UNIVERSITÉ DE REIMS-CHAMPAGNE-ARDENNE

Ecole doctorale Sciences Technologies Santé

THÈSE DE DOCTORAT

Spécialité :

Mathématiques appliquées

Sujet de la Thèse

Quelques aspects fonctionnels et non fonctionnels des grandes déviations et des déviations modérées en estimation non-paramétrique

présentée par

Sidi Mohamed OULD MAOULOUD

Soutenue publiquement le 14 décembre 2007, devant le jury composé de :

Denis BOSQ	Université de Paris VI	Président du Jury
Michel BRONIATOWSKI	Université de Paris VI	Rapporteur
Armelle GUILLOU	Université de Strasbourg I	Rapporteur
Djamal LOUANI	Université de Reims	Directeur de thèse
Abdelkader MOKKADEM	Université de Versailles	Membre du Jury
Liming WU	Université de Clermont II	Membre du Jury

Remerciements

Je tiens tout d'abord à remercier mon directeur de thèse Monsieur Djamal Louani pour l'encadrement de mon mémoire de DEA et de cette thèse de doctorat. Durant ces années, il a été porteur d'idées, de soutiens et d'encouragements permanents. Qu'il trouve l'expression de ma reconnaissance.

Je remercie très chaleureusement les professeurs Michel Broniatowski et Armelle Guillou qui ont rapporter cette thèse, pour leurs nombreuses suggestions constructives. Je leur suis très reconnaissant du temps qu'ils ont accordé à l'expertise de ce travail.

Mes remerciements s'adressent aussi au professeurs Denis Bosq, Abdelkader Mokkadem et Liming Wu pour l'intérêt qu'ils ont accordé à mon travail et pour m'avoir fait l'honneur de participer au jury.

Je remercie les doctorants du L.S.T.A. à Paris 6 pour leur aide de tous les jours. Je témoigne ma profonde sympathie.

Je remercie tous mes amis mauritaniens, étudiants à Reims, avec qui j'ai partagé des bon moments.

Mille merci aussi Abdel, Khattary et Amina, je leurs témoigne toute mon amitié.

Je remercie enfin mes parents et mes grands parents pour le soutien qu'ils m'ont apporté durant toutes mes années d'études.

Dédicace :
À mes grands parents et à mes parents,
modeste témoignage de ma très grande gratitude.

Liste des travaux

[1] Un principe fonctionnel de grandes déviations en estimation non paramétrique. *C. R. Math. Acad. Sci. Paris*, 344(10) : $645 - 650$. (2007). Avec Louani, D.

[2] Some uniform large deviation results in nonparametric function estimation. À paraître dans *"Journal of Nonparametric Statistics".*

[3] Some functional large deviations principles in nonparametric function estimation. Soumis.

[4] $L^1(\mathbb{R})$-functional moderate deviations principle for the histogram density estimate. Soumis.

Table des matières

Chapitre 1

Introduction

Cette thèse regroupe des travaux de recherches sur le domaine des grandes déviations et des déviations modérées en estimation fonctionnelle qui ont pour motivation l'obtention des principes de grandes deviations fonctionnels et non fonctionnels et des résultats de type Chernoff. Tout au long de ce thèse nous avons considéré le cadre des variables aléatoires i.i.d..

Dans le chapitre 2 nous introduisons un processus qui nous permet de déduire de façon unifiée un principe fonctionnel de grandes déviations pour l'estimateur par la méthode de noyau de la fonction de régression et un principe de grandes déviations pour l'estimateur de la fonction de densité. Nous utilisons par la suite Ces résultats pour définir un critère de selection de modèles.

Dans le chapitre 3 nous considérons les estimateurs par la méthode des histogrammes de la fonction de densité et de la fonction de régression généralisée indexée par une famille de fonctions. Nous obtenons des principes de grandes déviations et des résultats de type Chernoff ponctuels pour les deux estimateurs. Nous établissons ensuite des résultats de type Chernoff pour les déviations uniformes des estimateurs par rapport à la fonction estimée. Nous considérons aussi l'uniformité sur une classe de fonctions $\mathcal{F}$ pour l'estimateur de la fonction de régression généralisée indexée par la famille $\mathcal{F}$. En considérant une famille $\mathcal{F}_n$ de partitions, une famille d'estimateurs de la fonction de densité $\mathcal{G}_n$ et un famille d'estimateurs de la fonction de régression sont ainsi définies en associant à chaque estimateur une partition particulière de $\mathcal{F}_n$. En considérant, en outre, une classe de densités $\mathcal{G}$ et une classe de fonctions de régression $\mathcal{R}$, nous établissons des résultats de type minimax relativement à $\mathcal{G}_n$ et $\mathcal{G}$ et aussi par rapport à $\mathcal{R}_n$ et $\mathcal{R}$

Le chapitre 4 est consacré à l'estimation par la méthode des delta-suites de la fonction de densité et de la fonction de régression. Nous obtenons dans ce chapitre des principes fonctionnels de grandes déviations dans l'espace L^1 muni de sa topologie faible pour les estimateurs de la fonction de densité et de la fonction de régression.

Le dernier chapitre est consacré aux déviations modérées pour l'estimateur de la fonction de densité par la méthode des histogrammes. Plus précisément, nous établissons un principe fonctionnel de déviations modérées pour l'estimateur de la fonction de densité centré $f_n - Ef_n$. Comme corollaire, nous établissons un résultat de type Chernoff pour la déviation modérée en norme L^1 de l'estimateur f_n par rapport à son espérance Ef_n.

1.1 Grandes déviations

1.1.1 Principe de grandes déviations

Nous allons introduire ici quelques outils et définitions utilisés dans la théorie des grandes déviations. Dans un espace topologique $\mathcal{X}$, muni de sa tribu borélienne $\mathcal{B}$, les principes de grandes déviations donnent une caractérisation exponentielle du comportement asymptotique d'une famille de mesures de probabilité $\{\mu_\epsilon\}$ par une fonction I appelée fonction de taux. De façon plus explicite, pour tout borélien A, un principe de grandes déviations associé à la famille $\{\mu_\epsilon\}$ donne une majoration et une minoration en fonction de I de la limite quand $\epsilon \to 0$ de la quantité $\epsilon \log \mu_\epsilon(A)$. Les définitions ci-après sont relatives à la fonction de taux et au principe de grandes déviations.

Définition 1.1. *Une fonction de taux I est une fonction $I : \mathcal{X} \to [0, \infty]$ semi-continue inférieurement, i.e., pour tout $\alpha \in [0, \infty)$, l'ensemble de niveau $\Psi_I(\alpha) := \{x \in \mathcal{X} : I(x) \leq \alpha\}$ est fermé. On dira que la fonction I est une bonne fonction de taux si, en outre, $\Psi_I(\alpha)$ est compact pour tout $\alpha \in [0, \infty)$. On note par $\mathcal{D}_I$ le domaine effectif de la fonction de taux I, i.e.,*
$\mathcal{D}_I := \{x \in \mathcal{X} \ : \ I(x) < \infty\}$.

Dans la définition suivante les notations A^o et $\overline{A}$ désignent respectivement l'intérieur et la fermeture de A.

Définition 1.2. *On dit qu'une famille de mesures de probabilité $\{\mu_\epsilon\}$ satisfait un principe de grande déviations dans l'espace $\mathcal{X}$ avec la vitesse ϵ et une fonction de taux I si,* *pour tout $A \in \mathcal{B}$,*

$$-\inf_{x\in A^o} I(x) \;\leq\; \liminf_{\epsilon\rightarrow 0} \log \mu_\epsilon(A) \;\leq\; \limsup_{\epsilon\rightarrow 0} \log \mu_\epsilon(A) \;\leq\; -\inf_{x\in \overline{A}} I(x).$$

Remarques
1) Les membres à gauche et à droite de la dernière expression sont souvent appelés borne inférieure et borne supérieure.
3) On dit qu'une suite de variables aléatoires X_n à valeurs dans $\mathcal{X}$ satisfait un principe de grandes déviations si la suite des lois de X_n satisfait un principe de grandes déviations.
3) Le principe de grandes déviations donné dans la définition 1.2 est equivalent à :
(Borne inférieure). Pour tout ouvert $O \subset \mathcal{X}$,

$$\liminf_{\epsilon\rightarrow 0} \log \mu_\epsilon(O) \;\geq\; -\inf_{x\in O} I(x).$$

(Borne supérieure). Pour tout fermé $F \subset \mathcal{X}$,

$$\limsup_{\epsilon\rightarrow 0} \log \mu_\epsilon(F) \;\leq\; -\inf_{x\in F} I(x).$$

4) Soit $\{\mu_\epsilon\}$ une famille de mesure de probabilité satisfaisant un principe de grandes déviations et soit A est un ensemble de continuité de la fonction de taux I associée, i.e., $\inf_{x\in A^o} I(x) = \inf_{x\in\overline{A}} I(x)$, alors

$$\lim_{\epsilon\rightarrow 0} \epsilon \log \mu_\epsilon(A) \;=\; -\inf_{x\in A} I(x).$$

En particulier
- Si $\mathcal{X}$ est un espace vectoriel normé, la fonction I est convexe et si A est un ouvert convexe, alors A est un ensemble de de continuité de la fonction I.
- Si la fonction I est continue, alors toute partie $A \in \mathcal{B}$ est un ensemble de continuité de la fonction I.

Généralement, lorsque l'on veut démontrer un principe de grandes déviations, on utilise l'approche naturelle qui consiste à établir la borne supérieure pour les compacts, en d'autres termes on étudie le principe de grandes déviations faible donné dans la définition suivante.

Définition 1.3. *(**Principe de grandes déviations faible**) On dit qu'une famille de mesures de probabilité $\{\mu_\epsilon\}$ satisfait un principe de grandes déviations faible si,*
(Borne inférieure) Pour tout ouvert $O \subset \mathcal{X}$ et pour tout $x \in O$,

$$\liminf_{\epsilon\rightarrow \infty} \epsilon \log \mu_\epsilon(O) \;\geq\; -I(x).$$

(Borne supérieure) Pour tout compact $K \subset \mathcal{X}$,

$$\limsup_{\epsilon\rightarrow \infty} \epsilon \log \mu_\epsilon(K) \;\leq\; -\inf_{x\in K} I(x).$$

Une famille de mesure de probabilité satisfaisant un principe de grandes déviations faible ne vérifie pas toujours le principe de grandes déviations. Pour généraliser le principe faible, une condition supplémentaire, appelée tension exponentielle est suffisante.

Définition 1.4. *(**Tension exponentielle**) On dit qu'une famille de mesure de probabilité $\{\mu_\epsilon\}$ est exponentiellement tendue si, pour tout $\alpha \in [0, \infty)$, il existe un compact $K_\alpha \subset \mathcal{X}$ tel que*

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(K_\alpha^c) \;<\; -\alpha.$$

Dans le théorème suivant, la tension exponentielle permet de généraliser le principe de grandes déviations faible.

Théorème 1.1. *Soit $\{\mu_\epsilon\}$ une famille de mesures de probabilité satisfaisant un principe de grandes déviations faible avec une fonction de taux I. Si $\{\mu_\epsilon\}$ est, en outre, exponentiellement tendue, alors $\{\mu_\epsilon\}$ satisfait le principe de grandes déviations avec la fonction de taux I.*

Dans ce qui suit nous donnons quelques théorèmes importants dans le domaine des grandes déviations. Le théorème suivant montre que les principes de grandes déviations sont préservés par les applications continues.

Théorème 1.2. *(**Principe de contraction**) Soient $\mathcal{X}$ et $\mathcal{Y}$ deux espaces topologiques séparés, $f \;:\; \mathcal{X} \to \mathcal{Y}$ une fonction continue et soit $I \;:\; \mathcal{X} \to [0, \infty]$ une bonne fonction de taux. Alors la fonction définie sur $\mathcal{Y}$ par*

$$I'(y) \;=\; \inf\{I(x) \;:\; f(x) = y\}$$

est une bonne fonction de taux.
Par ailleurs, si $\{\mu_\epsilon\}$ satisfait un principe de grandes déviations dans $\mathcal{X}$ avec la fonction de taux I, alors $\{\mu_\epsilon \ o \ f^{-1}\}$ satisfait un principe de grandes déviations dans $\mathcal{Y}$ avec la fonction de taux I'.

Soit $(\mathcal{X}, d)$ un espace métrique complet et séparable (espace polonais) et soit $\{X_n\}$ et $\{Y_n^m\}$ deux suites de variables aléatoires à valeurs dans $\mathcal{X}$. Le théorème suivant permet de déduire un principe de grandes déviations pour $\{X_n\}$ à partir de celui de $\{Y_n^m\}$ dans le cas où $\{Y_n^m\}$ constitue une bonne approximation exponentielle de $\{X_n\}$, i.e.,

Définition 1.5. *On dit que deux suites de variables aléatoires $\{X_n\}$ et $\{Y_n\}$ sont exponentiellement équivalentes si, pour tout $\delta > 0$,*

$$\limsup_{n \to \infty} \epsilon_n \log P(d(X_n, Y_n) > \delta) \;=\; -\infty.$$

On dit que $\{Y_n^m\}$ est une bonne approximation exponentielle de $\{X_n\}$ si, pour tout $\delta > 0$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} \epsilon_n \log P(d(X_n, Y_n^m) > \delta) \ = \ -\infty.$$

Théorème 1.3. *(**Approximation exponentielle**) On suppose que $\{Y_n\}$ satisfait un principe de grandes déviations avec la bonne fonction de taux I et que $\{X_n\}$ et $\{Y_n\}$ sont exponentiellement équivalentes, alors $\{X_n\}$ satisfait le même principe de grandes déviations.*
On suppose que pour tout m, $\{Y_n^m\}$ satisfait un principe de grandes déviations avec la bonne fonction de taux I_m et que $\{Y_n^m\}$ est une bonne approximation exponentielle de $\{X_n\}$. Alors, $\{X_n\}$ satisfait un principe de grandes déviations avec la bonne fonction de taux donnée par

$$I(x) \ := \ \sup_{\delta>0} \liminf_{m\to\infty} \inf_{y\in B_{x,\delta}} I_m(y),$$

où

$$B_{x,\delta} \ = \ \{y \in \mathcal{X} \ : \ d(x,y) < \delta\}.$$

1.1.2 Théorèmes classiques en dimension finie

Il existe une littérature riche sur la théorie grandes déviations traitant de nombreux aspects dans les domaines des Probabilités et Statistiques. A ce sujet, nous renvoyons aux livres de Dembo & Zeitouni (1998)(23), Dupuis & Ellis (1997)(29) et Deuschel & Sroock (1989)(24) ainsi que les références qui y sont citées. Dans cette section nous allons rappeler quelque théorèmes majeurs dans le cas où $\mathcal{X}$ est espace vectoriel normé de dimension finie.

Théorème de Sanov

Soit $\Sigma = \{a_1, a_2, ..., a_N\}$ un ensemble fini. Soit $M_1(\Sigma)$ l'espace de toutes les mesures de probabilité sur Σ. L'espace $M_1(\Sigma)$ peut être identifié à l'ensemble des points $\mathbb{R}^N$ dont les coordonnées sont positives et de somme égale à 1 et $M_1(\Sigma)$ peut donc être considéré comme une partie de $\mathbb{R}^N$. Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées suivant une loi μ et à valeurs dans Σ. On suppose que $\mu(a) > 0$ pour tout $a \in \Sigma$. On considère le vecteur aléatoire L_n^X à valeurs dans $M_1(\Sigma)$ défini par

$$L_n^X(a_j) \ = \ \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{a_j}(X_i), \ \ j \ = \ 1, 2, ..., N,$$

où $\mathbb{1}_a$ désigne l'indicatrice de l'ensemble $\{a\}$. On définit l'entropie relative d'une loi de probabilité ν par rapport à la loi μ par

$$H(\nu|\mu) \ := \ \sum_{j=1}^{N} \nu(a_j) \log \frac{\nu(a_j)}{\mu(a_j)}.$$

Le théorème suivant, dû à Sanov (1961)(70), donne un principe de grandes deviations pour L_n^X dans $M_1(\Sigma)$ avec la fonction H comme fonction de taux.

Théorème 1.4. *(**Sanov**) Pour tout $\Gamma \in M_1(\Sigma)$,*

$$\begin{aligned} -\inf_{\nu \in \Gamma^o} H(\nu|\mu) \ \leq \ \liminf_{n\to\infty} \frac{1}{n} \log P_\mu \left(L_n^X \in \Gamma\right) \ &\leq \ \limsup_{n\to\infty} \frac{1}{n} \log P_\mu \left(L_n^X \in \Gamma\right) \\ &\leq \ -\inf_{\nu \in \Gamma} H(\nu|\mu). \end{aligned}$$

Théorème de Cramér

Soit $X_1, X_2, ..., X_n$ une suite de vecteurs aléatoires à valeurs dans $\mathbb{R}^d$, indépendants et identiquement distribués suivant une loi de probabilité μ. La fonction transformée de Laplace logarithmique associée à la loi de probabilité μ est définie, pour tout $\lambda \in \mathbb{R}^d$, par

$$\Lambda(\lambda) = \log E_\mu \left[e^{\langle \lambda, X_1 \rangle}\right],$$

où $\langle \ . \ , \ . \ \rangle$ désigne le produit scalaire usuel sur $\mathbb{R}^d$. On appelle transformée de Legendre de la fonction Λ, la fonction Λ^* définie sur $\mathbb{R}^d$ par

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} \left\{\langle x, \lambda \rangle - \Lambda(\lambda)\right\}.$$

Le théorème suivant établi par Cramér(1938)(17) montre que la moyenne empirique $S_n = \frac{1}{n}\sum_{i=1}^n X_i$ des vecteurs aléatoires $X_1, X_2, ...$ satisfait un principe de grandes déviations.

Théorème 1.5. *(**Cramér**) On suppose que $\mathcal{D}_\Lambda = \mathbb{R}^d$. Alors,*
(Borne inférieure) pour tout ouvert $O \subset \mathbb{R}^d$,

$$\liminf_{n\to\infty} \frac{1}{n} \log P(S_n \ \in \ O) \ \geq \ -\inf_{x \in O} \Lambda^*(x),$$

(Borne supérieure) pour tout fermé $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \frac{1}{n} \log P(S_n \ \in \ F) \ \leq \ -\inf_{x \in F} \Lambda^*(x).$$

La fonction Λ^* étant convexe alors, sous l'hypothèse du théorème de Cramér, tout ouvert convexe $O \subset \mathbb{R}^d$ vérifie la relation

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \ \in \ O) \ = \ -\inf_{x \in O} \Lambda^*(x).$$

Théorème de Chernoff

Comme conséquence au théorème de Cramér dans $\mathbb{R}$, on peut voir le théorème de Chernoff (1952)(15)

Théorème 1.6. *(**Chernoff**) Pour tout $y \in \mathbb{R}$,*

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq y) = -\inf_{x\geq y} \Lambda^*(x).$$

Dans la littérature on trouve des résultats de ce type, appelés d'ailleurs résultats de type-Chernoff, pour une large palette de statistiques. Ce type de résultat trouve une application directe dans la comparaison des performances des méthodes d'estimation et dans l'étude de l'efficacité des tests au sens de Bahadur. Dans ce registre, on cite les travaux de Bahadur (1971)(2), Nikitin (1995)(61), Louani (1998, 2000)(52; 54) et Berrahou et Louani (2006)(6).

Théorème de Gärtner-Ellis

Le théorème de Cramér est limité au cas de vecteurs aléatoires i.i.d. Cependant une extension au cas non i.i.d. est possible grace au théorème de Gärtner-Ellis. On considère une suite de vecteurs aléatoires X_n prenant leurs valeurs dans $\mathbb{R}^d$ de lois de probabilité μ_n. Pour tout $\lambda \in \mathbb{R}^d$, la transformée de Laplace logarithmique associée à X_n est donnée par

$$\Lambda_n(\lambda) \ := \ \log E\left[e^{\langle \lambda, X_n\rangle}\right].$$

Avant d'énoncer le théorème de Gärtner-Ellis, nous avons besoin de l'hypothèse et des définitions suivantes

Hypothèse de Gärtner-Ellis. Pour tout $\lambda \in \mathbb{R}^d$, la limite des transformées de Laplace logarithmiques

$$\Lambda(\lambda) \ := \ \lim_{n\to\infty} \epsilon_n \Lambda_n(\epsilon_n^{-1}\lambda)$$

existe dans $\overline{\mathbb{R}}$. De plus, l'origine appartient à l'intérieur de $\mathcal{D}_\Lambda$.

Définition 1.6. *Soit Λ^* la transformée de Legendre de la fonction Λ. On dit que $y \in \mathbb{R}^d$ est un point exposé pour la fonction Λ^* s'il existe $\lambda \in \mathbb{R}^d$, appelé hyperplan d'appui, tel que, pour tout $x \neq y$,*

$$\langle \lambda, y\rangle \ - \ \Lambda^*(y) \ > \ \langle \lambda, x\rangle \ - \ \Lambda^*(x).$$

Définition 1.7. *Une fonction* $\Lambda : \mathbb{R}^d \rightarrow (-\infty, \infty]$ *est dite essentiellement lisse si :*
(i) $\mathcal{D}_\Lambda^o$ *est non vide.*
(ii) Λ *est differentiable sur* $\mathcal{D}_\Lambda^o$.
(iii) $\lim_{n\to\infty} |\nabla\Lambda(\lambda_n)| = \infty$ *pour tout suite appartenant à* $\mathcal{D}_\Lambda^o$ *et convergent vers un point de la frontière de* $\mathcal{D}_\Lambda^o$.

Le théorème suivant est l'un des plus importants dans la théorie de grandes déviations en dimension finie.

Théorème 1.7. ***(Gärtner-Ellis)*** *On suppose que l'hypothèse de Gärtner-Ellis est vérifiée. Alors,*
(Borne supérieure) pour tout fermé $F \subset \mathbb{R}^d$,

$$\limsup_{n\to\infty} \epsilon_n \log P(X_n \in F) \;\leq\; -\inf_{x\in F} \Lambda^*(x),$$

(Borne supérieure) pour tout ouvert $O \subset \mathbb{R}^d$,

$$\liminf_{n\to\infty} \epsilon_n \log P(X_n \in O) \;\geq\; -\inf_{x\in O\cap\mathcal{H}} \Lambda^*(x),$$

où $\mathcal{H}$ *désigne l'ensemble des points exposés de* Λ^* *dont les hyperplans d'appui appartiennent à* $\mathcal{D}_\Lambda^o$.
Si en outre, Λ *est essentiellement lisse et semi-continue inférieurement, alors* X_n *satisfait un principe de grandes déviations avec une bonne fonction de taux* Λ^*.

