7 *Let Assumptions (H1), (H3), (H4)i), (H8)i), (H9)ii) and (H10) hold. Suppose either* $(v_n)\equiv 1$ *or (H6) and (H11) hold. For any* $b>0$*, set* $w_n=\exp\left(bnh_n^{d+2|\alpha|}/v_n^2\right)$ *; then, for any* $\delta>0$*, we have*

$$\limsup_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\|s_n\|\geq w_n \quad \textit{and} \quad v_n\left|\Psi_n^{[\alpha]}(s_n)\right|\geq\delta\right] \leq -b\xi.$$

Proof of Lemma 7

We first note that $s_n \in \overline{U}$ and therefore

$$\begin{aligned}
& \|s_n\| \geq w_n \text{ and } v_n \left|\Psi_n^{[\alpha]}(s_n)\right| \geq \delta \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \frac{v_n}{n}\left|\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \partial^{[\alpha]} K\left(\frac{s_n - X_i}{h_i}\right)\right| \\
& \quad + \frac{v_n}{n}\mathbb{E}\left|\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \partial^{[\alpha]} K\left(\frac{s_n - X_i}{h_i}\right)\right| \geq \delta \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \frac{v_n}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \left|\partial^{[\alpha]} K\left(\frac{s_n - X_i}{h_i}\right)\right| \geq \delta \\
& \quad - \sup_{\|x\|\geq w_n, x\in\overline{U}} \frac{v_n}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}\left|\partial^{[\alpha]} K\left(\frac{x - X_i}{h_i}\right)\right|
\end{aligned}$$

Set $\rho \in]0, \delta[$; the application of Lemma 6 ensures that, for n large enough,

$$\begin{aligned}
& \|s_n\| \geq w_n \text{ and } v_n \left|\Psi_n^{[\alpha]}(s_n)\right| \geq \delta \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \frac{v_n}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \left|\partial^{[\alpha]} K\left(\frac{s_n - X_i}{h_i}\right)\right| \geq \delta - \rho.
\end{aligned}$$

Set $\kappa = \sup_{x\in\mathbb{R}^d} \|x\|^\gamma |\partial^{[\alpha]}K(x)|$ (see Assumption (H9)ii)). We obtain, for n sufficiently large,

$$\begin{aligned}
& \|s_n\| \geq w_n \text{ and } v_n \left|\Psi_n^{[\alpha]}(s_n)\right| \geq \delta \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \exists i \in \{1, \ldots, n\} \text{ such that } \frac{v_n}{h_i^{d+|\alpha|}} \left|\partial^{[\alpha]} K\left(\frac{s_n - X_i}{h_i}\right)\right| \geq \delta - \rho \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \exists i \in \{1, \ldots, n\} \text{ such that } \kappa h_i^\gamma \geq \frac{h_i^{d+|\alpha|}}{v_n}\|s_n - X_i\|^\gamma(\delta - \rho) \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \exists i \in \{1, \ldots, n\} \text{ such that } \left|\|s_n\| - \|X_i\|\right| \leq \left[\frac{\kappa v_n h_i^{\gamma-d-|\alpha|}}{\delta - \rho}\right]^{\frac{1}{\gamma}} \\
& \Rightarrow \quad \|s_n\| \geq w_n \text{ and } \exists i \in \{1, \ldots, n\} \text{ such that } \|X_i\| \geq \|s_n\| - \left[\frac{\kappa v_n h_i^{\gamma-d-|\alpha|}}{\delta - \rho}\right]^{\frac{1}{\gamma}} \\
& \Rightarrow \quad \exists i \in \{1, \ldots, n\} \text{ such that } \|X_i\| \geq w_n(1 - u_{n,i}) \text{ with}
\end{aligned}$$

$$u_{n,i} = w_n^{-1} v_n^{\frac{1}{\gamma}} h_i^{\frac{\gamma-d-|\alpha|}{\gamma}} \left(\frac{\kappa}{\delta - \rho}\right)^{\frac{1}{\gamma}}.$$

Assume for the moment that

$$\lim_{n\to\infty} u_{n,i} = 0. \tag{2.19}$$

It then follows that $1-u_{n,i}>0$ for n sufficiently large ; therefore we can deduce that (see Assumption (H8)i)) :

$$\begin{aligned}
\mathbb{P}\left[\|s_n\|\geq w_n \quad\text{and}\quad v_n\left|\Psi_n^{[\alpha]}(s_n)\right|\geq\delta\right] &\leq \sum_{i=1}^{n}\mathbb{P}\left[\|X_i\|^{\xi}\geq w_n^{\xi}\left(1-u_{n,i}\right)^{\xi}\right]\\
&\leq \sum_{i=1}^{n}\mathbb{E}\left(\|X_i\|^{\xi}\right)w_n^{-\xi}\left(1-u_{n,i}\right)^{-\xi}\\
&\leq n\mathbb{E}\left(\|X_1\|^{\xi}\right)w_n^{-\xi}\max_{1\leq i\leq n}\left(1-u_{n,i}\right)^{-\xi}.
\end{aligned}$$

Consequently,

$$\begin{aligned}
&\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\|s_n\|\geq w_n \quad\text{and}\quad v_n\left|\Psi_n^{[\alpha]}(s_n)\right|\geq\delta\right]\\
&\leq \frac{v_n^2}{nh_n^{d+2|\alpha|}}\left[\log n+\log\mathbb{E}\left(\|X_1\|^{\xi}\right)-b\xi\frac{nh_n^{d+2|\alpha|}}{v_n^2}-\xi\log\max_{1\leq i\leq n}\left(1-u_{n,i}\right)\right],
\end{aligned}$$

and, thanks to assumption (H10), it follows that

$$\limsup_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\|s_n\|\geq w_n \quad\text{and}\quad v_n\left|\Psi_n^{[\alpha]}(s_n)\right|\geq\delta\right] \leq -b\xi.$$

Let us now prove relation (2.19). We expand

$$u_{n,i} = \exp\left(-b\frac{nh_n^{d+2|\alpha|}}{v_n^2}\left[1-\frac{1}{b\gamma}\frac{v_n^2\log v_n}{nh_n^{d+2|\alpha|}}-\frac{\gamma-d-|\alpha|}{b\gamma}\frac{v_n^2\log\left(h_i\right)}{nh_n^{d+2|\alpha|}}\right]\right)\left(\frac{\kappa}{\delta-\rho}\right)^{\frac{1}{\gamma}}$$

and assumptions (H6) and (H10) ensure that $\lim_{n\to\infty}u_{n,i}=0$ and thus Lemma 7 is proved. ■

Proof of Proposition 3

Let us at first note that the lower bound

$$\liminf_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\sup_{x\in U}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right]\geq-\tilde{g}_U(\delta) \tag{2.20}$$

follows from the application of Proposition 2 at a point $x_0\in\overline{U}$ such that $f(x_0)=\|f\|_{U,\infty}$.
In the case U is bounded, Proposition 3 is thus a straightforward consequence of (2.20) and of the first part of Lemma 5. Let us now consider the case U is unbounded.
Set $\delta>0$ and, for any $b>0$ set $w_n=\exp\left(bnh_n^{d+2|\alpha|}/v_n^2\right)$. Since, by definition of s_n,

$$\begin{aligned}
&\mathbb{P}\left[\sup_{x\in U}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right]\\
&\leq \mathbb{P}\left[\sup_{x\in U,\|x\|\leq w_n}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right]+\mathbb{P}\left[\|s_n\|\geq w_n \quad\text{and}\quad v_n\left|\Psi_n^{[\alpha]}(s_n)\right|\geq\delta\right]
\end{aligned}$$

it follows from Lemmas 5 and 7 that

$$\limsup_{n\to\infty}\frac{v_n^2}{nh_n^{d+2|\alpha|}}\log\mathbb{P}\left[\sup_{x\in U}v_n\left|\Psi_n^{[\alpha]}(x)\right|\geq\delta\right] \leq \max\{-b\xi;db-\tilde{g}_U(\delta)\}$$

and consequently

$$\limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\Psi_n^{[\alpha]}(x)\right| \geq \delta\right] \quad \leq \quad \inf_{b>0} \max\{-b\xi; db - \tilde{g}_U(\delta)\}.$$

Since the infimum in the right-hand-side of the previous bound is achieved for $b = \tilde{g}_U(\delta)/(\xi + d)$ and equals $-\xi\tilde{g}_U(\delta)/(\xi + d)$, we obtain the upper bound

$$\limsup_{n\to\infty} \frac{v_n^2}{nh_n^{d+2|\alpha|}} \log \mathbb{P}\left[\sup_{x\in U} v_n \left|\Psi_n^{[\alpha]}(x)\right| \geq \delta\right] \quad \leq \quad -\frac{\xi}{\xi + d}\tilde{g}_U(\delta)$$

which concludes the proof of Proposition 3. ■

2.3.4 Proof of Proposition 4

Let us set $g = \partial^{[\alpha]} f$, $D^j g$ $\big(j \in \{1, \ldots, q\}\big)$ the j-th differential of g, $y = (y_1, \ldots, y_d) \in \mathbb{R}^d$ and $y^{(j)} = (y, \ldots, y) \in (\mathbb{R}^d)^j$. With these notations,

$$D^j g(x)(y^{(j)}) = \sum_{\alpha_1+\cdots+\alpha_d=j} \frac{\partial^j g}{\partial y_1^{\alpha_1} \ldots \partial y_d^{\alpha_d}}(x) y_1^{\alpha_1} \ldots y_d^{\alpha_d}.$$

By successive integrations by parts (and using the fact that the partial derivatives of K vanish at infinity, see Assumption (H4)i)), we have

$$\begin{aligned}
\mathbb{E}\Big[\partial^{[\alpha]} f_n(x)\Big] &= \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \mathbb{E}\left[\partial^{[\alpha]} K\left(\frac{x - X_i}{h_i}\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{d+|\alpha|}} \int_{\mathbb{R}^d} \partial^{[\alpha]} K\left(\frac{x - y}{h_i}\right) f(y) dy \\
&= \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{d}} \int_{\mathbb{R}^d} K\left(\frac{x - y}{h_i}\right) g(y) dy \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} K(y) g(x - h_i y) dy.
\end{aligned}$$

Hence, using assumption (H7)i) and the fact that $\partial^{[\alpha]} f$ is q-times differentiable, it comes

$$\begin{aligned}
&\mathbb{E}\Big[\partial^{[\alpha]} f_n(x)\Big] - \partial^{[\alpha]} f(x) \\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} K(y)\big[g(x - h_i y) - g(x)\big] dy \\
&= \frac{1}{n}\sum_{i=1}^n h_i^q \int_{\mathbb{R}^d} K(y) \left[\frac{g(x - h_i y) - g(x) - \sum_{j=1}^{q-1} \frac{(-1)^j}{j!} h_i^j D^j g(x)(y^{(j)})}{h_i^q}\right] dy.
\end{aligned}$$

Let us set

$$\begin{aligned}
U_i(x) &= \int_{\mathbb{R}^d} K(y) \left[\frac{g(x - h_i y) - g(x) - \sum_{j=1}^{q-1} \frac{(-1)^j}{j!} h_i^j D^j g(x)(y^{(j)})}{h_i^q}\right] dy \quad \text{and} \\
U_\infty(x) &= \frac{(-1)^q}{q!} \int_{\mathbb{R}^d} D^q g(x)(y^{(q)}) K(y) dy.
\end{aligned}$$

We clearly have

$$\lim_{i\to\infty} U_i(x) \quad = \quad U_\infty(x) \tag{2.21}$$

and therefore, $\forall\varepsilon > 0$, $\exists i_0 \in \mathbb{R}$ such that $\forall i \geq i_0$, $|U_i(x) - U_\infty(x)| \leq \varepsilon$.

- If $\sum_i h_i^q = \infty$, then

$$\begin{aligned}
&\left|\frac{n}{\sum_{i=1}^n h_i^q}\left[\mathbb{E}\left(\partial^{[\alpha]} f_n(x)\right) - \partial^{[\alpha]} f(x)\right] - U_\infty(x)\right| \\
&= \left|\frac{\sum_{i=1}^n h_i^q U_i(x)}{\sum_{i=1}^n h_i^q} - U_\infty(x)\right| \\
&\leq \frac{\sum_{i=1}^{i_0-1} h_i^q\,|U_i(x) - U_\infty(x)| + \sum_{i=i_0}^n h_i^q\,|U_i(x) - U_\infty(x)|}{\sum_{i=1}^n h_i^q} \\
&\leq 2\varepsilon.
\end{aligned}$$

- If $\sum_i h_i^q < \infty$, we can write

$$\frac{n}{\sum_{i=1}^n h_i^q}\left[\mathbb{E}\left(\partial^{[\alpha]} f_n(x)\right) - \partial^{[\alpha]} f(x)\right] \quad = \quad \frac{\sum_{i=1}^n h_i^q U_i(x)}{\sum_{i=1}^n h_i^q}.$$

In view of (2.21), for x fixed and for all $i \in \mathbb{N}$, the sequence $(U_i(x))_i$ is bounded and thus Part 1 of Proposition 4 is completed. Let us now prove Part 2.
Since $(U_i(x))_i$ is bounded by $\sup_{x\in\mathbb{R}^d} \|D^q g(x)\| = M_q$ (see Assumption (H7)iii)), Part 2 follows. ■

2.3.5 Proof of Proposition 1

- Since $\left|e^t - 1\right| \leq |t|\, e^{|t|}$ $\forall t \in \mathbb{R}$, and thanks to the boundedness and integrability of K, we have

$$\int_{[0,1]\times\mathbb{R}^d} s^{-ad}\left|e^{s^{ad}uK(z)} - 1\right| dsdz \quad \leq \quad |u|e^{|u|\|K\|_\infty}\int_{[0,1]\times\mathbb{R}^d} |K(z)|\, dsdz < \infty$$

which ensures the existence of ψ_a. It is straightforward to check that ψ_a is twice differentiable, with

$$\begin{aligned}
\psi_a'(u) \quad &= \quad \int_{[0,1]\times\mathbb{R}^d} K(z) e^{s^{ad}uK(z)} dsdz, \\
\psi_a''(u) \quad &= \quad \int_{[0,1]\times\mathbb{R}^d} s^{ad}\left(K(z)\right)^2 e^{s^{ad}uK(z)} dsdz.
\end{aligned}$$

Since $\psi_a''(u) > 0$ $\forall u \in \mathbb{R}$, ψ_a' is increasing on $\mathbb{R}$, and ψ_a is strictly convex on $\mathbb{R}$. It follows that its Cramer transform I_a is a good rate function on $\mathbb{R}$ (see Dembo and Zeitouni [4]) and (i) of Proposition 1 is proved.

- Let us now assume that $\lambda(S_-) = 0$. We then have

$$\lim_{u\to-\infty} \psi_a'(u) = 0 \ \text{ and } \ \lim_{u\to+\infty} \psi_a'(u) = +\infty,$$

so that the range of ψ_a' is $]0, +\infty[$. Moreover $\lim_{u\to-\infty} \psi_a(u) = -\lambda(S_+)/(1 - ad)$ (which can be $-\infty$). This implies in particular that $I_a(0) = \lambda(S_+)/(1 - ad)$. Now, when $t < 0$, $\lim_{u\to-\infty}(ut - \psi_a(u)) = +\infty$, and $I_a(t) = +\infty$. Since ψ_a' is increasing with range $]0, +\infty[$, when $t > 0$, $\sup_u(ut - \psi_a(u))$ is reached for $u_0(t)$ such that $\psi_a'(u_0(t)) = t$, i.e. for $u_0(t) =$

$(\psi_a')^{-1}(t)$; this prove (2.3). (Note that, since $\psi_a''(t) > 0$, the function $t \mapsto u_0(t)$ is differentiable on $]0, +\infty[$). Now, differentiating (2.3), we have

$$\begin{aligned} I_a'(t) &= u_0(t) + tu_0'(t) - \psi_a'(u_0(t))u_0'(t) \\ &= (\psi_a')^{-1}(t) + tu_0'(t) - tu_0'(t) \\ &= (\psi_a')^{-1}(t). \end{aligned}$$

Since $(\psi_a')^{-1}$ is an increasing function on $]0, +\infty[$, it follows that I_a is strictly convex on $]0, +\infty[$ (and differentiable). Thus (ii) is proved.
Now, since $\lambda(S_-) = 0$, $\psi_a'(0) = \int_{[0,1]\times\mathbb{R}^d} K(z)dsdz = 1$; we have

$$I_a'(t) = 0 \quad \Leftrightarrow \quad (\psi_a')^{-1}(t) = 0 \quad \Leftrightarrow \quad \psi_a'(0) = t \quad \Leftrightarrow \quad t = 1.$$

Then $I_a'(1) = 0$, and $I_a(1) = 0$ is the unique global minimum of I_a on $]0, +\infty[$. This proves (iv) when $\lambda(S_-) = 0$.

- Assume that $\lambda(S_-) > 0$. In this case, ψ_a' can be rewritten as

$$\begin{aligned} \psi_a'(u) &= \int_{[0,1]\times(\mathbb{R}^d\cap S_+)} K(z)e^{s^{ad}uK(z)}dsdz \\ &\quad + \int_{[0,1]\times(\mathbb{R}^d\cap S_-)} K(z)e^{s^{ad}uK(z)}dsdz \end{aligned}$$

and we have

$$\lim_{u\to-\infty} \psi_a'(u) = -\infty \quad \text{and} \quad \lim_{u\to+\infty} \psi_a'(u) = +\infty$$

so that the range of ψ_a' is $\mathbb{R}$ in this case. The proof of (iii) and the case $\lambda(S_-) > 0$ of (iv) follows the same lines as previously, except that, in the present case, $(\psi_a')^{-1}$ is defined on $\mathbb{R}$, and not only on $]0, +\infty[$. ■

Bibliographie

[1] Ahmad, I.A. and Lin, P. (1976), Nonparametric sequential estimation of a multiple regression function, *Bull. Math. Statist.* **17**, pp. 63-75.

[2] Carroll, R.J. (1976), On sequential density estimation, *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **36**, pp. 137-151.

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

[4] Dembo, A. and Zeitouni, O. (1998), Large deviations techniques and applications, Springer, Applications of mathematics, New-York.

[5] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities, *Utilitas Math.* **15**, pp. 113-128.

[6] Feller, W. (1970), An introduction to probability theory and its applications, Second edition Volume II, Wiley.

[7] Gao, F. (2003), Moderate deviations and large deviations for kernel density estimators, *Journal of Theoritical Probab.* **16**, pp. 401-418.

[8] Lei, L. and Wu, L. (2004), The exponential convergence of kernel density estimator in L_1 for phi-mixing processes, *Ann. I.S.U.P.*, **48**, pp. 59-68.

[9] Lei, L. and Wu, L. (2005), Large deviations of kernel density estimator in $L_1(\mathbb{R}^d)$ for uniformly ergodic Markov processes, *Stoch. Process. Appl.* **115**, pp. 275-298.

[10] Lei, L., Wu, L. and Xie, B. (2003), Large deviations and deviation inequality for kernel density estimator in $L_1(\mathbb{R}^d)$-distance. *Ser. Biostat.*, **1**, pp. 89-97.

[11] Lei, L. (2006), Large deviations of the kernel density estimator in $L_1(\mathbb{R}^d)$ for reversible Markov processes, *Bernouilli*, **12**, pp. 65-83

[12] Louani, D. (1998), Large deviations limit theorems for the kernel density estimator, *Scand. J. Statist.* **25**, pp. 243-253.

[13] Louani, D. (2000), Large deviations for the L_1-distance in kernel density estimation, *J. Statist. Plann. Inference*, **90**, pp. 177-182.

[14] Menon, V.V., Prasad, B. and Singh R.S. (1984), Non-parametric recursive estimates of a probability density function and its derivatives, *Journal of Statistical Planning and inference* **9**, pp. 73-82.

[15] Mokkadem, A., Pelletier, M. and Worms, J. (2005), Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives, *Austral. J. Statist.*, **4**, pp. 489-502.

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

[17] Puhalskii, A.A. (1994), The method of stochastic exponentials for large deviations, *Stoch. Proc. and their Appl.* **54**, pp. 45-70.

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

[19] Roussas, G.G. (1992), Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimation, *J. Nonparametric Statist.* **1**, pp. 171-195.

[20] Wegman, E. J. (1972), Nonparametric probability density estimation. A comparison of density estimation methods, *J. Statist. Comput. Simulation* **1**, pp. 225-245.

[21] Wegman, E.J. and Davies, H. I. (1979), Remarks on some recursive estimators of a probability density, *Annals of Statistics* **7**, pp. 316-327.

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

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

[24] Worms, J. (2001), Moderate and large deviations of some dependent variables, Part II : some kernel estimators, *Math. Methods Statist.* **10**, pp. 161-193.

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

Chapitre 3

Large and moderate deviations principles for kernel estimators of the multivariate regression

3.1 Introduction

Let (X, Y), $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a sequence of independent and identically distributed $\mathbb{R}^d \times \mathbb{R}^q$-valued random variables with probability density $f(x, y)$ with $\mathbb{E}|Y| < \infty$. Moreover, let $g(x)$ be the marginal density of X and $r(x) = \mathbb{E}(Y|X = x) = m(x)/g(x)$ the regression of Y on X. The purpose of this paper is to establish large and moderate deviations principles for the Nadaraya-Watson estimator and for the semi-recursive kernel estimator of the regression.

Let us first recall the concept of large and moderate deviations. A speed is a sequence (ν_n) of positive numbers going to infinity. A good rate function on $\mathbb{R}^m$ is a lower semicontinuous function $I : \mathbb{R}^m \to [0, \infty]$ such that, for each $\alpha < \infty$, the level set $\{x \in \mathbb{R}^m, \quad I(x) \leq \alpha\}$ is a compact set. If the level sets of I are only closed, then I is said to be a rate function. A sequence $(Z_n)_{n\geq 1}$ of $\mathbb{R}^m$-valued random variables is said to satisfy a large deviations principle (LDP) with speed (ν_n) and rate function I if :

$$\begin{aligned}
\liminf_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in U\right] &\geq -\inf_{x\in U} I(x) \ \text{ for every open subset U of } \mathbb{R}^m, \\
\limsup_{n\to\infty} \nu_n^{-1} \log \mathbb{P}\left[Z_n \in V\right] &\leq -\inf_{x\in V} I(x) \ \text{ for every closed subset V of } \mathbb{R}^m.
\end{aligned}$$

Moreover, let (v_n) be a nonrandom sequence that goes to infinity; if $(v_n Z_n)$ satisfies a LDP, then (Z_n) is said to satisfy a moderate deviations principle (MDP).

The Nadaraya-Watson estimator ([15], [20]) of the regression function $r(x)$ is defined by

$$r_n(x) = \begin{cases} \dfrac{m_n(x)}{g_n(x)} & \text{if } g_n(x) \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.1}$$

with

$$m_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} Y_i K\left(\frac{x - X_i}{h_n}\right) \quad \text{and} \quad g_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h_n}\right),$$

where the bandwidth (h_n) is a positive sequence such that

$$\lim_{n\to\infty} h_n = 0 \quad \text{and} \quad \lim_{n\to\infty} nh_n^d = \infty, \tag{3.2}$$

and the kernel K a continuous function such that $\lim_{\|x\|\to\infty} K(x) = 0$ and $\int_{\mathbb{R}^d} K(x)dx = 1$. The weak and strong consistency of r_n has been widely discussed by many authors; let us cite, among many others, Collomb [4], Collomb and Härdle [5], Devroye [7], Mack and Silverman [12] and Senoussi [19]. For other works on the consistency of r_n, the reader is refered to the monographs of Bosq [3] and Prakasa Rao [16]. The large deviations behaviour of r_n has been studied at first by Louani [11], and then by Joutard [10] in the univariate framework. Moderate deviations principles have been obtained by Worms [21] in the particular case $Y = r(X) + \varepsilon$ with ε and X independent. The first aim of this paper is to generalize these large and moderate deviations results.

The approach used by Louani [11] and Joutard [10] to study the large deviations behaviour of r_n is to note that, if $d = q = 1$ and if the kernel is positive, then, for all $\delta > 0$,

$$\mathbb{P}\left[r_n(x) - r(x) \geq \delta\right] = \mathbb{P}\left[\frac{1}{nh_n} \sum_{j=1}^{n} \left[Y_j - r(x) - \delta\right] K\left(\frac{x - X_j}{h_n}\right) \geq 0\right].$$

Obviously, their approach can not be extended to the multivariate framework. Thus, to study the large deviations behaviour of r_n, our approach is totally different. We first establish a large deviations principle for the sequence $(m_n(x), g_n(x))$, and then show how the large deviations behaviour of r_n can be deduced. More precisely, for $x \in \mathbb{R}^d$, let Ψ_x be the function defined for any $(u, v) \in \mathbb{R}^q \times \mathbb{R}$ by

$$\Psi_x(u, v) \quad = \quad \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle u,y \rangle + v)K(z)} - 1 \right) f(x, y) dz dy,$$

(where $\langle u, y \rangle$ denotes the scalar product of u and y) and let I_x be the Fenchel-Legendre transform of Ψ_x. We give conditions ensuring that the sequence $(r_n(x))$ satisfies a LDP with speed (nh_n^d) and good rate function J defined, for any $s \in \mathbb{R}^q$, by

$$J(s) = \inf_{t \in \mathbb{R}} I_x(st, t).$$

Concerning the moderate deviations behaviour of the Nadaraya-Watson estimator, we prove that, for any positive sequence (v_n) such that

$$\lim_{n \to \infty} v_n = \infty, \quad \lim_{n \to \infty} \frac{v_n^2}{nh_n^d} = 0, \quad \text{and} \quad \lim_{n \to \infty} v_n h_n^p = 0, \tag{3.3}$$

(where p denotes the order of the kernel K) the sequence $(v_n [r_n(x) - r(x)])$ satisfies a LDP with speed $\left(nh_n^d / v_n^2\right)$ and good rate function G_x defined for all $v \in \mathbb{R}^q$ by

$$G_x(v) \quad = \quad \frac{g(x)}{2 \int_{\mathbb{R}^d} K^2(z) dz} v^T \Sigma_x^{-1} v, \tag{3.4}$$

where Σ_x denotes the $q \times q$ covariance matrix $V(Y|X = x)$. Let us note that, in the case the model $Y = r(X) + \varepsilon$ (with X and ε independent) is considered, the matrix Σ_x is the covariance matrix of ε and does depend on x ; we then find the MDP proved in Worms [21] again.

A semi-recursive version of the Nadaraya-Watson estimator (3.1) is defined as

$$\tilde{r}_n(x) = \begin{cases} \dfrac{\tilde{m}_n(x)}{\tilde{g}_n(x)} & \text{if } \tilde{g}_n(x) \neq 0 \\ 0 & \text{otherwise,} \end{cases} \tag{3.5}$$

where

$$\tilde{m}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{Y_i}{h_i^d} K\left(\frac{x - X_i}{h_i}\right) \quad \text{and} \quad \tilde{g}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right).$$

Weak conditions for various forms of consistency of $\tilde{r}_n$ have been obtained by Ahmad and Lin [1] and Devroye and Wagner [8]. Roussas [18] studied its almost sure convergence rate. The second aim of this paper is to establish the large and moderate deviations behaviour of $\tilde{r}_n$.

It turns out that the rate function that appears in the LDP is much more complex to explicit in the case the semi-recursive kernel regression estimator is considered than in the case the Nadaraya-Watson estimator is used. That is the reason why we only consider bandwidths defined as $(h_n) = (cn^{-a})$ with $c > 0$ and $a \in]0, 1/d[$ (instead of bandwidths satisfying (3.2)). For $x \in \mathbb{R}^d$, let $\tilde{\Psi}_{a,x}$ be the function defined for all $(u, v) \in \mathbb{R}^q \times \mathbb{R}$ by

$$\tilde{\Psi}_{a,x}(u, v) \quad = \quad \int_{[0,1] \times \mathbb{R}^d \times \mathbb{R}^q} s^{-ad} \left(e^{s^{ad}(\langle u,y \rangle + v)K(z)} - 1 \right) f(x, y) ds dz dy,$$

and let $\tilde{I}_{a,x}$ be the Fenchel-Legendre transform of $\tilde{\Psi}_{a,x}$. We give conditions ensuring that the sequence $(\tilde{r}_n(x))$ satisfies a LDP with speed (nh_n^d) and good rate function $\tilde{J}_a$ defined, for any $s \in \mathbb{R}^q$, by

$$\tilde{J}_a(s) = \inf_{t\in\mathbb{R}} \tilde{I}_{a,x}(st, t).$$

To establish the moderate deviations behaviour of $\tilde{r}_n$, we consider bandwidths (h_n) which vary regularly with exponent $(-a)$, $a \in]0, 1/d[$. We prove that, for any positive sequence (v_n) satisfying (3.3), the sequence $(v_n [\tilde{r}_n(x) - r(x)])$ satisfies a LDP with speed $\left(nh_n^d/v_n^2\right)$ and good rate function defined for all $v \in \mathbb{R}^q$ by

$$\tilde{G}_{a,x}(v) = \frac{(1+ad)g(x)}{2\int_{\mathbb{R}^d} K^2(z)dz} v^T \Sigma_x^{-1} v. \tag{3.6}$$

Let us underline that, because of the factor $(1+ad)$ which is present in (3.6) but not in (3.4), the rate function obtained in the MDP in the case the semi-recursive estimator is used is larger than the one which appears in the case the Nadaraya-Watson kernel estimator is considered ; this means that the semi-recursive estimator $\tilde{r}_n(x)$ is more concentrated around $r(x)$ than the Nadaraya-Watson estimator.

Our main results are stated in Section 3.2, whereas Section 3.3 is devoted to the proofs.

3.2 Assumptions and Main Results

We shall use the following notations.

- $\mathcal{D}(\mathcal{F}) = \{x,\ \mathcal{F}(x) < \infty\}$ denotes the domain of a function $\mathcal{F}$ and $\overset{\circ}{\mathcal{D}}(\mathcal{F})$ is the interior domain of $\mathcal{F}$.
- $\|x\|$ is the euclidean norm of x.
- λ is the Lebesgue measure.
- $a \wedge b = \min\{a, b\}$.
- $\vec{0} = (0, \ldots, 0) \in \mathbb{R}^q$.

The large and moderate deviations behaviours of the Nadaraya-Watson estimator r_n are given in Section 3.2.1, whereas the ones of the semi-recursive kernel estimator $\tilde{r}_n$ are stated in Section 3.2.2.

3.2.1 Large and moderate deviations principles for the Nadaraya-Watson estimator

The assumptions required for the LDP of the Nadaraya-Watson estimator are the following.

(A1) $K : \mathbb{R}^d \to \mathbb{R}$ is a bounded and integrable function, $\int_{\mathbb{R}^d} K(z)dz = 1$ and $\lim_{\|z\|\to\infty} K(z) = 0$.

(A2) For any $u \in \mathbb{R}^q$, $t \mapsto \displaystyle\int_{\mathbb{R}^q} e^{\langle u,y\rangle} f(t, y)dy$ is continuous at x and bounded.

Comments

- Notice that (A2) implies that the density g is continuous at x and bounded.

- In the model $Y = r(X) + \varepsilon$ with ε and X independent, let h be the probability density of ε. Then

$$\begin{aligned} f(t,y) &= g(t)h\left(y - r(t)\right) \\ \int_{\mathbb{R}^q} \|y\| f(t,y)dy &= g(t)\int_{\mathbb{R}^q} \|y + r(t)\| h(y)dy \\ \int_{\mathbb{R}^q} e^{\langle u,y\rangle} f(t,y)dy &= g(t)e^{\langle u,r(t)\rangle}\int_{\mathbb{R}^q} e^{\langle u,y\rangle} h(y)dy. \end{aligned}$$

Thus, (A2) can be translated as assumptions on g and r and on the moments of ε.
- As it can be seen from the proofs, the boundness assumption in (A2) is useless if K has a compact support.
- The boundness of the function $t \mapsto \int_{\mathbb{R}^q} e^{\langle u,y\rangle} f(t,y)dy$ for any $u \in \mathbb{R}^q$ implies that

$$\forall m \geq 0, \forall \rho \geq 0 \text{ the function } t \mapsto \int_{\mathbb{R}^q} \|y\|^m e^{\rho\|y\|} f(t,y)dy \quad \text{is bounded.} \tag{3.7}$$

Proof It suffices to prove that the function $t \mapsto \int_{\mathbb{R}^q} e^{\rho\|y\|} f(t,y)dy$ is bounded for any $\rho > 0$. Set $y = (y_1, \dots, y_q)$, we first note that

$$\begin{aligned} \int_{\mathbb{R}^q} e^{q\rho|y_j|} f(t,y)dy &\leq \int_{\{y_j \geq 0\}} e^{q\rho y_j} f(t,y)dy + \int_{\{y_j < 0\}} e^{-q\rho y_j} f(t,y)dy \\ &\leq \int_{\mathbb{R}^q} e^{q\rho y_j} f(t,y)dy + \int_{\mathbb{R}^q} e^{-q\rho y_j} f(t,y)dy. \end{aligned}$$

Now, we have

$$\begin{aligned} &\int_{\mathbb{R}^q} e^{\rho\|y\|} f(t,y)dy \\ &\leq \int_{\mathbb{R}^q} e^{\rho|y_1| + \cdots + \rho|y_q|} f(t,y)dy \\ &\leq \left(\int_{\mathbb{R}^q} e^{q\rho|y_1|} f(t,y)dy \dots \int_{\mathbb{R}^q} e^{q\rho|y_q|} f(t,y)dy\right)^{\frac{1}{q}} \quad \text{by the generalized Hölder inequality.} \\ &\leq \left(\left(\int_{\mathbb{R}^q} e^{q\rho y_1} f(t,y)dy + \int_{\mathbb{R}^q} e^{-q\rho y_1} f(t,y)dy\right) \dots \left(\int_{\mathbb{R}^q} e^{q\rho y_q} f(t,y)dy + \int_{\mathbb{R}^q} e^{-q\rho y_q} f(t,y)dy\right)\right)^{\frac{1}{q}} \end{aligned}$$

which is bounded.

Before stating our results, we need to introduce the rate function for the LDP of the Nadaraya-Watson estimator. Let $\Psi_x : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ and $I_x, \hat{I}_x : \mathbb{R}^q \times \mathbb{R} \to \mathbb{R}$ be the functions defined as follows :

$$\begin{aligned} \Psi_x(u,v) &= \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle u,y\rangle + v)K(z)} - 1\right) f(x,y)dzdy, &(3.8) \\ I_x(t_1,t_2) &= \sup_{(u,v)\in\mathbb{R}^q\times\mathbb{R}} \left\{\langle u,t_1\rangle + vt_2 - \Psi_x(u,v)\right\}, &(3.9) \\ \hat{I}_x(s,t) &= I_x(st,t). &(3.10) \end{aligned}$$

Moreover, for any $s \in \mathbb{R}^q$, set

$$\begin{aligned} J^*(s) &= \inf_{t \in \mathbb{R}^*} I_x(st, t) \\ &= \inf_{t \in \mathbb{R}^*} \hat{I}_x(s, t), \\ J(s) &= J^*(s) \wedge I_x(\vec{0}, 0) \\ &= \inf_{t \in \mathbb{R}} \hat{I}_x(s, t). \end{aligned}$$

To prove that J is a rate function, we need to assume that the following condition (C) is fulfilled.

(C) $\inf_{s \in \mathbb{R}^q} I_x(s, 0) = I_x(\vec{0}, 0)$.

Before stating the properties of the function J, let us give some cases when Condition (C) is satisfied (under Assumptions (A1) and (A2)).

Example 1 : Nonnegative kernel
Condition (C) is satisfied when K is nonnegative since, in this case, $I_x(s, 0) = +\infty$ for any $s \neq \vec{0}$, (this is stated in Proposition 7 of Section 3.3).

Example 2 : Model with symmetry
Condition (C) holds when f is symmetric in each coordinate of the second variable $y \in \mathbb{R}^q$. As a matter of fact, for a diagonal $q \times q$ matrix A such that $A_{ii} = \pm 1$, observe that

$$\begin{aligned} \Psi_x(Au, v) &= \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle Au, y \rangle + v) K(z)} - 1 \right) f(x, y) dz dy \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle u, Ay \rangle + v) K(z)} - 1 \right) f(x, y) dz dy \\ &= \int_{\mathbb{R}^d \times \mathbb{R}^q} \left(e^{(\langle u, y' \rangle + v) K(z)} - 1 \right) f(x, A^{-1} y') dz dy' \\ &= \Psi_x(u, v). \end{aligned}$$

For any given $s \in \mathbb{R}^q$, set

$$\mathcal{U}_s = \{ u \in \mathbb{R}^q, \ \langle u, s \rangle \geq 0 \}.$$

We have,

$$\sup_{u,v} \left(-\Psi_x(u, v) \right) = \sup_{u \in \mathcal{U}_s, \ v \in \mathbb{R}} \left(-\Psi_x(u, v) \right).$$

Now, for any $u \in \mathcal{U}_s$ and $v \in \mathbb{R}$,

$$\langle u, s \rangle - \Psi_x(u, v) \geq -\Psi_x(u, v),$$

so that

$$\sup_{u,v} \left\{ \langle u, s \rangle - \Psi_x(u, v) \right\} \geq \sup_{u \in \mathcal{U}_s, \ v \in \mathbb{R}} \left(-\Psi_x(u, v) \right),$$

and thus,

$$I_x(\vec{0}, 0) \leq I_x(s, 0) \quad \forall s \in \mathbb{R}^q,$$

so that Condition (C) follows.

Example 3 : A negative kernel without symmetry assumption on f, and for $d = q = 1$
If the kernel K can be written as $K = \mathbb{1}_D - \mathbb{1}_{D'}$ where D and D' are two subsets of $\mathbb{R}$ such that $D \cap D' = \emptyset$ and $\lambda(D) - \lambda(D') = 1$, then Condition (C) holds. As a matter of fact, we then have

$$\begin{aligned} \Psi_x(u,v) &= \int_{\mathbb{R}\times\mathbb{R}} \left(e^{(uy+v)K(z)} - 1\right) f(x,y)dzdy \\ &= \int_{D\times\mathbb{R}} \left(e^{uy+v} - 1\right) f(x,y)dzdy + \int_{D'\times\mathbb{R}} \left(e^{-uy-v} - 1\right) f(x,y)dzdy \\ &= e^v\lambda(D)\int_{\mathbb{R}} e^{uy} f(x,y)dy - \left[\lambda(D) + \lambda(D')\right] g(x) + e^{-v}\lambda(D')\int_{\mathbb{R}} e^{-uy} f(x,y)dy. \end{aligned}$$

Now, let M_x denote the Laplace transform of $f(x,\cdot)$, then

$$\Psi_x(u,v) = e^v\lambda(D)M_x(u) + e^{-v}\lambda(D')M_x(-u) - \left[\lambda(D) + \lambda(D')\right] g(x).$$

For any given u, it can easily be seen that the infimum of $\Psi_x(u,\cdot)$ is reached at

$$v_0 = \log\sqrt{\frac{\lambda(D')M_x(-u)}{\lambda(D)M_x(u)}},$$

and

$$\Psi_x(u,v_0) = 2\sqrt{\lambda(D)\lambda(D')}\sqrt{M_x(u)M_x(-u)} - \left[\lambda(D) + \lambda(D')\right] g(x).$$

Observe that

$$\Psi_x(u,v_0) = \Psi_x(-u,v_0),$$

and thus

$$\begin{aligned} \sup_u\left(-\Psi_x(u,v_0)\right) &= \sup_{u\geq 0}\left(-\Psi_x(u,v_0)\right) \\ &= \sup_{u\leq 0}\left(-\Psi_x(u,v_0)\right) \\ &= I_x(0,0). \end{aligned}$$

Now, if $s \geq 0$, we have for any $u \geq 0$

$$us - \Psi_x(u,v_0) \geq -\Psi_x(u,v_0),$$

and thus

$$I_x(s,0) \geq I_x(0,0)\ \forall s \geq 0.$$

Proceeding in the same way for $s < 0$, we obtain Condition (C).
Such an example of a four order kernel is $K = \mathbb{1}_{[-a,a]} - \mathbb{1}_{[-b,-a[\cup]a,b]}$, with

$$\begin{aligned} a &= \frac{1}{6}\sqrt[3]{2} + \frac{1}{12}\left(\sqrt[3]{2}\right)^2 + \frac{1}{3} \\ b &= \frac{1}{3}\sqrt[3]{2} + \frac{1}{6}\left(\sqrt[3]{2}\right)^2 + \frac{1}{6}. \end{aligned}$$

Let us now give the properties of the function J.

Proposition 5 *Assume that (A1), (A2) and (C) hold. Then,*

(i) J is a rate function on $\mathbb{R}^q$. More precisely, for $\alpha \in \mathbb{R}$,

- *if $\alpha < I_x(\vec{0}, 0)$, then $\{J(s) \leq \alpha\}$ is compact.*
- *if $\alpha \geq I_x(\vec{0}, 0)$, then $\{J(s) \leq \alpha\} = \mathbb{R}^q$.*

(ii) If $I_x(\vec{0}, 0) = \infty$, then J is a good rate function on $\mathbb{R}^q$ and $J = J^*$.

(iii) If $J^*(s) < \infty$, then $J(s) = J^*(s)$.

(iv) If $\alpha < I_x(\vec{0}, 0)$, then $\{J^*(s) \leq \alpha\} = \{J(s) \leq \alpha\}$.

Remark 6 *In view of the definition of J and J^*, and of Proposition 5 (iii), we have :*

$$J(s) \quad = \quad \begin{cases} J^*(s) & \text{if} \;\; J^*(s) < \infty \\ I_x(\vec{0}, 0) & \text{if} \;\; J^*(s) = \infty. \end{cases}$$

Let us now state the LDP for the Nadaraya-Watson estimator.

Theorem 6 ***(Pointwise LDP for the Nadaraya-Watson estimator)***
Assume that (A1), (A2) and (C) hold, and that (h_n) satisfies the conditions in (3.2). Then, for any open subset U of $\mathbb{R}^q$,

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in U\right] \geq - \inf_{s\in U} J^*(s),$$

and for any closed subset V of $\mathbb{R}^q$,

$$\limsup_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in V\right] \leq - \inf_{s\in V} J(s).$$

Comments.

1) Set $E = \{J^*(s) < \infty\}$. For any open subset U of $\mathbb{R}^q$ such that $U \cap E \neq \emptyset$, we have

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in U\right] \geq - \inf_{s\in U} J(s).$$

2) If I_x is finite in a neighbourhood of $(\vec{0}, 0)$, then J^* is finite everywhere and by Proposition 5 (iii), $J(s) = J^*(s) < \infty \;\; \forall s$; thus (r_n) satisfies a LDP with speed (nh_n^d) and rate function J. Of course, this does not hold for nonnegative kernel since in this case $I_x(s, 0) = +\infty$ for any s (see Proposition 7 in Section 3.3). However, it can hold for kernels which take negative values. For example, consider the previous Example 3, and assume $f(x, y)$ is symmetric in y ; in this case $M_x(u) = M_x(-u)$. The equation

$$\frac{\partial \Psi_x}{\partial v}(u, v) = \left[\lambda(D)e^v - \lambda(D')e^{-v}\right] M_x(u) = 0$$

has solution $v_0 = \log \sqrt{\dfrac{\lambda(D')}{\lambda(D)}}$ independent from u. Moreover, M' is continuous and has range $\mathbb{R}$, thus, there exists u_0 such that $M'(u_0) = 0$. This implies that the equation

$$\frac{\partial \Psi_x}{\partial u}(u, v) = \left[\lambda(D)e^v + \lambda(D')e^{-v}\right] M_x'(u) = 0$$

has a solution u_0 independent from v. Thus $(0, 0)$ is in the range of $\nabla \Psi_x$. It follows from Proposition 7 Section 3.3 that I_x is finite in a neighbourhood of $(0, 0)$.

3) When $I_x(\vec{0},0)=\infty$, it follows from Proposition 5 and Theorem 6 that (r_n) satisfies a LDP with speed (nh_n^d) and good rate function J.

In the case K is a nonnegative kernel whose support has an infinity measure, we will show in Proposition 7 that $I_x(\vec{0},0)=\infty$. We have thus the following corollary.

Corollary 2 *Let the assumptions of Theorem 6 hold. If K is a nonnegative kernel such that $\lambda\left(\left\{x\in\mathbb{R}^d, K(x)>0\right\}\right)=\infty$, then the sequence (r_n) satisfies a LDP with speed (nh_n^d) and good rate function J.*

This corollary is an extension of the results of Louani [11] and Joutard [10] to the multivariate framework (and to the case the kernel K may vanish). Moreover, it proves that the rate function that appears in their large deviations results is in fact a good rate function.

To establish pointwise MDP for the Nadaraya-Watson estimator, we need the following additionnal assumptions.

(A3) For any $u\in\mathbb{R}^q$, $t\mapsto\int_{\mathbb{R}^q}\langle u,y\rangle^2 f(t,y)dy$ and $t\mapsto\int_{\mathbb{R}^q}\langle u,y\rangle f(t,y)dy$ are continuous at x and $g(x)\neq 0$.

(A4) $\lim_{n\to\infty}v_n=\infty$ and $\lim_{n\to\infty}\frac{nh_n^d}{v_n^2}=\infty$.

(A5) i) There exists an integer $p\geq 2$ such that $\forall s\in\{1,\ldots,p-1\},\forall j\in\{1,\ldots,d\}$, $\int_{\mathbb{R}^d}y_j^sK(y)dy_j=0$, and $\int_{\mathbb{R}^d}\left|y_j^pK(y)\right|dy<\infty$.
ii) $\lim_{n\to\infty}v_nh_n^p=0$.
iii) m and g are p-times differentiable on $\mathbb{R}^d$, and their differentials of order p are bounded and continuous at x.

We can now state the MDP for the Nadaraya-Watson estimator.

Theorem 7 ***(Pointwise MDP for the Nadaraya-Watson kernel estimator of the regression)*** *Assume that (A1)-(A5) hold. Then, the sequence $(v_n\left(r_n(x)-r(x)\right))$ satisfies a LDP with speed $\left(\frac{nh_n^d}{v_n^2}\right)$ and good rate function G_x defined in* (3.4).

3.2.2 Large and moderate deviations principles for the semi-recursive estimator

For $a\in]0,1/d[$, let $\tilde{\Psi}_{a,x}:\mathbb{R}^q\times\mathbb{R}\to\mathbb{R}$ and $\tilde{I}_{a,x}:\mathbb{R}^q\times\mathbb{R}\to\mathbb{R}$ be the functions defined as follows :

$$\tilde{\Psi}_{a,x}(u,v) = \int_{[0,1]\times\mathbb{R}^d\times\mathbb{R}^q} s^{-ad}\left(e^{s^{ad}(\langle u,y\rangle+v)K(z)}-1\right)f(x,y)dsdzdy, \tag{3.11}$$

$$\tilde{I}_{a,x}(t_1,t_2) = \sup_{(u,v)\in\mathbb{R}^q\times\mathbb{R}}\left\{\langle u,t_1\rangle+vt_2-\tilde{\Psi}_x(u,v)\right\}. \tag{3.12}$$

Moreover, let $\tilde{J}_a$ and $\tilde{J}_a^*$ be defined as follows : for any $s\in\mathbb{R}^q$,

$$\tilde{J}_a^*(s) = \inf_{t\in\mathbb{R}^*}\tilde{I}_{a,x}(st,t) \tag{3.13}$$

$$\tilde{J}_a(s) = \tilde{J}_a^*(s)\wedge\tilde{I}_{a,x}(\vec{0},0). \tag{3.14}$$

Let us give the following additionnal hypotheses.

(A'1) For any $u\in\mathbb{R}^q$, $t\mapsto\int_{\mathbb{R}^q}e^{\alpha\langle u,y\rangle}f(t,y)dy$ is continuous at x uniformly with respect to $\alpha\in[0,1]$.

Condition (C) above is substituted by the following one,

(C') $\inf_{s\in\mathbb{R}^q} \tilde{I}_{a,x}(s,0) = \tilde{I}_{a,x}(\vec{0},0)$.

Examples for which Condition (C') holds are Examples 1 and 2 given for (C). The following proposition gives the properties of the function $\tilde{J}_a$.

Proposition 6 *Assume that (A1), (A2), (A'1) and (C') hold. Then,*

(i) $\tilde{J}_a$ is a rate function on $\mathbb{R}^q$. More precisely, for $\alpha \in \mathbb{R}$,

- *if $\alpha < \tilde{I}_{a,x}(\vec{0},0)$, then $\left\{\tilde{J}_a(s) \leq \alpha\right\}$ is compact.*
- *if $\alpha \geq \tilde{I}_{a,x}(\vec{0},0)$, then $\left\{\tilde{J}_a(s) \leq \alpha\right\} = \mathbb{R}^q$.*

(ii) If $\tilde{I}_{a,x}(\vec{0},0) = \infty$, then $\tilde{J}_a$ is a good rate function on $\mathbb{R}^q$ and $\tilde{J}_a = \tilde{J}_a^*$.

(iii) If $\tilde{J}_a^*(s) < \infty$, then $\tilde{J}_a(s) = \tilde{J}_a^*(s)$.

(iv) If $\alpha < \tilde{I}_{a,x}(\vec{0},0)$, then $\left\{\tilde{J}_a^*(s) \leq \alpha\right\} = \left\{\tilde{J}_a(s) \leq \alpha\right\}$.

Notice that, like for J and J^*, we have

$$\tilde{J}_a(s) = \begin{cases} \tilde{J}_a^*(s) & \text{if } \tilde{J}_a^*(s) < \infty \\ \tilde{I}_{a,x}(\vec{0},0) & \text{if } \tilde{J}_a^*(s) = \infty. \end{cases}$$

We can now state the LDP for the semi-recursive kernel estimator of the regression.

Theorem 8 ***(Pointwise LDP for the semi-recursive estimator of the regression)*** *Set $(h_n) = (cn^{-a})$ with $c > 0$ and $0 < a < 1/d$, and let (A1), (A2), (A'1) and (C') hold. Then, for any open subset U of $\mathbb{R}^q$,*

$$\liminf_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[\tilde{r}_n(x) \in U\right] \geq -\inf_{s\in U} \tilde{J}_a^*(s),$$

and for any closed subset V of $\mathbb{R}^q$,

$$\limsup_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[\tilde{r}_n(x) \in V\right] \leq -\inf_{s\in V} \tilde{J}_a(s).$$

The comments made for Theorem 6 are valid for Theorem 8. In particular, we have the following corollary.

Corollary 3 *Let the assumptions of Theorem 8 hold. If K is a nonnegative kernel such that $\lambda\left(\left\{x \in \mathbb{R}^d, K(x) > 0\right\}\right) = \infty$, then the sequence $(\tilde{r}_n)$ satisfies a LDP with speed (nh_n^d) and good rate function $\tilde{J}_a$.*

Before stating pointwise MDP for the semi-recursive estimator of the regression, let us recall that a sequence (u_n) is said to vary regularly with exponent α if there exists a function u which varies regularly with exponent α and such that $u_n = u(n)$ for all n (see, for example, Feller [9] page 275). We will use in the sequel the following property (see Bingham et al. [2] page 26). If (h_n) varies regularly with exponent $(-a)$ and if $\beta a < 1$, then

$$\lim_{n\to\infty} \frac{1}{nh_n^\beta} \sum_{i=1}^n h_i^\beta = \frac{1}{1-a\beta}. \tag{3.15}$$

We also consider the following condition.

$$\sup_{i\leq n} \frac{h_n}{h_i} < \infty. \tag{3.16}$$

(For example, this condition holds when h_n is nonincreasing).

Theorem 9 ***(Pointwise MDP for the semi-recursive kernel estimator of the regression)***
Assume that (h_n) *varies regularly with exponent* $(-a)$ *with* $a\in]0,1/d[$*, and satisfies* (3.16)*. Let (A1)-(A5) hold. Then, the sequence* $(v_n(\tilde{r}_n(x)-r(x)))$ *satisfies a LDP with speed* $\left(\frac{nh_n^d}{v_n^2}\right)$ *and good rate function* $\tilde{G}_{a,x}$ *defined in* (3.6)*.*

3.3 Proofs

The proofs of the results for the Nadaraya-Watson kernel estimator are in many cases similar to those of the semi-recursive kernel estimator of the regression, so we omit some details of the proofs for this last one.

First, let us state the following propositions which give the properties of the functions Ψ_x, $\tilde{\Psi}_{a,x}$, I_x and $\tilde{I}_{a,x}$. Set

$$S_+=\left\{x\in\mathbb{R}^d,\ K(x)>0\right\}\quad\text{and}\quad S_-=\left\{x\in\mathbb{R}^d,\ K(x)<0\right\}.$$

Proposition 7 ***(Properties of*** $\boldsymbol{\Psi_x}$ ***and*** $\boldsymbol{I_x}$***)***
Let Assumptions (A1) and (A2) hold. Then,

i) Ψ_x *is strictly convex, continuously differentiable on* $\mathbb{R}^q\times\mathbb{R}$*, and* I_x *is a good rate function on* $\mathbb{R}^q\times\mathbb{R}$*.*

ii) $\nabla\Psi_x$ *is an open map and the range of* Ψ_x *is* $\overset{\circ}{\mathcal{D}}(I_x)$*.* I_x *is strictly convex on* $\overset{\circ}{\mathcal{D}}(I_x)$ *and for any* $t\in\overset{\circ}{\mathcal{D}}(I_x)\subset\mathbb{R}^q\times\mathbb{R}$*,*

$$I_x(t)=\langle(\nabla\Psi_x)^{-1}(t),t\rangle-\Psi_x\left((\nabla\Psi_x)^{-1}(t)\right). \tag{3.17}$$

iii) If $\lambda(S_-)=0$*, then* $I_x(\vec{0},0)=g(x)\lambda(S_+)$*, and for any* $t_1\neq\vec{0}$*,* $I_x(t_1,0)=+\infty$*.*

Proposition 8 ***(Properties of*** $\boldsymbol{\tilde{\Psi}_{a,x}}$ ***and*** $\boldsymbol{\tilde{I}_{a,x}}$***)***
Let Assumptions (A1) and (A2) hold. Then,

i) $\tilde{\Psi}_{a,x}$ *is strictly convex, continuously differentiable on* $\mathbb{R}^q\times\mathbb{R}$*, and* $\tilde{I}_{a,x}$ *is a good rate function on* $\mathbb{R}^q\times\mathbb{R}$*.*

ii) $\nabla\tilde{\Psi}_{a,x}$ *is an open map and the range of* $\tilde{\Psi}_{a,x}$ *is* $\overset{\circ}{\mathcal{D}}(\tilde{I}_{a,x})$*.* $\tilde{I}_{a,x}$ *is strictly convex on* $\overset{\circ}{\mathcal{D}}(\tilde{I}_{a,x})$*, and for any* $t\in\overset{\circ}{\mathcal{D}}(\tilde{I}_{a,x})\subset\mathbb{R}^q\times\mathbb{R}$*,*

$$\tilde{I}_{a,x}(t)=\langle(\nabla\tilde{\Psi}_{a,x})^{-1}(t),t\rangle-\tilde{\Psi}_{a,x}\left((\nabla\tilde{\Psi}_{a,x})^{-1}(t)\right). \tag{3.18}$$

iii) If $\lambda(S_-)=0$*, then* $\tilde{I}_{a,x}(\vec{0},0)=g(x)\lambda(S_+)/(1-ad)$*, and for any* $t_1\neq\vec{0}$*,* $\tilde{I}_{a,x}(t_1,0)=+\infty$*.*

The two following lemmas are used for the proofs of Theorems 6 and 8.

Lemma 8 ***(Pointwise LDP for the sequence*** $\boldsymbol{(m_n(x),g_n(x))}$***)***
Let Assumptions (A1) and (A2) hold. Then, the sequence $(m_n(x),g_n(x))$ *satisfies a LDP with speed* $\left(nh_n^d\right)$ *and rate function* I_x *defined in* (3.9)*.*

Lemma 9 ***(Pointwise LDP for the sequence*** $\boldsymbol{(\tilde{m}_n(x),\tilde{g}_n(x))}$***)***
Set $h_n=cn^{-a}$ *with* $c>0$ *and* $a\in]0,1/d[$*, and let Assumptions (A1), (A2) and (A'1) hold. Then, the sequence* $(\tilde{m}_n(x),\tilde{g}_n(x))$ *satisfies a LDP with speed* $\left(nh_n^d\right)$ *and rate function* $\tilde{I}_{a,x}$ *defined in* (3.12)*.*

Our proofs are now organized as follows. Lemmas 8 and 9 are proved in Section 3.3.1, Theorems 6 and 8 in Section 3.3.2, Theorem 7 in Section 3.3.3, Theorem 9 is proved in Section 3.3.4. Section 3.3.5 is devoted to the proof of Propositions 7 and 8 on the rate functions I_x and $I_{a,x}$. Propositions 5 and 6 are proved in Section 3.3.6.