Remarque Si la fonction Λ est finie et differentiable sur $\mathbb{R}^d$ alors X_n satisfait un principe de grandes déviations avec une bonne fonction de taux Λ^*.

1.1.3 Théorèmes classiques en dimension infinie

La transformée de Laplace logarithmique joue un rôle majeur dans la théorie des grandes déviations. Au vu du théorème de Gärtner-Ellis, dans un espace vectoriel de dimension finie, la transformée de Legendre de la limite des transformées de Laplace logarithmiques associées à une suite de vecteurs aléatoires, lorsque celle-ci existe, est la candidate naturelle pour contrôler le principe de grandes déviations associé à cette suite de vecteurs. Dans cette partie nous rappelons des résultats qui généralise cette approche aux cas d'espaces vectoriels de dimensions infinie. La tension exponentielle y joue un rôle important.

Théorème de Baldi

Soient $\mathcal{X}$ un espace vectoriel séparé et $\mathcal{X}^*$ son dual topologique. On considère une suite de mesures de probabilité $\{\mu_n\}$ définies sur $\mathcal{X}$. Pour $\lambda \in \mathcal{X}^*$, on pose

$$\Lambda_n(\lambda) \ := \ \log \int_{\mathcal{X}} e^{\langle \lambda, x\rangle} \mu_n(dx),$$

$$\overline{\Lambda}(\lambda) \ := \ \limsup_{n\to\infty} \epsilon_n \Lambda_n(\epsilon_n^{-1}\lambda),$$

et $\overline{\Lambda}^*$ la transformée de Legendre de $\overline{\Lambda}$. On rappelle que $x \in \mathcal{X}$ est appelé point exposé pour $\overline{\Lambda}^*$ s'il existe $\lambda \in \mathcal{X}^*$, appelé hyperplan d'appui, tel que

$$\langle \lambda \ , \ x\rangle \ - \overline{\Lambda}^*(x) \ > \ \langle \lambda \ , \ y\rangle \ - \overline{\Lambda}^*(y), \quad \text{pour tout } y \neq x.$$

Théorème 1.8. *(**Baldi**) On suppose que $\{\mu_n\}$ est exponentiellement tendue. Alors Pour tout fermé $F \subset \mathcal{X}$,*

$$\limsup_{n\to\infty} \epsilon_n \log \mu_n(F) \ \leq \ - \inf_{x\in F} \overline{\Lambda}^*(x).$$

Pour tout ouvert $O \subset \mathcal{X}$,

$$\liminf_{n\to\infty} \epsilon_n \log \mu_n(O) \ \geq \ - \inf_{x\in F\cap\mathcal{H}} \overline{\Lambda}^*(x),$$

où $\mathcal{H}$ est l'ensemble des points exposés de $\overline{\Lambda}^*$, avec pour hyperplans d'appui λ vérifiant

$$\Lambda(\lambda) = \lim_{n\to\infty} \epsilon_n \Lambda_n(\epsilon_n^{-1}\lambda) \quad \textit{existe et } \overline{\Lambda}(\gamma\lambda) < \infty \quad \textit{pour un certain } \gamma > 1.$$

Si, pour tout ouvert $O \subset \mathcal{X}$,

$$\inf_{x\in O} \overline{\Lambda}^*(x) \ = \ \inf_{x\in O\cap\mathcal{H}} \overline{\Lambda}^*(x),$$

alors $\{\mu_n\}$ satisfait un principe de grandes déviations avec une bonne fonction de taux $\overline{\Lambda}^*$.

L'approche des systèmes projectifs

Soit $\mathcal{X}$ un espace topologique abstrait et soit $\{\mu_n\}$ une suite de mesures de probabilité sur $\mathcal{X}$. L'approche des systèmes projectifs consiste à établir des principes de grandes déviations pour les projections de $\{\mu_n\}$ sur des espaces plus "petits" et d'en déduire un principe de grandes déviations pour $\{\mu_n\}$. Plus explicitement, soit $(J, \leq)$ un ensemble partiellement ordonné, ce qui signifie que, pour tout $i, j \in J$, il existe $k \in J$, tel que $i \leq k$ et $j \leq k$. On dit que $(\mathcal{Y}_j, p_{ij})_{i \leq j \in J}$ est un système projectif si, pour tout $j \in J$, $\mathcal{Y}_j$ est un espace topologique séparé et les applications $p_{ij} \; : \; \mathcal{Y}_j \to \mathcal{Y}_i$ sont continues et vérifie $p_{ik} \; = \; p_{ij} \; o \; p_{jk}$, lorsque $i \leq j \leq k$. La limite projective de ce système, noté $\mathcal{X} = \varprojlim \mathcal{Y}_j$, est le sous espace de l'espace produit $\mathcal{Y} \; = \; \prod_{j \in J} \mathcal{Y}_j$ constitué des éléments $x = (y_j)_{j \in J}$ pour lesquels $y_i = p_{ij}(y_j)$ lorsque $i \leq j$. L'espace $\mathcal{X}$ est muni de la topologie induite par la topologie produit sur $\mathcal{Y}$. Les projections canoniques sur $\mathcal{X}$ sont définies, de manière naturelle, par

$$\begin{aligned} p_i \quad : \qquad & \mathcal{X} \longrightarrow \mathcal{Y}_i \\ & x = (y_j)_{j \in J} \longrightarrow p_i(x) := y_i \end{aligned}$$

Il est évident que ces projections sont continues sur $\mathcal{X}$.

Le théorème suivant montre qu'on peut établir un principe de grandes déviations associé à une suite de mesures de probabilité $\{\mu_n\}$ sur $\mathcal{X}$ à partir de ceux associés aux mesures images $\{\mu_n \; o \; p_i^{-1}\}$ sur $\mathcal{Y}_j$.

Théorème 1.9. ***(Dawson-Gärtner)*** *Soit $\{\mu_n\}$ une suite de mesures de probabilité sur $\mathcal{X}$, telle que pour tout $j \in J$, $\{\mu_n \; o \; p_j^{-1}\}$ satisfait un principe de grandes déviations dans $\mathcal{Y}_j$ avec une bonne fonction de taux I_j. Alors $\{\mu_n\}$ satisfait un principe de grandes déviations avec une bonne fonction de taux*

$$I(x) \;\; = \;\; \sup_{j \in J} I(p_j(x)), \quad x \in \mathcal{X}.$$

Comme conséquence du théorème de Dawson-Gärtner, Le corollaire suivant donne une condition sur la fonction limite des transformées de Laplace logarithmiques pour qu'une suite de mesures de probabilité, sur un espace vectoriel séparé localement convexe, vérifie un principe de grandes déviations.

Corollaire 1.1. *Soit $\{\mu_n\}$ une suite de mesures de probabilité exponentiellement tendue dans un espace vectoriel séparé et localement convexe $\mathcal{X}$. On suppose que $\Lambda(\; . \;) := \lim_{n \to \infty} \epsilon_n \Lambda_n(./\epsilon_n)$ est finie et Gâteaux differentiable. Alors $\{\mu_n\}$ satisfait un principe de grandes déviations avec une bonne fonction de taux convexe Λ^*.*

Nous citons, comme résultats de grandes déviations en dimension infinie, Moguls-kii (1976)(56) qui a établit un principe de grande déviations dans $L^{\infty}([0,1])$ pour les marches aléatoires. de Acosta (1994)(19) a utilisé l'approche des système projectifs pour établir un principe de grandes déviations pour le processus de Lévy. Bryc et Dembo

(1995)(12) on montré un principe de grandes déviations pour la mesure empirique associée a un processus gaussien stationnaire dans l'espace $\mathcal{M}(\mathbb{R})$ des mesures sur $\mathbb{R}$. Nous citons aussi Gamboa *et al.* (1999) (34) qui ont montré un principe de grandes déviations fonctionnel pour la forme quadratique d'un processus gaussien stationnaire dans les cas discret et continue. Ils ont utilisé comme outils de démonstration le théorème de Baldi.

1.2 Estimation non paramétrique

Il existe une multitude de méthodes en estimation fonctionnelle. Nous citons comme exemples, la méthode des histogrammes qui est la plus ancienne. Les propriétés des estimateurs par la méthode des histogrammes, comme nous le verrons dans la suite, ont été étudiées par de nombreux auteurs parmi lesquels nous citons Révész (1968, 1971, 1972)(67; 65; 66) , Lecoutre (1982)(43), Geffroy (1974)(37), Abou-Jaoudé (1976)(1) et Tukey (1961)(75). Introduite par Fix et Hodges (1951)(32), la méthode des points les plus proches pour estimer la fonction de densité a été étudiée par Loftsgaarden et Quesenberry (1965)(48). Royall (1966)(69) et Stone (1977)(72) ont étudié cette méthode dans le cadre de l'estimation de la fonction de régression. La méthode du noyau introduite, pour l'estimation de la fonction de densité, par Rozenblatt (1956) (68) a été développée ensuite par Parzen (1962) (63) et reprise par Nadaraja (1964)(59) et Watson (1964)(81) dans le cadre de l'estimation de la fonction de régression. Nous citons aussi la méthode des séries orthogonales pour l'estimation de la fonction de densité, appelée également méthode de projection, qui a été introduite par Cencov (1962)(13) et Tiago de Oliveira (1963)(73) et étudiée notamment par Bleuez et Bosq (1976,1979)(7; 8), Bosq (1980)(11) et Devroye et Gyorfi (1985)(27). La méthode des delta-suites permet d'unifier l'étude de plusieurs méthodes d'estimation fonctionnelle. Pour un aperçu global sur l'estimation fonctionnelle non paramétrique, nous renvoyons aux livres de Prakasa Rao (1983)(64), Devroye et Györfi (1985)(27), Devroye (1987)(26), Bosq et Lecoutre (1987)(10), et Scott (1992)(71).

Dans cette section nous allons aborder trois méthodes d'estimation de la fonction de densité et de la fonction de régression à savoir la méthode des histogrammes, la méthode du noyau et la méthode des delta-suites.

1.2.1 Méthode des histogrammes

La fonction de densité

Soit $(X_1, X_2, ..., X_n)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$ indépendantes et identiquement distribuées. Soit f la fonction de densité commune aux variables X_i. Soit $\mathcal{P}_n = \{A_{n,j} \ : \ j \in J\}$ une partition de $\mathbb{R}$. L'estimateur, noté f_n, de la fonction de densité f par la méthode des histogrammes basé sur la partition $\mathcal{P}_n$ est défini, pour $x \in A_{n,j}$, par

$$f_n(x) \ := \ \frac{1}{n\lambda(A_{n,j})} \sum_{i=1}^{n} \mathbb{1}_{A_{n,j}}(X_i),$$

où λ désigne la mesure de Lebesgue sur $\mathbb{R}$.

La méthode des histogrammes est la plus ancienne parmi les méthodes d'estimation et aussi la plus utilisé par les praticiens en raison de sa simplicité d'emploi. Nous citons Lecoutre qui a établi la convergence de l'estimateur f_n en moyenne quadratique et en moyenne quadratique intégrée. la convergence uniforme presque complète est dû à Geffroy (1974)(37) et la convergence en probabilité et presque complète en norme L^1 ont été établies par Abou-Jouadé (1976)(1).
Des résultats de grandes déviations de type Chernoff pour la déviation L^1 de l'estimateur de la densité par la méthode des histogrammes par rapport à la densité sous-jacente ont été établis par Beirlant *et al.* (2001) (3).

La fonction de régression

Soit (X, Y) un vecteur aléatoire à valeur dans $\mathbb{R}^2$. La fonction de régression, notée r, de la variable Y sur la variable X est définie au point x par

$$r(x) \ := \ E\left[Y \mid X = x\right].$$

Il est clair que si, $E[|Y|] < \infty$, la fonction de régression est bien définie. On suppose que (X, Y) admet une densité conjointe g et on note par g_X et g_Y les densités marginales de variables aléatoires X et Y respectivement. La fonction de régression peut alors s'écrire sous la forme

$$r(x) = \frac{1}{g_X(x)} \int_{-\infty}^{\infty} y \ g(x, y) dy.$$

L'estimateur par la méthode des histogrammes, noté r_n, de la fonction de regression r basé sur un échantillon i.i.d. $(X_1, Y_1), ..., (X_n, Y_n)$ du couple (X, Y) et la partition

$\mathcal{P}_n = \{A_{n,j} \ : \ j \in J\}$ est défini, pour $x \in A_{n,j}$, par

$$r_n(x) \ := \ \begin{cases} \frac{\sum_{i=1}^n Y_i \mathbb{1}_{A_{n,j}}(X_i)}{\sum_{i=1}^n \mathbb{1}_{A_{n,j}}(X_i)} & \text{si, } \sum_{i=1}^n \mathbb{1}_{A_{n,j}}(X_i) > 0, \\ 0 & \text{sinon.} \end{cases}$$

Cet estimateur, appelé régressogramme, a été introduit par Tukey (1961)(75). Sa convergence en moyenne quadratique a été établie par Lecoutre (1982) (43) et les résultats de convergence uniforme sur un compact en probabilité et presque complète sont dus à Geffroy (1980)(38).

1.2.2 Méthode du noyau

Soit K un noyau positif, i.e., K est une fonction réelle positive vérifiant $\int_{-\infty}^{\infty} K(x)dx = 1$ et soit $(h_n)_{n\geq 1}$ une suite de paramètres de lissage convergeant vers zéro.

La fonction de densité

Soit $(X_1, X_2, ..., X_n)$ une suite de variables aléatoires à valeurs dans $\mathbb{R}$ indépendantes et identiquement distribuées. Soit f la fonction de densité commune aux variables X_i. L'estimateur à noyau, noté f_n, de la fonction de densité f est défini par

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

Cet estimateur a été introduit par Rosenblatt (1956)(68) améliorant ainsi l'estimateur de la densité de Fix et Hodges (1951)(32) défini en un point par le nombre des observations situé dans intervalle de longueur $2h_n$ centré en ce point. Les convergences en moyenne quadratique et uniforme en probabilité de cet estimateur ont été établies par Parzen (1962)(63). Tiago de Oliveira (1963)(74) a établi la convergence en moyenne quadratique intégrée et la convergence uniforme presque complète est due à Nadaraja (1965)(60). Deheuvels (1974)(22) a obtenu une condition nécessaire et suffisante pour les convergence ponctuelle presque sûre et uniforme presque sûre de cet estimateur.

L'aspect des grandes déviations pour l'estimateur à noyau de la densité a été étudié en premier lieu par Louani (1997,1998) (51; 52). A ce propos, il a établi des principes de grandes déviations ponctuel et uniforme mais aussi des résultats ponctuels et uniformes de type Chernoff. Ces résultats ont été appliqués à l'étude de l'efficacité asymptotique de la statistique du test d'ajustement basée sur la déviation maximale entre l'estimateur à noyau de la densité par rapport à la densité sous-jacente. Louani (2000)(54) a obtenu un résultat de grandes pour la déviation de l'estimateur à noyau en norme L^1. Il a aussi obtenu, voir Louani (2005)(55), des résultats de grandes déviations uniformes par rapport à une familles de noyaux de type Vapnik-Chervonenkis, une classe de fonctions de densité et un intervalle de paramètres de lissage. Travaillant avec la méthode du noyau, Lei *et al.* (2003)(47) a établi un principe fonctionnel de grandes déviations dans l'espace L^1 muni de la topologie de la convergence faible pour l'estimateur de la densité. Lei et Wu (2005)(46) puis Lei (2006)(45) ont obtenu des résultats de grandes déviations en norme L^1 pour l'estimateur à noyau. Ils ont considéré des densités associées à des processus de Markov ergodiques et réversibles.

La fonction de régression

Soient K un noyau positif et h_n une suite de paramètres de lissage. L'estimateur à noyau r_n de la fonction de régression r, basé sur un échantillon $(X_1, Y_1), ..., (X_n, Y_n)$ du couple (X, Y), est défini par

$$r_n(x) \ := \ \begin{cases} \frac{\sum_{i=1}^n Y_i K\left(\frac{x-X_i}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right)} \text{ si, } \sum_{i=1}^n K\left(\frac{x-X_i}{h_n}\right) > 0, \\ 0 \qquad \text{sinon.} \end{cases}$$

La méthode du noyau introduite par Rozenblatt (1956)(68) pour estimer la fonction de densité a été reprise par Nadaraja (1964)(59) et Watson(1964)(81) pour estimer la fonction de régression. En raison de la forme de l'estimateur à noyau de la fonction de régression, qui est un rapport, les propriétés de convergence sont obtenues sous des conditions plus restrictives que pour l'estimateur de la densité. Notamment, la convergence uniforme n'est obtenue que sur des ensembles bornés. Nous citons comme résultats sur l'estimation par la méthode du noyau de la fonction de régression, Collomb (1979)(16) qui a donné une condition nécessaire et suffisante pour la convergence uniforme presque sûre sur un ensemble borné, généralisant ainsi les résultats antérieurs de Nadaraja (1964,1965)(59; 60), Greblick (1974) (39) et Devroye (1978)(28). On cite Louani (1999)(53) qui a obtenu un résultat de grandes déviations de type Chernoff ponctuel pour l'estimateur à noyau de la fonction de régression.

Einmahl et Mason (2000)(30) ont publié un travail dans lequel ils introduisent puis étudient les propriétés d'un processus permettant l'étude des propriétés de consistance de plusieurs estimateurs non paramétriques. Ce processus est défini par

$$W_n(x,f) \;:=\; \frac{1}{nh_n}\sum_{i=1}^{n}\left(c_f(z)f(Y_i)+d_f(z)\right)K\left(\frac{x-X_i}{h_n}\right)$$
$$-\;\; E\left[(c_f(x)f(Y_1)+d_f(x))K\left(\frac{x-X_1}{h_n}\right)\right],$$

pour x parcourant un intervalle compact I et f parcourant une classe de fonctions réelles $\mathcal{F}$ de type Vapnik-Chervonenkis qui est ponctuellement mesurable. Des choix particulier de la fonction f permettent de retrouver, entre autres, les estimateurs à noyau de la fonction de densité, de la fonction de régression et de la fonction de distribution conditionnelle. Plus précisément, l'étude du processus W_n a permis d'obtenir des vitesse de convergence uniforme sur la classe $\mathcal{F}$ et sur l'intervalle I pour l'estimateur à noyau de la fonction de régression ainsi qu'une vitesses de convergence uniforme sur l'intervalle I pour l'estimateur de la densité.

1.2.3 Méthode des delta-suites

La méthode des delta-suites est une technique permettant d'étudier simultanément les propriétés d'estimateurs fonctionnels définis par plusieurs méthodes non paramétriques. Parmi celles-ci, nous citons la méthode des histogrammes, la méthode du noyau et la méthode des séries orthogonales. Une suite de fonctions réelles δ_m définies sur $\mathbb{R}^2$ est appelée delta-suite si, pour toute fonction k de classe $\mathcal{C}^\infty$,

$$\lim_{m\to\infty}\int_{-\infty}^{\infty}\delta_m(x,y)k(x)dx \;=\; k(y), \qquad \text{pour tout } y\in\mathbb{R}.$$

La fonction de densité

Soit $(X_1, X_2, ..., X_n)$ une suite de variables aléatoire à valeurs dans $\mathbb{R}$ indépendantes et identiquement distribuées. Soit f la fonction de densité commune des variables X_i. L'estimateur par la méthode des delta-suites f_n de la fonction de densité f est défini par

$$f_n(x) \;=\; \frac{1}{n}\sum_{i=1}^{n}\delta(x,X_i).$$

Introduite par Watson et Leadbetter (1964) (79), la méthode des delta-suites été étudiée, entre autres, par Földes et Révész (1974) (33) et Walter et Blum (1979) (76).

Comme résultats de grandes déviations dédiés à l'estimateur de la fonction de densité par la méthode des delta-suites, nous citons Berrahou (2003, 2006) (5; 6) qui a établi un principe de grandes déviations ponctuel et un résultat de type chernoff uniforme pour la deviations de l'estimateur de la densité par rapport à la fonction de densité sous-jacente. Nous citons aussi Berrahou et Louani (2006) (4) où un résultat de grandes déviations pour la déviation L^1 a été établi.

La fonction de régression

L'estimateur de la fonction de régression par la méthode des delta-suites basé sur un échantillon $(X_1, Y_1), ...(X_n, Y_n)$ est défini, au point x, par

$$r_n(x) \ := \ \begin{cases} \frac{\sum_{i=1}^n Y_i \delta_{m_n}(x, X_i)}{\sum_{i=1}^n \delta_{m_n}(x, X_i)} & \text{si, } \sum_{i=1}^n \delta_{m_n}(x, X_i)) > 0, \\ 0 & \text{sinon.} \end{cases}$$

L'étude de cet estimateur n'a, à notre connaissance, pas été abordée dans la littérature.