3.3.1 Proof of Lemmas 8 and 9

Proof of Lemma 8

For any $w = (u,v) \in \mathbb{R}^q \times \mathbb{R}$, set

$$\begin{aligned}\Psi_n(x) &= (m_n(x), g_n(x)),\\ \Lambda_{n,x}(w) &= \frac{1}{nh_n^d}\log\mathbb{E}\left[\exp\left(nh_n^d\langle w, \Psi_n(x)\rangle\right)\right].\end{aligned}$$

Let us at first assume that the following lemma holds.

Lemma 10 ***(Convergence of $\boldsymbol{\Lambda_{n,x}}$)***
Assume that (A1) and (A2) hold, then

$$\lim_{n\to\infty}\Lambda_{n,x}(u,v) = \Psi_x(u,v), \tag{3.19}$$

where Ψ_x is defined in (3.8).

To prove Lemma 8, we apply Proposition 7, Lemma 10 and the Gärtner-Ellis Theorem (see Dembo and Zeitouni [6]). Proposition 7 ensures that Ψ_x is essentially smooth, lower semicontinuous function so that Lemma 8 follows from the Gärtner-Ellis Theorem.
Let us now prove Lemma 10. Set

$$Z_i = [\langle u, Y_i\rangle + v]\, K\left(\frac{x - X_i}{h_n}\right).$$

For any $(u,v) \in \mathbb{R}^q \times \mathbb{R}$, we have

$$\Lambda_{n,x}(u,v) = \frac{1}{nh_n^d}\log\mathbb{E}\left[\exp\left(\sum_{i=1}^n Z_i\right)\right],$$

and, since the random vectors (X_i, Y_i), $i = 1,\dots,n$ are independent and identically distributed, we get

$$\Lambda_{n,x}(u,v) = \frac{1}{h_n^d}\log\mathbb{E}\left[e^{Z_n}\right].$$

A Taylor's expansion implies that there exists c_n between 1 and $\mathbb{E}\left[e^{Z_n}\right]$ such that

$$\begin{aligned}\Lambda_{n,x}(u,v) &= \frac{1}{h_n^d}\mathbb{E}\left[e^{Z_n}-1\right] - \frac{1}{2c_n^2h_n^d}\left(\mathbb{E}\left[e^{Z_n}-1\right]\right)^2\\ &= \frac{1}{h_n^d}\int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K\left(\frac{x-s}{h_n}\right)}-1\right]f(s,y)dsdy - R_{n,x}^{(1)}(u,v)\\ &= \Psi_x(u,v) - R_{n,x}^{(1)}(u,v) + R_{n,x}^{(2)}(u,v),\end{aligned}$$

with

$$\begin{aligned} R^{(1)}_{n,x}(u,v) &= \frac{1}{2c_n^2 h_n^d}\left(\mathbb{E}\left[e^{Z_n}-1\right]\right)^2, \\ R^{(2)}_{n,x}(u,v) &= \int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_nz,y)-f(x,y)\right]dzdy. \end{aligned}$$

Let us prove that

$$\lim_{n\to\infty} R^{(2)}_{n,x}(u,v) \;=\; 0. \tag{3.20}$$

Set $A>0$ and $\epsilon>0$; we then have

$$\begin{aligned} R^{(2)}_{n,x}(u,v) &= \int_{\{\|z\|\le A\}\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_nz,y)-f(x,y)\right]dzdy \\ &\quad+\int_{\{\|z\|> A\}\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_nz,y)-f(x,y)\right]dzdy. \end{aligned} \tag{3.21}$$

Next, since for any $t\in\mathbb{R}$, $\left|e^t-1\right|\le|t|e^{|t|}$, we have

$$\begin{aligned} &\int_{\{\|z\|> A\}\times\mathbb{R}^q}\left|e^{(\langle u,y\rangle+v)K(z)}-1\right|\left|f(x-h_nz,y)-f(x,y)\right|dzdy \\ &\le \int_{\{\|z\|> A\}\times\mathbb{R}^q}\left|\langle u,y\rangle+v\right|\left|K(z)\right|e^{|\langle u,y\rangle+v||K(z)|}\left|f(x-h_nz,y)-f(x,y)\right|dzdy \\ &\le \int_{\{\|z\|> A\}\times\mathbb{R}^q}\left|\langle u,y\rangle+v\right|\left|K(z)\right|e^{|\langle u,y\rangle+v||K(z)|}f(x-h_nz,y)dzdy \\ &\quad+\int_{\{\|z\|> A\}\times\mathbb{R}^q}\left|\langle u,y\rangle+v\right|\left|K(z)\right|e^{|\langle u,y\rangle+v||K(z)|}f(x,y)dzdy \\ &\le e^{|v|\|K\|_\infty}\|u\|\int_{\{\|z\|>A\}}\left|K(z)\right|\left[\int_{\mathbb{R}^q}\|y\|e^{\|K\|_\infty\|u\|\|y\|}f(x-h_nz,y)dy\right]dz \\ &\quad+e^{|v|\|K\|_\infty}|v|\int_{\{\|z\|>A\}}\left|K(z)\right|\left[\int_{\mathbb{R}^q}e^{\|K\|_\infty\|u\|\|y\|}f(x-h_nz,y)dy\right]dz \\ &\quad+e^{|v|\|K\|_\infty}\|u\|\int_{\{\|z\|>A\}}\left|K(z)\right|dz\int_{\mathbb{R}^q}\|y\|e^{\|K\|_\infty\|u\|\|y\|}f(x,y)dy \\ &\quad+e^{|v|\|K\|_\infty}|v|\int_{\{\|z\|>A\}}\left|K(z)\right|dz\int_{\mathbb{R}^q}e^{\|K\|_\infty\|u\|\|y\|}f(x,y)dy \\ &\le B\int_{\{\|z\|>A\}}\left|K(z)\right|dz, \end{aligned} \tag{3.22}$$

where B is a constant ; this last inequality follows from (3.7) and from the fact that K is bounded. Now, since K is integrable, we can choose A such that

$$\int_{\{\|z\|> A\}\times\mathbb{R}^q}\left|e^{(\langle u,y\rangle+v)K(z)}-1\right|\left|f(x-h_nz,y)-f(x,y)\right|dzdy\le\frac{\epsilon}{2}. \tag{3.23}$$

Now, observe that

$$\begin{aligned} &\int_{\{\|z\|\le A\}\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_nz,y)-f(x,y)\right]dzdy \\ &= \int_{\{\|z\|\le A\}}e^{vK(z)}\left[\int_{\mathbb{R}^q}e^{\langle u,y\rangle K(z)}\left(f(x-h_nz,y)-f(x,y)\right)dy\right]dz \qquad (3.24) \\ &\quad-\int_{\{\|z\|\le A\}}\left[\int_{\mathbb{R}^q}\left(f(x-h_nz,y)-f(x,y)\right)dy\right]dz. \qquad (3.25) \end{aligned}$$

Assumption (A2) together with (3.7), and the dominated convergence theorem ensure that both integrals in (3.24) and (3.25) converge to 0. We deduce that for n large enough,

$$\left|\int_{\{\|z\|\leq A\}\times\mathbb{R}^q}\left[e^{(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_nz,y)-f(x,y)\right]dzdy\right|\leq\frac{\epsilon}{2}, \tag{3.26}$$

so that (3.20) follows from (3.23) and (3.26).
Let us now consider $R_{n,x}^{(1)}$; since c_n is between 1 and $\mathbb{E}\left[e^{Z_n}\right]$, we get

$$\frac{1}{c_n}\quad\leq\quad\max\left\{1,\frac{1}{\mathbb{E}\left(e^{Z_n}\right)}\right\}.$$

By Jensen's inequality, we obtain

$$\frac{1}{\mathbb{E}\left[e^{Z_n}\right]}\leq\frac{1}{e^{\mathbb{E}[Z_n]}}.$$

Observe that

$$\begin{aligned}
|\mathbb{E}\left(Z_n\right)| &= \left|\mathbb{E}\left[\left(\langle u,Y_n\rangle+v\right)K\left(\frac{x-X_n}{h_n}\right)\right]\right|\\
&\leq \int_{\mathbb{R}^d\times\mathbb{R}^q}\|u\|\|y\|\left|K\left(\frac{x-s}{h_n}\right)\right|f(s,y)dsdy+|v|\int_{\mathbb{R}^d\times\mathbb{R}^q}\left|K\left(\frac{x-s}{h_n}\right)\right|f(s,y)dsdy\\
&\leq h_n^d\left(\|u\|\int_{\mathbb{R}^d}|K(z)|\left[\int_{\mathbb{R}^q}\|y\|f(x-h_nz,y)dy\right]dz+|v|\int_{\mathbb{R}^d}|K(z)|\left[\int_{\mathbb{R}^q}f(x-h_nz,y)dy\right]dz\right),
\end{aligned}$$

which goes to 0 in view of (3.7) and since $\lim_{n\to\infty}h_n=0$. We deduce that there exists $c\in\mathbb{R}_+^*$ such that

$$\frac{1}{c_n^2}\leq c.$$

Noting that by (3.7),

$$\begin{aligned}
\mathbb{E}\left|e^{Z_n}-1\right| &\leq h_n^d\int_{\mathbb{R}^d\times\mathbb{R}^q}\left|\langle u,y\rangle+v\right||K(z)|e^{|\langle u,y\rangle+v||K(z)|}f(x-h_nz,y)dydz\\
&\leq Bh_n^d,
\end{aligned}$$

where B is a constant. It follows that

$$\lim_{n\to\infty}R_{n,x}^{(1)}(u,v)=0,$$

which proves Lemma 10. ■

Proof of Lemma 9

Similarly as the proof of Lemma 8, for any $w=(u,v)\in\mathbb{R}^q\times\mathbb{R}$, set

$$\begin{aligned}
\tilde{\Psi}_n(x) &= \left(\tilde{m}_n(x),\tilde{g}_n(x)\right),\\
\tilde{\Lambda}_{n,x}(w) &= \frac{1}{nh_n^d}\log\mathbb{E}\left[\exp\left(nh_n^d\langle w,\tilde{\Psi}_n(x)\rangle\right)\right].
\end{aligned}$$

When $h_n=cn^{-a}$, $c>0$ and $0<a<1/d$, assume for the moment that

$$\lim_{n\to\infty}\tilde{\Lambda}_{n,x}(u,w)\quad=\quad\tilde{\Psi}_{a,x}(u,v), \tag{3.27}$$

where $\tilde{\Psi}_{a,x}$ is defined in (3.11). The conclusion of Lemma 9 follows from Proposition 8 and again the Gärtner-Ellis Theorem.

Let us now prove (3.27). Set

$$M_i = [\langle u, Y_i\rangle + v]\, K\left(\frac{x - X_i}{h_i}\right),$$

then, for $(u,v) \in \mathbb{R}^q \times \mathbb{R}$,

$$\begin{aligned}\tilde{\Lambda}_{n,x}(u,v) &= \frac{1}{nh_n^d}\log \mathbb{E}\left[\exp\left(\sum_{i=1}^{n} M_i \frac{h_n^d}{h_i^d}\right)\right]\\ &= \frac{1}{nh_n^d}\sum_{i=1}^{n}\log \mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right)\right].\end{aligned}$$

By Taylor expansion, there exists $b_{i,n}$ between 1 and $\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right)\right]$ such that

$$\log \mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right)\right] = \mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right) - 1\right] - \frac{1}{2b_{i,n}^2}\left(\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right) - 1\right]\right)^2.$$

Noting that $h_n = cn^{-a}$ with $c > 0$ and $a \in]0, 1/d[$, $\tilde{\Lambda}_{n,x}$ can be rewritten as

$$\begin{aligned}\tilde{\Lambda}_{n,x}(u,v) &= \frac{1}{h_n^d}\sum_{i=1}^{n}\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right) - 1\right] - \frac{1}{2nh_n^d}\sum_{i=1}^{n}\frac{1}{b_{i,n}^2}\left(\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right) - 1\right]\right)^2\\ &= \frac{1}{n}\sum_{i=1}^{n}\left(\frac{i}{n}\right)^{-ad}\int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{\left(\frac{i}{n}\right)^{ad}(\langle u,y\rangle + v)K(z)} - 1\right] f(x,y)dzdy - R_{n,x}^{(1)}(u,v) + R_{n,x}^{(2)}(u,v),\end{aligned}$$

with

$$\begin{aligned}\tilde{R}_{n,x}^{(1)}(u,v) &= \frac{1}{2nh_n^d}\sum_{i=1}^{n}\frac{1}{b_{i,n}^2}\left(\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right) - 1\right]\right)^2\\ \tilde{R}_{n,x}^{(2)}(u,v) &= \frac{1}{nh_n^d}\sum_{i=1}^{n} h_i^d \int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle + v)K(z)} - 1\right][f(x - h_i z, y) - f(x,y)]\,dzdy.\end{aligned}$$

Since $b_{i,n}$ is between 1 and $\mathbb{E}\left[\exp\left(M_i \frac{h_n^d}{h_i^d}\right)\right]$, we have

$$\frac{1}{b_{i,n}} \leq \max\left\{1, \frac{1}{\mathbb{E}\left(e^{M_i \frac{h_n^d}{h_i^d}}\right)}\right\}.$$

By Jensen's inequality, we obtain

$$\frac{1}{\mathbb{E}\left[e^{M_i \frac{h_n^d}{h_i^d}}\right]} \leq \frac{1}{e^{\mathbb{E}\left[M_i \frac{h_n^d}{h_i^d}\right]}}.$$

Observe that

$$
\begin{aligned}
&\left|\mathbb{E}\left(M_i\frac{h_n^d}{h_i^d}\right)\right| \\
&\leq \frac{h_n^d}{h_i^d}\int_{\mathbb{R}^d}|\langle u,y\rangle+v|\left|K\left(\frac{x-z}{h_i}\right)\right|f(z,y)dzdy \\
&\leq h_n^d\left(\|u\|\int_{\mathbb{R}^d\times\mathbb{R}^q}|K(z)|\left[\int_{\mathbb{R}^q}\|y\|f(x-h_iz,y)dy\right]dz+|v|\int_{\mathbb{R}^d}|K(z)|\left[\int_{\mathbb{R}^q}f(x-h_iz,y)dy\right]dz\right),
\end{aligned}
$$

which goes to 0 in view of (3.7) and since $\lim_{n\to\infty}h_n=0$. We deduce that the sequence $\left(\mathbb{E}\left[\exp\left(M_i\frac{h_n^d}{h_i^d}\right)\right]\right)$ is bounded, so that there exists $c>0$ such that

$$
\frac{1}{b_{i,n}^2}\leq c,
$$

and thus

$$
\tilde{R}_{n,x}^{(1)}(u,v) \quad\leq\quad \frac{c}{2nh_n^d}\sum_{i=1}^n\left(\mathbb{E}\left[e^{M_i\frac{h_n^d}{h_i^d}}-1\right]\right)^2.
$$

Now, in view of (3.7), and since K is bounded integrable, we have

$$
\begin{aligned}
\mathbb{E}\left|e^{M_i\frac{h_n^d}{h_i^d}}-1\right| &\leq \mathbb{E}\left[\left|M_i\frac{h_n^d}{h_i^d}\right|e^{\left|M_i\frac{h_n^d}{h_i^d}\right|}\right] \\
&\leq \frac{h_n^d}{h_i^d}\int_{\mathbb{R}^d}|\langle u,y\rangle+v|\left|K\left(\frac{x-s}{h_i}\right)\right|e^{|\langle u,y\rangle+v|\left|K\left(\frac{x-s}{h_i}\right)\right|}f(s,y)dsdy \\
&\leq h_n^d\int_{\mathbb{R}^d\times\mathbb{R}^q}|\langle u,y\rangle+v|\left|K\left(z\right)\right|e^{|\langle u,y\rangle+v||K(z)|}f(x-h_iz,y)dzdy \\
&\leq Bh_n^d,
\end{aligned}
$$

where B is a constant. Thus

$$
\left|\tilde{R}_{n,x}^{(1)}(u,v)\right| \quad\leq\quad \frac{cB^2}{2}h_n^d,
$$

and

$$
\lim_{n\to\infty}\left|\tilde{R}_{n,x}^{(1)}(u,v)\right|=0.
$$

Let us now consider $R_{n,x}^{(2)}$. Set $A>0$ and $\epsilon>0$; we then have

$$
\begin{aligned}
&\tilde{R}_{n,x}^{(2)}(u,v) \\
&= \frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q}\left[e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_iz,y)-f(x,y)\right]dzdy \\
&\quad+\frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\{\|z\|>A\}\times\mathbb{R}^q}\left[e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}-1\right]\left[f(x-h_iz,y)-f(x,y)\right]dzdy \\
&= I+II.
\end{aligned}
$$

Since $|e^t - 1| \leq |t|e^{|t|}$, it follows that

$$|II| \leq \frac{1}{n}\sum_{i=1}^{n}\int_{\{\|z\|>A\}\times\mathbb{R}^q} |\langle u,y\rangle + v|\,|K(z)|\,e^{|\langle u,y\rangle+v||K(z)|}\,|f(x-h_iz,y)-f(x,y)|\,dzdy.$$

Using the same argument as in (3.22), it holds that

$$|II| \leq \frac{\epsilon}{2}.$$

Now, for I, we write

$$\begin{aligned} I &= \frac{1}{nh_n^d}\sum_{i=1}^{n} h_i^d \int_{\{\|z\|\leq A\}\times\mathbb{R}^q} e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}\,[f(x-h_iz,y)-f(x,y)]\,dzdy \\ &\quad -\frac{1}{nh_n^d}\sum_{i=1}^{n} h_i^d \int_{\{\|z\|\leq A\}\times\mathbb{R}^q} [f(x-h_iz,y)-f(x,y)]\,dzdy. \end{aligned}$$

On the one hand, Assumption (A2) with $u=0$ ensures that

$$\lim_{i\to\infty}\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} [f(x-h_iz,y)-f(x,y)]\,dzdy = 0.$$

Moreover, since $ad<1$, (3.15) ensures that

$$\lim_{n\to\infty}\frac{1}{nh_n^d}\sum_{i=1}^{n} h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} [f(x-h_iz,y)-f(x,y)]\,dzdy = 0,$$

so that for n large enough,

$$\left|\frac{1}{nh_n^d}\sum_{i=1}^{n} h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} [f(x-h_iz,y)-f(x,y)]\,dzdy\right| \leq \frac{\epsilon}{4}. \tag{3.28}$$

On the other hand, since for $i\leq n$, $0\leq \frac{h_n^d}{h_i^d}\leq 1$, by Assumption (A'1), there exists $n_0\in\mathbb{N}$ such that for any $i>n_0$,

$$\left|\int_{\mathbb{R}^q} e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}\,[f(x-h_iz,y)-f(x,y)]\,dy\right| \leq \frac{\epsilon}{8(2-ad)\int_{\|z\|\leq A}dz} \quad \forall n>i.$$

Noting that by (3.7), for any $\alpha\in[0,1]$,

$$\sup_t\left|\int_{\mathbb{R}^q} e^{\alpha\langle u,y\rangle} f(t,y)dy\right| \leq \sup_t\int_{\mathbb{R}^q} e^{\|u\|\|y\|} f(t,y)dy < \infty.$$

Since $ad<1$, by (3.15), we get for n sufficiently large

$$\left|\frac{1}{nh_n^d}\sum_{i=n_0+1}^{n} h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}\,[f(x-h_iz,y)-f(x,y)]\,dydz\right| \leq \frac{\epsilon}{8}.$$

Now, for n large enough, in view of (3.7),

$$\begin{aligned} &\left|\frac{1}{nh_n^d}\sum_{i=1}^{n_0} h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}\,[f(x-h_iz,y)-f(x,y)]\,dydz\right| \\ &\leq \frac{1}{nh_n^d}\sum_{i=1}^{n_0} h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} e^{|\langle u,y\rangle+v||K(z)|}\,|f(x-h_iz,y)-f(x,y)|\,dydz \\ &\leq \frac{\epsilon}{8}. \end{aligned}$$

It follows that for n large enough,

$$\left|\frac{1}{nh_n^d}\sum_{i=1}^n h_i^d\int_{\{\|z\|\leq A\}\times\mathbb{R}^q} e^{\frac{h_n^d}{h_i^d}(\langle u,y\rangle+v)K(z)}\left[f(x-h_iz,y)-f(x,y)\right]dydz\right| \leq \frac{\epsilon}{4}. \qquad (3.29)$$

The combination of (3.28) and (3.29) ensures that $|I|\leq\frac{\epsilon}{2}$, which ensures that

$$\lim_{n\to\infty}\left|\tilde{R}_{n,x}^{(2)}(u)\right| = 0.$$

Hence, (3.27) follows from analysis considerations. ■

3.3.2 Proof of Theorems 6 and 8

Let us consider the following functions defined as :

$$\begin{aligned} H_1:\mathbb{R}^q\times\mathbb{R}^* &\to \mathbb{R}^q \\ (\alpha,\beta) &\mapsto \frac{\alpha}{\beta}, \end{aligned}$$

and

$$\begin{aligned} H_2:\mathbb{R}^q\times\mathbb{R} &\to \mathbb{R}^q \\ (\alpha,\beta) &\mapsto \begin{cases}\frac{\alpha}{\beta} & \text{if } \beta\neq 0\\ 0 & \text{otherwise.}\end{cases} \end{aligned}$$

Proof of Theorem 6

i) Let U be an open subset of $\mathbb{R}^q$, we have

$$\frac{1}{nh_n^d}\log\mathbb{P}\left[r_n(x)\in U\right] = \frac{1}{nh_n^d}\log\mathbb{P}\left[(m_n(x),g_n(x))\in H_2^{-1}(U)\right]. \qquad (3.30)$$

Observe that $H_1^{-1}(U)\subset H_2^{-1}(U)$ and $H_1^{-1}(U)$ is an open subset on $\mathbb{R}^q\times\mathbb{R}^*$ which is open, it follows that $H_1^{-1}(U)$ is an open subset on $\mathbb{R}^q\times\mathbb{R}$. We deduce from (3.30) that

$$\frac{1}{nh_n^d}\log\mathbb{P}\left[r_n(x)\in U\right] \geq \frac{1}{nh_n^d}\log\mathbb{P}\left[(m_n(x),g_n(x))\in H_1^{-1}(U)\right].$$

The application of Lemma 8 ensures that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{nh_n^d}\log\mathbb{P}\left[r_n(x)\in U\right] &\geq \liminf_{n\to\infty}\frac{1}{nh_n^d}\log\mathbb{P}\left[(m_n(x),g_n(x))\in H_1^{-1}(U)\right]\\ &\geq -\inf_{(t_1,t_2)\in H_1^{-1}(U)} I_x(t_1,t_2) = -\inf_{s\in U}J^*(s),\end{aligned}$$

and the first part of Theorem 6 is proved.

ii) Let V be a closed subset of $\mathbb{R}^q$, we have

$$\begin{aligned}\frac{1}{nh_n^d}\log\mathbb{P}\left[r_n(x)\in V\right] &= \frac{1}{nh_n^d}\log\mathbb{P}\left[(m_n(x),g_n(x))\in H_2^{-1}(V)\right]\\ &\leq \frac{1}{nh_n^d}\log\mathbb{P}\left[(m_n(x),g_n(x))\in \overline{H_2^{-1}(V)}\right].\end{aligned}$$

Now, observe that $\overline{H_2^{-1}(V)} = H_1^{-1}(V) \cup A$ where $A \subset \mathbb{R}^q \times \{0\}$ and $\left(\vec{0}, 0\right) \in A$ (since for any $s \in \mathbb{R}^q$, $(\vec{0}, 0) \in \overline{H_2^{-1}(s)}$). The application of Lemma 8 again ensures that

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{nh_n^d} \log \mathbb{P}\left[r_n(x) \in V\right] &\leq -\inf_{(s,t)\in H_1^{-1}(V)\cup A} I_x(s,t) \\
&\leq -\inf_{(s,t)\in H_1^{-1}(V)\cup\{(\vec{0},0)\}} I_x(s,t) \\
&\leq -\inf_{s\in V,\ t\in\mathbb{R}} I_x(st,t) \\
&\leq -\inf_{s\in V,\ t\in\mathbb{R}} \hat{I}_x(s,t) \\
&\leq -\inf_{s\in V} J(s),
\end{aligned}$$

where the second inequality comes from Condition (C) ; this concludes the proof of Theorem 6. ■

Proof of Theorem 8

Applying Lemma 9, Theorem 8 is proved by following the same approach as for the proof of Theorem 6 with replacing m_n, g_n, J^* and J by $\tilde{m}_n$, $\tilde{g}_n$, $\tilde{J}_a^*$ and $\tilde{J}_a$ respectively. ■

3.3.3 Proof of Theorem 7

Set

$$B_n(x) = \frac{1}{g(x)}\left(m_n(x) - m(x)\right) - \frac{r(x)}{g(x)}\left(g_n(x) - g(x)\right).$$

Let us at first state the two following lemmas.

Lemma 11 *Under the assumptions of Theorem 7, the sequence $\left(v_n\left(B_n(x) - \mathbb{E}\left(B_n(x)\right)\right)\right)$ satisfies a LDP with speed $\left(\frac{nh_n^d}{v_n^2}\right)$ and good rate function G_x.*

Lemma 12 *Under the assumptions of Theorem 7,*

$$\lim_{n\to\infty} v_n \mathbb{E}(B_n(x)) = 0. \tag{3.31}$$

We first show that how Theorem 7 can be deduced from the application of Lemmas 11 and 12, and then prove Lemmas 11 and 12 successively.

Proof of Theorem 7

Lemmas 11 and 12 imply that the sequence $(v_n B_n(x))$ satisfies a LDP with speed $\left(\dfrac{nh_n^d}{v_n^2}\right)$ and good rate function G_x. To prove Theorem 7, we show that $(v_n(r_n - r))$ and $(v_n B_n)$ are exponentially contiguous.

Let us first note that, for x such that $g_n(x) \neq 0$, we have :

$$\begin{aligned}
r_n(x) - r(x) &= \frac{m_n(x)}{g_n(x)} - \frac{m(x)}{g(x)} \\
&= \frac{\left(m_n(x) - m(x)\right) g(x) + \left(g(x) - g_n(x)\right) m(x)}{g_n(x) g(x)} \\
&= B_n(x) \frac{g(x)}{g_n(x)}.
\end{aligned}$$

It follows that, for any $\delta > 0$, we have

$$\begin{aligned}
&\mathbb{P}\Big[v_n\| \left(r_n(x) - r(x)\right) - B_n(x)\| > \delta\Big] \\
&\leq \mathbb{P}\Big[v_n\|B_n(x)\left(\frac{g(x)}{g_n(x)} - 1\right)\| > \delta \text{ and } g_n(x) \neq 0\Big] + \mathbb{P}\left[g_n(x) = 0\right] \\
&\leq \mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] + \mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\left|g_n(x)\right|\right] + \mathbb{P}\left[\left|g(x) - g_n(x)\right| > \frac{g(x)}{2}\right] \\
&\leq \mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] + \mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\left|g_n(x)\right| \text{ and } \frac{g_n(x)}{g(x)} > \frac{1}{2}\right] \\
&\quad + \mathbb{P}\left[\frac{g_n(x)}{g(x)} \leq \frac{1}{2}\right] + \mathbb{P}\left[\left|g_n(x) - g(x)\right| > \frac{g(x)}{2}\right] \\
&\leq \mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] + \mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\frac{g(x)}{2}\right] + \mathbb{P}\left[g(x) - g_n(x) \geq \frac{g(x)}{2}\right] \\
&\quad + \mathbb{P}\left[\left|g_n(x) - g(x)\right| > \frac{g(x)}{2}\right].
\end{aligned}$$

Since $\lim_{n\to\infty} v_n = \infty$, it follows that, for n large enough,

$$\begin{aligned}
&\mathbb{P}\Big[v_n\| \left(r_n(x) - r(x)\right) - B_n(x)\| > \delta\Big] \\
&\leq 4\max\left\{\mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] \;;\; \mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\frac{g(x)}{2}\right]\right\},
\end{aligned}$$

and thus

$$\begin{aligned}
&\frac{v_n^2}{nh_n^d}\log\mathbb{P}\Big[v_n\| \left(r_n(x) - r(x)\right) - B_n(x)\| > \delta\Big] \\
&\leq \frac{v_n^2}{nh_n^d}\log 4 + \max\left\{\frac{v_n^2}{nh_n^d}\log\mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] \;;\; \frac{v_n^2}{nh_n^d}\log\mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\frac{g(x)}{2}\right]\right\}.
\end{aligned}$$

Now, since the sequence $(v_nB_n(x))$ satisfies a LDP with speed $\left(\frac{nh_n^d}{v_n^2}\right)$ and good rate function G_x, there exists $c_1 > 0$ such that

$$\limsup_{n\to\infty}\frac{v_n}{nh_n^d}\log\mathbb{P}\left[\sqrt{v_n}\|B_n(x)\| > \delta\right] \leq -c_1.$$

Moreover, the application of Theorem 1 in Mokkadem et al. [14] guarantees the existence of $c_2 > 0$ such that

$$\limsup_{n\to\infty}\frac{v_n}{nh_n^d}\log\mathbb{P}\left[\sqrt{v_n}\left|g(x) - g_n(x)\right| > \delta\frac{g(x)}{2}\right] < -c_2.$$

We thus deduce that

$$\lim_{n\to\infty}\frac{v_n^2}{nh_n^d}\log\mathbb{P}\Big[v_n\| \left(r_n(x) - r(x)\right) - B_n(x)\| > \delta\Big] = -\infty,$$

which means that the sequences $(v_n(r_n(x) - r(x)))$ and $(v_nB_n(x))$ are exponentially contiguous. Theorem 7 thus follows. ■

Proof of Lemma 11

For any $u \in \mathbb{R}^q$, set

$$\begin{aligned}\Gamma_{n,x}(u) &= \frac{v_n^2}{nh_n^d} \log \mathbb{E}\left[\exp\left(\frac{nh_n^d}{v_n}\langle u, B_n(x) - \mathbb{E}\left(B_n(x)\right)\rangle\right)\right], \\ \Phi_x(u) &= \frac{1}{2g^2(x)}\int_{\mathbb{R}^d\times\mathbb{R}^q} \langle u, y-r(x)\rangle^2 K^2(z) f(x,y)dzdy \\ &= \frac{u^T\Sigma_x u}{2g(x)}\int_{\mathbb{R}^d} K^2(z)dz.\end{aligned}$$

To prove Lemma 11, it suffices to show that, for all $u \in \mathbb{R}^q$,

$$\lim_{n\to\infty} \Gamma_{n,x}(u) = \Phi_x(u).$$

As a matter of fact, since Φ_x is a quadratic function, Lemma 11 then follows from the application of the Gärtner-Ellis Theorem. For $u \in \mathbb{R}^q$, set

$$\hat{Z}_i = \langle u, Y_i - r(x)\rangle K\left(\frac{x - X_i}{h_n}\right),$$

and note that

$$\Gamma_{n,x}(u) = \frac{v_n^2}{nh_n^d} \log \mathbb{E}\left[\exp\left(\frac{1}{v_n g(x)}\sum_{i=1}^n \left[\hat{Z}_i - \mathbb{E}(\hat{Z}_i)\right]\right)\right].$$

Since (X_i, Y_i), $i = 1, \ldots, n$ are independent and identically distributed, it holds that

$$\Gamma_{n,x}(u) = \frac{v_n^2}{h_n^d}\log \mathbb{E}\left[e^{\frac{\hat{Z}_n}{v_n g(x)}}\right] - \frac{v_n}{h_n^d g(x)}\mathbb{E}(\hat{Z}_n).$$

Now, we follow the same lines as in the proof of Lemma 10. A Taylor's expansion ensures that there exists $\hat{c}_n$ between 1 and $\mathbb{E}\left[e^{\frac{\hat{Z}_n}{v_n g(x)}}\right]$ such that

$$\begin{aligned}\Gamma_{n,x}(u) &= \frac{v_n^2}{h_n^d}\mathbb{E}\left[e^{\frac{\hat{Z}_n}{v_n g(x)}} - 1 - \frac{\hat{Z}_n}{v_n g(x)}\right] - \hat{R}_{n,x}^{(1)}(u) \\ &= v_n^2 \int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{\frac{1}{v_n g(x)}\langle u, y-r(x)\rangle K(z)} - 1 - \frac{1}{v_n g(x)}\langle u, y - r(x)\rangle K\left(z\right)\right] f(x,y)dzdy \\ &\quad -\hat{R}_{n,x}^{(1)}(u) + \hat{R}_{n,x}^{(2)}(u),\end{aligned}$$

with

$$\begin{aligned}&\hat{R}_{n,x}^{(1)}(u) \\ &\quad = \frac{v_n^2}{2\hat{c}_n^2 h_n^d}\left(\mathbb{E}\left[e^{\frac{\hat{Z}_n}{v_n g(x)}} - 1\right]\right)^2 \\ \hat{R}_{n,x}^{(2)}(u) &= v_n^2\int_{\mathbb{R}^d\times\mathbb{R}^q}\left[e^{\frac{1}{v_n g(x)}\langle u, y-r(x)\rangle K(z)} - 1 - \frac{1}{v_n g(x)}\langle u, y-r(x)\rangle K\left(z\right)\right]\left[f(x - h_n z, y) - f(x,y)\right]dzdy,\end{aligned}$$

and

$$\frac{1}{\hat{c}_n} \leq \max\left\{1, \frac{1}{\mathbb{E}\left(e^{\frac{\hat{Z}_n}{v_n g(x)}}\right)}\right\}.$$

Noting that

$$\begin{aligned}\left|\mathbb{E}\left(\frac{\hat{Z}_n}{v_n g(x)}\right)\right| &\leq \frac{h_n^d \|u\|}{v_n g(x)} \int_{\mathbb{R}^d} |K(z)| \left[\int_{\mathbb{R}^q} \|y\| f(x-h_n z, y) dy\right] dz \\ &\quad + \frac{h_n^d \|u\| \|r(x)\|}{v_n g(x)} \int_{\mathbb{R}^d} |K(z)| \left[\int_{\mathbb{R}^q} f(x-h_n z, y) dy\right] dz. \qquad (3.32)\end{aligned}$$

It follows from (3.7) that,

$$\mathbb{E}\left(\frac{\hat{Z}_n}{v_n g(x)}\right) \to 0.$$

We deduce that there exists $c' \in \mathbb{R}_+^*$ such that

$$\frac{1}{\hat{c}_n^2} \leq c',$$

and thus, in view of (3.7),

$$\begin{aligned}\hat{R}_{n,x}^{(1)}(u) &\leq \frac{c'}{2} \frac{v_n^2}{h_n^d} \left(\int_{\mathbb{R}^d \times \mathbb{R}^q} \left[e^{\frac{1}{v_n g(x)} \langle u, y-r(x) \rangle K\left(\frac{x-s}{h_n}\right)} - 1\right] f(s,y) ds dy\right)^2 \\ &\leq \frac{c'}{2g^2(x)} h_n^d \left(\int_{\mathbb{R}^d \times \mathbb{R}^q} |\langle u, y-r(x) \rangle K(z)| \, e^{\left|\frac{1}{g(x)} \langle u, y-r(x) \rangle K(z)\right|} f(x-h_n z, y) dz dy\right)^2 \\ &\leq \frac{c' e^{\frac{2\|K\|_\infty}{g(x)} \|u\| \|r(x)\|}}{2g^2(x)} h_n^d \left(\int_{\mathbb{R}^d \times \mathbb{R}^q} |\langle u, y-r(x) \rangle K(z)| \, e^{\frac{\|K\|_\infty}{g(x)} \|u\| \|y\|} f(x-h_n z, y) dz dy\right)^2 \\ &\leq B h_n^d,\end{aligned}$$

where B is a constant, so that

$$\lim_{n\to\infty} \left|\hat{R}_{n,x}^{(1)}(u)\right| = 0.$$

On the other hand, since $\forall x \in \mathbb{R}$, $e^x - 1 - x = \frac{x^2}{2} + \frac{x^3}{6} d(x)$, with $d(x) \leq e^{|x|}$, we get

$$\begin{aligned}&\hat{R}_{n,x}^{(2)}(u) \\ &= \frac{1}{2g^2(x)} \int_{\mathbb{R}^d \times \mathbb{R}^q} \langle u, y-r(x) \rangle^2 K^2(z) \left[f(x-h_n z, y) - f(x,y)\right] dz dy + \mathcal{R}_{n,x}(u), \qquad (3.33)\end{aligned}$$

with

$$|\mathcal{R}_{n,x}(u)| \leq \frac{1}{6 v_n g^3(x)} \int_{\mathbb{R}^d \times \mathbb{R}^q} \left|\langle u, y-r(x) \rangle^3 K^3(z)\right| e^{\frac{1}{g(x)} |\langle u, y-r(x) \rangle K(z)|} \left|f(x-h_n z, y) - f(x,y)\right| dz dy.$$

It follows from (3.7) that $\mathcal{R}_{n,x}$ converges to 0. Applying then (A2) and (A3), we find

$$\lim_{n\to\infty} \left|\hat{R}_{n,x}^{(2)}(u)\right| = 0.$$

Finally, we have

$$\begin{aligned}&\Gamma_{n,x}(u) \\ &= v_n^2 \int_{\mathbb{R}^d \times \mathbb{R}^q} \left[e^{\frac{1}{v_n g(x)} \langle u, y-r(x) \rangle K(z)} - 1 - \frac{1}{v_n g(x)} \langle u, y-r(x) \rangle K(z)\right] f(x,y) dz dy - \hat{R}_{n,x}^{(1)}(u) \\ &\quad + \hat{R}_{n,x}^{(2)}(u) \\ &= \Phi_x(u) - \hat{R}_{n,x}^{(1)}(u) + \hat{R}_{n,x}^{(2)}(u) + \hat{R}_{n,x}^{(3)}(u),\end{aligned}$$

with

$$\hat{R}_{n,x}^{(3)}(u)$$
$$= v_n^2 \int_{\mathbb{R}^d \times \mathbb{R}^q} \left[e^{\frac{1}{v_n g(x)} \langle u, y-r(x) \rangle K(z)} - 1 - \frac{\langle u, y-r(x) \rangle K(z)}{v_n g(x)} - \frac{\langle u, y-r(x) \rangle^2 K^2(z)}{2v_n^2 g^2(x)} \right] f(x,y) dz dy.$$

By the majoration $\left|e^x - 1 - x - \frac{x^2}{2}\right| \leq \left|\frac{x^3}{6} d(x)\right|$, we get

$$\left|\hat{R}_{n,x}^{(3)}(u)\right|$$
$$\leq \frac{1}{6 v_n g^3(x)} \int_{\mathbb{R}^d \times \mathbb{R}^q} \left|\langle u, y-r(x) \rangle^3 K^3(z)\right| e^{\frac{1}{v_n g(x)} |\langle u, y-r(x) \rangle K(z)|} f(x,y) dz dy,$$

and (3.7) ensures that

$$\lim_{n\to\infty} \left|\hat{R}_{n,x}^{(3)}(u)\right| = 0,$$

which concludes the proof of Lemma 11. ■

Proof of Lemma 12

Observe that

$$\mathbb{E}(B_n(x)) = \frac{1}{g(x)} \left[\mathbb{E}(m_n(x)) - m(x)\right] - \frac{r(x)}{g(x)} \left[\mathbb{E}(g_n(x)) - g(x)\right]. \quad (3.34)$$

Since

$$\begin{aligned}
\mathbb{E}(m_n(x)) - m(x) &= \frac{1}{h_n^d} \mathbb{E}\left(Y_1 K\left(\frac{x - X_1}{h_n}\right)\right) - m(x) \\
&= \frac{1}{h_n^d} \int_{\mathbb{R}^d \times \mathbb{R}^q} y K\left(\frac{x-z}{h_n}\right) f(z,y) dz dy - m(x) \\
&= \frac{1}{h_n^d} \int_{\mathbb{R}^d} m(z) K\left(\frac{x-z}{h_n}\right) dz - m(x) \\
&= \int_{\mathbb{R}^d} K(y) \left[m(x - h_n y) - m(x)\right] dy,
\end{aligned}$$

Assumptions (A5)i), (A5)iii) and a Taylor's expansion of m of order p ensure that

$$\mathbb{E}(m_n(x)) - m(x) = O(h_n^p). \quad (3.35)$$

Similarly, we have

$$\mathbb{E}(g_n(x)) - g(x) = O(h_n^p). \quad (3.36)$$

We deduce from (3.34), (3.35), and (3.36) that

$$\mathbb{E}(B_n(x)) = O(h_n^p),$$

and thus Lemma 12 follows from Assumption (A5)ii). ■

3.3.4 Proof of Theorem 9

Set

$$\tilde{B}_n(x) = \frac{1}{g(x)}\left(\tilde{m}_n(x) - m(x)\right) - \frac{r(x)}{g(x)}\left(\tilde{g}_n(x) - g(x)\right),$$

and, for any $u \in \mathbb{R}^q$,

$$\begin{aligned}
\tilde{\Gamma}_{n,x}(u) &= \frac{v_n^2}{nh_n^d}\log \mathbb{E}\left[\exp\left(\frac{nh_n^d}{v_n}\langle u, \tilde{B}_n(x) - \mathbb{E}\left(\tilde{B}_n(x)\right)\rangle\right)\right],\\
\tilde{\Phi}_{a,x}(u) &= \frac{1}{2(1+ad)g^2(x)}\int_{\mathbb{R}^d\times\mathbb{R}^q}\langle u, y-r(x)\rangle^2 K^2(z) f(x,y)dzdy\\
&= \frac{1}{1+ad}\frac{u^T\Sigma_x u}{2g(x)}\int_{\mathbb{R}^d} K^2(z)dz.
\end{aligned}$$

By following the steps of the proof of Lemma 11 and by using the property (3.15), we prove that

$$\lim_{n\to\infty}\tilde{\Gamma}_{n,x}(u) = \tilde{\Phi}_{a,x}(u). \tag{3.37}$$

We first show how (3.37) implies Theorem 9. The function $\tilde{\Phi}_{a,x}$ being quadratic, the application of the Gärtner-Ellis Theorem then ensures that

$$\begin{aligned}
&\text{the sequence } \left(v_n\left(\tilde{B}_n(x) - \mathbb{E}\left(\tilde{B}_n(x)\right)\right)\right) \text{ satisfies a LDP}\\
&\text{with speed } \left(\frac{nh_n^d}{v_n^2}\right) \text{ and good rate function } \tilde{G}_{a,x}.
\end{aligned} \tag{3.38}$$

Now, following the proof of Lemma 12, we have

$$\begin{aligned}
\mathbb{E}\left(\tilde{m}_n(x)\right) - m(x) &= \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^d}\int_{\mathbb{R}^d\times\mathbb{R}^q} yK\left(\frac{x-z}{h_i}\right) f(z,y)dzdy - m(x)\\
&= \frac{1}{n}\sum_{i=1}^n \int_{\mathbb{R}^d} K(y)\left[m(x-h_iy) - m(x)\right]dy.
\end{aligned}$$

Here again, Assumptions (A5)i), (A5)iii) and a Taylor's expansion of m of order p ensure that

$$\mathbb{E}\left(\tilde{m}_n(x)\right) - m(x) = O\left(\frac{1}{n}\sum_{i=1}^n h_i^p\right).$$

and similarly,

$$\mathbb{E}\left(\tilde{g}_n(x)\right) - g(x) = O\left(\frac{1}{n}\sum_{i=1}^n h_i^p\right),$$

thus

$$v_n\mathbb{E}\left(\tilde{B}_n(x)\right) = O\left(\frac{v_n}{n}\sum_{i=1}^n h_i^p\right).$$

- If $ap < 1$, since (h_n) varies regularly with exponent $(-a)$, we have, in view of (3.15) and Assumption (A5)ii),

$$\frac{v_n}{n}\sum_{i=1}^{n} h_i^p = O\left(\frac{v_n}{n}\left[nh_n^p\right]\right) = o(1).$$

- If $ap > 1$, we have $\sum_i h_i^p < \infty$ and thus, since $v_n = o(nh_n^d)$, we get

$$\frac{v_n}{n}\sum_{i=1}^{n} h_i^p = O\left(\frac{v_n}{n}\right) = o(1).$$

- In the case $ap = 1$, let $\mathcal{L}$ be the slowly varying function such that $h_n = n^{-a}\mathcal{L}(n)$, and set $\varepsilon > 0$ small enough. Since $a(p-\varepsilon) < 1$, we have $h_n^p = o(h_n^{p-\varepsilon})$, and in view of (3.15) and (A4),

$$\begin{aligned}\frac{v_n}{n}\sum_{i=1}^{n} h_i^p &= o\left(v_n h_n^{p-\varepsilon}\right) = o\left(nh_n^{d+p-\varepsilon}\right)\\ &= o\left(n^{1-a(d+p-\varepsilon)}\left[\mathcal{L}(n)\right]^{d+p-\varepsilon}\right)\\ &= o\left(n^{-a(d-\varepsilon)}\left[\mathcal{L}(n)\right]^{d+p-\varepsilon}\right) = o(1).\end{aligned}$$

We thus deduce that

$$\lim_{n\to\infty} v_n\mathbb{E}(\tilde{B}_n(x)) = 0. \tag{3.39}$$

To conclude the proof of Theorem 9, we follow the same lines as for the proof of Theorem 7 (see Section 3.3.3), except that we apply (3.38) instead of Lemma 11, (3.39) instead of Lemma 12, and Theorem 1 in Mokkadem et al. [13] instead of Theorem 1 in Mokkadem et al. [14]. ■
Let us now prove (3.37). For $u \in \mathbb{R}^q$, set

$$T_i = \langle u, Y_i - r(x)\rangle K\left(\frac{x - X_i}{h_i}\right),$$

and note that

$$\tilde{\Gamma}_{n,x}(u) = \frac{v_n^2}{nh_n^d}\log\mathbb{E}\left[\exp\left(\frac{h_n^d}{v_n g(x)}\sum_{i=1}^{n}\frac{1}{h_i^d}\left[T_i - \mathbb{E}(T_i)\right]\right)\right].$$

Since (X_i, Y_i), $i = 1, \ldots, n$ are independent and identically distributed, it holds that

$$\tilde{\Gamma}_{n,x}(u) = \frac{v_n^2}{h_n^d}\sum_{i=1}^{n}\log\mathbb{E}\left[e^{\frac{h_n^d T_i}{v_n g(x) h_i^d}}\right] - \frac{v_n}{ng(x)}\sum_{i=1}^{n}\frac{1}{h_i^d}\mathbb{E}(T_i).$$

By Taylor expansion, there exists $c_{i,n}$ between 1 and $\mathbb{E}\left[e^{\frac{h_n^d T_i}{v_n g(x)}h_i^d}\right]$ such that

$$\log\mathbb{E}\left[e^{\frac{h_n^d T_i}{v_n g(x)}h_i^d}\right] = \mathbb{E}\left[e^{\frac{h_n^d T_i}{v_n g(x)}h_i^d} - 1\right] - \frac{1}{2c_{i,n}^2}\left(\mathbb{E}\left[e^{\frac{h_n^d T_i}{v_n g(x)}h_i^d} - 1\right]\right)^2,$$

and $\tilde{\Gamma}_{n,x}$ can be rewritten as

$$\tilde{\Gamma}_{n,x}(u) = \frac{v_n^2}{nh_n^d}\sum_{i=1}^n \mathbb{E}\left[e^{\frac{T_i h_n^d}{v_n g(x) h_i^d}} - 1\right] - \frac{v_n^2}{2nh_n^d}\sum_{i=1}^n \frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[e^{\frac{T_i h_n^d}{v_n g(x) h_i^d}} - 1\right]\right)^2 - \frac{v_n}{ng(x)}\sum_{i=1}^n \frac{1}{h_i^d}\mathbb{E}(T_i).$$

A Taylor expansion implies again that there exists $c'_{i,n}$ between 0 and $\frac{T_i h_n^d}{v_n g(x) h_i^d}$ such that

$$\begin{aligned}
&\mathbb{E}\left[e^{\frac{T_i h_n^d}{v_n g(x) h_i^d}} - 1\right] \\
&= \frac{h_n^d}{v_n g(x) h_i^d}\mathbb{E}(T_i) + \frac{1}{2}\left(\frac{h_n^d}{v_n g(x) h_i^d}\right)^2 \mathbb{E}(T_i^2) + \frac{1}{6}\left(\frac{h_n^d}{v_n g(x) h_i^d}\right)^3 \mathbb{E}(e^{c'_{i,n}} T_i^3).
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\tilde{\Gamma}_{n,x}(u) \\
&= \frac{1}{2g^2(x)}\frac{1}{nh_n^{-d}}\sum_{i=1}^n \frac{1}{h_i^d}\int_{\mathbb{R}^d\times\mathbb{R}^q} \langle u, y - r(x)\rangle^2 K^2(z) f(x,y)dzdy + \ddot{R}_{n,x}^{(1)}(u) + \ddot{R}_{n,x}^{(2)}(u), \quad (3.40)
\end{aligned}$$

with

$$\begin{aligned}
\ddot{R}_{n,x}^{(1)}(u) &= \frac{1}{6}\frac{h_n^{2d}}{v_n g^3(x)}\frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^{2d}}\mathbb{E}(e^{c'_{i,n}} T_i^3) - \frac{v_n^2}{2nh_n^d}\sum_{i=1}^n \frac{1}{c_{i,n}^2}\left(\mathbb{E}\left[e^{\frac{T_i h_n^d}{v_n g(x) h_i^d}} - 1\right]\right)^2, \\
\ddot{R}_{n,x}^{(2)}(u) &= \frac{h_n^d}{2g^2(x)}\frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^d}\int_{\mathbb{R}^d\times\mathbb{R}^q} \langle u, y - r(x)\rangle^2 K^2(z)\left[f(x - h_i z, y) - f(x,y)\right] dzdy.
\end{aligned}$$

In view of (3.15), the first term in the right-hand-side of (3.40) converges to $\tilde{\Phi}_{a,x}$.
It remains to prove that $\ddot{R}_{n,x}^{(1)}$ and $\ddot{R}_{n,x}^{(2)}$ converge to 0. We have

$$\begin{aligned}
\left|\mathbb{E}\left[\frac{h_n^d T_i}{v_n g(x) h_i^d}\right]\right| &\leq \frac{h_n^d}{v_n g(x) h_i^d}\int_{\mathbb{R}^d\times\mathbb{R}^q} \left|\langle u, y - r(x)\rangle K\left(\frac{x-s}{h_i}\right)\right| f(s,y)dsdy \\
&\leq \frac{h_n^d}{v_n g(x)}\int_{\mathbb{R}^d\times\mathbb{R}^q} \left|\langle u, y - r(x)\rangle K(z)\right| f(x - h_i z, y)dzdy.
\end{aligned}$$

In view of (3.7), the integral is bounded, thus

$$\lim_{n\to\infty}\sup_{i\leq n}\mathbb{E}\left[\frac{h_n^d T_i}{v_n g(x) h_i^d}\right] = 0,$$

so that, there exists $c > 0$ such that

$$\frac{1}{c_{i,n}^2} \leq c.$$

Now, on the one hand, since $|e^t - 1| \leq |t|e^{|t|}$, and in view of (3.7) and (3.16), we have

$$\begin{aligned}
\mathbb{E}\left|e^{\frac{h_n^d T_i}{v_n g(x) h_i^d}} - 1\right| &\leq \frac{h_n^d}{v_n g(x)}\int_{\mathbb{R}^d\times\mathbb{R}^q} \left|\langle u, y - r(x)\rangle K(z)\right| e^{c\left|\frac{1}{g(x)}\langle u, y - r(x)\rangle K(z)\right|} f(x - h_i z, y)dzdy \\
&\leq B_1\frac{h_n^d}{v_n g(x)},
\end{aligned}$$

where B_1 and c are constants. We deduce that

$$\lim_{n\to\infty} \frac{v_n^2}{2nh_n^d} \sum_{i=1}^{n} \frac{1}{c_{i,n}^2} \left(\mathbb{E}\left[e^{\frac{T_i h_n^d}{v_n g(x) h_i^d}} - 1 \right] \right)^2 = 0.$$

On the other hand,

$$\begin{aligned} \mathbb{E}\left[T_i^3 e^{c'_{i,n}}\right] &\leq \mathbb{E}\left[|T_i|^3 e^{|c'_{i,n}|}\right] \\ &\leq h_i^d \int_{\mathbb{R}^d\times\mathbb{R}^q} |\langle u, y - r(x)\rangle K(z)|^3 \, e^{\frac{c}{g(x)}|\langle u,y-r(x)\rangle K(z)|} f(x - h_i z, y) dz dy \\ &\leq B_2 h_i^d, \end{aligned}$$

where B_2 is a constant. Thus,

$$\left| \frac{h_n^{2d}}{6nv_n g^3(x)} \sum_{i=1}^{n} \frac{1}{h_i^{2d}} \mathbb{E}(e^{c'_{i,n}} T_i^3) \right| \leq \frac{h_n^d}{6v_n g^3(x)} \frac{B_2}{nh_n^{-d}} \sum_{i=1}^{n} h_i^{-d}.$$

Since $\lim_{n\to\infty} \dfrac{h_n^d}{v_n} = 0$, (3.15) ensures that

$$\lim_{n\to\infty} \frac{h_n^{2d}}{6v_n g^3(x)} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i^{2d}} \mathbb{E}(e^{c'_{i,n}} T_i^3) = 0,$$

which proves that

$$\lim_{n\to\infty} \left| \ddot{R}_{n,x}^{(1)}(u) \right| = 0.$$

Finally, using (3.15), (A2) and (A3), we have

$$\begin{aligned} &\lim_{n\to\infty} \ddot{R}_{n,x}^{(2)}(u) \\ &= \lim_{n\to\infty} \frac{1}{2g^2(x)} \frac{\sum_{i=1}^n h_i^{-d}}{nh_n^{-d}} \frac{1}{\sum_{i=1}^n h_i^{-d}} \sum_{i=1}^{n} h_i^{-d} \int_{\mathbb{R}^d\times\mathbb{R}^q} \langle u, y - r(x)\rangle^2 K^2(z) \left[f(x - h_i z, y) - f(x, y) \right] dz dy \\ &= 0, \end{aligned}$$

which proves (3.37). ■

3.3.5 Proof of Propositions 7 and 8

Proof of Proposition 7

- The strict convexity of Ψ_x follows from its definition, since for any $\gamma \in]0, 1[$, and $(u, v) \neq (u', v')$,

$$\begin{aligned} \Psi_x\left(\gamma\,(u, v) + (1-\gamma)\left(u', v'\right)\right) &= \Psi_x\left(\left(\gamma u + (1-\gamma)u', \gamma v + (1-\gamma)v'\right)\right) \\ &= \int_{\mathbb{R}^d\times\mathbb{R}^q} \left(e^{[\langle \gamma u + (1-\gamma)u', y\rangle + \gamma v + (1-\gamma)v']K(z)} - 1 \right) f(x, y) dz dy \\ &< \gamma \int_{\mathbb{R}^d\times\mathbb{R}^q} \left(e^{(\langle u, y\rangle + v)K(z)} - 1 \right) f(x, y) dz dy \\ &\quad + (1-\gamma) \int_{\mathbb{R}^d\times\mathbb{R}^q} \left(e^{(\langle u', y\rangle + v')K(z)} - 1 \right) f(x, y) dz dy, \end{aligned}$$

where the last inequality follows from the fact that $x \mapsto e^x$ is strictly convex.
Since $|e^t - 1| \leq |t|e^{|t|}$ $\forall t \in \mathbb{R}$ and K is bounded and integrable, (3.7) imply that

$$\begin{aligned}
&\int_{\mathbb{R}^d\times\mathbb{R}^q} \left|\left(e^{(\langle u,y\rangle+v)K(z)} - 1\right) f(x,y)\right| dzdy \\
\leq \;& \int_{\mathbb{R}^d\times\mathbb{R}^q} |(\langle u,y\rangle + v)\, K(z)|\, e^{|(\langle u,y\rangle+v)K(z)|} f(x,y)dzdy \\
\leq \;& e^{|v|\|K\|_\infty}\|u\| \int_{\mathbb{R}^d} |K(z)|\, dz \int_{\mathbb{R}^q} \|y\| e^{\|u\|\|y\|\|K\|_\infty} f(x,y)dy \\
&+ e^{|v|\|K\|_\infty}|v| \int_{\mathbb{R}^d} |K(z)|\, dz \int_{\mathbb{R}^q} e^{\|u\|\|y\|\|K\|_\infty} f(x,y)dy < \infty,
\end{aligned}$$

which ensures the existence of Ψ_x.
Next, set

$$h_x(u,v,y,z) = \left[e^{(\langle u,y\rangle+v)K(z)} - 1\right] f(x,y).$$

Since h_x is differentiable with respect to (u,v) and

$$\nabla h_x(u,v,y,z) \;=\; \begin{pmatrix} ye^{(\langle u,y\rangle+v)K(z)}K(z)f(x,y) \\ e^{(\langle u,y\rangle+v)K(z)}K(z)f(x,y) \end{pmatrix},$$

using Assumption (A1) and (3.7), it can be seen that Ψ_x is differentiable on $\mathbb{R}^d \times \mathbb{R}^q$. Since Ψ_x is a smooth convex on $\mathbb{R}^d \times \mathbb{R}^q$, it follows that Ψ_x is essentially smooth so that I_x is a good rate function on $\mathbb{R}^q \times \mathbb{R}$ (see Dembo and Zeitouni [6]), which proves the first part of Proposition 7.
Now, observe that $\overset{\circ}{\mathcal{D}}(\Psi_x) = \mathbb{R}^d \times \mathbb{R}$, and since Ψ_x is strictly convex, it holds that the pair $\left(\overset{\circ}{\mathcal{D}}(\Psi_x), \Psi_x\right)$ is a convex function of Legendre type. It follows that $\left(\overset{\circ}{\mathcal{D}}(I_x), I_x\right)$ is a convex function of Legendre type (See Rockafellar [17]). Thus, Part 2 of Proposition 7 follows from Theorem 26.5 of Rockafellar [17].