1.3 Présentation des résultats

Cette thèse a pour motivation l'étude des propriétés de grandes déviations pour les estimateurs de la fonction de densité et de la fonction de régression. On considère une suite de vecteurs aléatoires (X_1, Y_1), (X_2, Y_2),...,(X_n, Y_n) prenant leurs valeurs dans $\mathbb{R}^2$, indépendants et identiquement distribués. On désigne par g la densité conjointe du vecteur (X_1, Y_1) et par g_X et par g_Y les densités marginales de X_1 et de Y_1 respectivement. Soit K un noyau positif. Dans le chapitre 2 nous étudions un principe de grandes déviations fonctionnel pour le processus $(W_n(f), W_n(\mathbb{1}))$ où

$$W_n(f) \ := \ W_n(f, z_0) = \frac{1}{nh_n} \sum_{i=1}^n f(Y_i) K\left(\frac{z_0 - X_i}{h_n}\right),$$

pour $z_0 \in \mathbb{R}$ fixé et f parcourant une classe de fonctions $\mathcal{F}$ compacte dans l'espace $\mathcal{C}([a, b])$ des fonctions continues sur $[a, b]$. Le principe de grandes déviations établi pour $(W_n(f), W_n(\mathbb{1}))$ permet via le principe de contraction de déduire un principe de grandes déviations fonctionnel pour l'estimateur à noyau

$$m_n(f) \ := \ m_n(f, z_0) = \frac{W_n(f)}{W_n(\mathbb{1})}$$

de la fonction de régression généralisée

$$m(f) \ := \ m(f, z_0) \ = \ E[f(Y_1)|X_1 = z_0],$$

un principe de grandes déviations pour l'estimateur à noyau

$$g_n(z_0) \ := \ W_n(\mathbb{1}) \ = \ \frac{1}{nh_n}\sum_{i=1}^{n} K\left(\frac{z_0 - X_i}{h_n}\right)$$

de la fonction de densité g_X. Les trajectoires du processus m_n englobent plusieurs estimateurs fonctionnels. Ainsi, l'estimateur de la fonction de régression classique $r(x) = E[Y|X = x]$ s'obtient en choisissant $f(y) = y$. L'estimateur la mesure conditionnelle sur un ensemble A est obtenu par le choix $f(y) = \mathbb{1}_A(y)$, alors que l'estimateur du moment conditionnel d'ordre $p > 0$ vient en choisissant $f(y) = y^p$.

Avant de donner un aperçu des résultats obtenus nous introduisons quelques fonctions et notations. On considère une partition $a = t_0 < t_1 <, ..., < t_m = b$ de l'intervalle $[a, b]$ tel que $t_j - t_{j-1} = (b - a)/m$. Pour $(\alpha, \beta, \gamma) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, on pose

$$l_1(\alpha, \beta) = \sum_{j=1}^{m} \int_{-\infty}^{\infty}\int_{t_{j-1}}^{t_j} (\exp\{(\alpha_j v + \beta_j)K(u)\} - 1)g(z_0, v)dvdu.$$

$$l_2(\alpha, \beta, \gamma) = \sum_{j=1}^{m} \int_{-\infty}^{\infty}\int_{t_{j-1}}^{t_j} (\exp\{(\alpha_j v + \beta_j + \gamma)K(u)\} - 1)g(z_0, v)dvdu.$$

et pour $(x, y, z) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, on pose

$$l_1^*(x, y) = \sup_{(\alpha,\beta)\in\mathbb{R}^m\times\mathbb{R}^m} \{< (x, y), (\alpha, \beta) > - l_1(\alpha, \beta)\},$$

$$l_2^*(x, y, z) = \sup_{(\alpha,\beta,\gamma)\in\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}} \{< (x, y, z), (\alpha, \beta, \gamma) > - l_2(\alpha, \beta, \gamma)\}.$$

Soit $\mathcal{C}(\mathcal{F})$ l'espace de fonctions Z continues sur $(\mathcal{F}, \| \ \|_\infty)$ et à valeurs dans $\mathbb{R}$. On considère la transformation

$$\begin{aligned} T_m \ : \ & \mathbb{R}^m \times \mathbb{R}^m \longrightarrow \mathcal{C}(\mathcal{F}) \\ & (\alpha, \beta) \longrightarrow T_m(x, y) \end{aligned}$$

définie par

$$T_m(x, y)(f) = \sum_{j=1}^{m}\left[\frac{f(t_{j-1}) - f(t_j)}{t_{j-1} - t_j}x_j \ + \ \frac{t_{j-1}f(t_j) - t_j f(t_{j-1})}{t_{j-1} - t_j}y_j\right].$$

Pour $Z \in \mathcal{C}(\mathcal{F})$ et $t \in \mathbb{R}$, on pose

$$J_m(Z, t) = \inf\{l_1^*(x, y) \ : \ T_m(x, y) \equiv Z\},$$

$$I_m(Z,t) = \inf \left\{ l_2^*(x,y,t) \ : \ T_m(x,y) \equiv Z \right\},$$

$$J(Z) = \sup_{\delta>0} \liminf_{m\to\infty} \inf \left\{ I_n(Z') \ : \ \sup_{f\in\mathcal{F}} |Z'(f) - Z(f)| < \delta \right\}, \tag{1.1}$$

$$I(Z,t) = \sup_{\delta>0} \liminf_{m\to\infty} \inf \left\{ I_n(Z',t') \ : \ \max\left\{ \sup_{f\in\mathcal{F}} |Z'(f) - Z(f)|, |t'-t| \right\} < \delta \right\}. \tag{1.2}$$

Sous des conditions de régularité sur le noyau K et sur les fonctions de densités g et g_X et des conditions de compacité (conditions de validité du théorème d'Arzèla-Ascoli) sur la famille $\mathcal{F}$, les processus W_n et $(W_n, W_n(\mathbb{1}))$ satisfont respectivement un principe de grandes déviations dans $\mathcal{C}(\mathcal{F})$ avec la vitesse $1/nh_n$ et une bonne fonction J définie dans (1.1) et un principe de grande déviations dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ avec la vitesse $1/nh_n$ et une bonne fonction de taux I definie dans (1.2). Plus explicitement,
pour tout borélien $A \subset \mathcal{C}(\mathcal{F})$,

$$\begin{aligned} -\inf_{Z\in A^o} J(Z) &\leq \liminf_{n\to\infty} \frac{1}{nh_n} \log P\left(W_n \in A\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(W_n \in A\right) \leq -\inf_{Z\in \overline{A}} J(Z), \end{aligned} \tag{1.3}$$

et pour tout borélien $B \subset \mathcal{C}(\mathcal{F}) \times \mathbb{R}$,

$$\begin{aligned} -\inf_{Z\in B^o} J(Z) &\leq \liminf_{n\to\infty} \frac{1}{nh_n} \log P\left((W_n, W_n(\mathbb{1}) \in B\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left((W_n, W_n(\mathbb{1}) \in B\right) \leq -\inf_{Z\in \overline{B}} J(Z). \end{aligned} \tag{1.4}$$

Ce résultat permet de déduire des corollaires pour les estimateurs de la fonction de densité et de la fonction de régression. Pour l'estimateur de la fonction de densité on utilise la projection : $(W_n, W_n(\mathbb{1})) \longrightarrow W_n(\mathbb{1})$ alors que pour l'estimateur de la fonction de régression on utilise la transformation

$$\begin{aligned} H \ : \ \mathcal{C}(\mathcal{F}) \times \mathbb{R}^* &\longrightarrow \mathcal{C}(\mathcal{F}) \\ (Z,t) \quad &\longrightarrow \frac{Z}{t} \end{aligned}$$

Pour $t \in \mathbb{R}$, on pose

$$R(t) = \sup_{\gamma\in\mathbb{R}} \left\{ t\gamma - g_X(z_0) \int_{-\infty}^{\infty} \left(\exp\left\{\gamma K(u)\right\} - 1\right) du \right\}.$$

Pour $Z \in \mathcal{C}(\mathcal{F})$ on pose

$$I_2(Z) \; := \; \inf \{ I(tZ, t) \; : \; t \in \mathbb{R}^* \}$$

On déduit du résultat précédent que les processus $g_n(z_0)$ et $m_n(f, z_0)$ satisfont respectivement un principe de grandes déviations dans $\mathbb{R}$ avec la vitesse $1/nh_n$ et une bonne fonction de taux R et un principe de grandes déviations dans $\mathcal{C}(\mathcal{F})$ avec la vitesse $1/nh_n$ et une bonne fonction de taux I_2, i.e.,
pour tout borélien $G \subset \mathbb{R}$

$$\begin{aligned} -\inf_{t \in G^o} R(t) &\leq \liminf_{n\to\infty} \frac{1}{nh_n} \log P\left(g_n(z_0) \in G\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(g_n(z_0) \in G\right) \leq -\inf_{t \in \overline{G}} R(t), \end{aligned}$$

et pour tout borélien $A \subset \mathcal{C}(\mathcal{F})$,

$$\begin{aligned} -\inf_{Z \in A^o} I_2(Z) &\leq \liminf_{n\to\infty} \frac{1}{nh_n} \log P\left(m_n \in A\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(m_n \in A\right) \leq -\inf_{Z \in \overline{A}} I_2(Z). \end{aligned}$$

Nous utilisons ces résultats pour définir un critère de grandes déviations pour la selection de modèles. Pour $\mathcal{G} \subset \mathcal{F}$ et $A \subset \mathbb{R}$, on pose

$$\Lambda_{\mathcal{G},A} = \left\{ Z \in \mathcal{C}(\mathcal{F}) \; : \; Z(f) \in A, \; \forall f \in \mathcal{G} \right\}.$$

Soient $\mathcal{G}_1$ et $\mathcal{G}_2$ deux parties de $\mathcal{F}$. On définit un critère de grandes déviations pour la comparaison des modèles $(\mathcal{G}_1, A)$ et $(\mathcal{G}_2, A)$ en étudiant le comportement asymptotique des quantités $P(W_n \in \Lambda_{\mathcal{G}_1,A})$ et $P(W_n \in \Lambda_{\mathcal{G}_2,A})$.

D'après les résultats précédents, on a

$$\begin{aligned} -\inf_{Z \in \Lambda^o_{\mathcal{G}_i,A}} l_1^*(Z) &\leq \liminf_{n\to\infty} \epsilon_n \log P(W_n \in \Lambda_{\mathcal{G}_i,A}) \\ &\leq \limsup_{n\to\infty} \epsilon_n \log P(W_n \in \Lambda_{\mathcal{G}_i,A}) \leq -\inf_{Z \in \overline{\Lambda}_{\mathcal{G}_i,A}} l_1^*(Z), \; i = 1, 2. \end{aligned}$$

Dans le cas où ces ensembles sont des ensembles de continuité de la fonction de taux L_1^*, i.e., $\inf_{Z \in \overline{\Lambda}_{\mathcal{G}_i,A}} l_1^*(Z) = \inf_{Z \in \Lambda^o_{\mathcal{G}_i,A}} l_1^*(Z) := l_1^*(\Lambda_{\mathcal{G}_i,A})$, $i = 1, 2$, la comparaison des performances des modèles $(\mathcal{G}_1, A)$ et $(\mathcal{G}_2, A)$ pour le critère des grandes déviations est alors possible en comparant les quantités $l_1^*(\Lambda_{\mathcal{G}_1,A})$ et $l_1^*(\Lambda_{\mathcal{G}_2,A})$ et on dit que le modèle

$(\mathcal{G}_1, A)$ est meilleur que le modèle $(\mathcal{G}_2, A)$ si $l_1^*(\Lambda_{\mathcal{G}_1,A}) \leq l_1^*(\Lambda_{\mathcal{G}_2,A})$. Nous donnerons un exemple particulier de classe de fonctions $\mathcal{F}$ où la comparaison des modèles est possible.

Dans le chapitre 3, on se propose d'étudier quelques résultats de grandes déviations pour les estimateurs par la méthode des histogrammes de la fonction de densité et de la fonction de régression. Soit $(X_1, Y_2), ..., (X_n, Y_n)$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^2$. Soient g la fonction de densité conjointe de (X_1, Y_1), g_X la densité marginale de la variable aléatoire X_1 et $r^f(x) = E[f(Y_1)|X_1 = x]$ la fonction de régression de la variable $f(Y_1)$ sachant l'événement $\{X_1 = x\}$. Soit $\mathcal{P}_n = \{A_{n,j} \ : \ j \in \mathbb{Z}\}$ une partition de $\mathbb{R}$. On définit l'estimateur de la fonction de densité g_X, pour $x \in A_{n,j}$, par

$$g_n(x) = \frac{1}{nm(A_{n,j})} \sum_{i=1}^{n} \mathbb{1}_{A_{n,j}}(X_i),$$

où $\mathbb{1}_{A_{n,j}}$ désigne la fonction indicatrice de l'ensemble $A_{n,j}$ et m désigne la mesure de Lebesgue sur $\mathbb{R}$. L'estimateur de la fonction de régression r^f est défini, pour $x \in A_{n,j}$, par

$$r_n^f(x) = \begin{cases} \frac{R_n^f(x)}{g_n(x)} & \text{si } g_n(x) > 0, \\ 0 & \text{sinon,} \end{cases}$$

où

$$R_n^f(x) = \frac{1}{nm(A_{n,j})} \sum_{i=1}^{n} f(Y_i) \mathbb{1}_{A_{n,j}}(X_i) \quad \text{si} \quad x \in A_{n,j},$$

avec f parcourant une famille de fonctions réelles mesurables $\mathcal{F}$.

Des principes de grandes déviations et des résultats de type Chernoff ont été établis pour les deux estimateur ci-dessus. Plusieurs aspects de grandes déviations ont été considérés. Outre le cas ponctuel, l'uniformité par rapport aux domaines de définition pour les deux estimateurs, l'uniformité sur une classe de Vapnik-Chervonenkis $\mathcal{F}$ a été étudiée pour l'estimateur de la fonction de régression alors que l'uniformité par rapport à une famille de partitions a été considérés pour les deux estimateurs. Le fait de considérer, en outre, l'uniformité par rapport à une classe de densités dans le premier cas et par rapport à une famille de fonctions de régression dans le second cas, cela a permis de déduire des résultats minimax dans les deux cas.

Pour pouvoir énoncer les résultats, nous allons introduire quelques notations. Pour

$x \in \mathbb{R}$ et $f \in \mathcal{F}$, on pose

$$h_x^f(t) = \frac{\int_{-\infty}^{\infty} f(y)e^{tf(y)}g(x,y)dy}{\int_{-\infty}^{\infty} e^{tf(y)}g(x,y)dy},$$

$$\theta_x^f(u) = \inf\left\{u \;:\; h_x^f(t) \geq u\right\},$$
$$H_x^f = int\left(\left\{h_x^f(t) \;: t \in \mathbb{R}\right\}\right),$$

où $int(A)$ désigne l'intérieur de A. On pose aussi

$$\Lambda_x^f(u,v) = \begin{cases} u\theta_x^f(\frac{u}{v}) + v\log v - v + g_X(x_0) & \\ \quad - v\log\left(\int_{-\infty}^{\infty} e^{\theta_x^f(\frac{u}{v})f(y)}g(x_0,y)dy\right) & \text{si, } (v,\frac{u}{v}) \in (0,\infty)\times H_x^f, \\ g_X(x_0) & \text{si, } (u,v) = (0,0), \\ \infty & \text{sinon.} \end{cases}$$

Sous des conditions portant sur les fonctions $f \in \mathcal{F}$ et sur la partition $\mathcal{P}_n$, nous avons le principe de grandes déviations ponctuel suivant

$$\begin{aligned} -\inf_{(u,v)\in A^o} \Lambda_x^f(u,v) &\leq \liminf_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left((R_n^f(x), g_n(x)) \in A\right) \\ &\leq \limsup_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left((R_n^f(x), g_n(x)) \in A\right) \\ &\leq -\inf_{(u,v)\in \overline{A}} \Lambda_x^f(u,v), \end{aligned}$$

pour tout $x \in \mathbb{R}$, tout $f \in \mathcal{F}$ et tout borélien $A \in \mathbb{R}^2$.

On pose maintenant

$$\delta_x(v) = \begin{cases} v\log\left(\frac{v}{g_X(x)}\right) + g_X(x) - v \text{ si, } v \geq 0, \\ \infty \qquad \text{sinon,} \end{cases}$$

et

$$\beta_x(a) = (g_X(x) + a)\log\left(1 + \frac{a}{g_X(x)}\right) - a.$$

En appliquant le principe de contraction, on obtient un principe de grandes déviations ponctuel et un résultat de type Chernoff ponctuel pour la déviation de l'estimateur de la fonction de densité $g_n(x)$. De manière plus explicite, on a

– Pour tout $x \in \mathbb{R}$ et tout borélien $A \subset \mathbb{R}$,

$$-\inf_{v \in A^o} \delta_x(v) \leq \liminf_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(g_n(x) \in A\right)$$
$$\leq \limsup_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(g_n(x) \in A\right) \leq -\inf_{v \in \overline{A}} \delta_x(v).$$

– Pour tout $x \in \mathbb{R}$ et tout $a > 0$,

$$\lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|g_n(x) - g_X(x)| > a\right) = -\beta_x(a).$$

Posons pour la suite

$$\Gamma_x^f(\lambda) = \begin{cases} \int_{-\infty}^{\infty} \left(1 - e^{\theta_x^f(\lambda)(f(y)-\lambda)}\right) g(x,y)dy & \text{si, } \lambda \in H_x^f, \\ \infty & \text{sinon,} \end{cases}$$

$$\gamma_x^{f,+}(a) = \begin{cases} \int_{-\infty}^{\infty} \left(1 - e^{\theta_x^f(r(x)+a)(f(y)-r(x)-a)}\right) g(x,y)dy & \text{si } r(x)+a \in H_x^f, \\ \infty & \text{sinon,} \end{cases}$$

$$\gamma_x^{f,-}(a) = \begin{cases} \int_{-\infty}^{\infty} \left(1 - e^{-\theta_x^f(r(x)-a)(f(y)-r(x)+a)}\right) g(x,y)dy & \text{si } r(x)-a \in H_x^f, \\ \infty & \text{sinon} \end{cases}$$

et

$$\gamma_x^f(a) = \min\left\{\gamma_x^{f,+}(a), \gamma_x^{f,-}(a)\right\}.$$

L'application du principe de contraction permet d'obtenir un principe de grandes déviations ponctuel et un résultat ponctuel de type Chernoff pour l'estimateur de la fonction de régression $r_n^f(x)$. Plus explicitement, on a

– Pour tout $x \in \mathbb{R}$, tout $f \in \mathcal{F}$ et tout borélien $A \subset \mathbb{R}$,

$$-\inf_{\lambda \in A^o} \Gamma_x^f(\lambda) \leq \liminf_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(r_n^f(x) \in A\right)$$
$$\leq \limsup_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(r_n^f(x) \in A\right) \leq -\inf_{\lambda \in \overline{A}} \Gamma_x^f(\lambda).$$

– Pour tout $x \in \mathbb{R}$, tout $f \in \mathcal{F}$ et tout $a > 0$,

$$\lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|r_n^f(x) - r^f(x)| > a\right) = -\gamma_x^f(a).$$

Sous des conditions supplémentaires portant sur la partition $\mathcal{P}_n$ et sur les queues de distribution de la variable X_1, nous obtenons en utilisant le résultat ponctuel et un maillage approprié de $\mathbb{R}$, le résultat de grandes déviations uniforme suivant

$$\lim_{n\to\infty} \frac{1}{n \max_{j\in\mathbb{Z}} m(A_{n,j})} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) = -\beta(a),$$

où

$$\beta(a) = (M+a)\log\left(1+\frac{a}{M}\right) - a, \qquad \text{avec } M = \sup_{x\in\mathcal{R}} g_X(x).$$

En supposant que le support S de la variable aléatoire X_1 est compact dans $\mathbb{R}$, nous obtenons, de manière similaire, un résultat de type Chernoff pour la déviation uniforme sur le support S de l'estimateur de la fonction de régression, exprimé dans les termes qui suivent,
pour tout $f \in \mathcal{F}$ et tout $a > 0$,

$$\lim_{n\to\infty} \frac{1}{n \max_{j\in\mathbb{Z}} m(A_{n,j})} \log P\left(\sup_{x\in S} |r_n^f(x) - r^f(x)| > a\right) = -\gamma^f(a),$$

avec

$$\gamma^f(a) = \inf_{x\in S} \gamma_x^f(a).$$

Considérant l'uniformité sur une classe de Vapnik-Chervonenkis de fonctions $\mathcal{F}$ et sur un support compact S, nous obtenons le résultat suivant de type Chernoff pour l'estimateur de la fonction de régression. Pour tout $a > 0$,

$$\lim_{n\to\infty} \frac{1}{n \max_{j\in\mathbb{Z}} m(A_{n,j})} \log P\left(\sup_{f\in\mathcal{F}} \sup_{x\in S} |r_n^f(x) - r^f(x)| > a\right) = -\gamma(a),$$

avec

$$\gamma(a) = \inf_{f\in\mathcal{F}} \gamma^f(a).$$

Soit maintenant $\mathcal{P}_n^s = \left\{A_{n,j}^s \ : \ j \in \mathbb{Z}\right\}$, $s = 0, 1$ deux partitions de $\mathbb{R}$ telles que

$$\lim_{n\to\infty} \frac{\min_{j\in\mathbb{Z}} m(A_{n,j}^0)}{\max_{j\in\mathbb{Z}} m(A_{n,j}^1)} = \alpha > 0$$

et vérifiant certaines conditions de régularité. On considère la famille $\mathcal{F}_n$ de partitions $\mathcal{P}_n = \{A_{n,j} \ : \ j \in \mathbb{Z}\}$ de $\mathbb{R}$ tel que, pour tout $j \in \mathbb{Z}$, il existe $k, l \in \mathbb{Z}$ tel que, $A_{n,j}^0 \subset A_{n,k} \subset A_{n,l}^1$. Soit $\mathcal{G}_n$ la classe des estimateurs, notés $g_n^{\mathcal{P}_n}(x)$, de la fonction de densité basés sur les partitions $\mathcal{P}_n \in \mathcal{F}_n$ et soit $\mathcal{R}_n$ la classe des estimateurs notés $r_n^{\mathcal{P}_n}(x)$ où $\mathcal{P}_n$

parcours $\mathcal{F}_n$. Soient Θ une classe de fonctions de densité et $\mathcal{R}$ une classe de fonctions de régression. On considère les fonctions suivantes

$$\beta^+_{\alpha,t}(a) = (t+a)\log\left(\frac{t+a}{\alpha t}\right) - (1-\alpha)t - a,$$

$$\beta^-_{\alpha,t}(a) = \begin{cases} \alpha(t-a)\log\left(\alpha\frac{t-a}{t}\right) - (1-\alpha)t + \alpha a & \text{si } 0 \le a \le t, \\ \infty & \text{sinon,} \end{cases}$$

$$\beta_{\alpha,t}(a) = \min\left\{\beta^+_{\alpha,t}(a), \beta^-_{\alpha,t}(a)\right\}, \qquad \beta^g_\alpha(a) = \inf_{x\in\mathbb{R}} \beta_{\alpha,g_X(x)}(a),$$

et

$$\beta^\Theta_\alpha(a) = \inf_{g_X\in\Theta} \beta^g_\alpha(a).$$

Sous quelques conditions portant sur les partitions $\mathcal{P}^0_n$ et $\mathcal{P}^1_n$ et des conditions portant sur la classe de fonctions de densité Θ, on obtient les résultats suivants :