- Let us now assume that $\lambda(S_-) = 0$. Thus

$$\Psi_x(u,v) = \int_{\mathbb{R}^d\times\mathbb{R}^q} \left(e^{u^T yK(z)}e^{vK(z)} - 1\right) \mathbb{1}_{S_+}(z)f(x,y)dzdy.$$

For each $u \in \mathbb{R}^q$, the function $v \mapsto \left(e^{u^T yK(z)}e^{vK(z)} - 1\right) \mathbb{1}_{S_+}(z)f(x,y)$ is increasing in v and goes to $-f(x,y)$ when $v \to -\infty$. Thus $\lim_{v\to-\infty} \Psi_x(u,v) = -g(x)\lambda(S_+)$ and $I_x(\vec{0},0) = g(x)\lambda(S_+)$. Now, when $t_1 \neq \vec{0}$, let us show that

$$I_x(t_1, 0) = +\infty.$$

Let $M > 0$, $\epsilon > 0$ and set $u = (M+\epsilon)t_1/\|t_1\|^2$. Let $v \in \mathbb{R}$ such that

$$\begin{cases} -\Psi_x(u,v) > g(x)\lambda(S_+) - \epsilon \;\text{ if }\; \lambda(S_+) < \infty \\ -\Psi_x(u,v) > M \;\text{ if }\; \lambda(S_+) = \infty. \end{cases}$$

Then, on the one hand, when $\lambda(S_+) < \infty$, we have

$$u^T t_1 - \Psi_x(u,v) \geq M + \epsilon + g(x)\lambda(S_+) - \epsilon > M.$$

On the other hand, when $\lambda(S_+) = \infty$, we get

$$u^T t_1 - \Psi_x(u,v) \geq M + \epsilon + M > M.$$

It follows that $\sup_{u,v}\left(u^T t_1 - \Psi_x(u,v)\right) = +\infty$. ■

Proof of Proposition 8

Following the same lines of the proof of Proposition 7, we prove Proposition 8. When $\lambda(S_-) = 0$, for each $u \in \mathbb{R}^q$ and $s \in]0,1]$, the map $v \mapsto s^{-ad}\left(e^{s^{ad}u^T yK(z)}e^{vK(z)} - 1\right)\mathbb{1}_{S_+}(z)f(x,y)$ is increasing in v and goes to $-s^{-ad}f(x,y)$ when $v \to -\infty$. We deduce that $\lim_{v\to-\infty}\tilde{\Psi}_{a,x}(u,v) = -g(x)\lambda(S_+)\int_0^1 s^{-ad} = -g(x)\lambda(S_+)/(1-ad)$ and $I_x(\vec{0},0) = g(x)\lambda(S_+)/(1-ad)$. ■

3.3.6 Proof of Propositions 5 and 6

Proof of Proposition 5

(i) Let us prove the first part of Proposition 5.

- If $\alpha < I_x(\vec{0},0)$, set

$$G = \{(a,b) \in \mathbb{R}^q \times \mathbb{R},\ \ I_x(a,b) \leq \alpha\} \ \text{ and } \ \hat{G} = \{(s,t) \in \mathbb{R}^q \times \mathbb{R},\ \ \hat{I}_x(s,t) \leq \alpha\}.$$

We first show that $\hat{G}$ is a compact subset of $\mathbb{R}^q \times \mathbb{R}$.
First, observe that since I_x is a good rate function, G is a compact subset of $\mathbb{R}^q \times \mathbb{R}$. Let us define the following function

$$\begin{aligned} F : \mathbb{R}^q \times \mathbb{R} &\to \mathbb{R}^q \times \mathbb{R} \\ (s,t) &\mapsto (st,t). \end{aligned}$$

Observe that F is continuous and $\hat{G} = F^{-1}(G)$. We deduce that $\hat{G}$ is a closed subset of $\mathbb{R}^q \times \mathbb{R}$.
Now, let (s_n,t_n) be a sequence of real numbers of $\hat{G}$, there exists $(x_n,y_n) \in G$ such that $(x_n,y_n) = F(s_n,t_n) = (s_nt_n,t_n) \in G$.
The compactness of G on $\mathbb{R}^q \times \mathbb{R}$ ensures that there exists a sequence of real numbers $(x_{n_k},y_{n_k}) \in G$ such that $(x_{n_k},y_{n_k}) \to (x_0,y_0)$ as $k \to \infty$, where $(x_0,y_0) \in G$. Therefore, $(s_{n_k}t_{n_k},t_{n_k}) \to (x_0,y_0)$ as $k \to \infty$.
Noting that Condition (C) ensures that $\forall s \in \mathbb{R}^q$, $I_x(s,0) \geq I_x(\vec{0},0) > \alpha$ so that $(s,0) \notin G$. It follows that $y_0 \neq 0$, and thus $t_{n_k} \to y_0$ and $s_{n_k} \to s_0$ as $k \to \infty$, where $s_0 = x_0/y_0$. We deduce that $(s_{n_k},t_{n_k}) \to (s_0,y_0)$ as $k \to \infty$, so that $(s_0,y_0) \in \hat{G}$. Thus $\hat{G}$ is a compact set.
Now we claim that the set $A = \{s,\ J(s) \leq \alpha\}$ is the image of $\hat{G}$ by the continuous map $\pi : (s,t) \mapsto s$, and thus it is a compact.
Indeed, clearly $\pi(\hat{G}) \subset A$. For the opposite inclusion, consider $\alpha < \alpha' < I_x\left(\vec{0},0\right)$; the set $\hat{G}' = \left\{\hat{I}_x(s,t) \leq \alpha'\right\}$ is compact. Let $s_0 \in A$, since $J(s_0) \leq \alpha$, we have $J(s_0) = \inf_{(s_0,t)\in\hat{G}'}\hat{I}_x(s_0,t)$; by compacity, there exists t_0 such that $J(s_0) = \hat{I}_x(s_0,t_0)$; $(s_0,t_0) \in \hat{G}$ and $\pi(s_0,t_0) = s_0$, thus $A \subset \pi(\hat{G})$.
- If $\alpha \geq I_x(\vec{0},0)$, let $s \in \mathbb{R}^q$, we have

$$\begin{aligned} J(s) &\leq I_x(st,t) \ \ \forall t \\ &\leq I_x(\vec{0},0) \\ &\leq \alpha. \end{aligned}$$

We deduce that $\mathbb{R}^q \subseteq \{J(s) \leq \alpha\}$ and the second part of Proposition 5 (i) follows.

(ii) It is an obvious consequence of (i) and the definitions of J and J^*.

(iii) Assume that $J^*(s) < \infty$.

- If $I_x(\vec{0},0) > \inf_t I_x(st,t)$, then

$$\inf_t I_x(st,t) \quad = \inf_{t\neq 0} I_x(st,t),$$

  so that $J(s) = J^*(s)$.
- If $I_x(\vec{0},0) = \inf_t I_x(st,t)$, since $J^*(s) < \infty$, there exists $t_0 \neq 0$ such that $I_x(st_0,t_0) < \infty$. By the convexity of I_x, we have for any $\nu \in]0,1]$, $\quad I_x(st_0\nu, t_0\nu) < \infty$ and

$$I_x(st_0\nu, t_0\nu) \leq \nu I_x(st_0,t_0) + (1-\nu) I_x(\vec{0},0).$$

  We deduce that

$$0 \leq I_x(st_0\nu, t_0\nu) - I_x(\vec{0},0) \leq \nu \left(I_x(st_0,t_0) - I_x(\vec{0},0) \right),$$

  and if we take $\nu \to 0$, the third part of Proposition 5 follows.

(iv) Let us suppose that $\alpha < I_x(\vec{0},0)$ and let $s \in \{J^*(s) \leq \alpha\}$, then we have $J^*(s) < \infty$. We deduce from (iii) that $J(s) = J^*(s)$. It follows that $J(s) \leq \alpha$, which ensures that $s \in \{J(s) \leq \alpha\}$. Conversely, if $s \in \{J(s) \leq \alpha\}$, then

$$I_x(\vec{0},0) > \inf_t I_x(st,t),$$

so that

$$\inf_t I_x(st,t) = \inf_{t\neq 0} I_x(st,t).$$

That is $J(s) = J^*(s)$. Therefore, $J^*(s) \leq \alpha$, which ensures that $s \in \{J^*(s) \leq \alpha\}$, and thus Proposition 5 is proved. ■

Proof of Proposition 6

Proposition 6 is proved by following the same approach as for the proof of Proposition 5 with replacing I_x, J and J^* by $\tilde{I}_{a,x}$, $\tilde{J}_a$ and $\tilde{J}_a^*$ respectively. ■

Bibliographie

[1] Ahmad, I.A. and Lin, P. (1976), Nonparametric sequential estimation of a multiple regression function. *Bull. Math. Statist.*, **17**, pp. 63-75.

[2] Bingham, N. H., Goldie, C. M. and Teugels, J. L. (1987), Regular variation. *Cambridge University Press.*

[3] Bosq, D. (1985), Nonparametric statistics for stochastic processes. *Lectures notes in Control and Inform. Sc. Springer.*

[4] Collomb, G. (1984), Proprietés de convergence presque-complète du prédicateur à noyau. *Z. Wahrsch. verw. Gebiete*, **66**, pp. 441-460.

[5] Collomb, G. and Härdle W. (1986), Strong uniform convergence rates in robust nonparametric time series analysis and prediction : kernel regression estimation from dependent observations. *Stoch. Proc. and their Appl.* **23**, pp. 77-89.

[6] Dembo, A. and Zeitouni, O. (1998), Large deviations techniques and applications. Springer, Applications of mathematics, New-York.

[7] Devroye, L. (1979), The uniform convergence of the Nadaraya-Watson regression function estimate. *Can. J. Stat.* **6**, pp. 179-191.

[8] Devroye, L. and Wagner, T.J. (1980), On the L_1 convergence of kernel estimators of regression function with applications in discrimination. *Wahrsch. verw. Gebiete*, **51**, pp. 15-25.

[9] Feller, W. (1970), An introduction to probability theory and its applications, Second edition Volume II, Wiley.

[10] Joutard, C. (2006), Sharp large deviations in nonparametric estimation. *J. of Nonparam. Stat.*, **18**, pp. 293-306.

[11] Louani, D. (1999), Some large deviations limit theorems in conditionnal nonparametric statistics. *Statistics*, **33**, pp. 171-196.

[12] Mack, Y. P. and Silverman, B.W. (1982), Weak and strong uniform consistency of kernel regression estimates. *Z. Wahrsch. verw. Gebiete*, **61**, pp. 405-415.

[13] Mokkadem, A., Pelletier, M. and Thiam, B. (2005), Large and moderate deviations principles for recursive kernel estimators of a multivariate density and its partial derivatives. arXiv :math.ST/0601429.

[14] Mokkadem, A., Pelletier, M. and Worms, J. (2005), Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives. *Austral. J. Statist.*, **4**, pp. 489-502.

[15] Nadaraya, E. A. (1964), On estimating regression. *Theory Probab. Appl.* **10**, pp. 186-190.

[16] Prakasa Rao, B. L. S. (1983), Nonparametric functionnal estimation. *Academic presse New York.*

[17] Rockafellar, R. T. (1970), Convex analysis. Princeton University Press.

[18] Roussas, G. (1992), Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimate. *J. of Nonparam. Statist.* **3**, pp. 171-195.

[19] Senoussi, R. (1991), Loi du log itéré et identification, Thèse. *Université Paris-Sud.*

[20] Watson, G. S. (1964), Smooth regression analysis. *Sankhya Ser. A.* **26**, pp. 359-372.

[21] Worms, J. (2000), Principes de déviations modérées pour des martingales et applications statistiques, Thèse. *Université Marne-la-vallée.*

Chapitre 4

Joint behaviour of the semi-recursive kernel estimator of the mode and the modal value of a probability density function

4.1 Introduction

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed $\mathbb{R}^d$-valued random variables with unknown probability density f. We assume f has a unique mode θ, that is, we suppose there exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for any $x \neq \theta$. Moreover we assume θ is nondegenerate, that is, the second order differential $D^2 f(\theta)$ at the point θ is nonsingular (in the sequel, $D^m g$ will denote the differential of order m of a multivariate function g).

The well known kernel mode estimator has been introduced by Parzen [16]. It is defined as a random variable θ_n^* such that

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

where f_n^* is the Rosenblatt estimator of the density f. Recall that f_n^* is defined as

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

where the bandwidth (h_n) is a sequence of positive real numbers going to zero and where the kernel K is a continuous function satisfying $\lim_{\|x\| \to +\infty} K(x) = 0$, $\int_{\mathbb{R}^d} K(x)dx = 1$.

Abraham et al. ([1], [2]) proposed another estimator $\hat{\theta}_n^*$ of the mode defined by

$$\hat{\theta}_n^* = \underset{1 \leq i \leq n}{\operatorname{argmax}} f_n^*(X_i),$$

and investigated the asymptotic properties of this new estimator ; in particular, they proved an asymptotic proximity theorem between θ_n^* and $\hat{\theta}_n^*$, which implies that θ_n^* and $\hat{\theta}_n^*$ have the same weak asymptotic behaviour.

The asymptotic behaviour of the estimator θ_n^* has been widely studied ; see, among others, Parzen [16], Nadaraya [15], Van Ryzin [23], Rüschendorf [21], Konakov [10], Samanta [22], Eddy ([5], [6]), Romano [18], Vieu [24], and Mokkadem and Pelletier [14].

To estimate $f(\theta)$, the size of the mode, it is usual to set $\mu_n^* = f_n^*(\theta_n^*)$, but, at our knowledge, the asymptotic properties of the estimator μ_n^* have note really been investigated.

The purpose of this paper is the study of the joint convergence rate of kernel estimators of the location and of the size of the mode, and the construction of confidence regions for the couple $(\theta, f(\theta))$. However, the estimators we consider are not θ_n^* and μ_n^*. The reasons why we do not use these estimators are the following.

First, let us mention that, as soon as $D^2 f_n^*$ converges uniformly in a neighbourhood of θ, the asymptotic behaviour of $\left([\theta_n^* - \theta]^T, \mu_n^* - \mu\right)^T$ is given by that of

$$\left(\left[[D^2 f(\theta)]^{-1} \nabla f_n^*(\theta)\right]^T, [f_n^*(\theta) - f(\theta)]\right)^T.$$

Hall [9] shows that, to minimize the coverage error of the confidence interval for a probability density, avoiding bias estimation by a slight undersmoothing is more efficient than explicit bias correction. So, to construct confidence regions for $(\theta, f(\theta))$, it seems more appropriate to use density estimators with small asymptotic variances, rather than density estimators with small mean squared error. The recursive kernel estimator of the density introduced by Wolverton and Wagner [27] (and discussed, among others, by Yamato [28], Davies [3], Devroye [4], Menon et al. [12], Wertz [26], Wegman and Davies [25], Roussas [20], and Mokkadem et al. [13]) being known to have a smaller variance than the Rosenblatt estimator, we construct our estimators θ and $f(\theta)$ by using this recursive density estimator.

More precisely, let the bandwidth (h_n) be a regularly varying sequence with exponent $(-a)$, $a > 0$; we set

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_i^d} K\left(\frac{x - X_i}{h_i}\right) \tag{4.1}$$

and consider the kernel mode estimator defined as a random variable θ_n such that

$$f_n(\theta_n) = \sup_{y\in\mathbb{R}^d} f_n(y). \tag{4.2}$$

Since K is continuous and vanishing at infinity, the choice of θ_n as a random variable satisfying (4.2) can be made with the help of an order on $\mathbb{R}^d$. For example, one can consider the following lexicographic order : $x \le y$ if the first nonzero coordinate of $x - y$ is negative. The definition

$$\theta_n = \inf\left\{ y \in \mathbb{R}^d \text{ such that } f_n(y) = \sup_{x\in\mathbb{R}^d} f_n(x) \right\},$$

where the infimum is taken with respect the lexicographic order on $\mathbb{R}^d$, ensures the mesurability of the kernel mode estimator.

Now, let us explain why the natural estimator

$$\mu_n = f_n(\theta_n) \tag{4.3}$$

of the size of the mode is not the most appropriate.

Similarly to the case when Rosenblatt's kernel density estimator is used, the asymptotic behaviour of $\left([\theta_n - \theta]^T, \mu_n - \mu\right)^T$ is given by that of

$$\left(\left[[D^2 f(\theta)]^{-1}\nabla f_n(\theta)\right]^T, \left[f_n(\theta) - f(\theta)\right]\right)^T,$$

as soon as $D^2 f_n$ converges uniformly in a neighbourhood of θ. However, the condition on the bandwidth (h_n) necessary to ensure the strong uniform consistency of $D^2 f_n$ is

$$\lim_{n\to\infty} \frac{n h_n^{d+4}}{\log n} = \infty. \tag{4.4}$$

Now, the weak convergence rate of $f_n(\theta) - f(\theta)$ to zero is governed by the one of the variance term $f_n(\theta) - \mathbb{E}[f_n(\theta)]$ on the one hand and by the one of the bias term $\mathbb{E}[f_n(\theta)] - f(\theta)$ on the other hand. Let us recall that

$$\begin{aligned}
\sqrt{n h_n^d}\left[f_n(\theta) - \mathbb{E}(f_n(\theta))\right] &\xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{f(\theta)}{1 + ad}\int_{\mathbb{R}^d} K^2(x)dx\right),\\
\lim_{n\to\infty} \frac{1}{h_n^q}\left(\mathbb{E}[f_n(\theta)] - f(\theta)\right) &= B_\mu^q,
\end{aligned}$$

where q is the order of the kernel K and B_μ^q is usually nonzero. So, in the usual case when $q = 2$, the condition (4.4) implies that the variance term $f_n(\theta) - \mathbb{E}[f_n(\theta)]$ is negligeable in front of the bias term, and thus

$$\frac{1}{h_n^q}\left(\mu_n - \mu\right) \xrightarrow{\mathbb{P}} B_\mu^q.$$

This is the first main drawback of the modal value kernel estimator μ_n defined by (4.3) : the use of a higher order kernel K is necessary for $(\mu_n - \mu)$ $\left(\text{and thus } \left((\theta_n - \theta)^T, \mu_n - \mu\right)^T\right)$ to satisfy a central limit theorem (CLT), and thus for the construction of confidence regions of $\left(\theta^T, \mu\right)^T$.

Let us also underline that, when higher order kernels of order q are used for the estimation of θ_n and μ_n, the optimal convergence rate of $\left(\nabla f_n(\theta)\right)$ and thus of $\left(\theta_n - \theta\right)$ is reached when the bandwidth is chosen such that $(h_n) \equiv \left(cn^{-\frac{1}{(d+2+2q)}}\right)$, whereas the optimal convergence rate of $\left(f_n(\theta) - f(\theta)\right)$ and thus of $\left(\mu_n - \mu\right)$ is attained when $(h_n) \equiv \left(cn^{-\frac{1}{(d+2q)}}\right)$. This is the second main drawback of the estimator μ_n defined by (4.3) : it is not possible to choose a bandwidth for which both estimators θ_n and μ_n simultaneously converge at the optimal rate.

These two constatations lead us to introduce a second bandwidth, and thus a second recursive kernel density estimator, to estimate the modal value μ. More precisely, let $(\tilde{h}_n)$ be a regularly varying sequence with exponent $(-\tilde{a})$, $\tilde{a} > 0$, and let

$$\tilde{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\tilde{h}_i^d} K\left(\frac{x - X_i}{\tilde{h}_i}\right)$$

be the kernel density estimator computed with the bandwidth $(\tilde{h}_n)$; we define the modal value kernel estimator $\tilde{\mu}_n$ by setting

$$\tilde{\mu}_n = \tilde{f}_n(\theta_n) \tag{4.5}$$

(where θ_n is the kernel mode estimator computed with the help of the first bandwidth (h_n)). We first prove the strong consistency of θ_n and $\tilde{\mu}_n$, and then study the joint asymptotic behaviour of $\left((\theta_n - \theta)^T, (\tilde{\mu}_n - \mu)\right)^T$. We prove in particular that, whatever the order q of the kernel K is, it is possible to choose two bandwidths (h_n) and $(\tilde{h}_n)$ such that

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\left(\theta_n - \theta\right) \\ \sqrt{n\tilde{h}_n^d}\left(\tilde{\mu}_n - \mu\right) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, f(\theta) \begin{pmatrix} \frac{\left[D^2 f(\theta)\right]^{-1} G \left[D^2 f(\theta)\right]^{-1}}{1+a(d+2)} & 0 \\ 0 & \frac{\int_{\mathbb{R}^d} K^2(x)dx}{1+\tilde{a}d} \end{pmatrix}\right) \tag{4.6}$$

where G is the matrix $d \times d$ defined by $G^{(i,j)} = \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_i}(x) \frac{\partial K}{\partial x_j}(x)dx$.

Of course, in the case the Rosenblatt kernel density estimator is used, we can also use two different bandwidths (h_n) and $(\tilde{h}_n)$ to estimate θ and μ. Under our assumptions and following the same lines of our proof, it can be shown that

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\left(\theta_n^* - \theta\right) \\ \sqrt{n\tilde{h}_n^d}\left(\tilde{\mu}_n^* - \mu\right) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, f(\theta) \begin{pmatrix} \left[D^2 f(\theta)\right]^{-1} G \left[D^2 f(\theta)\right]^{-1} & 0 \\ 0 & \int_{\mathbb{R}^d} K^2(x)dx \end{pmatrix}\right),$$

where $\tilde{\mu}_n^* = \tilde{f}_n^*(\theta_n^*)$. It turns out that the asymptotic variance obtained in the recursive case is smaller than the one obtained when the Rosenblatt kernel estimator is used.
Let us mention that, the estimator of Abraham et al. [1] can be defined with recursive kernel estimation of f ; more precisely we can set $\hat{\theta}_n = \text{argmax}_{1 \le i \le n} f_n(X_i)$ and $\tilde{\hat{\mu}}_n = \tilde{f}_n(\hat{\theta}_n)$; it follows from the work of Abraham et al. [2] that under suitable assumptions, $(\hat{\theta}_n, \tilde{\hat{\mu}}_n)$ and $(\theta_n, \tilde{\mu}_n)$ have the same weak convergence rate.

To complete the study of the asymptotic behaviour of our couple of estimators $\left(\theta_n^T, \tilde{\mu}_n\right)^T$, we establish a compact law of the iterated logarithm. We prove in particular that, for adequate choices of the bandwidths, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix}\sqrt{nh_n^{d+2}}\left(\theta_n-\theta\right)\\ \sqrt{n\tilde{h}_n^d}\left(\tilde{\mu}_n-\mu\right)\end{pmatrix}$$

is relatively compact and its limit set is the ellipsoid $\mathcal{E}$ which is explicitly defined in Section 4.2.

In the case when Rosenblatt kernel estimator is used, such a compact law of the iterated logarithm has been proved in Mokkadem and Pelletier [14] for the mode estimator θ_n^*. However, the technics used in Mokkadem and Pelletier [14] are totally different from the ones employed here. Moreover these lattest do not apply in the case when Rosenblatt kernel estimator is used, and establishing a law of the iterated logarithm for the couple $\left((\theta_n^*-\theta)^T, (\tilde{\mu}_n^*-\mu)\right)^T$ remains an open question.

4.2 Assumptions and Results

$X_1, \ldots, X_n$ are independent and identically distributed random variables with bounded probability density function f. Let h and $\tilde{h}$ be two positive functions, and (h_n) and $(\tilde{h}_n)$ be the sequences defined as $h_n = h(n)$ and $\tilde{h}_n = \tilde{h}(n)$ for all $n \geq 1$. The recursive kernel estimators of f are

$$f_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{h_i^d}K\left(\frac{x-X_i}{h_i}\right) \quad \text{and} \quad \tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^n \frac{1}{\tilde{h}_i^d}K\left(\frac{x-X_i}{\tilde{h}_i}\right).$$

The kernel mode estimator is

$$\theta_n = \inf\left\{y \in \mathbb{R}^d \text{ such that } f_n(y) = \sup_{x\in\mathbb{R}^d} f_n(x)\right\},$$

and the kernel estimator of the size of the mode is

$$\tilde{\mu}_n = \tilde{f}_n(\theta_n).$$

The conditions we require for the strong consistency of θ_n and $\tilde{\mu}_n$ are the following.

(A1) i) K is an integrable differentiable even function and $\int_{\mathbb{R}^d} K(z)dz = 1$.
ii) K is a kernel of order $q \geq 2$ i.e. $\forall s \in \{1, \ldots, q-1\}$, $\forall j \in \{1, \ldots, d\}$, $\int_{\mathbb{R}^d} y_j^s K(y) dy_j = 0$ and $\int_{\mathbb{R}^d} \left|y_j^q K(y)\right| dy < \infty$.
iii) K is Hölder continuous.
iv) There exists $\gamma > 0$ such that $z \mapsto \|z\|^\gamma |K(z)|$ is a bounded function.

(A2) i) f is uniformly continuous on $\mathbb{R}^d$, q-times differentiable and $\sup_{x\in\mathbb{R}^d}\|D^q f(x)\| < \infty$, where q is the order of K.
ii) There exists $\xi > 0$ such that $\int_{\mathbb{R}^d} \|x\|^\xi f(x) dx < \infty$.
iii) There exists $\eta > 0$ such that $z \mapsto \|z\|^\eta f(z)$ is a bounded function.
iv) There exists $\theta \in \mathbb{R}^d$ such that $f(x) < f(\theta)$ for all $x \neq \theta$.

(A3) The functions h and $\tilde{h}$ are locally bounded and vary regularly with exponent $(-a)$ and $(-\tilde{a})$ respectively, where $a \in]0, 1/(d+4)[$, $\tilde{a} \in]0, 1/(d+2)[$.

Remark 7 *Note that (A1)iv) implies that K is bounded.*

Remark 8 *A positive function (not necessarily monotone) $\mathcal{L}$ defined on $]0,\infty[$ is slowly varying if $\lim_{t\to\infty}\mathcal{L}(tx)/\mathcal{L}(t)=1$ and a function G is said to vary regularly with exponent ρ, $\rho\in\mathbb{R}$, if and only if it is of the form $G(x)=x^\rho\mathcal{L}(x)$ with $\mathcal{L}$ slowly varying (see, for example, Feller [8] page 275). Typical examples of regularly varying functions are x^ρ, $x^\rho\log x$, $x^\rho\log\log x$, $x^\rho\log x/\log\log x$, and so on.*

Remark 9 *An important consequence of (A3) which will be used in the sequel is :*

$$\text{if } \beta a<1, \quad \text{then} \quad \lim_{n\to\infty}\frac{1}{nh_n^\beta}\sum_{i=1}^n h_i^\beta \;=\; \frac{1}{1-a\beta}. \tag{4.7}$$

Moreover, observe that for all $\varepsilon>0$ small enough,

$$\frac{1}{n}\sum_{i=1}^n h_i^q = O\left(h_n^{q-\varepsilon}+\frac{1}{n}\right). \tag{4.8}$$

As a matter of fact :

- If $aq<1$, (4.8) follows easily from (4.7).
- If $aq>1$, since $\sum_i h_i^q$ is summable, (4.8) holds.
- If $aq=1$, since $a(q-\varepsilon)<1$, using (4.7) again, we have $\frac{1}{n}\sum_{i=1}^n h_i^q = O\left(h_n^{q-\varepsilon}\right)$, and thus (4.8) follows.

Of course (4.7) and (4.8) also hold when (h_n) and a are replaced by $(\tilde{h}_n)$ and $\tilde{a}$.

Proposition 9 *Let Assumptions (A1)-(A3) hold, then*

$$\lim_{n\to\infty}\theta_n=\theta \quad a.s. \quad \text{and} \quad \lim_{n\to\infty}\tilde{\mu}_n=\mu \quad a.s.$$

From now on, we set

$$h_n=n^{-a}\mathcal{L}_\theta(n) \quad \text{and} \quad \tilde{h}_n=n^{-\tilde{a}}\mathcal{L}_\mu(n),$$

where $\mathcal{L}_\theta$ and $\mathcal{L}_\mu$ are positive slowly varying functions (the existence of such functions is ensured by (A3), see Remark 8).

Moreover, we introduce the following notations :

$$B_q(\theta)=\begin{pmatrix} \frac{(-1)^q}{q!(1-aq)}\nabla\left(\sum_{j=1}^d\beta_j^q\frac{\partial^q f}{\partial x_j^q}(\theta)\right) \\ \frac{(-1)^q}{q!(1-\tilde{a}q)}\sum_{j=1}^d\beta_j^q\frac{\partial^q f}{\partial x_j^q}(\theta)\end{pmatrix} \quad \text{with} \quad \beta_j^q=\int_{\mathbb{R}^d}y_j^qK(y)dy, \quad aq\neq 1 \quad \text{and} \quad \tilde{a}q\neq 1, \tag{4.9}$$

$$A=\begin{pmatrix} -\left[D^2f(\theta)\right]^{-1} & 0 \\ 0 & 1\end{pmatrix}, \quad \Sigma=\begin{pmatrix}\frac{f(\theta)G}{1+a(d+2)} & 0 \\ 0 & \frac{f(\theta)\int_{\mathbb{R}^d}K^2(z)dz}{1+\tilde{a}d}\end{pmatrix}, \tag{4.10}$$

and, for any $c,\tilde{c}\geq 0$, we set $D(c,\tilde{c})=\begin{pmatrix}\sqrt{c}I_d & 0\\ 0 & \sqrt{\tilde{c}}\end{pmatrix}$ where I_d is the $d\times d$ identity matrix.

To state the weak convergence rate of $(\theta_n^T,\tilde{\mu}_n)^T$, we need to introduce the following conditions.

(C1) Either

i) $\left(\frac{1}{d+4} < \tilde{a} < \frac{q}{d+2q+2} \text{ and } \frac{\tilde{a}}{q} < a < \frac{1-2\tilde{a}}{d+2}\right)$ or

ii) $\left(\frac{1}{d+2q} < \tilde{a} \leq \frac{1}{d+4} \text{ and } \frac{1}{d+2q+2} < a < \frac{1+\tilde{a}d}{2\,(d+2)}\right)$ or

iii) $\tilde{a} = \frac{1}{d+2q}$, $a = \frac{1}{d+2q+2}$, $\lim_{n\to\infty} \mathcal{L}_\mu(n) = \tilde{\gamma} < \infty$ and $\lim_{n\to\infty} \mathcal{L}_\theta(n) = \gamma < \infty$.

(C2) Either

i) $\left(0 < \tilde{a} < \frac{1}{d+2q} \text{ and } \frac{\tilde{a}}{2} < a < \frac{1}{d+2q+2}\right)$ or

ii) $\tilde{a} = \frac{1}{d+2q}$, $\lim_{n\to\infty} \mathcal{L}_\mu(n) = \infty$ and $\frac{1}{2(d+2q)} < a < \frac{1}{d+2q+2}$.

(A4) i) K is twice differentiable on $\mathbb{R}^d$.
ii) $z \mapsto z\nabla K(z)$ is integrable.
iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\partial^2 K/\partial x_i \partial x_j$ is bounded integrable and Hölder continuous.

(A5) i) $D^2 f(\theta)$ is nonsingular.
ii) $D^2 f$ is q-times differentiable and ∇f is bounded.
iii) For any $(i,j) \in \{1,\ldots,d\}^2$, $\sup_{x\in\mathbb{R}^d} \|D^q\left(\partial^2 f/\partial x_i \partial x_j\right)\| < \infty$, and for any $k \in \{1,\ldots,d\}$, $\sup_{x\in\mathbb{R}^d} \|D^q\left(\partial f/\partial x_k\right)\| < \infty$.

Remark 10 *Note that (A4)ii) and (A4)iii) imply that ∇K is Lipschitz-continuous and integrable ; it is thus straightforward to see that $\lim_{\|x\|\to\infty} \|\nabla K(x)\| = 0$ (and in particular ∇K is bounded).*

Theorem 10 *Assume (A1)-(A5) hold.*

i) When (C1)i) or (C1)ii) is satisfied, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, A\Sigma A\right).$$

ii) When (C1)iii) is satisfied, set $c = \gamma^{d+2q+2}$ and $\tilde{c} = \tilde{\gamma}^{d+2q}$; then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(D(c,\tilde{c})AB_q(\theta), A\Sigma A\right).$$

iii) When (C2) is satisfied, then

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n - \theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathbb{P}} AB_q(\theta).$$

Comments. Part 1 (respectively Part 3) of Theorem 10 corresponds to the case when the bias (respectively the variances) of both estimators θ_n and $\tilde{\mu}_n$ are negligeable in front of their respective variances (respectively bias). When $\gamma, \tilde{\gamma} > 0$, Part 2 of Theorem 10 corresponds to the case when the bias and the variance of each estimator θ_n and $\tilde{\mu}_n$ have the same convergence rate. Other possible conditions lead to different combinations ; these ones have been omitted for sake of simplicity.

Both estimators θ_n and $\tilde{\mu}_n$ converge simultaneously with their optimal rate when the bandwidths are chosen such that $\lim_{n\to\infty} nh_n^{d+2q+2} = \gamma \in]0,\infty[$ and $\lim_{n\to\infty} n\tilde{h}_n^{d+2q} = \tilde{\gamma} \in]0,\infty[$ (which leads to Part 2 of Theorem 10). This is the first advantage of using the estimator $\tilde{\mu}_n$ defined in

(4.5) rather than the estimator μ_n defined in (4.3). As a matter of fact, when this lattest (which corresponds to the case $(\tilde{h}_n) = (h_n)$) is used, it is not possible to make the estimators of the location and of the size of the mode converge with their optimal rate simultaneously.
For the construction of confidence regions for the couple of parameters (θ, μ), Part 1 or Part 2 of Theorem 10 must be applied. Let us underline that condition (C1)ii) is empty in the case $q = 2$, and that the choice $a = \tilde{a}$ (and a fortiori, the choice $(h_n) = (\tilde{h}_n)$) is forbidden by conditions (C1)i) and (C1)iii). This enlightens the second advantage of using $\tilde{\mu}_n$ instead of μ_n : Kernels of order 2 can be used to construct confidence regions for (θ, μ) only when different bandwidths are used for the estimations of θ and of μ.

A result similar to Theorem 10 can be obtained when the nonrecursive Rosenblatt kernel estimator is used instead of the recursive estimator f_n. Recall that θ_n^* (respectively $\tilde{\mu}_n^*$) denotes the estimator of the location (respectively of the size) of the mode computed with the help of the bandwidth h_n (respectively $\tilde{h}_n$) and the associated Rosenblatt kernel density estimator f_n^* (respectively $\tilde{f}_n^*$). Moreover, set

$$\Sigma^* = \begin{pmatrix} f(\theta)G & 0 \\ 0 & f(\theta)\int_{\mathbb{R}^d} K^2(z)dz \end{pmatrix} \quad \text{and} \quad B_q^*(\theta) = \begin{pmatrix} \frac{(-1)^q}{q!}\nabla\left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\right) \\ \frac{(-1)^q}{q!}\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix}.$$

Following the same lines of our proofs, it can be shown that θ_n, $\tilde{\mu}_n$, Σ, and $B_q(\theta)$ can be replaced by θ_n^*, $\tilde{\mu}_n^*$, Σ^*, and $B_q^*(\theta)$ respectively in Theorem 10. Note that

$$\Sigma = \begin{pmatrix} [1 + a(d+2)]^{-1} I_d & 0 \\ 0 & [1 + \tilde{a}d]^{-1} \end{pmatrix} \Sigma^*.$$

To construct confidence regions for (θ, μ) by slightly undersmoothing, it is thus better to use the estimators θ_n and $\tilde{\mu}_n$ computed with the help of the recursive kernel density estimator f_n rather than the ones defined via Rosenblatt's estimator. Simulations results comparing the empirical levels obtained when confidence regions are constructed by using either the recursive estimator f_n or Rosenblatt's estimator are given in Section 4.3.

To complete the study of the asymptotic behaviour of $(\theta_n^T, \tilde{\mu}_n)^T$, we state now its almost sure convergence rate. To this end, we need the following additionnal assumptions.

(A6) i) h and $\tilde{h}$ are differentiables, their derivatives vary regularly with exponent $(-a-1)$ and $(-\tilde{a}-1)$ respectively.
ii) There exists $n_0 \in \mathbb{N}$ such that

$$n \geq m \geq n_0 \quad \Rightarrow \quad \max\left\{\frac{mh_m^{-(d+2)}}{nh_n^{-(d+2)}}; \frac{m\tilde{h}_m^{-d}}{n\tilde{h}_n^{-d}}\right\} \quad = \quad \frac{\min\left\{mh_m^{-(d+2)}; m\tilde{h}_m^{-d}\right\}}{\min\left\{nh_n^{-(d+2)}; n\tilde{h}_n^{-d}\right\}}.$$

Remark 11 *Assumption (A6)ii) holds when $a \neq \tilde{a}$, and in the case $a = \tilde{a}$, it is satisfied when $\mathcal{L}_\theta(n) = (\mathcal{L}_\mu(n))^{\frac{d}{d+2}}$ for n large enough.*

Moreover, conditions (C1) and (C2) are replaced by the following ones.

(C'1) Either

i) $\left(\frac{1}{d+4} < \tilde{a} < \frac{q}{d+2q+2} \text{ and } \frac{\tilde{a}}{q} < a < \frac{1-2\tilde{a}}{d+2}\right)$ or

ii) $\left(\frac{1}{d+2q} < \tilde{a} \leq \frac{1}{d+4} \text{ and } \frac{1}{d+2q+2} < a < \frac{1+\tilde{a}d}{2(d+2)}\right)$ or

iii) $\tilde{a} = \dfrac{1}{d+2q}$, $a = \dfrac{1}{d+2q+2}$, $\lim_{n\to\infty} \dfrac{(\mathcal{L}_\mu(n))^{d+2q}}{2\log\log n} = \tilde{\alpha} < \infty$ and $\lim_{n\to\infty} \dfrac{(\mathcal{L}_\theta(n))^{d+2q+2}}{2\log\log n} = \alpha < \infty$.

(C'2) Either

i) $\left(0 < \tilde{a} < \dfrac{1}{d+2q} \text{ and } \dfrac{\tilde{a}}{2} < a < \dfrac{1}{d+2q+2}\right)$ or

ii) $\tilde{a} = \dfrac{1}{d+2q}$, $\lim_{n\to\infty} \dfrac{(\mathcal{L}_\mu(n))^{d+2q}}{2\log\log n} = \infty$ and $\dfrac{1}{2(d+2q)} < a < \dfrac{1}{d+2q+2}$.

Before stating the almost sure convergence rate of $(\theta_n^T, \tilde{\mu}_n)^T$, let us remark that Proposition 2.3 in Mokkadem and Pelletier [14] ensures that the matrix G (and thus the matrix Σ) is nonsingular.

Theorem 11 *Let Assumptions (A1)-(A6) hold.*

i) When (C'1)i) or (C'1)ii) is fulfilled, then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix}$$

is relatively compact and its limit set is the ellipsoid

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T A^{-1}\Sigma^{-1}A^{-1}\nu \leq 1\right\}.$$

ii) When (C'1)iii) is fulfilled, set $c = \alpha$ and $\tilde{c} = \tilde{\alpha}$; then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}} \begin{pmatrix} \sqrt{nh_n^{d+2}}(\theta_n - \theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{\mu}_n - \mu) \end{pmatrix}$$

is relatively compact and its limit set is the ellipsoid

$$\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ such that } \left(A^{-1}\nu - D(c,\tilde{c})\, B_q(\theta)\right)^T \Sigma^{-1} \left(A^{-1}\nu - D(c,\tilde{c})\, B_q(\theta)\right) \leq 1\right\}.$$

iii) When (C'2) is satisfied, then

$$\begin{pmatrix} \frac{1}{h_n^q}(\theta_n - \theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{a.s.} AB_q(\theta).$$

Comments. When (C'1) is satisfied, then

$$\lim_{n\to\infty} \frac{nh_n^{d+2q+2}}{2\log\log n} = \alpha \quad\text{and}\quad \lim_{n\to\infty} \frac{n\tilde{h}_n^{d+2q}}{2\log\log n} = \tilde{\alpha},$$

with $\alpha = \tilde{\alpha} = 0$ in the framework of Part 1 of Theorem 11, and $\alpha, \tilde{\alpha} \in [0, \infty[$ in the framework of Part 2 of Theorem 11. In the case (C'2) is satisfied,

$$\lim_{n\to\infty} \frac{nh_n^{d+2q+2}}{2\log\log n} = \infty \quad\text{and}\quad \lim_{n\to\infty} \frac{n\tilde{h}_n^{d+2q}}{2\log\log n} = \infty.$$

Let us mention that, on the opposite to the weak convergence rate, the almost sure convergence rate of the couple of estimators of θ and μ computed with the help of the nonrecursive Rosenblatt kernel estimator cannot be established by following the lines of our proof made in the case the recursive estimator f_n is used. This is due to the fact that the recursive estimator f_n is, up to a factor, a sum of independent variables, whereas Rosenblatt estimator f_n^* equals, up to a factor, a triangular sum of independent variables.

4.3 Simulation study of confidence region

The aim of our simulation study is double. On the one hand, we provide a comparison between the case $(h_n) = (\tilde{h}_n)$ and the case $(h_n) \neq (\tilde{h}_n)$. On the other hand, we make a comparison between the use of the recursive density estimator and that of Rosenblatt density estimator.

We consider the unidimensional framework ($d = 1$), and, since we construct our confidence regions for (θ, μ) by undersmoothing, we choose bandwidths (h_n) and $(\tilde{h}_n)$ satisfying (C1) with $\gamma = \tilde{\gamma} = 0$.

Let us first consider the case when the recursive estimator f_n is used to compute the estimators of θ and μ. In order to construct our asymptotic confidence regions, it is necessary to estimate Σ consistently. This can be done by replacing $f''(\theta)$ and $f(\theta)$ by $f_n''(\theta_n)$ and $\tilde{f}_n(\theta_n)$ respectively. In this case, we have,

$$\begin{pmatrix} \frac{\sqrt{1+3a}\sqrt{nh_n^3}|f_n''(\theta_n)|}{\sqrt{\left|\tilde{f}_n(\theta_n)\right| \int_{\mathbb{R}} K'^2(x)dx}}(\theta_n - \theta) \\ \frac{\sqrt{1+\tilde{a}}\sqrt{n\tilde{h}_n}}{\sqrt{\left|\tilde{f}_n(\theta_n)\right| \int_{\mathbb{R}} K^2(x)dx}}(\tilde{\mu}_n - \mu) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right).$$

Now, set

$$P_n = \frac{\sqrt{1+3a}\sqrt{nh_n^3}\,|f_n''(\theta_n)|}{\sqrt{\left|\tilde{f}_n(\theta_n)\right| \int_{\mathbb{R}} K'^2(x)dx}} \quad \text{and} \quad Q_n = \frac{\sqrt{1+\tilde{a}}\sqrt{n\tilde{h}_n}}{\sqrt{\left|\tilde{f}_n(\theta_n)\right| \int_{\mathbb{R}} K^2(x)dx}}.$$

The application of Part 1 or 2 of Theorem 10 ensures that

$$P_n^2(\theta_n - \theta)^2 + Q_n^2(\tilde{\mu}_n - \mu)^2 \xrightarrow{\mathcal{D}} \chi^2(2)$$

where $\chi^2(2)$ is the chi-squared distribution with 2 degrees of freedom. It follows that the confidence region $\mathcal{E}_\alpha$ such that

$$\mathbb{P}\left(\begin{pmatrix} \theta_n - \theta \\ \tilde{\mu}_n - \mu \end{pmatrix} \in \mathcal{E}_\alpha\right) = 1 - \alpha$$

is the ellipsoid $\mathcal{E}_\alpha = \left\{u \in \mathbb{R}^2, \quad u^t \Sigma_n u \leq t_\alpha\right\}$ where

$$\Sigma_n = \begin{pmatrix} P_n^2 & 0 \\ 0 & Q_n^2 \end{pmatrix}$$

and t_α is the quantile of the chi-squared distribution corresponding to the asymptotic level $1-\alpha$. When Rosenblatt density estimator is used, in order to construct confidence band with asymptotic level $1-\alpha$, P_n and Q_n above are replaced by P_n' and Q_n' defined as :

$$P_n' = \frac{\sqrt{nh_n^3}\,\left|f_n^{*\prime\prime}(\theta_n^*)\right|}{\sqrt{\left|\tilde{f}_n^*(\theta_n^*)\right| \int_{\mathbb{R}} K'^2(x)dx}} \quad \text{and} \quad Q_n' = \frac{\sqrt{n\tilde{h}_n}}{\sqrt{\left|\tilde{f}_n^*(\theta_n^*)\right| \int_{\mathbb{R}} K^2(x)dx}}.$$

In the theoretical framework, let us mention that the axes of the ellipsoid in the recursive case are smaller than those for the Rosenblatt case. More precisely, let E_r and F_r be respectively the

θ-axis and the μ-axis of the ellipsoid in the recursive case, E_s and F_s be respectively the axes in the Rosenblatt case, then

$$\begin{aligned} E_r &= \frac{1}{\sqrt{1+3a}} E_s \\ Fr &= \frac{1}{\sqrt{1+\tilde{a}}} F_s. \end{aligned}$$

We consider five sample sizes $n = 25$, 50, 100, 200, and 300, and four densities f : standard normal, normal mixture, standard Cauchy, and Student with 2 degrees of freedom. In each case the number of simulations is $N = 500$. We set $1 - \alpha = 0.95$ (and thus $t_\alpha = 5.99$). In order to allow the choice $(h_n) = (\tilde{h}_n)$, we use a four order kernel $(q = 4)$: $K(x) = \frac{1}{\sqrt{8\pi}} \left(3 - x^2\right) e^{-\frac{x^2}{2}}$.

For $(h_n, \tilde{h}_n)$, we have considered the four following choices, which satisfy (C1) with $\gamma = \tilde{\gamma} = 0$.

A. $h_n = n^{-1/10}$ and $\tilde{h}_n = n^{-1/8}$.

B. $h_n = \dfrac{n^{-1/11}}{(\log n)^{1/3}}$ and $\tilde{h}_n = \dfrac{n^{-1/9}}{\log n}$.

C. $h_n = \tilde{h}_n = n^{-1/6}$.

D. $h_n = \tilde{h}_n = n^{-1/8}$.

The tables give the proportion of simulations for which $(\theta, f(\theta))$ belongs to the confidence ellipsoid. For each case in the table, the first line corresponds to the Rosenblatt kernel estimator and the second one to the recursive estimator. The choice B corresponds to the best choice of a and $\tilde{a}$.

For each density, we have a table giving the percentage of times when $(\theta, f(\theta))$ is in the confidence region for the four choices A, B, C and D.

The tables confirm our theoretical results : the choices A and B seem better than the choices C and D. The best choice of a and $\tilde{a}$ is B (that is when (C1)iii) is satisfied). It seems also that the estimator $(\theta_n, \tilde{\mu}_n)$ performs better than $(\theta_n^*, \tilde{\mu}_n^*)$.

Estimation of normal standard distribution

n	25	50	100	200	300
A	97.8%	96.4%	96.4%	95.4%	94.8%
	98.2%	97.4%	96.8%	97%	95.8%
B	98%	98%	96.6%	95.4%	95%
	99.6%	99%	98.6%	97.8%	97.2%
C	86.4%	82%	83.6%	81.8%	77.8%
	89.8%	85.6%	88.4%	87.6%	84%
D	87.8%	90.8%	89.8%	92%	91%
	89.4%	92.2%	92.4%	93.8%	93.2%

Estimation of normal mixture $1/2\mathcal{N}(0,1) + 1/2\mathcal{N}(1,1)$

n	25	50	100	200	300
A	98.8%	98.8%	96.2 %	96.2%	94.6%
	99%	99.6%	97.2%	97%	96.2%
B	97.2%	97.2%	96.6%	96.4%	96.8%
	98.4%	99.2%	98.8%	97.8%	98.8%

C	85.4%	84.6%	78.4%	78.4%	76.6%
	87%	88%	83.4%	81.6%	80%
D	88.6%	85.8%	86.6%	86.4%	85.2%
	90.6%	88.2%	89.4%	90.2%	89.8%

Estimation of standard Cauchy

n	25	50	100	200	300
A	98.4%	98.4%	98.6%	99.2%	98%
	98.6%	98.4 %	98.6%	98.8%	98%
B	96.2%	97.4%	97.6%	98%	98.6%
	97%	98.6%	98.4%	99.4%	99.4%

C	84.8%	87.6%	88.8%	84.8%	90.6%
	87.4%	90%	91.2%	89.2%	92.8%
D	89.2%	89%	92.4%	92.2%	96.2%
	89.8%	89.8%	93.2%	93%	96.8%

Estimation of Student with 2 degrees of freedom

n	25	50	100	200	300
A	98.4%	95.8%	96.4%	97%	96.8%
	99.2 %	96.6%	96.6 %	97.4%	96.8%
B	97%	97.8%	98.8%	98.6%	97.8%
	97.6%	98.2%	99.8%	99.6%	99.6%

C	83.4%	87.4%	85.8%	86.2%	85.6%
	86.8%	90.6%	87.2%	89.8%	89.8%
D	89.6%	91.4%	91.8%	92.6%	93.6%
	92.8%	93.8%	94.2%	95.2%	95.4%

4.4 Proofs

Our proofs are organized as follows. Section 4.4.1 is devoted to the proof of the strong consistency of θ_n and $\tilde{\mu}_n$. In Section 4.4.2, we give the convergence rate of the derivatives of f_n. In Section 4.4.3, we show how the study of the weak and strong convergence rate of $\left((\theta_n-\theta)^T, \tilde{\mu}_n-\mu\right)^T$ can be related to the one of $\Big([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta)-f(\theta)\Big)^T$. In Section 4.4.4 (respectively in Section 4.4.5), we establish the weak convergence rate (respectively the strong convergence rate) of $\Big([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta)-f(\theta)\Big)^T$. Finally, Section 4.4.6 is devoted to the proof of Theorems 1 and 2.

4.4.1 Proof of Proposition 9

Since θ_n is the mode of f_n and θ the mode of f, we have :

$$\begin{aligned}
0 &\leq f(\theta) - f(\theta_n) \\
&\leq [f(\theta) - f(\theta_n)] + [f_n(\theta_n) - f_n(\theta)] \\
&\leq \left|f(\theta) - f_n(\theta)\right| + \left|f_n(\theta_n) - f(\theta_n)\right| \\
&\leq 2\|f_n - f\|_\infty. \qquad (4.11)
\end{aligned}$$

The application of Theorem 5 in Mokkadem et al. [13] with $|\alpha| = 0$ and $v_n = \log n$ ensures that for any $\delta > 0$, there exists $c(\delta) > 0$ such that

$$\mathbb{P}\left[(\log n)\|f_n - f\|_\infty \geq \delta\right] \leq \exp\left(-c(\delta)\frac{\sum_{i=1}^n h_i^d}{(\log n)^2}\right).$$

In view of (4.7), since $ad < 1$, we can write

$$\begin{aligned}
n^2 \exp\left(-c(\delta)\frac{\sum_{i=1}^n h_i^d}{(\log n)^2}\right) &= n^2 \exp\left(-c(\delta)\frac{nh_n^d}{(\log n)^2}\frac{\sum_{i=1}^n h_i^d}{nh_n^d}\right) \\
&= o(1).
\end{aligned}$$

Borell-Cantelli's Lemma ensures that

$$\lim_{n\to\infty} \|f_n - f\|_\infty = 0 \text{ a.s.}$$

and thus by (4.11),

$$\lim_{n\to\infty} f(\theta_n) = f(\theta) \text{ a.s.}$$

Since f is continuous, $\lim_{\|z\|\to+\infty} f(z) = 0$ and θ is the unique mode of f, it follows that $\lim_{n\to\infty} \theta_n = \theta$ a.s.
Now,

$$\begin{aligned}
|\tilde{\mu}_n - \mu| &\leq |\tilde{f}_n(\theta_n) - f(\theta_n)| + |f(\theta_n) - f(\theta)| \\
&\leq \|\tilde{f}_n - f\|_\infty + 2\|f_n - f\|_\infty
\end{aligned}$$

where the last inequality follows from (4.11) ; As previously, one can show that $\lim_{n\to\infty} \|\tilde{f}_n - f\|_\infty = 0$ and thus $\lim_{n\to\infty} \tilde{\mu}_n = \mu$ a.s. ■

4.4.2 Convergence rate of the derivatives of the density

For any d-uplet $[\alpha] = \left(\alpha_1, \dots, \alpha_d\right) \in \mathbb{N}^d$, we set $|\alpha| = \alpha_1 + \cdots + \alpha_d$ and, for any function g, let

$$\partial^{[\alpha]} g(x) = \frac{\partial^{|\alpha|} g}{\partial x_1^{\alpha_1} \dots \partial x_d^{\alpha_d}}(x)$$

denote the $[\alpha]$-th partial derivative of g. The following Lemma is proved in Mokkadem et al. [13].