– Pour tout $a > 0$,

$$\lim_{n\to\infty} \frac{1}{n\max_{j\in\mathbb{Z}} m(A^1_{n,j})} \log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n} \sup_{x\in\mathbb{R}} |g^{\mathcal{P}_n}_n(x) - g_X(x)| > a\right) \ge -\beta^g_\alpha(a).$$

– (Résultat de type Minimax) Pour tout $a > 0$,

$$\lim_{n\to\infty} \inf_{g_n\in\mathcal{G}_n} \sup_{g_X\in\Theta} \frac{1}{n\max_{j\in\mathbb{Z}} m(A^1_{n,j})} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) \ge -\beta^\Theta_\alpha(a).$$

De façon similaire, on obtient des résultats pour l'estimateur de la fonction de régression. Pour cela, considérons d'abord les fonctions suivantes

$$I^+_{\alpha,x,a}(t) = \alpha\int_{-\infty}^{\infty}\left(e^{t(f(y)-r_f(x)-a)} - 1\right) g(x,y)dy + (1-\alpha)g_X(x)\left(e^{-t(r_f(x)+a)} - 1\right),$$

$$I^-_{\alpha,x,a}(t) = \alpha\int_{-\infty}^{\infty}\left(e^{t(f(y)-r_f(x)+a)} - 1\right) g(x,y)dy + (1-\alpha)\int_{-\infty}^{\infty}\left(e^{tf(y)} - 1\right) g(x,y)dy,$$

$$\gamma^\pm_{\alpha,x}(a) = -\inf_{t\in\mathbb{R}} I^\pm_{\alpha,x,a}(t),$$

$$\gamma_{\alpha,x}(a) = \min\left\{\gamma^+_{\alpha,x}(a), \gamma^-_{\alpha,x}(a)\right\} \qquad \text{et} \qquad \gamma^r_\alpha(a) = \inf_{x\in\mathbb{R}} \gamma_{\alpha,x}(a).$$

Sous quelques conditions portant sur les partitions $\mathcal{P}^0_n$ et $\mathcal{P}^1_n$ et sur la classe de fonctions de régression, on obtient les résultats minimax suivants

– Pour tout $a > 0$,

$$\liminf_{n\to\infty} \frac{1}{n\max_{j\in\mathbb{Z}} m(A^1_{n,j})} \log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n} \sup_{x\in\mathbb{R}} |r^{\mathcal{P}_n}_n(x) - r(x)| > a\right) \ge -\gamma^r_\alpha(a).$$

– (Résultat Minimax) Pour tout $a > 0$,

$$\liminf_{n\to\infty} \inf_{r_n\in\mathcal{R}_n} \sup_{r\in\mathcal{R}} \frac{1}{n \max_{j\in\mathbb{Z}} m(A^1_{n,j})} \log P\left(\sup_{x\in\mathbb{R}} |r_n(x) - r(x)| > a\right)$$
$$\geq -\inf_{r\in\mathcal{R}} \gamma^r_\alpha(a).$$

Le chapitre 4 est consacré à l'étude des principes fonctionnels de grandes déviations dans le cadre de l'estimation de la fonction de densité et de la fonction de régression par la méthode des delta-suites. Plus précisément, nous établissons des principes de grandes déviations pour ces estimateurs dans l'espace L^1 muni de la topologie faible $\sigma(L^1, L^\infty)$, où L^∞ désigne l'espace des fonctions réelles bornées. Nous commençons tout d'abord par définir notre cadre de travail. Soit $(X_1, Y_1), ..., (X_n, Y_n)$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R}^2$ de fonction de densité conjointe g. On note g_X la densité marginale de la variable aléatoire X_1. On rappelle que l'on appelle delta-suite toute suite de fonctions réelles $\delta_m(x, y)$, définies sur $\mathbb{R}^2$, et vérifiant la condition suivante

$$\lim_{m\to\infty} \int_{-\infty}^{\infty} \delta_m(x, y) k(x) dx = k(y),$$

pour tout $y \in \mathbb{R}$ et toute fonction k de classe $\mathcal{C}^\infty$.
Pour une fonction réelle et mesurable f, on pose

$$g_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{m_n}(x, X_i),$$

$$R^f_n(x) = \frac{1}{n} \sum_{i=1}^{n} f(Y_i) \delta_{m_n}(x, X_i),$$

$$r^f_n(x) = \frac{R^f_n(x)}{g_n(x)}.$$

Soient φ et ϕ deux fonctions réelles positives et mesurables. Nous établissons dans un premier temps un principe de grandes déviations dans l'espace $\mathcal{X} = L^1 \times L^1 \times L^1$ pour le processus $(R^\varphi_n, R^\phi_n, g_n)$ et nous déduisons ensuite des résultats pour les estimateurs de la fonction de densité g_X et de la fonction de régression généralisée $r^f(x) = E[f(Y)|X = x]$ qui englobe plusieurs fonctions particulières comme la fonction de régression classique $r(x) = E[Y|X = x]$, la fonction des moments conditionnels $r(x) = E[Y^p|X = x]$, $p > 0$, et la mesure conditionnelle sur un ensemble A $r(x) = P(Y \in A|X = x)$. Introduisons maintenant quelques notations permettant de présenter nos résultats. Pour une fonction $c \in L^1$, on définit l'information de Kulback entre c et g_X par

$$R(c|g_X) = \begin{cases} \int c(x) \log \frac{c(x)}{g_X(x)} dx \text{ si } c \in \mathcal{AC}(g_X), \\ \infty \qquad \text{sinon}, \end{cases}$$

où $\mathcal{AC}(g_X)$ désigne l'espace des fonctions de densité absolument continues par rapport à g_X. Pour $k_1, k_2, k_3 \in L^\infty$, on pose

$$\Lambda_{\phi,\varphi}(k_1, k_2, k_3) = \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{k_1(x)\varphi(y) + k_2(x)\phi(y) + k_3(x)\right\} g(x, y)dxdy,$$

et pour $a, b, c \in L^1$, on pose

$$\Lambda^*_{\phi,\varphi}(a, b, c) = \sup_{k_1,k_2,k_3 \in L^\infty} \left\{ \int_{-\infty}^{\infty} (k_1(x)a(x) + k_2(x)b(x) + k_3(x)c(x))dx - \Lambda_{\phi,\varphi}(k_1, k_2, k_3) \right\}.$$

Nous donnons ici quelques conditions portants sur la delta-suite δ_{m_n} suffisantes pour établir des principes de grandes déviations fonctionnels pour les estimateurs de la fonction de densité et la fonction de régression et nous discutons ces conditions dans les cas particuliers des delta-suites définies par la méthode du noyau et par la méthode des histogrammes.
(H.1) Pour toute fonction $k \in L^\infty$,

$$\lim_{m\to\infty} \int_{-\infty}^{\infty} \delta_m(x, X_1)k(x)dx = k(X_1) \qquad \text{en probabilité.}$$

(H.2) Pour tout $l \in \mathbb{N}$ et tout m assez grand, il existe un réel $M_l > 0$, indépendant de m, tel que

$$\mathbb{1}_{\{|z|\le M_l\}} \int_{|x|>2M_l} \delta_m(x, z)dx \le \frac{1}{2l^2}.$$

Nous montrons que si K est une fonction positive d'intégrale égale à 1 et h_n une suite de réels positifs convergeant vers zéro, la delta-suite K_n définie par

$$K_n(x, y) = \frac{1}{h_n} K\left(\frac{x - y}{h_n}\right)$$

vérifie les conditions (H.1) et (H.2).

Considérons maintenant une partition $\mathcal{P}_n = \{A_{n,j}\ j \in J\}$ (l'ensemble J peut être $\mathbb{N}$ ou $\mathbb{Z}$) de $\mathbb{R}$. On pose

$$P_n(x, y) = \sum_{j\in J} \frac{\mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y)}{\lambda(A_{n,j})},$$

où $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A et λ désigne la mesure de Lebesgue sur $\mathbb{R}$. La suite de fonctions P_n est delta-suite et vérifie les conditions (H.1) et (H.2) si $\lim_{n\to\infty} \max_{j\in J} \lambda(A_{n,j}) = 0$

Sous les conditions (H.1) et (H.2) et des conditions supplémentaires sur les fonctions φ et ϕ, on établit un principe de grandes déviations sur $\mathcal{X}$ pour le processus $(R_n^\varphi, R_n^\phi, g_n)$ avec la vitesse n^{-1} et la bonne fonction de taux $\Lambda^*_{\phi,\varphi}$.

Des corollaires du résultat précédent sont déduits pour les estimateurs de la fonction de densité et la fonction de régression. En effet, en considérant la transformation : $(R_n^\varphi, R_n^\phi, g_n) \longrightarrow g_n$, nous montrons par l'application du principe de contraction que l'estimateur g_n de la fonction de densité satisfait un principe de grandes déviations dans L^1 avec la vitesse n^{-1} et la fonction de taux R définie par

$$R(c) = R(c\|g_X) := \begin{cases} \int_{-\infty}^{\infty} c(x)\log\left(\frac{c(x)}{g_X(x)}\right)dx & \text{si} \quad c \in \mathcal{AC}(g_X), \\ \infty & \text{sinon,} \end{cases}$$

où $\mathcal{AC}(g_X)$ désigne l'ensemble des densités absolument continues par rapport à g_X. Pour l'estimateur de la fonction de régression, avec la transformation

$$(R_n^{f^+}, R_n^{f^-}, g_n) \longrightarrow \frac{R_n^{f^+} - R_n^{f^-}}{g_n} = r_n^f$$

nous établissons un principe de grandes déviations avec la vitesse n^{-1} et la fonction de taux Γ^*.

Le dernier chapitre est consacré à l'étude des déviations modérées fonctionnelles de l'estimateur de la fonction de densité par la méthode des histogrammes dans l'espace L^1 muni de la topologie faible. Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires et soit f la fonction densité commune des variables X_i. Soit $\mathcal{P}_n = \{A_{n,j} : j \in J\}$ (où J est un ensemble index dénombrable, par exemple $J = \mathbb{N}, \mathbb{Z}$) une suite de partition de $\mathbb{R}$. On considère l'estimateur f_n de la fonction de densité f, basé sur la partition $\mathcal{P}_n$, défini par

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_n(x, X_i),$$

$$\delta_n(x,y) = \sum_{j\in J} \frac{\mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y)}{\lambda(A_{n,j})}.$$

Ici $\lambda(A_{n,j})$ et $\mathbb{1}_{A_{n,j}}$ désignent respectivement la mesure de Lebesgue et l'indicatrice de l'ensemble $A_{n,j}$. Le but de ce chapitre est d'établir que le processus centré $f_n - Ef_n$ vérifie un principe de déviations modérées dans l'espace L^1. Plus explicitement, nous identifions une classe de suite $(b_n)_n$ (vérifiant des conditions classiques de la théorie des déviations modérées, à savoir $b_n = o(n)$ et $n = o(b_n^2)$) ainsi que une fonction I semi-continue inférieurement pour lesquelles le processus $nb_n^{-1}(f_n - Ef_n)$ satisfait un principe de grandes déviations avec la vitesse nb_n^{-2} et la fonction de taux I. Nous déduisons ensuite un résultat de type Chernoff pour la déviation modérée en norme L^1 de l'estimateur f_n par rapport à son espérance Ef_n

Avant d'énoncer les résultats de ce chapitre nous donnons quelques notations et définitions. Les fonctions suivantes sont respectivement la limite des fonctions logarithmiques génératrices des moments associée à $nb_n^{-1}(f_n - Ef_n)$ et sa tranformée de Legendre. Pour $k \in L^\infty$, on pose

$$\Lambda(k) = \frac{1}{2} \int k^2(x) f(x) dx,$$

et pour $g \in L^1$, on pose

$$\Lambda^*(g) = \sup_{k \in L^\infty} \left\{ \int g(x) k(x) dx - \Lambda(k) \right\}.$$

On désigne par μ la mesure de probabilité associée à la fonction de densité f et par $L^2(\mu)$ l'espace des fonctions de carré intégrable par rapport à la mesure μ. En considérant l'application H de $L^2(\mu)$ dans L^1 qui à une fonction associe la fonction $H(h) = hf$, l'espace $\mathcal{H}_\mu := H(L^2(\mu))$ est un sous espace vectoriel qu'on peut munir d'un produit scalaire, noté $\langle \ , \ \rangle_{\mathcal{H}_\mu}$, et de la norme sous-jacente $\| \ \|_{\mathcal{H}_\mu}$ ($\|g\|_{\mathcal{H}_\mu} = \langle g, g \rangle_{\mathcal{H}_\mu}^{1/2}$) définis comme suit, pour $g_1 = H(h_1), g_2 = H(h_2) \in \mathcal{H}_\mu$, $\langle g_1, g_2 \rangle_{\mathcal{H}_\mu} = \langle h_1, h_2 \rangle_\mu$, où $\langle \ , \ \rangle_\mu$ désigne le produit scalaire usuel sur $L^2(\mu)$ i.e., $\langle h_1, h_2 \rangle_\mu = \int h_1 h_2 d\mu$. On montre que

$$\Lambda^*(g) = \begin{cases} \frac{1}{2} \|g\|^2_{\mathcal{H}_\mu} & \text{si } g \in \mathcal{H}_\mu, \\ \infty & \text{sinon.} \end{cases}$$

Nous établissons, sous des conditions sur la suite de normalisation $(b_n)_n$, la suite de partitions $\mathcal{P}_n$ et les queues de distribution de la variable X_1, le résultat suivant

– Pour tout ouvert $O \subset L^1$,

$$\liminf_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}(f_n - Ef_n) \in O\right) \geq -\inf_{g \in O} \Lambda^*(g).$$

– Pour tout fermé $F \subset L^1$,

$$\limsup_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}(f_n - Ef_n) \in F\right) \leq -\inf_{g \in F} \Lambda^*(g).$$

Sous les même conditions, nous établissons un résultat de type Chernoff pour La déviation modérée en norme L^1 de l'estimateur de la fonction de densité par la méthode des histogrammes par rapport à son espérance. Plus explicitement, pour tout $r > 0$, on a

$$\lim_{n\to\infty} \frac{n}{b_n^2} \log P\left(nb_n^{-1} \|f_n - Ef_n\|_{L^1} > r\right) = -\frac{r^2}{2}.$$

Chapitre 2

Un principe fonctionnel de grandes déviations en estimation non-paramétrique

L'objet de ce chapitre est d'établir un principe fonctionnel de grandes déviations pour le processus $U_n(f)$ indexé par une famille de fonctions $\mathcal{F}$ et de déduire de façon unifiée des résultats de grandes déviations pour les estimateurs de la fonction de régression et de la fonctions de densité. Nos résultats sont alors appliqués au problème de selection de modèles.

2.1 Introduction

Soit $(X_1, Y_1), ..., (X_n, Y_n), ...$ une suite de vecteurs aléatoires i.i.d. à valeurs dans $\mathbb{R} \times [a, b]$ ayant pour densité jointe la fonction g et pour densités marginales les fonctions g_X et g_Y. Soit $(h_n)_{n\geq 1}$ une suite de paramètres de lissage (i.e., une suite de réels positifs convergeant vers zéro). Dans toute la suite de ce travail K désigne une fonction réelle positive telle que $\int K(u)du = 1$. Soit $\mathcal{F}$ une partie compacte de l'espace $\mathcal{C}([a, b])$ des fonctions réelles continue sur l'intervalle $[a, b] \subset \mathbb{R}$. Pour $z_0 \in \mathbb{R}$ et $f \in \mathcal{F}$, on considère le processus suivant indexé par la famille de fonctions $\mathcal{F}$

$$W_n(f) := W_n(f)(z_0) = \frac{1}{nh_n} \sum_{i=1}^{n} f(Y_i) K\left(\frac{X_i - z_0}{h_n}\right).$$

Ce processus a été introduit par Einmahl et Mason (2000)(30) sous une forme plus complexe. Il leur a permis d'unifier l'étude de la consistance uniforme par rapport à la famille des fonctions $\mathcal{F}$ de l'estimateur du type Nadaraya-Watson de la fonction de régression et de l'estimateur à noyau de la fonction de densité. Ils ont aussi obtenu des vitesses de convergence pour ces estimateurs.

Le processus $W_n(f)$ nous permet dans le cadre de ce travail d'établir de façon unifiée des principes de grandes déviations pour l'estimateur de type Nadaraya-Watson de la fonction de régression et de l'estimateur à noyau de la fonction de densité. Pour cela, on considère le processus vectoriel indexé par la famille de fonctions $\mathcal{F}$ défini par

$$U_n(f) = (W_n(f), W_n(\mathbb{1}))$$

où $\mathbb{1}$ désigne la fonction identiquement égale à 1. Notons $\mathcal{C}(\mathcal{F})$ l'espace des fonctions réelles continues sur $(\mathcal{F}, \| \|_\infty)$, où $\| \|_\infty$ désigne la norme infinie, que l'on munit de la norme

$$\|Z\|_{\mathcal{C}(\mathcal{F})} = \sup_{f\in\mathcal{F}} |Z(f)|.$$

L'espace $(\mathcal{C}(\mathcal{F}), \| \|_{\mathcal{C}(\mathcal{F})})$ ainsi formé est un espace vectoriel normé. Considérons maintenant l'espace $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ muni de la norme

$$\|(Z,t)\|_{\mathcal{C}(\mathcal{F})\times\mathbb{R}} = \max\left\{\|Z\|_{\mathcal{C}(\mathcal{F})}, |t|\right\}.$$

Pour $(Z,t) \in \mathcal{C}(\mathcal{F}) \times \mathbb{R}$ et δ un réel positif, la boule ouverte $B_{(Z,t),\delta}$ dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ de centre (Z,t) et de rayon δ est définie par

$$B_{(Z,t),\delta} = \left\{(Z',t') \in \mathcal{C}(\mathcal{F}) \times \mathbb{R} : \|(Z,t) - (Z',t')\|_{\mathcal{C}(\mathcal{F})\times\mathbb{R}} < \delta\right\}.$$

Établir un principe de grandes déviations (PGD) pour le processus U_n indexé par la famille de fonctions $\mathcal{F}$ consiste à trouver une suite ϵ_n de réels positifs convergeant vers zéro et une fonction I semi-continue inférieurement sur $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ telles que
(Borne supérieure) Pour tout fermé $F \subset \mathcal{C}(\mathcal{F}) \times \mathbb{R}$

$$\limsup_{n\to\infty} \epsilon_n \log P(U_n \in F) \leq -\inf_{(Z,t)\in F} I(Z,t).$$

(Borne inférieure) Pour tout ouvert $G \subset \mathcal{C}(\mathcal{F}) \times \mathbb{R}$

$$\liminf_{n\to\infty} \epsilon_n \log P(U_n \in G) \geq -\inf_{(Z,t)\in G} I(Z,t).$$

On dira alors que le processus U_n satisfait le PGD dans $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ avec la vitesse ϵ_n^{-1} et la fonction de taux I.

Le résultat obtenu pour le processus U_n permet alors de déduire des résultats de grandes déviations à la fois pour l'estimateur à noyau de la fonction de densité et pour l'estimateur de type Nadaraya-Watson de la fonction de régression. Ces résultats ont été obtenu séparément par Louani (1997)(51) et Louani (1999) (53). On rappelle que la fonction de densité g_X est estimée par $W_n(\mathbb{1})$ et que l'estimateur de la fonction de régression $m(f, z_0) := E(f(Y)|X = z_0)$ est

$$m_n(f, z_0) := \frac{W_n(f)}{W_n(\mathbb{1})}.$$

Les implications résultant relatifs à ces deux estimateurs sont donnés dans les corollaire 1 et 2 respectivement.

La méthode de démonstration classique de résultats de grandes déviations fonctionnels consiste à établir en premier lieu des résultats pour la version fini-dimensionnelle du processus pour ensuite réaliser sa tension exponentielle. La démarche entreprise dans cette Note est différente. Elle consiste à établir un principe des grandes déviations dans $\mathbb{R}^{2m+1}$, pour le processus vectoriel $(A_n, B_n, W_n(\mathbb{1}))$ où les vecteurs $A_n = (A_n^1, ..., A_n^m)$ et $B_n = (B_n^1, ..., B_n^m)$ sont définis par où

$$\begin{aligned} A_n^j &= \frac{1}{nh_n} \sum_{i=1}^{n} Y_i \mathbb{1}_{[t_{j-1},t_j]}(Y_i) K\left(\frac{X_i - z_0}{h_n}\right), \\ B_n^j &= \frac{1}{nh_n} \sum_{i=1}^{n} \mathbb{1}_{[t_{j-1},t_j]}(Y_i) K\left(\frac{X_i - z_0}{h_n}\right), \end{aligned} \tag{2.1}$$

pour $j = 1, 2, .., m$. Les t_j formant ici une partition uniforme de l'intervalle $[a, b]$, i.e., $a = t_0 < t_1 <, ..., < t_m = b$ et $t_j - t_{j-1} = (b-a)/m$, $1 \leq j \leq m$ et $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A. Nous construisons ensuite une approximation exponentielle U_n^m du processus U_n définie par

$$U_n^m = (W_n^m, W_n(\mathbb{1}))$$

où,

$$W_n^m(f) = \frac{1}{nh_n} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} (\alpha_j(f) Y_i + \beta_j(f)) \mathbb{1}_{[t_{j-1},t_j]}(Y_i) \right) K\left(\frac{X_i - z_0}{h_n}\right)$$

avec $\alpha_j(f) = \dfrac{f(t_{j-1}) - f(t_j)}{t_{j-1} - t_j}$ et $\beta_j(f) = \dfrac{t_{j-1} f(t_j) - t_j f(t_{j-1})}{t_{j-1} - t_j}; j = 1, 2, .., m$.
Un PGD pour U_n^m est alors établi à l'aide du résultat obtenu pour le processus $(A_n, B_n, W_n(\mathbb{1}))$ et du principe de contraction. En utilisant le théorème 1.3 (Approximation exponentielle), on conclut notre preuve en montrant que le processus U_n^m est une bonne approximation exponentielle de U_n, i.e, pour tout $\tau > 0$

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(\|U_n - U_n^m\|_{\mathcal{C}(\mathcal{F})\times R} > \tau\right) = -\infty. \tag{2.2}$$

2.2 Résultats

2.2.1 Hypothèses et notations

Nous donnons maintenant l'ensemble des hypothèses nécessaires à l'établissement de nos résultats.

(H.1) $h_n \to 0$ et $nh_n \to \infty$ quand $n \to \infty$

(K.1) Pour tout $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} |\exp\{tK(u)\} - 1| du < \infty.$$

(K.2) Pour tout $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} |u(\exp\{tK(u)\} - 1)| du < \infty.$$

(K.3) Pour tout $t \in \mathbb{R}$

$$\int_{-\infty}^{\infty} K(u) \exp\{tK(u)\} du < \infty.$$

(G.1) g est derivable par rapport à la première composante.

(G.2) $\sup_{-\infty<u<\infty} \int_a^b \left| \frac{\partial}{\partial u} g(u,v) \right| dv < \infty.$

(F.1) $\sup_{f \in \mathcal{F}} |f(t)| < \infty$ pour un certain $t \in [a,b]$.

(F.2) $\sup_{f \in \mathcal{F}} \sup_{|t-s| \leq \delta} |f(t) - f(s)| \to 0$ quand $\delta \to 0$.

Discussion des conditions. Les conditions (K) sont des conditions classiques rencontrées dans les résultats des grandes déviations en estimation non paramétrique. Les conditions (F) assurent la compacité de la famille $\mathcal{F}$ et constituent des conditions d'application du théorème d'Arzela-Ascoli.

Introduisons maintenant des fonctions qui serviront à exposer nos résultats. Pour $(\alpha, \beta, \gamma) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, posons

$$l_1(\alpha, \beta) = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} (\exp\{(\alpha_j v + \beta_j) K(u)\} - 1) g(z_0, v) dv du \text{ et}$$

$$l_2(\alpha, \beta, \gamma) = \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} (\exp\{(\alpha_j v + \beta_j + \gamma) K(u)\} - 1) g(z_0, v) dv du.$$

Pour $(x, y, z) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$, posons

$$l_1^*(x, y) = \sup_{(\alpha,\beta) \in \mathbb{R}^m \times \mathbb{R}^m} \{< (x,y), (\alpha, \beta) > - l_1(\alpha, \beta)\} \text{ et}$$

$$l_2^*(x,y,z) = \sup_{(\alpha,\beta,\gamma)\in\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}} \left\{ < (x,y,z),(\alpha,\beta,\gamma) > -l_2(\alpha,\beta,\gamma) \right\}.$$

Les fonctions l_1 et l_2 sont les fonctions limites des suites de transformées de Laplace associées au vecteurs (A_n, B_n) et $(A_n, B_n, W_n(\mathbb{1}))$ respectivement alors que les fonctions l_1^* et l_2^* sont les transformées de Legendre des fonctions l_1 et l_2 respectivement.

2.2.2 PGDs fonctionnels généraux

Nos résultats sont présentés sous la forme de théorèmes exposant ses PGDs pour les processus $W_n(f)$ et $U_n(f)$ et de corollaires présentant les déductions faites pour les estimateurs de la fonctions de densité et de la fonction de régression.

Théorème 2.1. *On suppose vérifiées les conditions (H.1), (K.1)-(K.3), (G.1)-(G.2) et (F.1)-(F.2). Alors le processus W_n indexé par la famille de fonctions $\mathcal{F}$ satisfait un PGD dans $\mathcal{C}(\mathcal{F})$ avec la vitesse nh_n et une bonne fonction de taux*

$$J(Z) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{y\in B_{Z,\delta}} J_m(Z')$$

où

$$J_m(Z') = \inf \left\{ l_1^*(x,y) : \sum_{j=1}^m x_j\alpha_j(f) + y_j\beta_j(f) = Z'(f) \text{ pour tout } f \in \mathcal{F} \right\}.$$

Théorème 2.2. *Sous les hypothèses du théorème 2.1, le processus vectoriel U_n satisfait un PGD dans $\mathcal{C}(\mathcal{F})\times\mathbb{R}$ avec la vitesse nh_n et une bonne fonction de taux*

$$I(Z,t) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{y\in B_{(Z,t),\delta}} I_m(Z',t'),$$

où

$$I_m(Z',t') = \inf \left\{ l_2^*(x,y,t') : \sum_{j=1}^m x_j\alpha_j(f) + y_j\beta_j(f) = Z'(f) \text{ pour tout } f \in \mathcal{F} \right\}.$$

2.2.3 Corollaires pour les estimateurs de la fonction de densité et de la fonction de régression

Les corollaires suivant donnent respectivement un principe de grandes déviations ponctuel pour l'estimateur à noyau de la fonction de densité et un principe de grandes déviations fonctionnel par rapport à la famille de fonctions $\mathcal{F}$.

Corollaire 2.1. *On suppose vérifiées les conditions (H.1) et (K.1)-(K.3) et (G.1)-(G.2). Alors* $W_n(\mathbb{1})$ *satisfait un PGD dans* $\mathbb{R}$ *avec la vitesse* nh_n *et la bonne fonction de taux* J *définie par*

$$R(t) = \sup_{\gamma \in \mathbb{R}} \left\{ t\gamma - g_X(z_0) \int_{-\infty}^{\infty} (\exp\{\gamma K(u)\} - 1)\, du \right\}.$$

Corollaire 2.2. *On suppose vérifiées les conditions du théorème 2.1. Alors le processus* $m_n(f, z_0)$ *indexé par la famille des fonctions* $\mathcal{F}$ *satisfait un PGD dans* $\mathcal{C}(\mathcal{F})$ *avec la vitesse* nh_n *et une fonction de taux définie par*

$$I_2(Z) = \inf\{I(tZ, t) : t \in \mathbb{R}^*\}.$$

2.3 Application à la sélection de modèles pour le critère des grandes déviations

On considère le processus $V_n : \Omega \times \mathcal{F} \times T \to \mathbb{R}$, où T est une partie compacte de $\mathbb{R}$ telle que pour tout $n \geq 1$ et pour presque tout $\omega \in \Omega$, l'application $(f, z) \to V_n(\omega, f, z)$ soit continue. On suppose que ce processus vérifie un PGD fonctionnel dans $\mathcal{C}(\mathcal{F} \times T)$. Pour une partie A de $\mathcal{C}(T)$ et une sous famille $\mathcal{G}$ de $\mathcal{F}$, posons

$$\Omega_{n,A,\mathcal{G}} = \{\omega \in \Omega : V_n(\omega, ., .) \in \Lambda_{A,\mathcal{G}}\}$$

où

$$\Lambda_{A,\mathcal{G}} = \{Z \in \mathcal{C}(\mathcal{F} \times T) : \forall f \in \mathcal{G}, Z(f, .) \in A\}.$$

La proposition suivante est un outil permettant de faire de la sélection de modèles. Avant de l'énoncer, on définit d'abord ce que l'on entend par ensemble de continuité. Un ensemble O est dit ensemble de continuité d'une fonction I si l'on a $\inf_{x \in \mathring{O}} I(x) = \inf_{x \in \overline{O}} I(x)$.

Proposition 2.1. *On suppose que* $\Lambda_{A,\mathcal{G}_1}$ *et* $\Lambda_{A,\mathcal{G}_2}$ *sont des ensembles de continuité pour la fonction* I*. Alors on a*

$$\lim_{n \to \infty} \epsilon_n \log P(\Omega_{n,A,\mathcal{G}_i}) = -\inf\{I(Z) : Z \in \Lambda_{A,\mathcal{G}_i}\} := -I_{A,\mathcal{G}_i}, i = 1, 2.$$

La comparaison des performances des modèles $(A, \mathcal{G}_1)$ et $(A, \mathcal{G}_2)$ pour le critère des grandes déviations se fait en comparant les quantités $I_{A,\mathcal{G}_1}$ et $I_{A,\mathcal{G}_2}$. On dira alors que le modèle $(A, \mathcal{G}_1)$ est meilleur que le modèle $(A, \mathcal{G}_2)$ si $I_{A,\mathcal{G}_1} \leq I_{A,\mathcal{G}_2}$.

Comme les résultats de grandes déviations obtenus dans ce chapitre sont restraints au cas ponctuel, la comparaison des performances des modèles qui peut être faite ici suppose l'ensemble T réduit au point z_0. En utilisant le théorème 1, on peut comparer les modèles si $\Lambda_{A,\mathcal{G}1}$ et $\Lambda_{A,\mathcal{G}_2}$ sont des ensemble de continuité de J. L'exemple ci-dessous traite le cas d'une famille de fonctions affines et aboutit à la comparaison explicite de modèles.

Exemple

On considère la famille de l'ensemble des fonctions affines donnée par $\mathcal{F} = \{f : [a,b] \to \mathbb{R} : f(t) = xt + y, (x,y) \in \mathbb{R}^2\}$. Pour $f \in \mathcal{F}$, on a $W_n(f) = xA_n + yB_n$, avec

$$A_n = \frac{1}{nh_n}\sum_{i=1}^{n} Y_i K(\frac{X_i - z_0}{h_n}) \text{ et } B_n = \frac{1}{nh_n}\sum_{i=1}^{n} K(\frac{X_i - z_0}{h_n}).$$

Soit $h : \mathbb{R}^2 \to \mathcal{L}(\mathbb{R}^2)$ la bijection canonique qui à (x', y') associe $< (x', y'), (.,.) >$ où $\mathcal{L}(\mathbb{R}^2)$ désigne l'espace des formes linéaires sur $\mathbb{R}^2$. On remarque alors que $W_n = h(A_n, B_n)$ satisfait le même PGD que le processus (A_n, B_n). Au vu de ce qui a précédé, il est clair que le processus (A_n, B_n) satisfait un PGD dans $\mathbb{R}^2$ avec une bonne fonction de taux donnée par

$$l^*(x,y) = \sup_{(\alpha,\beta)\in\mathbb{R}^2} \{x\alpha + y\beta - l(\alpha,\beta)\} \text{ où}$$

$$l(\alpha,\beta)\} = \int_{-\infty}^{\infty}\int_{a}^{b} (\exp\{(\alpha v + \beta)K(u)\} - 1)g(z_0, v)dudv.$$

Soit A une partie de $\mathbb{R}$ et soient $\mathcal{G}_1$ et $\mathcal{G}_2$ deux partie de $\mathbb{R}^2$. On a alors

$$\Lambda_{A,\mathcal{G}_i} = \{(x,y) \in \mathbb{R}^2 : \forall(\alpha,\beta) \in \mathcal{G}_i, x\alpha + y\beta \in A\}, i = 1,2.$$

Pour A un ouvert convexe de $\mathbb{R}$ et $\mathcal{G}_1$ et $\mathcal{G}_2$ deux compactes dans $\mathbb{R}^2$, il suit que $\Lambda_{A,\mathcal{G}_1}$ et $\Lambda_{A,\mathcal{G}_2}$ sont ouverts convexes de $\mathbb{R}^2$. Comme la fonction de taux l^* est convexe alors ces deux ensembles sont des ensembles de continuité de la fonction de taux l^*. Cela nous permet alors de comparer ces deux modèles.

2.4 Preuves

Preuve du Théorème 2.1 Pour montrer ce théorème nous avons besoin du lemme suivant.

Lemme 2.1. *On suppose vérifiées les conditions (H.1), (K.1)-(K.3) et (G.1)-(G.2). Alors le processus (A_n, B_n) défini dans (2.1) satisfait un principe de grandes déviations dans $\mathbb{R}^m \times \mathbb{R}^m$ avec la vitesse nh_n et la bonne fonction de taux l_2^*.*

Preuve du lemme 2.1 Pour la preuve de ce lemme nous allons utiliser le théorème de Gärtner-Ellis. Nous allons calculer d'abord la limite des fonction génératrices des moments et montrer ensuite, en utilisant nos conditions, que cette limite est définie et differentiable partout. Pour $(\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^m$, nous allons calculer

$$\lim_{n\to\infty} \frac{1}{nh_n} \log E\left[\exp\left\{nh_n \left\langle(\alpha,\beta),(A_n,B_n)\right\rangle\right\}\right].$$

Comme les (X_i, Y_i) sont i.i.d., alors on a

$$\begin{aligned}
E\left[\exp\left\{nh_n \left\langle(\alpha,\beta),(A_n,B_n)\right\rangle\right\}\right] &= E\left[\exp\left\{nh_n \sum_{j=1}^{m}\left\{(\alpha_j A_n^j + \beta_j B_n^j)\right\}\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^{n}\sum_{j=1}^{m}(\alpha_j Y_i + \beta_j)\mathbb{1}_{[t_{j-1},t_j]}(Y_i) K\left(\frac{X_i - z_0}{h_n}\right)\right\}\right] \qquad (2.3)\\
&= \left[E\left[\exp\left\{\sum_{j=1}^{m}(\alpha_j Y + \beta_j)\mathbb{1}_{[t_{j-1},t_j]}(Y) K\left(\frac{X - z_0}{h_n}\right)\right\}\right]\right]^n.
\end{aligned}$$

En prenant $u = (x - z_0)/h_n$, $v = y$, il en résulte que

$$\begin{aligned}
&E \exp \sum_{j=1}^{m} (\alpha_j Y + \beta_j)\, \mathbb{1}_{[t_{j-1},t_j]}(Y) K\left(\frac{X - z_0}{h_n}\right) \\
&= 1 + \int_{-\infty}^{\infty}\int_{a}^{b}\left(\exp\left\{\sum_{j=1}^{m}(\alpha_j y + \beta_j)\,\mathbb{1}_{[t_{j-1},t_j]}(y) K\left(\frac{x - z_0}{h_n}\right)\right\} - 1\right) \\
&\qquad \times g(x,y)dxdy \\
&= 1 + h_n \int_{-\infty}^{\infty}\int_{a}^{b}\left(\exp\left\{\sum_{j=1}^{m}(\alpha_j v + \beta_j)\mathbb{1}_{[t_{j-1},t_j]}(v) K(u)\right\} - 1\right) \\
&\qquad \times g(h_n u + z_0, v)dudv
\end{aligned}$$

La condition (G.1) implique que g est derivable par rapport à la première variable, le développement en séries de Taylor donne

$$g(h_n u + z_0, v) = g(z_0, v) + h_n u \frac{\partial g}{\partial u}(u,v)|_{u=z_0^*(u)}$$

où $z_0^*(u)$ est un point entre z_0 et $z_0 + h_n u$.

Par conséquent,

$$
\begin{aligned}
&E \exp \sum_{j=1}^{m} \left(\alpha_j Y + \beta_j\right) \mathbb{1}_{[t_{j-1},t_j]}(Y) K\left(\frac{X - z_0}{h_n}\right) \\
&= 1 + h_n \int_{-\infty}^{\infty} \int_a^b \left(\exp\left\{ \sum_{j=1}^{m} (\alpha_j v + \beta_j) \mathbb{1}_{[t_{j-1},t_j]}(v) K(u) \right\} - 1 \right) \\
&\qquad \times g(z_0, v) du dv \\
&\quad + h_n^2 \int_{-\infty}^{\infty} \int_a^b u \left(\exp\left\{ \sum_{j=1}^{m} (\alpha_j v + \beta_j) \mathbb{1}_{[t_{j-1},t_j]}(v) K(u) \right\} - 1 \right) \\
&\qquad \times \frac{\partial g}{\partial u}(u, v)|_{u = z_0^*(u)} du dv \\
&= 1 + h_n \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} \left(\exp\left\{ (\alpha_j v + \beta_j) K(u) \right\} - 1 \right) g(z_0, v) du dv \\
&+ h_n^2 \sum_{j=1}^{m} \int_{-\infty}^{\infty} \int_{t_{j-1}}^{t_j} u \left(\exp\left\{ (\alpha_j v + \beta_j) K(u) \right\} - 1 \right) \frac{\partial g}{\partial u}(u, v)|_{u = z_0^*(u)} du dv.
\end{aligned}
$$

En utilisant les condition (K.2) et (G.2), on obtient

$$
E \exp \sum_{j=1}^{m} \left(\alpha_j Y + \beta_j\right) \mathbb{1}_{[t_{j-1},t_j]}(Y) K\left(\frac{X - z_0}{h_n}\right) = 1 + h_n l_1(\alpha, \beta) + O(h_n^2). \tag{2.4}
$$

Tenant compte de la condition (K.1), (2.3), (2.4) et le fait que $\log(1 + u) \simeq u$ au voisinage de 0, on obtient

$$
\lim_{n \to \infty} \frac{1}{n h_n} \log E \exp n h_n < (\alpha, \beta), (A_n, B_n) > = l_1(\alpha, \beta) \tag{2.5}
$$

La condition (K.1) assure que la fonction l_1 est finie sur $\mathbb{R}^m \times \mathbb{R}^m$. Pour pouvoir utiliser le théorème de Gärtner-Ellis il suffit de montrer qu'elle est aussi differentiable sur $\mathbb{R}^m \times \mathbb{R}^m$. En utilisant le théorème de convergence dominée et la condition (K.3) on montre la differentiabilité. En outre, ona

$$
\frac{\partial l_1}{\partial \alpha_j}(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{t_j}^{t_{j-1}} v K(u) \left(\exp\left\{ (\alpha_j v + \beta_j) K(u) \right\} \right) g(z_0, v) du dv
$$

et

$$
\frac{\partial l_1}{\partial \beta_j}(\alpha, \beta) = \int_{-\infty}^{\infty} \int_{t_j}^{t_{j-1}} K(u) \left(\exp\left\{ (\alpha_j v + \beta_j) K(u) \right\} \right) g(z_0, v) du dv.
$$

Ceci nous permet d'appliquer le théorème de Gärtner-Ellis. Il en résulte alors que (A_n, B_n) satisfait un principe de grandes déviations avec la vitesse $n h_n$ et la bonne fonction de taux

$$
l_1^*(x, y) = \sup_{(\alpha, \beta) \in \mathbb{R}^m \times \mathbb{R}^m} \left\{ < (x, y), (\alpha, \beta) > - l_1(\alpha, \beta) \right\}.
$$

Ce qui achève la preuve du lemme. $\square$

Retournons maintenant à la preuve du Théorème. Nous allons construire pour $m \geq 1$ un processus $W_n^m(f)$ indexé par la famille de fonctions $\mathcal{F}$ et montrer que, pour tout $m \geq 1$, ce processus satisfait un principe de grandes déviations avec la bonne fonction de taux J_m en utilisant le principe de contraction et le lemme précédent.

Soit $T_m : \mathbb{R}^m \times \mathbb{R}^m \to \mathcal{C}(\mathcal{F})$ une fonction définie par

$$T_m(x,y)(f) = \sum_{j=1}^{m} (x_j.\alpha_j(f) + y_j.\beta_j(f)) =< (x,y), (\alpha(f), \beta(f)) > .$$

Nous allons montrer que T_m est continue. En utilisant l'inégalité de Cauchy-Schwartz on obtient

$$|T_m(x,y)(f)| \leq \|(\alpha(f), \beta(f))\| \|(x,y)\|$$

Il en résulte que

$$\|T_m(x,y)(f)\|_{\mathcal{C}(\mathcal{F})} \leq \sup_{f\in\mathcal{F}} \|(\alpha(f), \beta(f))\| \|(x,y)\|.$$

Comme T_m est linéaire il suffit donc de montrer que $\sup_{f\in\mathcal{F}} \|(\alpha(f), \beta(f))\| < \infty$. cela suit facilement en utilisant les conditions (F).

Posons

$$W_n^m(f) = \frac{1}{nh_n} \sum_{i=1}^{n} \left(\sum_{j=1}^{m} (\alpha_j(f) Y_i + \beta_j(f)) \, \mathbb{1}_{[t_{j-1}, t_j]}(Y_i) \right) K\left(\frac{X_i - z_0}{h_n} \right).$$

On observe que

$$T_m(A_n, B_n) = W_n^m.$$

Par conséquent, pour $m \geq 1$ fixé W_n^m est l'image de (A_n, B_n) par l'application continue T_m. Il est clair alors, en utilisant le principe de contraction et le lemme précédent, que W_n^m satisfait le principe de grandes déviations avec la vitesse nh_n et une bonne fonction de taux définie, pour $Z \in \mathcal{C}(\mathcal{F})$, par

$$\begin{aligned} J_m(Z) &= \inf \{ l_1^*(x,y) : T_m(x,y) = Z \} \\ &= \inf \left\{ l_1^*(x,y) : \sum_{j=1}^{m} (\alpha_j(f) x_j + \beta_j(f) y_j) = Z(f) \text{ pour tout } f \in \mathcal{F} \right\}. \end{aligned}$$

Nous allons montrer maintenant que W_n^m est une bonne approximation exponentielle de W_n. i.e. Pour tout $\tau > 0$

$$\lim_{m\to\infty} \limsup_{n\to\infty} \frac{1}{nh_n} \log P\left(\|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})} \geq \tau\right) = -\infty. \tag{2.6}$$

Observons que

$$\|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})} \leq \sup_{f\in\mathcal{F}} \|f - f_m\|_\infty \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - z_0}{h_n}\right)$$

où

$$f_m(t) = \sum_{j=1}^{m} (\alpha_j(f)t + \beta_j(f))\, \mathbb{1}_{[t_{j-1},t_j]}(t).$$

On peut montrer facilement que

$$\sup_{f\in\mathcal{F}} \|f - f_m\|_\infty \leq 2w_\mathcal{F}((b-a)m^{-1}),$$

où

$$w_\mathcal{F}(\delta) = \sup_{f\in\mathcal{F}} w_f(\delta) \text{ et } w_f(\delta) = \sup_{|t-s|<\delta} |f(t) - f(s)|.$$

En effet

$$\begin{aligned}
\|f - f_m\|_\infty &\leq \max_{1\leq j\leq m} \sup_{t_{j-1}\leq t\leq t_j} |f(t) - f_m(t)| \\
&\leq \max_{1\leq j\leq m} \sup_{t_{j-1}\leq t\leq t_j} |f(t) - f(t_{j-1})| + \max_{1\leq j\leq m} \sup_{t_{j-1}\leq t\leq t_j} |f(t_{j-1}) - f_m(t)| \\
&\leq 2w_f((b-a)m^{-1})
\end{aligned}$$

Puisque

$$\max_{1\leq j\leq m} \sup_{t_{j-1}\leq t\leq t_j} |f(t) - f(t_{j-1})| \leq w_f((b-a)m^{-1})$$

et

$$\begin{aligned}
|f(t_{j-1}) - f_m(t)| &= |\alpha_j(f)t_{j-1} + \beta_j(f) - f_m(t)| \\
&\leq |f(t_{j-1}) - f(t_j)| \frac{|t - t_{j-1}|}{|t_j - t_{j-1}|} \leq w_f((b-a)m^{-1}).
\end{aligned}$$

Par conséquent

$$\|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})\times\mathbb{R}} \leq 2w_\mathcal{F}((b-a)m^{-1}) \frac{1}{nh_n} \sum_{i=1}^{n} K\left(\frac{X_i - z_0}{h_n}\right).$$

Il en résulte, en utilisant l'inégalité de Tchebychev et le fait que les variables X_i sont i.i.d., que

$$P\left(\|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})} \geq \tau\right) \leq P\left(K\left(\sum_{i=1}^{n}\frac{X_i - z_0}{h_n}\right) \geq \frac{nh_n\tau}{2w_{\mathcal{F}}((b-a)m^{-1})}\right)$$
$$\leq \exp\left\{-\frac{nh_n\tau}{2w_{\mathcal{F}}((b-a)m^{-1})}\right\}\left[E\exp\left\{K\left(\frac{X_i - z_0}{h_n}\right)\right\}\right]^n.$$

On a alors

$$\limsup_{n\to\infty}\frac{1}{nh_n}\log P\left(\|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})} \geq \tau\right) \leq -\frac{\tau}{2w_{\mathcal{F}}((b-a)m^{-1})} + \lim_{n\to\infty}\frac{1}{h_n}\log E\exp\left\{K\left(\frac{X - z_0}{h_n}\right)\right\}. \tag{2.7}$$

D'autre part, en utilisant les conditions (K.1), (K.2) et (G.1)-(G.2), on montre facilement que

$$\lim_{n\to\infty}\frac{1}{h_n}\log E\exp\left\{K\left(\frac{X - z_0}{h_n}\right)\right\} = g_X(z_0)\int_{-\infty}^{\infty}(\exp\{K(u)\} - 1)du < \infty.$$

Ce qui est fait en utilisant les mêmes arguments que ceux du lemme précédent.