Lemma 13 *Assume (A3)-(A5) hold. Let (g_n) and (b_n) be defined as follows :*

$$\left\{\begin{array}{l} g_n = f_n \text{ and } b_n = h_n \text{ or} \\ g_n = \tilde{f}_n \text{ and } b_n = \tilde{h}_n. \end{array}\right. \qquad (4.12)$$

For $|\alpha| \in \{0, 1, 2\}$, we have *i)*

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^n b_i^q}\left[\mathbb{E}\left[\partial^{[\alpha]} g_n(x)\right] - \partial^{[\alpha]} f(x)\right] \quad = \quad \frac{(-1)^q}{q!}\partial^{[\alpha]}\left(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}\right)(x)$$

where β_j^q is defined in (4.9).

ii) Moreover, if we set $M_q = \sup_{x\in\mathbb{R}^d} \|D^q \partial^{[\alpha]} f(x)\|$, then

$$\lim_{n\to\infty} \frac{n}{\sum_{i=1}^n b_i^q} \sup_{x\in\mathbb{R}^d}\left|\mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right) - \partial^{[\alpha]} f(x)\right| \quad \leq \quad \frac{M_q}{q!}\int_{\mathbb{R}^d} \|z\|^q \, |K(z)| \, dz.$$

Lemma 14 *Let U be a compact set of $\mathbb{R}^d$ and assume that (A1)iii), (A3), (A4) and (A5)ii) hold. Let (g_n) and (b_n) be defined as in* (4.12)*. Then, for all $\gamma > 0$ and $|\alpha| = 1, 2$, we have*

$$\sup_{x\in U}\left|\partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right)\right| \quad = \quad O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^n b_i^{d+2|\alpha|}}}\right) \qquad a.s.$$

Proof of Lemma 14

Set

$$v_n \quad = \quad \sqrt{\frac{\sum_{i=1}^n b_i^{d+2|\alpha|}}{(\log n)^{1+\gamma}}}.$$

Applying Proposition 3 in Mokkadem et al. [13], it holds that for any $\delta > 0$, there exists $c(\delta) > 0$ such that

$$\mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right)\right| \geq \delta\right] \quad \leq \quad \exp\left(-c(\delta)\frac{\sum_{i=1}^n b_i^{d+2|\alpha|}}{2v_n^2}\right).$$

Since $\lim_{n\to\infty} \sum_{i=1}^n b_i^{d+2|\alpha|}/(v_n^2 \log n) = \infty$ we have, for n large enough,

$$c(\delta)\frac{\sum_{i=1}^n b_i^{d+2|\alpha|}}{2v_n^2} \quad \geq \quad 2\log n.$$

We deduce that

$$\sum_n \mathbb{P}\left[\sup_{x\in U} v_n \left|\partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right)\right| \geq \delta\right] \quad = \quad O\left(\sum_n \exp\left(-2\log n\right)\right) \quad < \quad \infty.$$

Borel-Cantelli's Lemma then gives that

$$\lim_{n\to\infty} \sup_{x\in U} v_n \left|\partial^{[\alpha]} g_n(x) - \mathbb{E}\left(\partial^{[\alpha]} g_n(x)\right)\right| \quad = \quad 0 \quad \text{a.s.}$$

and thus Lemma 14 follows. ■

4.4.3 Relationship between $\left((\theta_n - \theta)^T, (\tilde{\mu}_n - \mu)\right)^T$ and $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$

By definition of θ_n, we have $\nabla f_n(\theta_n) = 0$ so that

$$\nabla f_n(\theta_n) - \nabla f_n(\theta) = -\nabla f_n(\theta). \tag{4.13}$$

For each $i \in \{1, \ldots, d\}$, a Taylor expansion applied to the real valued application $\frac{\partial f_n}{\partial x_i}$ implies the existence of $\varepsilon_n(i) = \left(\varepsilon_n^{(1)}(i), \ldots, \varepsilon_n^{(d)}(i)\right)^t$ such that

$$\begin{cases} \frac{\partial f_n}{\partial x_i}(\theta_n) - \frac{\partial f_n}{\partial x_i}(\theta) = \sum_{j=1}^{d} \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\varepsilon_n(i))\left(\theta_n^{(j)} - \theta^{(j)}\right), \\ \left|\varepsilon_n^{(j)}(i) - \theta^{(j)}\right| \leq \left|\theta_n^{(j)}(i) - \theta^{(j)}\right| \quad \forall j \in \{1, \ldots, d\}. \end{cases}$$

Define the $d \times d$ matrix $H_n = \left(H_n^{(i,j)}\right)_{1 \leq i,j \leq d}$ by setting

$$H_n^{(i,j)} = \frac{\partial^2 f_n}{\partial x_i \partial x_j}(\varepsilon_n(i));$$

Equation (4.13) can be then rewritten as

$$H_n(\theta_n - \theta) = -\nabla f_n(\theta).$$

On the other hand, we have

$$\begin{aligned} \tilde{\mu}_n - \mu &= \tilde{f}_n(\theta_n) - f(\theta) \\ &= \tilde{f}_n(\theta) - f(\theta) + R_n \end{aligned}$$

with

$$R_n = \tilde{f}_n(\theta_n) - \tilde{f}_n(\theta). \tag{4.14}$$

We can then write :

$$\begin{pmatrix} \left[D^2 f(\theta)\right]^{-1} H_n(\theta_n - \theta) \\ \tilde{\mu}_n - \mu \end{pmatrix} = \begin{pmatrix} \left[D^2 f(\theta)\right]^{-1} \nabla f_n(\theta) \\ \tilde{f}_n(\theta) - f(\theta) \end{pmatrix} + \begin{pmatrix} 0 \\ R_n \end{pmatrix}. \tag{4.15}$$

Let U be a compact set of $\mathbb{R}^d$ containing θ. The combination of Lemma 13 and Lemma 14 with $|\alpha| = 2$, $g_n = f_n$ and $b_n = h_n$ ensures that for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\begin{aligned} \sup_{x \in U} \left|\partial^{[\alpha]} f_n(x) - \partial^{[\alpha]} f(x)\right| &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{\sum_{i=1}^{n} h_i^{d+4}}} + \frac{\sum_{i=1}^{n} h_i^q}{n}\right) \quad \text{a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n h_n^{d+4}}} + h_n^{q-\varepsilon} + \frac{1}{n}\right) \quad \text{a.s.} \\ &= o(1) \quad \text{a.s.} \end{aligned} \tag{4.16}$$

Since $D^2 f$ is continuous in a neighbourhood of θ and since $\lim_{n\to\infty} \theta_n = \theta$ a.s., (4.16) ensures that $\lim_{n\to\infty} H_n = [D^2 f(\theta)]$ a.s. It follows that the weak and a.s. behaviours of $\left((\theta_n - \theta)^T, (\tilde{\mu}_n - \mu)\right)^T$ are given by the one of the right-hand-sided term of (4.15).

4.4.4 Weak convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$

Let us at first assume that the following lemma holds.

Lemma 15 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i) and (A4)ii) hold. Then*

$$W_n = \begin{pmatrix} \sqrt{nh_n^{d+2}}\Big[\nabla f_n(\theta) - \mathbb{E}\big(\nabla f_n(\theta)\big)\Big] \\ \sqrt{n\tilde{h}_n^d}\Big[\tilde{f}_n(\theta) - \mathbb{E}\big(\tilde{f}_n(\theta)\big)\Big] \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\Big(0, \Sigma\Big).$$

The application of Lemma 13 gives

$$\lim_{n\to\infty} \begin{pmatrix} \frac{n}{\sum_{i=1}^n h_i^q}\mathbb{E}\big(\nabla f_n(\theta)\big) \\ \frac{n}{\sum_{i=1}^n \tilde{h}_i^q}\Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta)\Big] \end{pmatrix} = \begin{pmatrix} \frac{(-1)^q}{q!}\nabla\Big(\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta)\Big) \\ \frac{(-1)^q}{q!}\sum_{j=1}^d \beta_j^q \frac{\partial^q f}{\partial x_j^q}(\theta) \end{pmatrix}. \tag{4.17}$$

1) If $aq < 1$ and $\tilde{a}q < 1$, by using (4.7), it is straightforward to see that

$$\lim_{n\to\infty} \begin{pmatrix} \frac{1}{h_n^q}\mathbb{E}\big(\nabla f_n(\theta)\big) \\ \frac{1}{\tilde{h}_n^q}\Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta)\Big] \end{pmatrix} = B_q(\theta). \tag{4.18}$$

2) Let us now consider the case $aq \geq 1$ and $\tilde{a}q \geq 1$. We have

$$\sqrt{nh_n^{d+2}}\mathbb{E}\left(\nabla f_n(\theta)\right) = \sqrt{nh_n^{d+2}}\frac{\sum_{i=1}^n h_i^q}{n}\frac{n}{\sum_{i=1}^n h_i^q}\mathbb{E}\left(\nabla f_n(\theta)\right),$$

with, in view of (4.8), for all $\varepsilon > 0$ small enough,

$$\begin{aligned} \sqrt{nh_n^{d+2}}\frac{\sum_{i=1}^n h_i^q}{n} &= O\left(n^{\frac{1}{2}(1-(a-\varepsilon)(d+2))}n^{-aq+a\varepsilon}\right) \\ &= O\left(n^{-\frac{1}{2}(1+ad)+\varepsilon\left(\frac{1}{2}(d+2)+a\right)}\right) \\ &= o(1). \end{aligned}$$

Applying (4.17), it follows that

$$\lim_{n\to\infty}\sqrt{nh_n^{d+2}}\mathbb{E}\big(\nabla f_n(\theta)\big) = 0.$$

Proceeding in the same way for $\mathbb{E}\big(\tilde{f}_n(\theta)\big)$, we obtain

$$\lim_{n\to\infty} \begin{pmatrix} \sqrt{nh_n^{d+2}}\mathbb{E}\big(\nabla f_n(\theta)\big) \\ \sqrt{n\tilde{h}_n^d}\Big[\mathbb{E}\big(\tilde{f}_n(\theta)\big) - f(\theta)\Big] \end{pmatrix} = 0. \tag{4.19}$$

The combination of either (4.18) or (4.19) and of Lemma 15 gives the weak convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$:

- If (C1)i) or (C1)ii) holds, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \Sigma\right). \tag{4.20}$$

- If (C1)iii) holds, then

$$\begin{pmatrix} \sqrt{nh_n^{d+2}}\nabla f_n(\theta) \\ \sqrt{n\tilde{h}_n^d}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(D(c,\tilde{c})B_q(\theta), \Sigma\right). \tag{4.21}$$

- If (C2) holds, since $aq < 1$ and $\tilde{a}q < 1$, (4.7) implies that

$$\begin{pmatrix} \frac{1}{h_n^q}\nabla f_n(\theta) \\ \frac{1}{\tilde{h}_n^q}(\tilde{f}_n(\theta) - f(\theta)) \end{pmatrix} \xrightarrow{\mathbb{P}} B_q(\theta). \tag{4.22}$$

Proof of Lemma 15

To prove Lemma 15, we first prove that

$$\lim_{n\to\infty} \mathbb{E}\Big(W_n W_n^T\Big) = \Sigma, \tag{4.23}$$

and then check that (W_n) satisfies Lyapounov's condition. Set

$$\begin{aligned} Y_{k,n} &= \frac{1}{\sqrt{nh_n^{-d-2}}} h_k^{-d-1}\left[\nabla K\left(\frac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta - X_k}{h_k}\right)\right)\right] \\ Z_{k,n} &= \frac{1}{\sqrt{n\tilde{h}_n^{-d}}} \tilde{h}_k^{-d}\left[K\left(\frac{\theta - X_k}{\tilde{h}_k}\right) - \mathbb{E}\left(K\left(\frac{\theta - X_k}{\tilde{h}_k}\right)\right)\right], \end{aligned}$$

and note that

$$\mathbb{E}\Big(W_n W_n^T\Big) = \sum_{k=1}^{n} \begin{pmatrix} \mathbb{E}\Big(Y_{k,n}Y_{k,n}^T\Big) & \mathbb{E}\Big(Y_{k,n}Z_{k,n}\Big) \\ \mathbb{E}\Big(Y_{k,n}^T Z_{k,n}\Big) & \mathbb{E}\Big(Z_{k,n}^2\Big) \end{pmatrix}.$$

Now, for any $s, t \in \{1, \ldots, d\}$, we have

$$\begin{aligned} \mathbb{E}\left[\frac{\partial K}{\partial x_s}\left(\frac{\theta - X_k}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - X_k}{h_k}\right)\right] &= \int_{\mathbb{R}^d} \frac{\partial K}{\partial x_s}\left(\frac{\theta - y}{h_k}\right)\frac{\partial K}{\partial x_t}\left(\frac{\theta - y}{h_k}\right) f(y)dy \\ &= h_k^d f(\theta)\int_{\mathbb{R}^d} \frac{\partial K}{\partial x_s}(z)\frac{\partial K}{\partial x_t}(z)dz + o(h_k^d) \\ &= h_k^d f(\theta) G_{s,t} + o(h_k^d), \end{aligned}$$

and since, $\mathbb{E}\left[\frac{\partial K}{\partial x_s}\left(\frac{\theta - X_k}{h_k}\right)\right] = O(h_k^d)$, we deduce that

$$\begin{aligned} \mathbb{E}\left(\left[\nabla K\left(\tfrac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\tfrac{\theta - X_k}{h_k}\right)\right)\right]\left[\nabla K\left(\tfrac{\theta - X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\tfrac{\theta - X_k}{h_k}\right)\right)\right]^T\right) \\ = f(\theta) G h_k^d\Big[1 + o(1)\Big] \end{aligned} \tag{4.24}$$

which implies that

$$\begin{aligned} \lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}\Big(Y_{k,n}Y_{k,n}^T\Big) &= \lim_{n\to\infty}\frac{1}{nh_n^{-d-2}}\sum_{k=1}^{n} h_k^{-2d-2}\left(f(\theta)Gh_k^d\Big[1 + o(1)\Big]\right) \\ &= \frac{f(\theta)}{1 + a(d+2)}G. \end{aligned}$$

In the same way, we have

$$\mathbb{E}\left(\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)-\mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right]^2\right) = \tilde{h}_k^d f(\theta)\int_{\mathbb{R}} K^2(z)dz\Big[1+o(1)\Big] \qquad (4.25)$$

and thus

$$\lim_{n\to\infty}\sum_{k=1}^{n}\mathbb{E}\left(Z_{k,n}^2\right)=\frac{f(\theta)}{1+\tilde{a}d}\int_{\mathbb{R}^d}K^2(z)dz.$$

Moreover, set $h_n^*=\min(h_n,\tilde{h}_n)$; we have

$$\begin{aligned}\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right] &= \int_{\mathbb{R}^d}\nabla K\left(\frac{\theta-y}{h_k}\right)K\left(\frac{\theta-y}{\tilde{h}_k}\right)f(y)dy\\ &= h_k^{*d}\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)f(\theta-h_k^*z)dz.\end{aligned}$$

Noting that $f(\theta-h_k^*z)=f(\theta)+h_k^*R_k(\theta,z)$ with $|R_k(\theta,z)|\leq\|\nabla f\|_\infty\|z\|$, we get

$$\begin{aligned}&\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right]\\ &= h_k^{*d}\left[f(\theta)\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)dz+h_k^*\int_{\mathbb{R}^d}\nabla K\Big(\frac{h_k^*}{h_k}z\Big)K\Big(\frac{h_k^*}{\tilde{h}_k}z\Big)R_k(\theta,z)dz\right].\end{aligned}$$

Since the function $z\mapsto\big[\nabla K(z)\big]K(z$ is odd (in each coordinate), the first right-handed integral is zero, and, since h_k^* equals either h_k or $\tilde{h}_k$, we get

$$\begin{aligned}&\left\|\mathbb{E}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right]\right\|\\ &\leq h_k^{*(d+1)}\|\nabla f\|_\infty\left[\|K\|_\infty\int_{\mathbb{R}^d}\|z\|\|\nabla K(z)\|dz+\|\nabla K\|_\infty\int_{\mathbb{R}^d}\|z\||K(z)|dz\right]\\ &= O\left(h_k^{*(d+1)}\right).\end{aligned}$$

We then deduce that

$$\begin{aligned}&\mathbb{E}\left(\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right)-\mathbb{E}\left(\nabla K\left(\frac{\theta-X_k}{h_k}\right)\right)\right]\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)-\mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right]\right)\\ &= O\left(\Big[\min(h_k,\tilde{h}_k)\Big]^{d+1}\right)+O\left(h_k^d\tilde{h}_k^d\right)\\ &= O\left(h_k^{\frac{d+1}{2}}\tilde{h}_k^{\frac{d+1}{2}}\right), \qquad (4.26)\end{aligned}$$

and thus, in view of (4.7),

$$\begin{aligned}
\sum_{k=1}^{n}\mathbb{E}\Big(Y_{k,n}Z_{k,n}\Big) &= O\left(\frac{1}{\sqrt{(nh_n^{-d-2})(n\tilde{h}_n^{-d})}}\sum_{k=1}^{n}h_k^{-\frac{d+1}{2}}\tilde{h}_k^{\frac{1-d}{2}}\right)\\
&= O\left(\frac{1}{\sqrt{(nh_n^{-d-2})(n\tilde{h}_n^{-d})}}\left(nh_n^{-\frac{d+1}{2}}\tilde{h}_n^{\frac{1-d}{2}}\right)\right)\\
&= O\left(h_n^{1/2}\tilde{h}_n^{1/2}\right)\\
&= o(1),
\end{aligned}$$

which concludes the proof of (4.23). Now we check that (W_n) satisfies the Lyapounov's condition. Set $p > 2$. Since K and ∇K are bounded and integrable, we have $\int_{\mathbb{R}^d}\|\nabla K(z)\|^p dz < \infty$ and $\int_{\mathbb{R}^d}|K(z)|^p dz < \infty$. On the one hand,

$$\begin{aligned}
\sum_{k=1}^{n}\mathbb{E}\left(\|Y_{k,n}\|^p\right) &= O\left(\frac{1}{\left(nh_n^{-d-2}\right)^{\frac{p}{2}}}\sum_{k=1}^{n}h_k^{(-d-1)p}\int_{\mathbb{R}^d}\left\|\nabla K\left(\frac{\theta-y}{h_k}\right)\right\|^p f(y)dy\right)\\
&= O\Bigg(\frac{1}{\left(nh_n^{-d-2}\right)^{\frac{p}{2}}}\sum_{k=1}^{n}h_k^{(-d-1)p}h_k^d\Bigg)\\
&= O\Bigg(\frac{1}{(nh_n^{-d-2})^{\frac{p}{2}}}nh_n^{(-d-1)p+d}\Bigg)\\
&= O\left(\frac{1}{(nh_n^d)^{\frac{p}{2}-1}}\right)\\
&= o(1),
\end{aligned}$$

and, on the other hand, we get

$$\begin{aligned}
\sum_{k=1}^{n}\mathbb{E}\left(\left|Z_{k,n}\right|^p\right) &= O\left(\frac{1}{(n\tilde{h}_n^{-d})^{\frac{p}{2}}}\sum_{k=1}^{n}\tilde{h}_k^{-dp}\int_{\mathbb{R}^d}\left|K\left(\frac{\theta-y}{\tilde{h}_k}\right)\right|^p f(y)dy\right)\\
&= O\Bigg(\frac{1}{(n\tilde{h}_n^{-d})^{\frac{p}{2}}}\sum_{k=1}^{n}\tilde{h}_k^{-dp}\tilde{h}_k^d\Bigg)\\
&= O\Bigg(\frac{1}{(n\tilde{h}_n^{-d})^{\frac{p}{2}}}n\tilde{h}_n^{-dp+d}\Bigg)\\
&= O\left(\frac{1}{(n\tilde{h}_n^d)^{\frac{p}{2}-1}}\right)\\
&= o(1),
\end{aligned}$$

which concludes the proof of Lemma 15. ■

4.4.5 A.s. convergence rate of $\left([\nabla f_n(\theta)]^T, \tilde{f}_n(\theta) - f(\theta)\right)^T$

Let us at first assume that the following Lemma holds.

Lemma 16 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i), (A4)ii) and (A6) hold. With probability one, the sequence*

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix}\sqrt{nh_n^{d+2}}\left[\nabla f_n(\theta)-\mathbb{E}\left(\nabla f_n(\theta)\right)\right]\\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta)-\mathbb{E}\left(\tilde{f}_n(\theta)\right)\right]\end{pmatrix}$$

is relatively compact and its limit set is $\mathcal{E}=\left\{\nu\in\mathbb{R}^{d+1}\text{ such that }\nu^T\Sigma^{-1}\nu\leq 1\right\}$.

The combination of either (4.18) or (4.19) and of Lemma 16 gives the almost sure convergence rate of $\left([\nabla f_n(\theta)]^T,\tilde{f}_n(\theta)-f(\theta)\right)^T$:

- If (C'1)i) or (C'1)ii) holds, then, with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix}\sqrt{nh_n^{d+2}}\nabla f_n(\theta)\\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta)-f(\theta)\right]\end{pmatrix} \tag{4.27}$$

  is relatively compact and its limit set is

$$\mathcal{E}=\left\{\nu\in\mathbb{R}^{d+1}\text{ such that }\nu^T\Sigma^{-1}\nu\leq 1\right\}.$$

- If (C'1)iii) holds, then with probability one, the sequence

$$\frac{1}{\sqrt{2\log\log n}}\begin{pmatrix}\sqrt{nh_n^{d+2}}\nabla f_n(\theta)\\ \sqrt{n\tilde{h}_n^d}\left[\tilde{f}_n(\theta)-f(\theta)\right]\end{pmatrix} \tag{4.28}$$

  is relatively compact and its limit set is

$$\mathcal{E}=\left\{\nu\in\mathbb{R}^{d+1}\text{ such that }\left(\nu-D(c,\tilde{c})B_q(\theta)\right)^T\Sigma^{-1}\left(\nu-D(c,\tilde{c})B_q(\theta)\right)\leq 1\right\}.$$

- If (C'2) holds, then

$$\begin{pmatrix}\frac{1}{h_n^q}\nabla f_n(\theta)\\ \frac{1}{\tilde{h}_n^q}[\tilde{f}_n(\theta)-f(\theta)]\end{pmatrix}\xrightarrow{\text{a.s.}}B_q(\theta). \tag{4.29}$$

We now prove Lemma 16. Set

$$\Delta_n=\begin{pmatrix}\frac{1}{\sqrt{nh_n^{-d-2}}}I_d & 0\\ 0 & \frac{1}{\sqrt{n\tilde{h}_n^{-d}}}\end{pmatrix}\quad\text{and}\quad Q_n=\begin{pmatrix}\sqrt{h_n^{-d-2}}I_d & 0\\ 0 & \sqrt{\tilde{h}_n^{-d}}\end{pmatrix},$$

let (ε_n) be a sequence of $\mathbb{R}^{d+1}$-valued, independent and $\mathcal{N}(0,\Sigma)$-distributed random vectors, and set $S_n=\sum_{k=1}^n Q_k\varepsilon_k$. In order to prove Lemma 16, we first establish the following Lemma 17, and then show how Lemma 16 can be deduced from Lemma 17.

Lemma 17 *Let Assumptions (A1)i), (A1)iv), (A3), (A4)i), (A4)ii) and (A6)ii) hold. With probability one, the sequence* $(T_n)\equiv\left(\frac{\Sigma^{-1/2}\Delta_n S_n}{\sqrt{2\log\log n}}\right)$ *is relatively compact and its limit set is the unit ball* $\overline{\mathcal{B}}_{d+1}(0,1)=\left\{\nu\in\mathbb{R}^{d+1}\text{ such that }\|\nu\|_2\leq 1\right\}$.

Proof of Lemma 17

Set

$$B_n = \mathbb{E}(S_n S_n^T) \quad = \quad \begin{pmatrix} \sum_{k=1}^n h_k^{-(d+2)} f(\theta) G & 0 \\ 0 & \sum_{k=1}^n \tilde{h}_k^{-d} f(\theta) \int_{\mathbb{R}^d} K^2(x)dx \end{pmatrix},$$

let $\|x\|_2$ (respectively $\|\!|A|\!\|_2$) denote the euclidean norm (respectively the spectral norm) of the vector x (respectively of the matrix A). The application of Theorem 2 in Koval [11] ensures that

$$\limsup_{n\to\infty} \frac{\|\Sigma^{-1/2}\Delta_n S_n\|_2}{\sqrt{2\,\|\!|\Sigma^{-1/2}\Delta_n B_n \Delta_n \Sigma^{-1/2}|\!\|_2 \log\log \|\!|B_n|\!\|_2}} \leq 1 \quad \text{a.s.}$$

Since $\lim_{n\to\infty}\Delta_n B_n \Delta_n = \Sigma$ and $\log\log \|\!|B_n|\!\|_2 \sim \log\log n$, we deduce that

$$\limsup_{n\to\infty} \|T_n\|_2 \leq 1 \quad \text{a.s.} \tag{4.30}$$

Thus, the sequence (T_n) is relatively compact and its limit set $\mathcal{U}$ is included in $\overline{\mathcal{B}}_{d+1}(0,1)$. Now, set $\mathcal{S}_{d+1} = \left\{w \in \mathbb{R}^{d+1},\ \|w\|_2 = 1\right\}$, and let us at first assume that

$$\forall w \in \mathcal{S}_{d+1}, \quad \limsup_{n\to\infty} w^T T_n \geq 1 \quad \text{a.s.} \tag{4.31}$$

The combination of (4.30) and (4.31) ensures that, with probability one, $\forall\varepsilon > 0, \forall n_0 \geq 1, \exists n \geq n_0$ such that $w^T T_n > 1-\varepsilon$ and $\|T_n\|_2^2 \leq 1+\varepsilon$. Noting that $\|T_n - w\|_2^2 = \|T_n\|_2^2 + \|w\|_2^2 - 2w^T T_n$, it follows that, with probability one,

$$\forall\varepsilon > 0, \forall n_0 \geq 1, \exists n \geq n_0 \quad \text{such that} \quad \|T_n - w\|_2^2 \leq 1+\varepsilon+1-2(1-\varepsilon) = 3\varepsilon.$$

Thus, with probability one, $\mathcal{S}_{d+1} \subset \mathcal{U}$. To deduce that $\overline{\mathcal{B}}_{d+1}(0,1) \subset \mathcal{U}$, we introduce (e_k), a sequence of real-valued, independent, and $\mathcal{N}(0,1)$-distributed random variables such that (e_k) is independent of (ε_k). Moreover, we set

$$\tilde{Q}_n = \begin{pmatrix} \sqrt{h_n^{-d-2}} I_{d+1} & 0 \\ 0 & \sqrt{\tilde{h}_n^{-d}} \end{pmatrix}, \quad \tilde{S}_n \quad = \quad \sum_{k=1}^n \tilde{Q}_k \begin{pmatrix} e_k \\ \varepsilon_k \end{pmatrix}$$

$$\tilde{\Delta}_n = \begin{pmatrix} \frac{1}{\sqrt{n h_n^{-d-2}}} I_{d+1} & 0 \\ 0 & \frac{1}{\sqrt{n\tilde{h}_n^{-d}}} \end{pmatrix}, \quad \text{and} \quad \tilde{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & \Sigma \end{pmatrix}.$$

We then note that the previous result applied to $(\tilde{T}_n) \equiv \left(\frac{\tilde{\Sigma}^{-1/2}\tilde{\Delta}_n\tilde{S}_n}{\sqrt{2\log\log n}}\right)$ ensures that, with probability one, $\mathcal{S}_{d+2} = \left\{w \in \mathbb{R}^{d+2},\ \|w\|_2 = 1\right\}$ is included in the limit set of $\tilde{T}_n$. Now let $\pi : \mathbb{R}^{d+2} \longrightarrow \mathbb{R}^{d+1}$ be the projection map defined by $\pi\big((x_1,\ldots,x_{d+2})^T\big) = (x_2,\ldots,x_{d+2})^T$. We clearly have $\pi(\mathcal{S}_{d+2}) = \overline{\mathcal{B}}_{d+1}(0,1)$ and $\pi(\tilde{T}_n) = T_n$, and thus deduce that, with probability one, $\overline{\mathcal{B}}_{d+1}(0,1)$ is included in the limit set of T_n.