Compte tenu donc de la condition (F.2) et de (2.7) on obtient (2.6) en faisant tendre m vers l'infini dans (2.7) puisque $(b-a)m^{-1} \to 0$ quand $m \to \infty$.
Comme l'espace $\mathcal{C}(\mathcal{F}) \times \mathbb{R}$ est un espace polonais le reste de la demonstration est une conséquence du théorème 1.3 (Approximation exponentielle). □

Preuve du Théorème 2.2 La démonstration du théorème 2.2 est similaire à celle du théorème 2.1. En effet on montre d'abord que pour tout $(\alpha,\beta,\gamma) \in \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}$

$$\lim_{n\to\infty}\frac{1}{n_h n}\log E\exp nh_n < (\alpha,\beta,\gamma),(A_n,B_n,W_n(\mathbf{1})) >= l_2(\alpha,\beta,\gamma).$$

Avec les mêmes arguments utilisés dans la démonstration du lemme précédent, on montre que cette fonction est finie et differentiable sur $\mathbb{R}^m\times\mathbb{R}^m\times\mathbb{R}$. Ensuite, en utilisant le théorème de Gärtner-Ellis, on montre que le processus $(A_n, B_n, W_n(\mathbf{1}))$ satisfait le PGD avec la vitesse nh_n et la bonne fonction de taux l_2^*.

On pose

$$U_n^m = (W_n^m, W_n(\mathbf{1}))$$

et soit la fonction $T'_m : \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R} \to \mathcal{C}(\mathcal{F}) \times \mathbb{R}$ définie par

$$T'_m(x,y,z) = (T_m(x,y),z)$$

avec T_m la fonction définie dans la démonstration du théorème 1. En remarquant que la fonction T'_mest continue et que $U_n^m = T'_m(A_n, B_n, W_n(\mathbf{1}))$, on montre, en utilisant le principe de contraction, que U_n^m satisfait un PGD avec la vitesse nh_n et la bonne fonction de taux I_m.

Il reste alors à montrer que U_n^m est une bonne approximation exponentielle de U_n. Cela suit du fait que

$$\|U_n - U_n^m\|_{\mathcal{C}(\mathcal{F})\times\mathbb{R}} = \|W_n - W_n^m\|_{\mathcal{C}(\mathcal{F})}.$$

$\square$

Preuve du corollaire 2.1 La démonstration utilise les mêmes arguments que celle du lemme précédent. On montre que pour tout $\gamma \in \mathbb{R}$

$$\lim_{n\to\infty} \frac{1}{nh_n} \log E \exp nh_n \gamma W_n(\mathbf{1}) = g_X(z_0) \int_{-\infty}^{\infty} (\exp\{\gamma K(u)\} - 1) du.$$

En utilisant les arguments développés dans la démonstration du lemme, on montre que $W_n(\mathbf{1})$ satisfait un PGD avec la vitesse nh_n et la bonne fonction de taux

$$R(t) = \sup_{\gamma\in\mathbb{R}} \left\{ t\gamma - g_X(z_0) \int_{-\infty}^{\infty} (\exp\{\gamma K(u)\} - 1)\, du \right\}.$$

$\square$

Preuve du corollaire 2.2 La démonstration de ce résultat s'effectue de façon directe en utilisant le principe de contraction. La fonction continue utilisée ici est celle définie par

$$\begin{aligned} H : \mathcal{C}(\mathcal{F}) \times \mathbb{R}^* &\to \mathcal{C}(\mathcal{F}) \\ (Z, t) &\to H(Z, t) \end{aligned}$$

Où

$$H(Z, t)(f) = \frac{Z(f)}{t} \text{ pour tout } f \in \mathcal{F}.$$

$\square$

Preuve de la proposition 2.1 Comme le processus V_n satisfait un principe de grandes déviations alors on a

$$\begin{aligned}
-\inf\{I(Z) : Z \in \overset{o}{\Lambda}_{A,\mathcal{G}_i}\} &\leq \liminf_{n\to\infty} \epsilon_n \log P(V_n(.,.) \in \overset{o}{\Lambda}_{A,\mathcal{G}_i}) \\
&\leq \liminf_{n\to\infty} \epsilon_n \log P(\Omega_{n,A,\mathcal{G}_i}) \\
&\leq \limsup_{n\to\infty} \epsilon_n \log P(\Omega_{n,A,\mathcal{G}_i}) \\
&\leq \limsup_{n\to\infty} \epsilon_n \log P(V_n(.,.) \in \overline{\Lambda}_{A,\mathcal{G}_i}) \\
&\leq -\inf\{I(Z) : Z \in \overline{\Lambda}_{A,\mathcal{G}_i}\}
\end{aligned}$$

Puisque $\Lambda_{A,\mathcal{G}_i}$ est un ensemble de continuité de I alors

$$\inf\{I(Z) : Z \in \overset{o}{\Lambda}_{A,\mathcal{G}_i}\} = \inf\{I(Z) : Z \in \overline{\Lambda}_{A,\mathcal{G}_i}\}.$$

Par conséquent

$$\lim_{n\to\infty} \epsilon_n \log P(V_n(.,.) \in \Lambda_{A,\mathcal{G}_i}) = -\inf\{I(Z) : Z \in \Lambda_{A,\mathcal{G}_i}\} := -I_{A,\mathcal{G}_i}, i = 1, 2.$$

$\square$

Chapitre 3

Some uniform large deviation results in nonparametric function estimation

In this chapter we investigate large deviation asymptotics involving classes of nonparametric estimates. We introduce a random process, say Z_n, which allows to derive, using the contraction principle, results for several nonparametric estimates from the large deviations principle stated for Z_n. The usual examples of nonparametric estimates include the histogram density estimate as well as the regressogram. Note that uniform behaviours over classes of density and regression functions as well as over classes of their estimates have been considered.

3.1 Introduction

Let (X, Y) be a $\mathbb{R}^2$-valued random vector defined on a probability space $(\Omega, \mathcal{A}, P)$. Let g be the joint density function of the vector (X, Y). Denote by g_X the marginal density of the random variable X with respect to the Lebesgue measure m. Assume that we observe (X_i, Y_i), $1 \leq i \leq n$, a sequence of i.i.d. random vectors with the same distribution as (X, Y). Consider $\mathcal{P}_n = \{A_{n,j}; j \in \mathbb{Z}\}$ a partition of $\mathbb{R}$ satisfying some conditions that will be specified later on. In this paper, we are concerned by large deviation study of a process that allows us to deduce as a particular cases results pertaining to various estimates in nonparametric function estimation.

In first place, consider a vector process taking values in $\mathbb{R} \times \mathbb{R}_+$ and defined by

$$Z_n(x) = (R_n(x), g_n(x)),$$

where

$$R_n(x) = R_n^f(x) = \frac{1}{nm(A_{n,j})} \sum_{i=1}^{n} f(Y_i) \mathbb{1}_{A_{n,j}}(X_i) \quad \text{if} \quad x \in A_{n,j},$$

$$g_n(x) = \frac{1}{nm(A_{n,j})} \sum_{i=1}^{n} \mathbb{1}_{A_{n,j}}(X_i) \quad \text{if} \quad x \in A_{n,j},$$

$\mathbb{1}_{A_{n,j}}$ denotes the indicator of the set A and f is a measurable real function defined upon $\mathbb{R}$. Taking f in a class of functions $\mathcal{F}$, the f-indexed regression function, denoted by r_f, is defined, for any $x \in \mathbb{R}$, by

$$r_f(x) = E[f(Y)|X = x].$$

It is well-known that r_f exists whenever $E[|f(Y)|] < \infty$. Choices of functions in the class $\mathcal{F}$ allow to perform in a unified way the study of several usual functionals. The first example is given by the most classical regression function for which f is taken as the identity. Whenever $f(x) = x^p$ with $p > 0$, the functional r_f reduces to the conditional moment function. If $f(x) = \mathbb{1}_A(x)$ is the indicator function of a set A, then r_f gives the conditional distribution function of the set A. The estimator (called regressogram) of the f-indexed regression function r_f based on the sample $(X_1, Y_1), ..., (X_n, Y_n)$ and the partition $\mathcal{P}_n$, is defined, for $x \in A_{n,j}$, by

$$r_n(x) = \frac{\sum_{i=1}^n f(Y_i) \mathbb{1}_{A_{n,j}}(X_i)}{\sum_{i=1}^n \mathbb{1}_{A_{n,j}}(X_i)}.$$

The histogram estimator of the density function g_X is defined, for $x \in A_{n,j}$, by $g_n(x)$ defined above.

After the pointwise large deviation principle stated for the process $Z_n(x)$ and its direct implications in stating results for a class of nonparametric function estimates, the main goal of this work involves uniform large deviation results pertaining to nonparametric estimation of some functionals. More precisely, taking the histogram estimation method, we investigate in the first part the impact of uniformity over classes of densities and classes of their estimates. Note, behind this, that uniform results over support of the densities are stated. When dealing with the regressogram, uniformity with respect to the class of functions $\mathcal{F}$ has been also considered.

From partition to partition, the estimates may be very different. It is then worthy to consider uniformity over families of partitions. This has been considered for both the histogram the regressogram. Taking also uniformity over classes of densities or classes of regression functions, these results allow to deduce minimax lower bound in both cases.

The proof of results involving the process $(Z_n(x))$ uses mainly the Gärtner-Ellis Theorem (see, Dembo & Zeitouni (1998)(23)). Straight implications in nonparametric function estimation are deduced by making use of the contraction principle. Chernoff-type results similar to the ones obtained in Louani (1998,1999)(52; 53) have been established under less restrictive conditions since the differentiability hypotheses is relaxed in the pointwise case.

There exists an extensive large deviations literature involving a lot of areas in probability and statistics. We refer to the books of Dembo & Zeitouni (1998)(23) and Deuschel & Stroock (1989) (24) and the references therein for an account of results and methods on the subject. In nonparametric function estimation, several results have been obtained recently. We cite Louani (1998,2003)(52; 49) who stated for the kernel and orthogonal series density estimators respectively pointwise and uniform results. Beirlant *et al.* (2001) (3) obtained L_1 results for the histogram density estimate. In Gao (2003)(35), the author studied moderate and large deviations for the kernel estimator based on independent and identically distributed samples. Berrahou (2003) (5) stated pointwise large deviations principle for the delta-sequence method density estimator. The pointwise and uniform large deviation behaviours of the multivariate kernel density estimator have been studied by Mokkadem *et al.* (2005)(58). They considered the case where the kernel takes negative values. Moreover, Mokkadem *et al.* (2006)(57) established a large deviations upper bound for the kernel mode estimator when the mode corresponds to a strict maximum of a density function. Esaulova (2004)(31) obtained a Chernoff large deviation result for the L_1-error of the empirical measure on partitions and derived, as a corollary, large deviations asymptotics for the L_1-error of the histogram density estimator. We cite also the paper of Worms (2001)(82) which gives Chernoff-type bounds as well as a large deviations principle for the pointwise and uniform errors of the kernel density estimator related to the stationary distribution of a stable Markov chain. Lei & Wu (2005) (46) and Lei (2006)(45) established some large deviation results pertaining to the L_1-view of the kernel density estimation. They considered densities associated to ergodic and reversible Markov processes.
In the regression estimation setting, Louani (1999)(53) established analogous results for the conditional empirical process and the Nadaraya-Watson estimate of the regression function. We cite also the work of Hu (1993)(41) where large deviations asymptotics for the least squares estimator in a nonlinear model with dependent errors are provided. An exponential rate for the large deviation probability of the nearest neighbor estimate of a regression function has been obtained by Wang & Bing Zhang (1989)(78). Osmou-

khina (2001)(62) established a large deviations result for the kernel density estimator with applications to the symmetry testing. In the same framework, Gyorfi (2004)(40) considered the problem of testing an unknown density function by using the Hellinger distance.

3.2 Results

3.2.1 Hypothesis and notations

Introduce now hypotheses and notations necessary to establish our results.
(H.1) $\max_{j\in\mathbb{Z}} m(A_{n,j}) \to 0$ and $n \min_{j\in\mathbb{Z}} m(A_{n,j}) \to \infty$ as $n \to \infty$.
(H.2) For any $x \in \mathbb{R}$, there exists $\epsilon > 0$ such, that for any $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} e^{tf(y)} \sup_{|z-x|<\epsilon} g(z,y)dy < \infty.$$

(H'.2) g is differentiable with respect to the first coordinate and for any $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} e^{tf(y)} \sup_{x\in\mathbb{R}} \left| \frac{\partial g}{\partial x}(x,y) \right| dy < \infty.$$

(H.3) For any $x \in \mathbb{R}$ and any $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} |f(y)| e^{tf(y)} g(x,y)dy < \infty.$$

(H.4) For any $x \in \mathbb{R}$ and any $t \in \mathbb{R}$,

$$\int_{-\infty}^{\infty} |f(y)|^2 e^{tf(y)} g(x,y)dy < \infty.$$

Set now

$$I_x(t,s) = \int_{-\infty}^{\infty} \left(e^{tf(y)+s} - 1 \right) g(x,y)dy,$$

and

$$\Lambda_x(u,v) = \sup_{t,s\in\mathbb{R}} \left\{ tu + sv - I_x(t,s) \right\}.$$

Comments. Conditions (H.2) and (H.3) are of classical nature in large deviations results. Notice that the condition (H'.2) together with a condition less restrictive than the condition (H.2), namely the condition
(H) $\forall t \in \mathbb{R}$ and $\forall x \in \mathbb{R}$

$$\int_{-\infty}^{\infty} e^{tf(y)} g(x,y)dy < \infty,$$

imply that the condition (H.2) holds. The condition (H.4) which is required to establish Chernoff type results is stronger than the condition (H.3). (H.4) is required to establish Chernoff-type results. The function I_x, defined above, is the limit of the normalized logarithm of the Laplace transform associated with the process $Z_n(x)$ and Λ_x is its Legendre transform which actually is a rate function.

If h is a real function, then h^{-1} denotes its inverse which is defined on the range of h by

$$h^{-1}(t) = \inf\{s : h(s) \geq t\}.$$

Below, we use the following notation

$$\theta_{x_0}(u) = h_{x_0}^{-1}(u)$$

where

$$h_x(u) = \frac{\int_{-\infty}^{\infty} f(y)e^{tf(y)}g(x,y)dy}{\int_{-\infty}^{\infty} e^{tf(y)}g(x,y)dy}$$

and denote by H_{x_0} the interior of the range of h_{x_0}.

3.2.2 Pointwise LPDs

For the pointwise results given in theorems 3.1, 3.2, 3.3 and corollaries 3.1, 3.2 we suppose that x_0 belongs to $A_{n,j}$ an element of the partition $\mathcal{P}_n$.

Theorem 3.1 establishes a pointwise large deviation principle for the vector process $Z_n(x_0)$.

Theorem 3.1. *Suppose that conditions (H.1), (H.2) and (H.3) are satisfied. Then the process $Z_n(x_0)$ satisfies a large deviations principle in $\mathbb{R}^2$ with the rate function Λ_{x_0} and the rate $nm(A_{n,j})$, i.e., for any Borel set $A \subset \mathbb{R}^2$, we have*

$$\begin{aligned}
-\inf_{(u,v)\in A^o} \Lambda_{x_0}(u,v) &\leq \liminf_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(Z_n(x_0) \in A\right) \\
&\leq \limsup_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(Z_n(x_0) \in A\right) \\
&\leq -\inf_{(u,v)\in \overline{A}} \Lambda_{x_0}(u,v),
\end{aligned}$$

where

$$\Lambda_{x_0}(u,v) = \begin{cases} u\theta_{x_0}(\frac{u}{v}) + v\log v - v + g_X(x_0) \\ \quad - v\log\left(\int_{-\infty}^{\infty} e^{\theta_{x_0}(\frac{u}{v})f(y)} g(x_0,y)dy\right) & \text{if, } (v,\frac{u}{v}) \in (0,\infty)\times H_{x_0}, \\ g_X(x_0) & \text{if}(u,v) = (0,0), \\ \infty & \text{elsewhere.} \end{cases}$$

We denote by A^o and $\overline{A}$ the interior and the closure of the set A respectively.

Theorem 3.2 and Corollary 3.2, below, give respectively a large deviations principle and a Chernoff-type result for the histogram density estimator.

Theorem 3.2. *Suppose that the condition (H.1) is satisfied. Then $g_n(x_0)$ satisfies a large deviation principle with the rate $nm(A_{n,j})$ and the rate function*

$$\delta_{x_0}(v) = \begin{cases} v\log\left(\frac{v}{g_X(x_0)}\right) + g_X(x_0) - v & \text{if } v \geq 0, \\ \infty & \text{elsewhere.} \end{cases}$$

Corollary 3.1. *Suppose that hypotheses of Theorem 3.2 are satisfied. Then, for any $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|g_n(x_0) - g_X(x_0)| > a\right) = -\beta_{x_0}(a),$$

where

$$\beta_{x_0}(a) = (g_X(x_0) + a)\log\left(1 + \frac{a}{g_X(x_0)}\right) - a.$$

The next theorem and corollary give respectively a large deviations principle and a Chernoff-type result for the regressogram.

Theorem 3.3. *Suppose that conditions of Theorem 3.1 are satisfied. Then the process $r_n(x_0)$ satisfies a large deviations principle in $\mathbb{R}$ with the rate $nm(A_{n,j})$ and the rate function Γ given by*

$$\Gamma_{x_0}(\lambda) = \begin{cases} \int_{-\infty}^{\infty}\left(1 - e^{\theta_{x_0}(\lambda)(f(y)-\lambda)}\right) g(x_0,y)dy & \text{if } \lambda \in H_{x_0}, \\ \infty & \text{elsewhere.} \end{cases}$$

Corollary 3.2. *Assume that hypotheses (H.1), (H.2) and (H.4) are satisfied. Then, for any $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|r_n(x_0) - r(x_0)| > a\right) = -\gamma_{x_0}(a)$$

where

$$\gamma_{x_0}(a) = \min\left\{\gamma^{+}_{x_0}(a), \gamma^{-}_{x_0}(a)\right\},$$

and

$$\gamma^{+}_{x_0}(a) = \begin{cases} \int_{-\infty}^{\infty}\left(1 - e^{\theta_{x_0}(r(x_0)+a)(f(y)-r(x_0)-a)}\right) g(x_0,y)dy & \text{if } r(x_0) + a \in H_{x_0}, \\ \infty & \text{elsewhere,} \end{cases}$$

$$\gamma^{-}_{x_0}(a) = \begin{cases} \int_{-\infty}^{\infty}\left(1 - e^{-\theta_{x_0}(r(x_0)-a)(f(y)-r(x_0)+a)}\right) g(x_0,y)dy & \text{if } r(x_0) - a \in H_{x_0}, \\ \infty & \text{elsewhere.} \end{cases}$$

3.2.3 Uniform large deviations results

The uniform large deviation result for the density estimator given in Theorem 3.4 is stated under some additional hypotheses. Set

$$\inf_{j\in\mathbb{Z}} m(A_{n,j}) = h_{n,1} \quad \text{and} \quad \sup_{j\in\mathbb{Z}} m(A_{n,j}) = h_{n,2}.$$

Define

$$\beta(a) = (M + a)\log\left(1 + \frac{a}{M}\right) - a,$$

where $M = \sup_{x\in\mathbb{R}} g_X(x)$. Introduce now the following hypotheses,

(A.1) There exists a sequence of positive real numbers (H_n) tending to infinity such that $P(|X| > H_n) \le \epsilon_n$ and that for some $\tau > \beta(a)$,

$$\lim_{n\to\infty} \frac{\epsilon_n exp\{\tau n h_{n,2}\}}{nh_{n,2}} = 0.$$

(A.2) $\lim_{n\to\infty} \frac{1}{nh_{n,2}} \log\left(\frac{H_n}{h_{n,2}}\right) = 0.$

(A.3) $\lim_{n\to\infty} \frac{h_{n,1}}{h_{n,2}} = 1.$

Theorem 3.4. *Assume that conditions (H.1) and (A.1)-(A.3) are satisfied and that the function g_X is continuous. Then, for all $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{nh_{n,2}} \log P(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a) = -\beta(a).$$

In order to give in Theorem 3.5 below a uniform large deviation result for the regressogram, we have to suppose that X is a random variable with bounded support S and that $\mathcal{P}_n = \{A_{n,j} : 1 \leq j \leq m_n\}$ is a partition of S satisfying the following condition

(B.1) $\lim_{n\to\infty} \frac{1}{nh_{n,2}} \log m_n = 0.$

Set

$$\gamma(a) = \inf_{x\in S} \gamma_x(a).$$

Theorem 3.5. *Suppose that conditions (H.1), (H.2) ,(H'.2), (H.4), (A.3) and (B.1) are satisfied and that the functions r_f and γ are continuous. Then, for any $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{nh_n} \log P(\sup_{x\in S} |r_n(x) - r_f(x)| > a) = -\gamma(a).$$

Consider now $\mathcal{F}$ as a class of functions f and, for any $\epsilon > 0$, define the following number

$$\mathcal{H}_{\mathcal{F}}(\epsilon) = \min\{m : \text{ there exist } f_1, ..., f_m, \text{ such that } \forall f \in \mathcal{F}, \\ \|f - f_j\|_\infty < \epsilon, \text{ for some } 1 \leq j \leq m\}.$$

Set now $F(y) = \sup_{f\in\mathcal{F}} |f(y)|$. Introduce now the following condition upon the class $\mathcal{F}$.

(B.2) $\epsilon \log \mathcal{H}_{\mathcal{F}}(\epsilon) \to 0$ as $\epsilon \to 0$.

3.2.4 Uniformity over the class $\mathcal{F}$

The uniformity upon the class $\mathcal{F}$ is treated in Theorem 3.6. Actually, the function γ defined above depend on f and may be denoted by γ^f. Set now

$$\gamma_{\mathcal{F}}(a) = \inf_{f\in\mathcal{F}} \gamma^f(a).$$

Theorem 3.6. *Suppose that (H.1), (A.3), (B.1) and (B.2) are satisfied and that the function F satisfies the conditions (H.2), (H'.2) and (H.4). Moreover Assume that, the functions r^F and $\gamma_{\mathcal{F}}$ are continuous. Then, for all $a > 0$, we have*

$$\lim_{n\to\infty} \frac{1}{nh_n} \log P\left(\sup_{f\in\mathcal{F}} \sup_{x\in S} |r_n(x) - r_f(x)| > a\right) = -\gamma_{\mathcal{F}}(a).$$

3.2.5 Minimax theorems

In the sequel we will construct a family of partitions which allows to define a class of histogram density estimators. In this respect, let $\mathcal{P}_n^0 = \left\{A_{n,j}^0 : j \in \mathbb{Z}\right\}$ and $\mathcal{P}_n^1 = \left\{A_{n,j}^1 : j \in \mathbb{Z}\right\}$ be two partitions of $\mathbb{R}$ such that, for any $j \in \mathbb{Z}$ there exists $l \in \mathbb{Z}$ such that $A_{n,j}^0 \subset A_{n,l}^1$. Introduce now some conditions necessary to establish minimax results for the density and the regression functions.

(F.1) $\lim_{n\to\infty} \max_{j\in\mathbb{Z}} m(A_{n,j}^1) = 0$ and $\lim_{n\to\infty} n \min_{j\in\mathbb{Z}} m(A_{n,j}^0) = \infty$.

(F.2) $\lim_{n\to\infty} \dfrac{\min_{j\in\mathbb{Z}} m(A_{n,j}^0)}{\max_{j\in\mathbb{Z}} m(A_{n,j}^0)} = 1$ and $\lim_{n\to\infty} \dfrac{\min_{j\in\mathbb{Z}} m(A_{n,j}^1)}{\max_{j\in\mathbb{Z}} m(A_{n,j}^1)} = 1$.

(F.3) $\lim_{n\to\infty} \dfrac{\min_{j\in\mathbb{Z}} m(A_{n,j}^0)}{\max_{j\in\mathbb{Z}} m(A_{n,j}^1)} = \alpha > 0$.

For any $n \geq 1$, let $\mathcal{F}_n$ be the family of all partitions $\mathcal{P}_n = \{A_{n,j} : j \in \mathbb{Z}\}$ of $\mathbb{R}$ such that, for any $j \in \mathbb{Z}$, there exist $k, l \in \mathbb{Z}$ such that, $A_{n,j}^0 \subset A_{n,k} \subset A_{n,l}^1$.

Define the functions

$$\beta_{\alpha,t}^+(a) = (t+a)\log\left(\frac{t+a}{\alpha t}\right) - (1-\alpha)t - a,$$

$$\beta_{\alpha,t}^-(a) = \begin{cases} \alpha(t-a)\log\left(\alpha\frac{t-a}{t}\right) - (1-\alpha)t + \alpha a & \text{if } 0 \leq a \leq t, \\ \infty & \text{elswhere,} \end{cases}$$

$$\beta_{\alpha,t}(a) = \min\left\{\beta_{\alpha,t}^+(a), \beta_{\alpha,t}^-(a)\right\} \qquad \text{and} \qquad \beta_\alpha^g(a) = \inf_{x\in\mathbb{R}} \beta_{\alpha,g_X(x)}(a).$$

The following Theorem establishes an uniform result over the family $\mathcal{F}_n$ of partitions $\mathcal{P}_n$ of $\mathbb{R}$.

Theorem 3.7. *Set* $\beta_n^1 = \max_{j\in\mathbb{Z}} m(A_{n,j}^1)$*. Suppose that* $\sup_{x\in\mathbb{R}} g_X(x) := M < \infty$ *and that the assumptions (F.1)-(F.3) are satisfied. Then, for any* $a > 0$*, we have*

$$\lim_{n\to\infty} \frac{1}{n\beta_n^1} \log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n} \sup_{x\in\mathbb{R}} |g_n^{\mathcal{P}_n}(x) - g_X(x)| > a\right) \geq -\beta_\alpha^g(a),$$

where $g_n^{\mathcal{P}_n}$ *is the histogram based on the partition* $\mathcal{P}_n$*.*

Consider now a class of density functions Θ. Let $\mathcal{G}_n$ be the class of estimators g_n defined by taking a partition $\mathcal{P}_n$ from the family $\mathcal{F}_n$. The following corollary gives a minimax result over the class Θ of densities and the class $\mathcal{G}_n$ of histogram density estimators.

Corollary 3.3. *Suppose that* $\sup_{g_X\in\Theta} \sup_{x\in\mathbb{R}} g_X(x) := M_\Theta < \infty$ *and that the conditions (F.1)-(F.3) are satisfied. Then, for all* $a > 0$*, we have*

$$\lim_{n\to\infty} \inf_{g_n\in\mathcal{G}_n} \sup_{g_X\in\Theta} \frac{1}{n\beta_n^1} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) \geq -\beta_\alpha^\Theta(a)$$

where

$$\beta_\alpha^\Theta(a) = \inf_{g_X\in\Theta} \beta_\alpha^g(a).$$

Similarly, we obtain versions of the last two results for the regressogram. In order to be more explicit on the matter, introduce the following functions

$$I_{\alpha,x,a}^+(t) = \alpha \int_{-\infty}^{\infty} \left(e^{t(f(y)-r_f(x)-a)} - 1\right) g(x,y)dy + (1-\alpha)g_X(x)\left(e^{-t(r_f(x)+a)} - 1\right),$$

$$I_{\alpha,x,a}^-(t) = \alpha \int_{-\infty}^{\infty} \left(e^{t(f(y)-r_f(x)+a)} - 1\right) g(x,y)dy + (1-\alpha) \int_{-\infty}^{\infty} \left(e^{tf(y)} - 1\right) g(x,y)dy,$$

$$\gamma_{\alpha,x}^\pm(a) = -\inf_{t\in\mathbb{R}} I_{\alpha,x,a}^\pm(t),$$

$$\gamma_{\alpha,x}(a) = \min\left\{\gamma_{\alpha,x}^+(a), \gamma_{\alpha,x}^-(a)\right\} \qquad \text{and} \qquad \gamma_\alpha^r(a) = \inf_{x\in\mathbb{R}} \gamma_{\alpha,x}(a).$$

Theorem 3.8. *If the function* f *is positive, and if (F.1)-(F.3), (H.2), (H'.2) and (H.3) hold. Then, we have for any* $a > 0$

$$\liminf_{n\to\infty} \frac{1}{n\beta_n^1} \log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n} \sup_{x\in\mathbb{R}} |r_n^{\mathcal{P}_n}(x) - r(x)| > a\right) \geq -\gamma_\alpha^r(a),$$

where $r_n^{\mathcal{P}_n}$ *is the regressogram based on the partition* $\mathcal{P}_n$*.*

Consider now a class $\mathcal{R}$ of regression functions r. Denote by $\mathcal{R}_n$ the class of regressograms r_n defined by taking a partition $\mathcal{P}_n$ from the family $\mathcal{F}_n$. The following result gives a minimax lower bound.

Corollary 3.4. *Under hypotheses of Theorem 3.8, we have for any $a > 0$*

$$\liminf_{n\to\infty} \inf_{r_n\in\mathcal{R}_n} \sup_{r\in\mathcal{R}} \frac{1}{n\beta_n^1} \log P\left(\sup_{x\in\mathbb{R}} |r_n(x) - r(x)| > a\right) \geq - \inf_{r\in\mathcal{R}} \gamma_\alpha^r(a).$$

3.3 Proofs

Proof of Theorem 3.1
The proof uses the Gärtner-Ellis theorem. For any $(t,s) \in \mathbb{R}^2$, the limit of the normalized logarithm of the Laplace transform of the process $Z_n(x_0)$ is given by

$$I_{x_0}(t,s) = \lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log E\left[\exp\left\{n\lambda A_{n,j} \left\langle (t,s), Z_n(x_0)\right\rangle\right\}\right]. \tag{3.1}$$

Set

$$\psi_{x_0}(t,s) = E\left[\exp\left\{n\lambda A_{n,j} \left\langle (t,s), Z_n(x_0)\right\rangle\right\}\right].$$

Since $(X_i, Y_i)_{i\geq 1}$ are i.i.d., we have clearly

$$\begin{aligned}
\psi_{x_0}(t,s) &= E\left[\exp\left\{n\lambda A_{n,j} \left\langle (t,s), Z_n(x_0)\right\rangle\right\}\right] \\
&= E\left[\exp\left\{\sum_{i=1}^n (tf(Y_i)+s)\,\mathbb{1}_{A_{n,j}}(X_i)\right\}\right] \\
&= E\left[\exp\left\{(tf(Y)+s)\,\mathbb{1}_{A_{n,j}}(X)\right\}\right]^n = \left[\varphi_{x_0}(t,s)\right]^n.
\end{aligned}$$

Thus,

$$\begin{aligned}
\varphi_{x_0}(t,s) &= 1 + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \left(\exp\left\{(tf(y)+s)\mathbb{1}_{A_{n_j}}(x)\right\} - 1\right) g(x,y)dxdy \\
&= 1 + \int_{-\infty}^{\infty} \left(\exp\left\{tf(y)+s\right\} - 1\right)\left(\int_{A_{n,j}} g(x,y)dx\right) dy.
\end{aligned}$$

By the conditions (H.1)-(H.2), we have

$$\int_{-\infty}^{\infty} \left(\exp\left\{tf(y)+s\right\} - 1\right)\left(\int_{A_{n,j}} g(x,y)dx\right) dy = o(1).$$

Taylor series expansion of $\log(1+u)$ in a neighborhood of $u = 0$ gives

$$\frac{1}{nm(A_{n,j})} \log \psi_{x_0}(t,s) = \int_{-\infty}^{\infty} \left(\exp\left\{tf(y)+s\right\} - 1\right)\left(\frac{1}{m(A_{n,j})}\int_{A_{n,j}} g(x,y)dx\right) dy.$$

mean value theorem implies that, for almost all $y \in \mathbb{R}$, we have

$$\lim_{n\to\infty} \frac{1}{m(A_{n,j})} \int_{A_{n,j}} g(x,y)dx = g(x_0,y).$$

It follows by the dominated convergence Theorem that

$$\lim_{n\to\infty} \frac{1}{nm(A_{n,j})} \log \psi_{x_0}(t,s) = \int_{-\infty}^{\infty} \left(\exp\{tf(y)+s\} - 1\right) g(x_0,y)dy.$$

This proves the statement (3.1). In order to use the Gärtner-Ellis Theorem, we need to prove that the normalized logarithm of the Laplace transform has an everywhere finite and differentiable limit. The condition (H.2) implies that this function is finite everywhere. The dominated convergence Theorem implies the differentiability of I_{x_0} and gives

$$\frac{\partial I_{x_0}}{\partial t}(t,s) = e^s \int_{-\infty}^{\infty} f(y)e^{tf(y)}g(x_0,y)dy$$

and

$$\frac{\partial I_{x_0}}{\partial s}(t,s) = e^s \int_{-\infty}^{\infty} e^{tf(y)}g(x_0,y)dy.$$

Then, by Gärtner-Ellis Theorem the process $Z_n(x_0)$ satisfies a large deviation principle with the rate $nm(A_{n,j})$ and the rate function

$$\Lambda_{x_0}(u,v) = \sup_{(t,s)\in\mathbb{R}^2} \left\{ut + vs - I_{x_0}(t,s)\right\}.$$

Now let us show that

$$\begin{aligned}\Lambda_{x_0}(u,v) &= \sup_{(t,s)\in\mathbb{R}^2} \left\{ut + vs - I_{x_0}(t,s)\right\} \\ &= \begin{cases} u\theta_{x_0}(\frac{u}{v}) + v\log v - v + g_X(x_0) & \\ \quad - v\log\left(\int_{-\infty}^{\infty} e^{\theta_{x_0}(\frac{u}{v})f(y)}g(x_0,y)dy\right) & \text{if, } (v,\frac{u}{v}) \in (0,\infty)\times H_{x_0} \\ g_X(x_0) & \text{if, } (u,v) = (0,0) \\ \infty & \text{elsewhere.}\end{cases}\end{aligned} \tag{3.2}$$

For $v > 0$ and $u/v \in H_{x_0}$ fixed, set

$$\Phi_{u,v}(t,s) = ut + vs - I_{x_0}(t,s).$$

Clearly the function $\Phi_{u,v}$ is concave and differentiable. Therefore, the supremum of $\Phi_{u,v}$ is reached at the point (t,s) which realizes

$$u = e^s \int_{-\infty}^{\infty} f(y)e^{tf(y)}g(x_0,y)dy \quad \text{and} \quad v = e^s \int_{-\infty}^{\infty} e^{tf(y)}g(x_0,y)dy.$$

This is equivalent to

$$\frac{u}{v} = h_{x_0}(t) \quad \text{and} \quad v = e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0, y) dy. \tag{3.3}$$

Therefore, the only point (t, s) which satisfies the equation (3.3) is

$$t = \theta_{x_0}(\frac{u}{v}) \quad \text{and} \quad s = \log \left(\frac{v}{\int_{-\infty}^{\infty} e^{\theta_{x_0}(\frac{u}{v}) f(y)} g(x_0, y) dy} \right).$$

It follows, for $v > 0$ and $u/v \in H_{x_0}$ fixed, that

$$\begin{aligned}\sup_{(t,s)\in\mathbb{R}^2} \{ut + vs - I_{x_0}(t, s)\} = u\theta_{x_0}(\frac{u}{v}) + v \log v - v + g_X(x_0) \\ - v \log \left(\int_{-\infty}^{\infty} e^{\theta_{x_0}(\frac{u}{v}) f(y)} g(x_0, y) dy \right).\end{aligned}$$

Clearly, we have

$$\Lambda_{x_0}(0, 0) = g_X(x_0).$$

Indeed,

$$g_X(x_0) - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0, y) dy \leq \Lambda_{x_0}(0, 0) \leq g_X(x_0).$$

We obtain the result by tending $s \to -\infty$ in this first member.

At this stage, it remains to show that for $(v, u/v) \notin (0, \infty) \times H_{x_0} \cup \{(0, 0)\}$,

$$\sup_{(t,s)\in\mathbb{R}^2} \{ut + vs - I_{x_0}(t, s)\} = \infty.$$

We have to consider several cases.
1) For $v = 0$ and $u \neq 0$

$$\Lambda_{x_0}(u, 0) = \sup_{(t,s)\in\mathbb{R}^2} \{ut - I_{x_0}(t, s)\} \geq ut - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0, y) dy + g_X(x_0).$$

Let us make $s \to -\infty$ and $ut \to \infty$ in the second member. Then,

$$\begin{aligned}\Lambda_{x_0}(u, 0) &= \sup_{(t,s)\in\mathbb{R}^2} \{ut - I_{x_0}(t, s)\} \\ &\geq \lim_{ut\to\infty} \lim_{s\to-\infty} \left(ut - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0, y) dy + g_X(x_0) \right) \\ &= \infty.\end{aligned}$$

2) For $v < 0$, observe that

$$\begin{aligned}
\Lambda_{x_0}(u,v) &\geq ut + g_X(x_0) + \sup_{s\in\mathbb{R}} \left\{ vs - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0,y)dy \right\} \\
&\geq ut + g_X(x_0) + \lim_{s\to-\infty} \left\{ vs - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0,y)dy \right\} \\
&= \infty
\end{aligned}$$

3) For $v > 0$ and $u/v \notin H_{x_0}$, set $H_{x_0} = (m, M)$ where

$$m = \inf_{t\in\mathbb{R}} h_{x_0}(t) \text{ and } M = \sup_{t\in\mathbb{R}} h_{x_0}(t)$$

we limit ourselves to the case $v > 0$ and $u/v > M$. We have then

$$\begin{aligned}
\Lambda_{x_0}(u,v) &= \sup_{(t,s)\in\mathbb{R}^2} \{ut + vs - I_{x_0}(t,s)\} \\
&\geq ut + vs - e^s \int_{-\infty}^{\infty} e^{tf(y)} g(x_0,y)dy + g_X(x_0).
\end{aligned}$$

In the second member, we substitute s by

$$\log\left(\frac{v}{\int_{-\infty}^{\infty} e^{tf(y)} g(x_0,y)dy}\right).$$

Thus,

$$\begin{aligned}
\Lambda_{x_0}(u,v) &\geq v\log v - v + g_X(x_0) + ut - v\log\left(\int_{-\infty}^{\infty} e^{tf(y)} g(x_0,y)dy\right) \\
&\geq v\log v - v + g_X(x_0) + (u - vM)t - \log g_X(x_0).
\end{aligned} \tag{3.4}$$

We achieve the proof of the statement (3.2) by making t go to the infinite in (3.4). □

Proof of Theorem 3.2
Taking the function f as the null function in Theorem 3.1, it follows that, the process $(0, g_n(x))$ satisfies a large deviation principle with the rate $nm(A_{n,j})$ and the rate function

$$\begin{aligned}
\Lambda_{x_0}(u,v) &= \sup_{(t,s)\in\mathbb{R}^2} \{ut + vs - g_X(x_0)(e^s - 1)\} \\
&= \begin{cases} v\log\left(\frac{v}{g_X(x_0)}\right) + g_X(x_0) - v & \text{if } v \geq 0 \text{ and } u = 0 \\ \infty & \text{elsewhere.} \end{cases}
\end{aligned}$$

Using the continuous function from $\mathbb{R}^2$ into $\mathbb{R}$ given by $(u,v) \to v$, the contraction principle into that $g_n(x_0)$ satisfies a large deviation principle with the rate $nm(A_{n,j})$

and the rate function

$$\begin{aligned}\delta_{x_0}(v) &= \inf\left\{\Lambda_{x_0}(u,v)\ : u \in \mathbb{R}\right\} \\ &= \begin{cases} v\log\left(\frac{v}{g_X(x_0)}\right) + g_X(x_0) - v & \text{if } v \geq 0 \\ \infty & \text{elsewhere.}\end{cases}\end{aligned}$$

This achieves the proof. □

Proof of Corollary 3.1
Theorem 3.2 gives

$$\begin{aligned} -\inf_{|v-g_X(x_0)|>a}\delta_{x_0}(v) &\leq \liminf_{n\to\infty}\frac{1}{nm(A_{n,j})}\log P\left(|g_n(x_0)-g_X(x_0)>a\right) \\ &\leq \limsup_{n\to\infty}\frac{1}{nm(A_{n,j})}\log P\left(|g_n(x_0)-g_X(x_0)>a\right) \\ &\leq -\inf_{|v-g_X(x_0)|\geq a}\delta_{x_0}(v).\end{aligned}$$

It is easy to show that

$$\inf_{|v-g_X(x_0)|>a}\delta_{x_0}(v) = \inf_{|v-g_X(x_0)|\geq a}\delta_{x_0}(v) = \delta_{x_0}(g_X(x_0)+a) = \beta_{x_0}(a)$$

Indeed, the function δ_{x_0} is continuous on $[0,\infty)$ and is decreasing in $[0, g_X(x_0)]$ and increasing in $[g_X(x_0),\infty)$. Therefore,

$$\inf_{|v-g_X(x_0)|>a}\delta_{x_0}(v) = \inf_{|v-g_X(x_0)|\geq a}\delta_{x_0}(v) = \min\left\{\delta_{x_0}(g_X(x_0)+a), \delta_{x_0}(g_X(x_0)-a)\right\}.$$

We can easily check that

$$\min\left\{\delta_{x_0}(g_X(x_0)+a), \delta_{x_0}(g_X(x_0)-a)\right\} = \delta_{x_0}(g_X(x_0)+a).$$

Thus, the result is established. □

Proof of Theorem 3.3
Using the function $\mathbb{R}\times\mathbb{R}^* \to \mathbb{R}$, $(u,v)\to\frac{u}{v}$ and the contraction principle, Theorem 3.1 implies that $r_n(x_0)$ satisfies a large deviations principle with the rate $nm(A_{n,j})$ and the rate function

$$\Gamma_{x_0}(\lambda) := \inf\left\{\Lambda_{x_0}(u,v) : u = \lambda v\right\} = \inf_{v\in\mathbb{R}}\Lambda_{x_0}(\lambda v, v).$$

Afterwards, a simple calculation gives

$$\Gamma_{x_0}(\lambda) = \begin{cases} -\int_{-\infty}^{\infty}\left(e^{\theta_{x_0}(\lambda)(f(y)-\lambda)}-1\right)g(x_0,y)dy & \text{if } \lambda\in H_{x_0} \\ \infty & \text{elsewhere.}\end{cases}$$