To conclude the proof of Lemma 17, it remains to prove (4.31). In fact, we shall prove that,

$$\forall w \neq 0, \quad \limsup_{n\to\infty} \frac{w^T\Delta_n S_n}{\sqrt{2\log\log n}} \geq \sqrt{w^T\Sigma w} \quad \text{a.s.} \tag{4.32}$$

Set $v_n = \min\left\{\sqrt{nh_n^{-(d+2)}}; \sqrt{n\tilde{h}_n^{-d}}\right\}$, $A_n = v_n w^T \Delta_n$ and $V_n = \mathbb{E}\big(A_n S_n S_n^T A_n^T\big)$; we follow a method used by Petrov [17] in the proof of his Theorems 7.1 and 7.2.
Since $\lim_{n\to\infty} V_n = \infty$, $\forall \tau > 0$, there exists a non-decreasing sequence of integers n_k such that $n_k \to \infty$ as $k \to \infty$ and

$$V_{n_k - 1} \leq (1+\tau)^k \leq V_{n_k}, \quad (k = 1, 2, \ldots).$$

Since $\lim_{n\to\infty} \frac{V_{n-1}}{V_n} = 1$, we obtain $V_{n_k} \sim (1+\tau)^k$. Moreover, we have

$$V_{n_k} - V_{n_{k-1}} = V_{n_k}\big(1 - \frac{V_{n_{k-1}}}{V_{n_k}}\big) \sim V_{n_k}\frac{\tau}{\tau+1}. \tag{4.33}$$

Set

$$\begin{aligned} \chi(n) &= \sqrt{2V_n \log\log V_n} \\ \psi(n_k) &= \sqrt{2(V_{n_k} - V_{n_{k-1}})\log\log(V_{n_k} - V_{n_{k-1}})}. \end{aligned}$$

It follows from (4.33) that $\psi(n_k) \sim \tau^{1/2}\chi(n_{k-1})$. Then for any $\gamma \in]0,1[$ and k sufficiently large, we have

$$\begin{aligned} &\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \\ &\geq \mathbb{P}\left(\left\{A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\psi(n_k)\right\} \cap \left\{A_{n_k}S_{n_{k-1}} < \frac{\gamma\psi(n_k)}{2}\right\}\right) \\ &\geq \mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\psi(n_k)\right) - \mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\psi(n_k)}{2}\right) \\ &\geq \mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\chi(n_k)\right) - \mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\right). \end{aligned} \tag{4.34}$$

Since $A_{n_k}S_{n_k}$ is $\mathcal{N}\big(0, V_{n_k}\big)$-distributed, we have

$$\begin{aligned} \mathbb{P}\left(A_{n_k}S_{n_k} \geq (1-\frac{\gamma}{2})\chi(n_k)\right) &= \frac{1}{\sqrt{2\pi}}\int_{(1-\frac{\gamma}{2})\sqrt{2V_{n_k}\log\log V_{n_k}}}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt \\ &\geq [\log V_{n_k}]^{-(1+\mu)(1-\frac{\gamma}{2})^2} \end{aligned} \tag{4.35}$$

for every μ and sufficiently large k.
Set $\tilde{V}_{n_k} = v_{n_k}^2 w^T \Delta_{n_k} B_{n_{k-1}} \Delta_{n_k} w$; since $A_{n_k}S_{n_{k-1}}$ is $\mathcal{N}\big(0, \tilde{V}_{n_k}\big)$-distributed, we have

$$\mathbb{P}\Big(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\Big) = \frac{1}{\sqrt{2\pi}}\int_{\frac{\gamma\sqrt{\tau}}{3}\sqrt{2\frac{V_{n_{k-1}}}{\tilde{V}_{n_k}}\log\log V_{n_{k-1}}}}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt.$$

Let $\rho_{min}(A)$ (respectively $\rho_{max}(A)$) denote the smallest (respectively the largest) eigenvalue of a matrix A, set $\Sigma_n = \Delta_n B_n \Delta_n$, and note that

$$\begin{aligned} \frac{V_{n_{k-1}}}{\tilde{V}_{n_k}} &= \frac{v_{n_{k-1}}^2 w^T \Delta_{n_{k-1}} B_{n_{k-1}} \Delta_{n_{k-1}} w}{v_{n_k}^2 w^T \Delta_{n_k} B_{n_{k-1}} \Delta_{n_k} w} \\ &\geq \frac{v_{n_{k-1}}^2 \rho_{min}(\Sigma_{n_{k-1}})}{v_{n_k}^2 \rho_{max}(\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k})} \end{aligned} \tag{4.36}$$

with

$$\begin{aligned}\rho_{max}(\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k}) &\leq \left\|\Sigma_{n_{k-1}}\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2 \\ &\leq \left\|\Sigma_{n_{k-1}}\right\|_2 \left\|\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2. \end{aligned} \tag{4.37}$$

It follows from (4.7) that

$$\begin{aligned}\left\|\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2 &= \max\left\{\frac{n_{k-1}h_{n_{k-1}}^{-(d+2)}}{n_k h_{n_k}^{-(d+2)}}; \frac{n_{k-1}\tilde{h}_{n_{k-1}}^{-d}}{n_k\tilde{h}_{n_k}^{-d}}\right\} \\ &\sim \frac{v_{n_{k-1}}^2}{v_{n_k}^2} \quad \text{by Assumption A6)ii).} \end{aligned} \tag{4.38}$$

From (4.36), (4.37) and (4.38), we deduce that, for sufficiently large k,

$$\begin{aligned}\frac{V_{n_{k-1}}}{\tilde{V}_{n_k}} &\geq \frac{\rho_{min}(\Sigma_{n_{k-1}})}{2\rho_{max}(\Sigma_{n_{k-1}})} \\ &\geq \frac{\rho_{min}(\Sigma)}{4\rho_{max}(\Sigma)}\end{aligned}$$

and therefore, for sufficiently large k,

$$\begin{aligned}\mathbb{P}\left(A_{n_k}S_{n_{k-1}} \geq \frac{\gamma\sqrt{\tau}}{3}\chi(n_{k-1})\right) &\leq \frac{1}{\sqrt{2\pi}}\int_{\frac{\gamma\sqrt{\tau}}{6}\sqrt{\frac{\rho_{min}(\Sigma)}{\rho_{max}(\Sigma)}}\sqrt{2V_{n_{k-1}}\log\log V_{n_{k-1}}}}^{\infty} \exp\left(-\frac{t^2}{2}\right)dt \\ &\leq \left[\log V_{n_{k-1}}\right]^{\frac{-\gamma^2\tau\rho_{min}(\Sigma)}{36\rho_{max}(\Sigma)}}. \end{aligned} \tag{4.39}$$

The inequalities (4.34), (4.35) and (4.39) imply that

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq \left[\log V_{n_k}\right]^{-(1+\mu)(1-\frac{\gamma}{2})^2} - \left[\log V_{n_{k-1}}\right]^{\frac{-\gamma^2\tau\rho_{min}(\Sigma)}{36\rho_{max}(\Sigma)}}.$$

Thus, for sufficiently large k and τ, there exists $c > 0$ such that c does not depend on k and

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq c\Big[k^{-(1+\mu)(1-\frac{\gamma}{2})^2} - k^{-1}\Big].$$

Choosing μ such that $(1+\mu)\left(1-\frac{\gamma}{2}\right)^2 < 1$, we get

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) \geq \frac{c}{2}k^{-(1+\mu)(1-\frac{\gamma}{2})^2}$$

and thus

$$\sum_k \mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k)\right) = \infty.$$

Applying Borel-Cantelli's Lemma, we obtain

$$\mathbb{P}\left(A_{n_k}S_{n_k} - A_{n_k}S_{n_{k-1}} \geq (1-\gamma)\psi(n_k) \quad \text{i.o.}\right) = 1. \tag{4.40}$$

Now,

$$\begin{aligned}\limsup_{k\to\infty}\frac{|A_{n_k}S_{n_{k-1}}|}{\chi(n_{k-1})} &= \limsup_{k\to\infty}\frac{v_{n_k}|w^T\Delta_{n_k}S_{n_{k-1}}|}{\sqrt{2V_{n_{k-1}}\log\log V_{n_{k-1}}}}\\ &\le \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left\|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2\|\Delta_{n_{k-1}}S_{n_{k-1}}\|_2}{\sqrt{2v_{n_{k-1}}^2(w^T\Delta_{n_{k-1}}B_{n_{k-1}}\Delta_{n_{k-1}}w)\log\log V_{n_{k-1}}}}\\ &\le \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left\|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2\|\Delta_{n_{k-1}}S_{n_{k-1}}\|_2}{\sqrt{2v_{n_{k-1}}^2(w^T\Sigma w)\log\log V_{n_{k-1}}}}.\end{aligned}$$

Applying Theorem 2 in Koval [11] again, and using the fact that $\lim_{n\to\infty}\Delta_nB_n\Delta_n=\Sigma$, we obtain

$$\limsup_{n\to\infty}\frac{\|\Delta_nS_n\|_2}{\sqrt{2\,\|\Sigma\|_2\log\log n}}\le 1 \quad \text{a.s.}$$

and, therefore

$$\limsup_{k\to\infty}\frac{|A_{n_k}S_{n_{k-1}}|}{\chi(n_{k-1})} \le \limsup_{k\to\infty}\frac{v_{n_k}\|w\|_2\left\|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2\sqrt{\|\Sigma\|_2}}{v_{n_{k-1}}\sqrt{w^T\Sigma w}} \quad \text{a.s.}$$

Since $\left\|\Delta_{n_k}\Delta_{n_{k-1}}^{-1}\right\|_2=\sqrt{\rho_{max}(\Delta_{n_{k-1}}^{-1}\Delta_{n_k}\Delta_{n_k}\Delta_{n_{k-1}}^{-1})}\le\frac{2v_{n_{k-1}}}{v_{n_k}}$, for sufficiently large k, we obtain

$$\limsup_{k\to\infty}\frac{|A_{n_k}S_{n_{k-1}}|}{\chi(n_{k-1})}\le\frac{2\|w\|_2\sqrt{\|\Sigma\|_2}}{\sqrt{w^T\Sigma w}} \quad \text{a.s.}$$

Set $\varepsilon\in]0,1[$ and $\kappa=\frac{2\|w\|_2\sqrt{\|\Sigma\|_2}}{\sqrt{w^T\Sigma w}}$. Noting that

$$(1-\gamma)\psi(n_k)-2\kappa\chi(n_{k-1})\sim\left[(1-\gamma)\sqrt{\tau}(1+\tau)^{-1/2}-2\kappa(1+\tau)^{-1/2}\right]\chi(n_k),$$

and noting that γ can be chosen sufficiently small and τ sufficiently large so that

$$\left[(1-\gamma)\sqrt{\tau}(1+\tau)^{-1/2}-2\kappa(1+\tau)^{-1/2}\right]>1-\varepsilon,$$

we obtain

$$\mathbb{P}\left(A_{n_k}S_{n_k}>(1-\varepsilon)\chi(n_k)\quad\text{i.o.}\right)\ge\mathbb{P}\left(A_{n_k}S_{n_k}>(1-\gamma)\psi(n_k)-2\kappa\chi(n_{k-1})\quad\text{i.o.}\right).$$

Taking (4.40) into account, we then obtain

$$\mathbb{P}\left(A_{n_k}S_{n_k}>(1-\varepsilon)\chi(n_k)\quad\text{i.o.}\right)=1.$$

Thus we get

$$\limsup_{n\to\infty}\frac{A_nS_n}{\chi(n)}\ge 1 \quad \text{a.s.},$$

which proves (4.32), and concludes the proof of Lemma 17. ■

Proof of Lemma 16

Set

$$\tilde{V}_k = \begin{pmatrix} h_k^{-d/2}\left[\nabla K\left(\frac{\theta-X_k}{h_k}\right) - \mathbb{E}\left(\nabla K\left(\frac{\theta-X_k}{h_k}\right)\right)\right] \\ \tilde{h}_k^{-d/2}\left[K\left(\frac{\theta-X_k}{\tilde{h}_k}\right) - \mathbb{E}\left(K\left(\frac{\theta-X_k}{\tilde{h}_k}\right)\right)\right] \end{pmatrix}$$

and $\Gamma_k = \mathbb{E}\left(\tilde{V}_k\tilde{V}_k^T\right)$. In view of (4.24), (4.25) and (4.26), we have $\lim_{k\to\infty}\Gamma_k = \Sigma$. It follows that $\exists k_0 \geq 1$ such that $\forall k \geq k_0$, Γ_k is inversible ; without loss of generality, we assume $k_0 = 1$, and set $\tilde{U}_k = \Gamma_k^{-1/2}\tilde{V}_k$.
Set $p \in]2, 4[$ and let $\mathcal{L}$ be a slowly varying function, we have :

$$\begin{aligned}
\frac{\mathbb{E}(\|\tilde{U}_k\|^p)}{(k\log\log k)^{p/2}} &= O\left(\frac{h_k^{-dp/2}\mathbb{E}\left[\|\nabla K\left(\frac{\theta-X_k}{h_k}\right)\|^p\right] + \tilde{h}_k^{-dp/2}\mathbb{E}\left[\left|K\left(\frac{\theta-X_k}{h_k}\right)\right|^p\right]}{(k\log\log k)^{p/2}}\right) \\
&= O\left(\frac{h_k^{d-dp/2} + \tilde{h}_k^{d-dp/2}}{(k\log\log k)^{p/2}}\right) \\
&= O\left(\mathcal{L}(k)\left[k^{-\left[1+(\frac{p}{2}-1)(1-ad)\right]} + k^{-\left[1+(\frac{p}{2}-1)(1-\tilde{a}d)\right]}\right]\right)
\end{aligned}$$

so that

$$\sum_k \frac{\mathbb{E}(\|\tilde{U}_k\|^p)}{(k\log\log k)^{p/2}} < \infty.$$

By application of Theorem 2 of Einmahl [7], we deduce that

$$\sum_{k=1}^n \tilde{U}_k - \sum_{k=1}^n \eta_k = o(\sqrt{n\log\log n}) \quad \text{a.s.}$$

where η_k are independent, and $\mathcal{N}(0, I_{d+1})$-distributed random vectors.
It follows that

$$\sum_{k=1}^n \Sigma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n \varepsilon_k = o(\sqrt{n\log\log n}) \quad \text{a.s.} \tag{4.41}$$

Now,

$$\begin{aligned}
&\Delta_n\left[\sum_{k=1}^n Q_k\Sigma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \sum_{k=1}^n Q_k\varepsilon_k\right] \\
&= \Delta_n\sum_{k=1}^n Q_k\left[\Sigma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k - \varepsilon_k\right] \\
&= \Delta_n\sum_{k=1}^n Q_k\left(\sum_{j=1}^k\left[\Sigma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right] - \sum_{j=1}^{k-1}\left[\Sigma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right]\right) \quad (\text{with } \sum_{j=1}^0 = 0) \\
&= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left(\sum_{j=1}^k\left(\Sigma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right)\right) + \Delta_nQ_n\sum_{j=1}^n\left(\Sigma^{1/2}\Gamma_j^{-1/2}\tilde{V}_j - \varepsilon_j\right) \\
&= \Delta_n\sum_{k=1}^{n-1}(Q_k - Q_{k+1})\left[o\left(\sqrt{k\log\log k}\right)\right] + \Delta_nQ_n\left[o\left(\sqrt{n\log\log n}\right)\right] \quad \text{a.s.}
\end{aligned}$$

Moreover,

$$\Delta_n \sum_{k=1}^{n-1} (Q_k - Q_{k+1}) \left[o\left(\sqrt{k\log\log k}\right)\right]$$

$$= \begin{pmatrix} \sqrt{\frac{h_n^{d+2}}{n}} \sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}}\right) o\left(\sqrt{k\log\log k}\right) & 0 \\ 0 & \sqrt{\frac{\tilde{h}_n^{d}}{n}} \sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}}\right) \end{pmatrix}$$

$$= \begin{pmatrix} o\left(\sqrt{h_n^{d+2}\log\log n}\right) \sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}}\right) & 0 \\ 0 & o\left(\sqrt{\tilde{h}_n^{d}\log\log n}\right) \sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}}\right) \end{pmatrix}.$$

Set $\phi(s) = [h(s)]^{-\frac{d+2}{2}}$ and $\tilde{\phi}(s) = \left[\tilde{h}(s)\right]^{-\frac{d}{2}}$; since ϕ' and $\tilde{\phi}'$ vary regularly with exponent $(a(d+2)/2-1)$ and $(\tilde{a}d/2-1)$ respectively, we have

$$\begin{aligned} \sum_{k=1}^{n-1} \left(h_k^{-\frac{d+2}{2}} - h_{k+1}^{-\frac{d+2}{2}}\right) &= O\left(\sum_{k=1}^{n-1} \phi'(u_k)\right) \quad \text{with} \quad u_k \in [k, k+1] \\ &= O\left(\int_1^n \phi'(s)ds\right) \\ &= O\left(h_n^{-\frac{d+2}{2}}\right) \end{aligned}$$

and

$$\begin{aligned} \sum_{k=1}^{n-1} \left(\tilde{h}_k^{-\frac{d}{2}} - \tilde{h}_{k+1}^{-\frac{d}{2}}\right) &= O\left(\sum_{k=1}^{n-1} \tilde{\phi}'(u_k)\right) \quad \text{with} \quad u_k \in [k, k+1] \\ &= O\left(\int_1^n \tilde{\phi}'(s)ds\right) \\ &= O\left(\tilde{h}_n^{-\frac{d}{2}}\right) \end{aligned}$$

so that

$$\Delta_n \sum_{k=1}^{n-1} (Q_k - Q_{k+1}) \left[o\left(\sqrt{k\log\log k}\right)\right] = o\left(\sqrt{\log\log n}\right).$$

Since

$$\begin{aligned} \Delta_n Q_n \left[o\left(\sqrt{n\log\log n}\right)\right] &= \begin{pmatrix} \sqrt{\frac{h_n^{d+2}}{n}} h_n^{-\frac{d+2}{2}} o\left(\sqrt{n\log\log n}\right) & 0 \\ 0 & \sqrt{\frac{\tilde{h}_n^{d}}{n}} \tilde{h}_n^{-\frac{d}{2}} o\left(\sqrt{n\log\log n}\right) \end{pmatrix} \\ &= o\left(\sqrt{\log\log n}\right), \end{aligned}$$

we deduce that

$$\frac{\Delta_n \sum_{k=1}^n Q_k \Sigma^{1/2} \Gamma_k^{-1/2} \tilde{V}_k}{\sqrt{2\log\log n}} - \frac{\Delta_n \sum_{k=1}^n Q_k \varepsilon_k}{\sqrt{2\log\log n}} = o(1) \quad \text{a.s.}$$

The application of Lemma 17 then ensures that, with probability one, the sequence

$$\left(\frac{\Delta_n \sum_{k=1}^n Q_k \Sigma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k}{\sqrt{2\log\log n}}\right)$$

is relatively compact and its limit set is $\mathcal{E} = \left\{\nu \in \mathbb{R}^{d+1} \text{ such that } \nu^T\Sigma^{-1}\nu \leq 1\right\}$. Since

$$\frac{\Delta_n \sum_{k=1}^n Q_k \tilde{V}_k}{\sqrt{2\log\log n}} = \frac{\Delta_n \sum_{k=1}^n Q_k \Sigma^{1/2}\Gamma_k^{-1/2}\tilde{V}_k}{\sqrt{2\log\log n}} + \frac{\Delta_n \sum_{k=1}^n Q_k \left(I_{d+1} - \Sigma^{1/2}\Gamma_k^{-1/2}\right)\tilde{V}_k}{\sqrt{2\log\log n}}$$

with $\lim_{k\to\infty}\left(I_{d+1} - \Sigma^{1/2}\Gamma_k^{-1/2}\right) = 0$, Lemma 16 follows. ■

4.4.6 Weak and a.s. convergence rate of $\left((\boldsymbol{\theta_n} - \boldsymbol{\theta})^T, \tilde{\boldsymbol{\mu}}_n - \boldsymbol{\mu}\right)^T$; Proof of Theorems 10 and 11

In view of (4.15) (and the comment below), Theorem 1 (respectively Theorem 2) is a straightforward consequence of the combination of (4.20), (4.21) and (4.22) (respectively (4.27), (4.28) and (4.29)) together with the following lemma, which establishes that the residual term R_n (defined as in (4.14)) is negligeable.

Lemma 18 *Let Assumptions (A1)-(A5) hold. Moreover,*

1. *if (C1) holds, then*

$$\sqrt{n\tilde{h}_n^d}R_n \stackrel{a.s.}{\to} 0.$$

2. *If (C2) holds, then*

$$\frac{1}{\tilde{h}_n^q}R_n \stackrel{a.s.}{\to} 0.$$

Proof of Lemma 18

We first note that a Taylor's expansion implies the existence of ζ_n such that $\|\zeta_n - \theta_n\| \leq \|\theta_n - \theta\|$ and

$$\begin{aligned} R_n &= (\theta_n - \theta)^T \nabla\tilde{f}_n(\zeta_n) \\ &= (\theta_n - \theta)^T \left[\nabla\tilde{f}_n(\zeta_n) - \nabla f(\zeta_n) + \nabla f(\zeta_n) - \nabla f(\theta)\right]. \end{aligned}$$

Let $\mathcal{V}$ be a compact set that contains θ; for n large enough, we get

$$\begin{aligned} \|R_n\| &= O\left(\|\theta_n - \theta\| \left[\sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| + \|\zeta_n - \theta\|\right]\right) \\ &= O\left(\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| + \|\theta_n - \theta\|^2\right). \end{aligned}$$

On the one hand, let us recall that the a.s. convergence rate of $(\theta_n - \theta)$ is given by the one of $\left[D^2 f(\theta)\right]^{-1}\nabla f_n(\theta)$ (see (4.15) and the comment below). One can apply (4.27)), (4.28), and (4.29) and obtain the exact a.s. convergence rate of $\theta_n - \theta$. However, to avoid assuming (A6), we apply here Lemmas 13 and 14 (with $|\alpha| = 1$ and $(g_n, b_n) = (\tilde{f}_n, \tilde{h}_n)$), and get the following upper bound of the a.s. convergence rate of $\theta_n - \theta$: for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\begin{aligned} \|\theta_n - \theta\| &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{nh_n^{d+2}}} + \frac{\sum_{i=1}^n h_i^q}{n}\right) \quad \text{a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{nh_n^{d+2}}} + h_n^{q-\varepsilon} + \frac{1}{n}\right) \quad \text{a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{nh_n^{d+2}}} + h_n^{q-\varepsilon}\right) \quad \text{a.s.,} \end{aligned} \tag{4.42}$$

where the last equality follows from the fact that

$$\frac{1}{n} = o\left(\sqrt{\frac{(\log n)^{1+\gamma}}{nh_n^{d+2}}}\right).$$

On the other hand, we have

$$\sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| \leq \sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \mathbb{E}\left(\nabla\tilde{f}_n(x)\right)\| + \sup_{x\in\mathcal{V}} \|\mathbb{E}\left(\nabla\tilde{f}_n(x)\right) - \nabla f(x)\|.$$

The application of Lemmas 13 and 14 with $|\alpha| = 1$, $(g_n, b_n) = (\tilde{f}_n, \tilde{h}_n)$ ensures that, for any $\gamma > 0$ and $\varepsilon > 0$ small enough,

$$\begin{aligned} \sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n\tilde{h}_n^{d+2}}} + \frac{\sum_{i=1}^n \tilde{h}_i^q}{n}\right) \quad \text{a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n\tilde{h}_n^{d+2}}} + \tilde{h}_n^{q-\varepsilon} + \frac{1}{n}\right) \quad \text{a.s.} \\ &= O\left(\sqrt{\frac{(\log n)^{1+\gamma}}{n\tilde{h}_n^{d+2}}} + \tilde{h}_n^{q-\varepsilon}\right) \quad \text{a.s.} \end{aligned} \tag{4.43}$$

Let $\mathcal{L}$ denotes a generic slowly varying function that may vary from line to line.

- Let us first assume that (C1) holds. The application of (4.42) and (4.43) ensures that for any $\varepsilon > 0$ small enough,

$$\begin{aligned} &\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| \\ &= O\left(\sqrt{n\tilde{h}_n^d}\left[\frac{\mathcal{L}(n)}{\sqrt{n^2 h_n^{d+2}\tilde{h}_n^{d+2}}} + \frac{\mathcal{L}(n)h_n^{q-\varepsilon}}{\sqrt{n\tilde{h}_n^{d+2}}} + \tilde{h}_n^{q-\varepsilon}\left(\frac{\mathcal{L}(n)}{\sqrt{nh_n^{d+2}}} + h_n^{q-\varepsilon}\right)\right]\right) \quad \text{a.s.} \\ &= O\left(\frac{\mathcal{L}(n)}{\sqrt{nh_n^{d+2}\tilde{h}_n^2}} + \frac{\mathcal{L}(n)h_n^{q-\varepsilon}}{\tilde{h}_n} + \sqrt{n\tilde{h}_n^{d+2q-2\varepsilon}}\left[\frac{\mathcal{L}(n)}{\sqrt{nh_n^{d+2}}} + h_n^{q-\varepsilon}\right]\right) \quad \text{a.s.} \\ &= O\left(\mathcal{L}(n)\left[n^{-\frac{1}{2}(1-a(d+2)-2\tilde{a})} + n^{\tilde{a}-a(q-\varepsilon)}\right]\right) + o(1) \quad \text{a.s.} \end{aligned}$$

  Observe that by (C1)i), it is straightforward to see that $2\tilde{a} + a(d+2) < 1$ and $\tilde{a} < a(q-\varepsilon)$ for any $\varepsilon > 0$ small enough, so that

$$\sqrt{n\tilde{h}_n^d}\|\theta_n - \theta\| \sup_{x\in\mathcal{V}} \|\nabla\tilde{f}_n(x) - \nabla f(x)\| = o(1) \quad \text{a.s.}$$

Moreover, the application of (4.42) ensures that

$$\begin{aligned}\sqrt{n\tilde{h}_n^d}\|\theta_n-\theta\|^2 &= O\left(\sqrt{n\tilde{h}_n^d}\left[\frac{\mathcal{L}(n)}{nh_n^{d+2}}+h_n^{2(q-\varepsilon)}\right]\right) \quad \text{a.s.}\\ &= O\left(\mathcal{L}(n)\left[n^{-\frac{1}{2}(1-2a(d+2)+\tilde{a}d)}+n^{\frac{1}{2}(1-\tilde{a}d-4a(q-\varepsilon))}\right]\right) \quad \text{a.s.}\end{aligned}$$

Now, by (C1)ii) we have $2a(d+2)-\tilde{a}d<1$ and $\tilde{a}d+4a(q-\varepsilon)>1$ for any $\varepsilon>0$ small enough, and thus it follows that

$$\sqrt{n\tilde{h}_n^d}\|\theta_n-\theta\|^2=o(1) \quad \text{a.s.}$$

which ensures the first part of Lemma 18.

- We now assume that (C2) holds. Since $\tilde{a}q\leq\dfrac{q}{d+2q}<1$, using (4.7), (4.42) and (4.43), we have

$$\begin{aligned}&\frac{1}{\tilde{h}_n^q}\|\theta_n-\theta\|\sup_{x\in\mathcal{V}}\|\nabla\tilde{f}_n(x)-\nabla f(x)\|\\ &= O\left(\frac{1}{\tilde{h}_n^q}\left[\frac{\mathcal{L}(n)}{\sqrt{n^2h_n^{d+2}\tilde{h}_n^{d+2}}}+\frac{h_n^{q-\varepsilon}\mathcal{L}(n)}{\sqrt{n\tilde{h}_n^{d+2}}}+\tilde{h}_n^{q-\varepsilon}\left(\frac{\mathcal{L}(n)}{\sqrt{nh_n^{d+2}}}+h_n^{q-\varepsilon}\right)\right]\right) \quad \text{a.s.}\\ &= O\left(\mathcal{L}(n)\left[n^{-1+\frac{a(d+2)}{2}+\frac{\tilde{a}(d+2q+2)}{2}}+n^{-\frac{1}{2}-a(q-\varepsilon)+\frac{\tilde{a}(d+2q+2)}{2}}\right]\right)+o(1) \quad \text{a.s.} \qquad (4.44)\end{aligned}$$

On the one hand, for any $\varepsilon>0$ small enough, it is straightforward to see that condition (C2) implies the two following inequalities :

$$a(d+2)+\tilde{a}(d+2q+2)<2 \quad \text{and} \quad \tilde{a}(d+2+2q)<1+2a(q-\varepsilon), \qquad (4.45)$$

$$\tilde{a}q+a(d+2)<1 \quad \text{and} \quad \tilde{a}q<2a(q-\varepsilon). \qquad (4.46)$$

Therefore, it follows from (4.44) and (4.45) that

$$\frac{1}{\tilde{h}_n^q}\|\theta_n-\theta\|\sup_{x\in\mathcal{V}}\|\nabla\tilde{f}_n(x)-\nabla f(x)\| = o(1) \quad \text{a.s.}$$

On the other hand, observe again that by (4.42) and (4.46), we have

$$\begin{aligned}\frac{1}{\tilde{h}_n^q}\|\theta_n-\theta\|^2 &= O\left(\frac{1}{\tilde{h}_n^q}\left[\frac{\mathcal{L}(n)}{nh_n^{d+2}}+h_n^{2(q-\varepsilon)}\right]\right) \quad \text{a.s.}\\ &= O\left(\mathcal{L}(n)\left[n^{-(1-\tilde{a}q-a(d+2))}+n^{\tilde{a}q-2a(q-\varepsilon)}\right]\right) \quad \text{a.s.}\\ &= o(1) \quad \text{a.s.}\end{aligned}$$

which concludes the proof of Lemma 18. ■

Bibliographie

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

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

[3] Davies, H.L. (1973), Strong consistency of a sequential estimator of a probability density function. *Bull. Math. Statist.* **15**, pp. 49-54.

[4] Devroye, L. (1979), On the pointwise and integral convergence of recursive kernel estimates of probability densities. *Utilitas Math.* **15**, pp. 113-128.

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

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

[7] Einmahl, U. (1987), A useful estimate in the multidimensional invariance principle. *Probability theory and related fields*, **76**, pp. 81-101.

[8] Feller (1970), An introduction to probability theory and its applications. Second edition Volume II, Wiley.

[9] Hall, P. (1992), Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability density, *Ann. Statist.* **20**, pp. 675-694.

[10] Konakov, V.D. (1973), On asymptotic normality of the sample mode of multivariate distributions, *Theory Probab. Appl.* **18**, pp. 836-842.

[11] Koval, V. (2002), A new law of the iterated logarithm in $\mathbb{R}^d$ with application to matrix-normalized sums of randoms vectors. *Journal of Theoretical Probability* **15**, pp. 249-257.

[12] Menon, V.V., Prasad, B. and Singh, R.S. (1984), Non-parametric recursive estimates of a probability density function and its derivatives. *Journal of Statistical Planning and inference* **9**, pp. 73-82.

[13] Mokkadem, A., Pelletier, M. and Thiam, B. (2005), Large and moderate deviations principles for recursive kernel estimators of a multivariate density and its partial derivatives. arXiv :math.ST/0601429.

[14] Mokkadem, A. and Pelletier, M. (2003), The law of the iterated logarithm for the multivariate kernel mode estimator. *ESAIM : Probab. Statist.* **7**, pp. 1-21.

[15] Nadaraya, E.A. (1965), On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.* **10**, pp. 186-190.

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

[17] Petrov, V.V. (1995), Limit theorems in probability theory, Clarendon Press, Oxford.

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

[19] Rosenblatt, M. (1956), Remarks on some-non-parametric estimates of density function. *Ann. Math. Statist.* **27**, pp. 832-837.

[20] Roussas, G. (1992), Exact rates of almost sure convergence of a recursive kernel estimate of a probability density function : Application to regression and hazard rate estimate. *J. of Nonparam. Statist.* **3**, pp. 171-195.

[21] Rüschendorf, L. (1977), Consistency of estimators for multivariate density functions and for the mode. *Sankhya Ser. A*, **39**, pp. 243-250.

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

[23] Van Ryzin, J. (1969), On strong consistency of density estimates. *Ann. Math. Statist.* **40**, pp 1765-1772.

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

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

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

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

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