□

Proof of Corollary 3.2
Theorem 3.3 gives, for any $a > 0$,

$$\begin{aligned}
-\inf_{|\lambda - r(x_0)|>a} \Gamma_{x_0}(\lambda) &\leq \liminf_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|r_n(x_0) - r(x_0)| > a\right) \\
&\leq \limsup_{n\to\infty} \frac{1}{nm(A_{n,j})} \log P\left(|r_n(x_0) - r(x_0)| > a\right) \\
&\leq -\inf_{|\lambda - r(x_0)|\geq a} \Gamma_{x_0}(\lambda).
\end{aligned}$$

Conditions (H.2)-(H.4) guarantee the differentiability of θ_{x_0} in H_{x_0} and consequently, Γ_{x_0} is also differentiable. Therefore, by a simple calculation we obtain the result. □

Proof of Theorem 3.4
Let us begin by the proof of the lower bound. For any $x \in \mathbb{R}$, we have

$$P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) \geq P\left(|g_n(x) - g_X(x)| > a\right).$$

Remark that Condition (A.3) implies that, for any $j \in \mathbb{Z}$,

$$\lim_{n\to\infty} \frac{m(A_{n,j})}{h_{n,2}} = 1.$$

Using Corollary 3.1, we obtain, for $x \in A_{n,j}$,

$$\begin{aligned}
\liminf_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) &\geq \liminf_{n\to\infty} \frac{m(A_{n,j})}{h_{n,2}} \frac{1}{nm(A_{n,j})} \\
&\quad \times \log P\left(|g_n(x) - g_X(x)| > a\right) \\
&= -\beta_x(a).
\end{aligned}$$

Therefore,

$$\liminf_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) \geq -\inf_{x\in\mathbb{R}} \beta_x(a).$$

In order to derive the upper bound, we proceed as follows. Let H_n be as in the condition (A.1). We have

$$\begin{aligned}
P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) &\leq P\left(\sup_{|x|\leq H_n} |g_n(x) - g_X(x)| > a\right) \\
&\quad + P\left(\sup_{|x|> H_n} |g_n(x) - g_X(x)| > a\right).
\end{aligned}$$

Using the fact that $\log(1+u) \leq u$ for $u \geq 0$, we obtain

$$\frac{1}{nh_{n,2}} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a\right) \leq \frac{1}{nh_{n,2}} \log P\left(\sup_{|x|\leq H_n} |g_n(x) - g_X(x)| > a\right) + \frac{1}{nh_{n,2}} \frac{P\left(\sup_{|x|>H_n} |g_n(x) - g_X(x)| > a\right)}{P\left(\sup_{|x|\geq H_n} |g_n(x) - g_X(x)| > a\right)}. \tag{3.5}$$

In addition, we have

$$\sup_{|x|>H_n} |g_n(x) - g_X(x)| \leq \sup_{|x|>H_n} g_n(x) + \sup_{|x|>H_n} g_X(x) = \sup_{|x|>H_n} g_n(x) + \mu_n,$$

where

$$\mu_n = \sup_{|x|>H_n} g_X(x).$$

Observe that

$$\sup_{|x|>H_n} g_n(x) \leq \frac{1}{nh_{n,1}} \sum_{i=1}^{n} \mathbb{1}_{\{|X_i|>H_n\}}.$$

Consequently,

$$P\left(\sup_{|x|>H_n} |g_n(x) - g_X(x)| > a\right) \leq P\left(\frac{1}{nh_{n,1}} \sum_{i=1}^{n} \mathbb{1}_{\{|X_i|>H_n\}} > a - \mu_n\right).$$

Therefore, the Markov's inequality gives

$$P\left(\sup_{|x|>H_n} |g_n(x) - g_X(x)| > a\right) \leq \frac{P(|X|>H_n)}{h_{n,1}(a-\mu_n)} \leq \frac{\epsilon_n}{h_{n,1}(a-\mu_n)}.$$

Since $\lim_{|x|\to\infty} g_X(x) = 0$, then, for any $\epsilon > 0$, there exists $n_0 \geq 0$ such that, for any $n \geq n_0$, $\mu_n < \epsilon$. It follows then that

$$P\left(\sup_{|x|>H_n} |g_n(x) - g_X(x)| > a\right) \leq \frac{\epsilon_n}{h_{n,1}(a-\epsilon)}. \tag{3.6}$$

Set now $J_n = \{j \in \mathbb{Z} : A_{n,j} \cap [-H_n, H_n] \neq \emptyset\}$. We can, without loss of generality, suppose that $J_n = \{1, 2, ..., m_n\}$ and that, for any $j \in J_n$, $[-H_n, H_n] \cap A_{n,j} = [a_{j-1}^n, a_j^n)$. Set $x_j^n = a_{j-1}^n + (a_j^n - a_{j-1}^n)/2$. Therefore, we have

$$\sup_{|x|\leq H_n} |g_n(x) - g_X(x)| \leq \max_{1\leq j\leq m_n} |g_n(x_j^n) - g_X(x_j^n)| + \max_{1\leq j\leq m_n} \sup_{a_{j-1}^n \leq x < a_j^n} |g_X(x) - g_X(x_j^n)|.$$

The continuity of the function g_X which is actually the uniform continuity on the compact set $[-H_n, H_n]$, implies that, for any $\epsilon > 0$, there exist n_0 such that, for any $n \geq n_0$,

$$\max_{1\leq j\leq m_n} \sup_{a_{j-1}^n \leq x < a_j^n} |g_X(x) - g_X(x_j^n)| < \epsilon.$$

Thus, we have

$$P\left(\sup_{|x|\leq H_n}|g_n(x)-g_X(x)|>a\right)\leq\sum_{j=1}^{m_n}P\left(|g_n(x_j^n)-g_X(x_j^n)|>a-\epsilon\right). \tag{3.7}$$

Using the fact that

$$P\left(|g_n(x_j^n)-g_X(x_j^n)|>a-\epsilon\right)\leq 2\max\left\{P\left(g_n(x_j^n)-g_X(x_j^n)>a-\epsilon\right),\right.$$
$$\left.P\left(-g_n(x_j^n)+g_X(x_j^n)|>a-\epsilon\right)\right\}$$

and the Tchebychev's inequality, we obtain, for any $t>0$,

$$P\left(g_n(x_j^n)-g_X(x_j^n)>a-\epsilon\right)\leq\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon+g_X(x_j^n))t\right.\right.$$
$$\left.\left.-\frac{1}{(a_j^n-a_{j-1}^n)}\log\left(1+(e^t-1)\int_{a_{j-1}^n}^{a_j^n}g(x)dx\right)\right]\right\},$$

and

$$P\left(-g_n(x_j^n)+g_X(x_j^n)>a-\epsilon\right)\leq\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon-g_X(x_j^n))t\right.\right.$$
$$\left.\left.-\frac{1}{(a_j^n-a_{j-1}^n)}\log\left(1+(e^{-t}-1)\int_{a_{j-1}^n}^{a_j^n}g(x)dx\right)\right]\right\}.$$

Since $\log(1+u)\leq u$ for any $u>-1$, we obtain

$$\begin{aligned}P\left(g_n(x_j^n)-g_X(x_j^n)>a-\epsilon\right)&\leq\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon+g_X(x_j^n))t\right.\right.\\&\qquad\left.\left.-\frac{(e^t-1)}{(a_j^n-a_{j-1}^n)}\int_{a_{j-1}^n}^{a_j^n}g(x)dx\right]\right\}\\&=\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon)t-\rho^+_{x_j^n,n}(t)\right]\right\},\end{aligned}$$

and

$$\begin{aligned}P\left(-g_n(x_j^n)+g_X(x_j^n)>a-\epsilon\right)&\leq\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon-g_X(x_j^n))t\right.\right.\\&\qquad\left.\left.-\frac{(e^{-t}-1)}{(a_j^n-a_{j-1})^n}\int_{a_{j-1}^n}^{a_j^n}g(x)dx\right]\right\}\\&=\exp\left\{-n(a_j^n-a_{j-1}^n)\left[(a-\epsilon)t-\rho^-_{x_j^n,n}(t)\right]\right\}\end{aligned}$$

where

$$\rho^+_{x,n}(t)=-g_X(x)t+\frac{(e^t-1)}{m(A_{n,j})}\int_{A_{n,j}}g(x)dx$$

and

$$\rho^-_{x,n}(t)=g_X(x)t+\frac{(e^{-t}-1)}{m(A_{n,j})}\int_{A_{n,j}}g(x)dx,$$

for $x \in A_{n,j}$.
Set

$$\rho_{x,n}(t) = \max\left\{\rho^+_{x,n}(t), \rho^-_{x,n}(t)\right\}$$

Thus,

$$\begin{aligned}\frac{1}{nh_{n,2}} \log P\left(\left|g_n(x_j^n) - g_X(x_j^n)\right| > a - \epsilon\right) \leq & \frac{\log 2}{nh_{n,2}} \\ & + \frac{aj^n - a^n_{j-1}}{h_{n,2}} \left\{(a-\epsilon)t - \sup_{x\in\mathbb{R}} \rho_{x,n}(t)\right\}.\end{aligned} \tag{3.8}$$

Since g_X is continuous and $g_X(x) \to 0$ as $|x| \to \infty$, then

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \left| \frac{1}{m(A_{n,j})} \int_{A_{n,j}} g(x)dx - g(x)\right| = 0.$$

It follows then that, for any $t \geq 0$,

$$\lim_{n\to\infty} \sup_{x\in\mathbb{R}} \rho_{x,n}(t) = \sup_{x\in\mathbb{R}} \rho_x(t),$$

where

$$\rho_x(t) = \max\left\{\rho^+_x(t), \rho^-_x(t)\right\},$$

with

$$\rho^+_x(t) = g_X(x)(e^t - 1 - t) \qquad \text{and} \qquad \rho^-_x(t) = g_X(x)(e^{-t} - 1 + t).$$

Easily, we have $\rho_x(t) = \rho^+_x(t)$. Therefore, combining the statements (3.7), (3.8) together with the condition (A.3), we obtain

$$\limsup_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{|x|\leq H_n} |g_n(x) - g_X(x)| > a\right) \leq -\sup_{t\geq 0} \inf_{x\in\mathbb{R}} \left\{(a-\epsilon)t - \rho^+_x(t)\right\}. \tag{3.9}$$

Using the minimax theorem, we can show that

$$\sup_{t\geq 0} \inf_{x\in\mathbb{R}} \left\{(a-\epsilon)t - \rho^+_x(t)\right\} = \inf_{x\in\mathbb{R}} \sup_{t\geq 0} \left\{(a-\epsilon)t - \rho^+_x(t)\right\}. \tag{3.10}$$

Indeed, we have

$$\sup_{t\geq 0} \inf_{x\in\mathbb{R}} \left\{(a-\epsilon)t - \rho^+_x(t)\right\} = \sup_{t\geq 0} \inf_{y\in[0,M]} H(y,t),$$

and

$$\inf_{x\in\mathbb{R}} \sup_{t\geq 0} \left\{(a-\epsilon)t - \rho^+_x(t)\right\} = \inf_{y\in[0,M]} \sup_{t\geq 0} H(y,t),$$

where

$$H(y,t) = (a-\epsilon)t - y(e^t - 1 - t).$$

and $M = \sup_{x\in\mathbb{R}} g_X(x)$. The function $H : [0,\infty) \times [0, M] \to \mathbb{R}$ satisfies the property that for a fixed t, $H(y,t)$ is convex as a function of y, and for a fixed y, $H(y,t)$ is concave as a function of t. Therefore, the minimax theorem implies that

$$\inf_{y\in[0,M]} \sup_{t\geq 0} H(y,t) = \sup_{t\geq 0} \inf_{y\in[0,M]} H(y,t).$$

The statement (10) then holds. On an other hand, it is easy to show that

$$\beta(a-\epsilon) = \inf_{y\in[0,M]} \sup_{t\geq 0} H(y,t).$$

Subsequently, the statements (3.9) and (3.10) give

$$\limsup_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{|x|\leq H_n} |g_n(x) - g_X(x)| > a \right) \leq -\beta(a-\epsilon). \tag{3.11}$$

Combining now the statements (3.5), (3.6), (3.11), the condition (A.1) together with the lower bound, we obtain

$$\limsup_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{x\in\mathbb{R}} |g_n(x) - g_X(x)| > a \right) \leq -\beta(a-\epsilon).$$

We achieve the proof by making use of the continuity of β and by making ϵ go to Zero. □

Proof of Theorem 3.5
Similarly as in the proof of the lower bound in Theorem 3.4, Corollary 3.2 and Condition (A.3) imply that, for any $x \in S$,

$$-\gamma_x(a) \leq \liminf_{n\to\infty} \frac{1}{nh_{n,2}} \log \left(\sup_{x\in S} |r_n(x) - r(x)| > a \right).$$

Therefore we have,

$$-\inf_{x\in S} \gamma_x(a) \leq \liminf_{n\to\infty} \frac{1}{nh_{n,2}} \log \left(\sup_{x\in S} |r_n(x) - r(x)| > a \right). \tag{3.12}$$

This proves the lower bound.

In order to prove the upper bound, let $\mathcal{P}_n = \{A_{n,1}, A_{n,2}, ..., A_{n,m_n}\}$ be a partition of the support S of the random variable X which is assumed to be compact. Without loss of generality, we suppose that $S = [a,b]$ and that, for any $1 \leq j \leq m_n - 1$, $A_{n,j} = [a^n_{j-1}, a^n_j)$ and $A_{n,m_n} = [a^n_{m_n}, b]$. Set $x^n_j = a^n_{j-1} + (a^n_j - a^n_{j-1})/2$. Therefore, we have

$$\sup_{x\in S} |r_n(x) - r(x)| \leq \max_{1\leq j\leq m_n} |r_n(x^n_j) - r(x^n_j)| + \max_{1\leq j\leq m_n} \sup_{a^n_{j-1}\leq x< a^n_j} |r(x) - r(x^n_j)|.$$

Similarly as in the proof of Theorem 3.4, for an $\epsilon > 0$ and n large enough, we obtain

$$P\left(\sup_{x\in S}|r_n(x)-r(x)|>a\right) \le \sum_{j=1}^{m_n} P\left(r_n(x_j^n)-r(x_j^n)>a-\epsilon\right) + \sum_{j=1}^{m_n} P\left(-r_n(x_j^n)+r(x_j^n)>a-\epsilon\right). \tag{3.13}$$

For $x \in A_{n,j}$ and any $t \ge 0$, set

$$\varrho^{+}_{x,n,a}(t) = \frac{1}{m(A_{n,j})}\int_{-\infty}^{\infty}\left(1-e^{t(f(y)-r(x)-a)}\right)\left(\int_{A_{n,j}} g(u,y)du\right)dy,$$

and

$$\varrho^{-}_{x,n,a}(t) = \frac{1}{m(A_{n,j})}\int_{-\infty}^{\infty}\left(1-e^{-t(f(y)-r(x)+a)}\right)\left(\int_{A_{n,j}} g(u,y)du\right)dy.$$

Analogous steps in the proof of Theorem 3.4, give

$$\log P\left(r_n(x_j^n)-r(x_j^n)>a-\epsilon\right) \le -n(a_j^n-a_{j-1}^n)\inf_{x\in S}\sup_{t\ge 0}\varrho^{+}_{x,n,a-\epsilon}(t),$$

and

$$\log P\left(-r_n(x_j^n)+r(x_j^n)>a-\epsilon\right) \le -n(a_j^n-a_{j-1}^n)\inf_{x\in S}\sup_{t\ge 0}\varrho^{-}_{x,n,a-\epsilon}(t).$$

Therefore, conditions (B.1) and (A.3) imply that

$$\limsup_{n\to\infty}\frac{1}{nh_{n,2}}\log\sum_{j=1}^{m_n} P\left(r_n(x_j^n)-r(x_j^n)>a-\epsilon\right) \le -\liminf_{n\to\infty}\inf_{x\in S}\sup_{t\ge 0}\varrho^{+}_{x,n,a-\epsilon}(t), \tag{3.14}$$

and

$$\limsup_{n\to\infty}\frac{1}{nh_{n,2}}\log\sum_{j=1}^{m_n} P\left(-r_n(x_j^n)+r(x_j^n)>a-\epsilon\right) \le -\liminf_{n\to\infty}\inf_{x\in S}\sup_{t\ge 0}\varrho^{-}_{x,n,a-\epsilon}(t). \tag{3.15}$$

Now combining (3.13), (3.14) and (3.15) we obtain

$$\limsup_{n\to\infty}\frac{1}{nh_{n,2}}\log P\left(\sup_{x\in S}|r_n(x)-r(x)|>a\right) \le -\liminf_{n\to\infty}\inf_{x\in S}\sup_{t\ge 0}\varrho^{+}_{x,n,a-\epsilon}(t) \bigvee -\liminf_{n\to\infty}\inf_{x\in S}\sup_{t\ge 0}\varrho^{-}_{x,n,a-\epsilon}(t), \tag{3.16}$$

where $u \vee u = \max\{u, v\}$.

Moreover, the continuity of the function r and the function g with respect to x, together with the conditions (H.2)-(H'.2) imply that, for any $t \geq 0$, $\varrho^+_{x,n,a}(t) \to \varrho^+_{x,a}(t)$ and $\varrho^-_{x,n,a}(t) \to \varrho^-_{x,a}(t)$ as $n \to \infty$ uniformly over the compact set S, where

$$\varrho^+_{x,a}(t) = \int_{-\infty}^{\infty} \left(1 - e^{t(f(y)-r(x)-a)}\right) g(x,y)dy$$

and

$$\varrho^-_{x,a}(t) = \int_{-\infty}^{\infty} \left(1 - e^{-t(f(y)+r(x)-a)}\right) g(x,y)dy,$$

and that the functions $x \to \varrho^{\pm}_{x,a}(t)$ are continuous on S. Now let us show that

$$\liminf_{n\to\infty} \inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,n,a-\epsilon}(t) \geq \inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,a-\epsilon}(t), \tag{3.17}$$

and

$$\liminf_{n\to\infty} \inf_{x\in S} \sup_{t\geq 0} \varrho^-_{x,n,a-\epsilon}(t) \geq \inf_{x\in S} \sup_{t\geq 0} \varrho^-_{x,a-\epsilon}(t). \tag{3.18}$$

The proof of both statements is the same, we limit ourself to the first one only. We have to consider two cases :
If $\liminf_{n\to\infty} \inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,n,a-\epsilon}(t) = +\infty$, the result is obvious.
Now if

$$\liminf_{n\to\infty} \inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,n,a-\epsilon}(t) = l < \infty,$$

then, we can extract from the sequence $\left\{\inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,n,a-\epsilon}(t)\right\}_{n\geq 1}$ a subsequence $\left\{\inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,\varphi(n),a-\epsilon}(t)\right\}_{n\geq 1}$ converging to l. Let $\epsilon > 0$ and $n_0 \geq 1$, be such that, for any $n \geq n_0$,

$$\inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,\varphi(n),a-\epsilon}(t) \leq l + \epsilon.$$

For any $n \geq n_0$, let $x_n \in S$ such that

$$\sup_{t\geq 0} \varrho^+_{x_n,\varphi(n),a-\epsilon}(t) \leq \inf_{x\in S} \sup_{t\geq 0} \varrho^+_{x,\varphi(n),a-\epsilon}(t) + \epsilon.$$

Therefore, for a such sequence $(x_n)_n$, we have

$$\sup_{t\geq 0} \varrho^+_{x_n,\varphi(n),a-\epsilon}(t) \leq l + 2\epsilon.$$

Since the sequence $(x_n)_{n\geq n_0}$ belongs to the compact set S, we can extract from it a subsequence $x_{\phi(n)}$ converging in S. Set $x_0 = \lim_{n\to\infty} x_{\phi(n)}$. Consequently, we have

$$\sup_{t\geq 0} \lim_{n\to\infty} \varrho^+_{x_{\phi(n)},\varphi(\phi(n)),a-\epsilon}(t) \leq \lim_{n\to\infty} \sup_{t\geq 0} \varrho^+_{x_n,\varphi(n),a-\epsilon}(t) \leq l + 2\epsilon.$$

The uniform convergence of $\varrho^{+}_{x,n,a-\epsilon}(t)$ to $\varrho^{+}_{x,a-\epsilon}(t)$ on S, and the continuity of the function $x \longrightarrow \varrho^{+}_{x,a-\epsilon}(t)$, imply that, for any $t \geq 0$,

$$\lim_{n\to\infty} \varrho^{+}_{x_{\phi(n)},\varphi(\phi(n)),a-\epsilon}(t) = \varrho^{+}_{x_0,a-\epsilon}(t).$$

Finally, we have

$$\sup_{t\geq 0} \varrho^{+}_{x_0,a-\epsilon}(t) \leq l + 2\epsilon$$

and

$$\inf_{x\in S} \sup_{t\geq 0} \varrho^{+}_{x,a-\epsilon}(t) \leq l + 2\epsilon.$$

Making ϵ go to zero allows to achieve the proof of the statement (3.17).

Now combining the statements (3.16), (3.17) and (3.18) we obtain

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{x\in S} |r_n(x) - r(x)| > a\right) &\leq -\min\left\{\inf_{x\in S}\sup_{t\geq 0} \varrho^{+}_{x,a-\epsilon}(t),\right.\\
&\qquad \left.\inf_{x\in S}\sup_{t\geq 0} \varrho^{-}_{x,a-\epsilon}(t)\right\}\\
&= -\inf_{x\in S} \min\left\{\sup_{t\geq 0} \varrho^{+}_{x,a-\epsilon}(t),\right.\\
&\qquad \left.\sup_{t\geq 0} \varrho^{-}_{x,a-\epsilon}(t)\right\}\\
&= -\inf_{x\in S}\min\left\{\gamma^{+}_{x}(a-\epsilon), \gamma^{-}_{x}(a-\epsilon)\right\}\\
&= -\gamma(a-\epsilon),
\end{aligned} \tag{3.19}$$

since $\gamma^{\pm}_{x}(a-\epsilon) = \sup_{t\geq 0} \varrho^{\pm}_{x,a-\epsilon}(t)$. We complet the proof by making ϵ go to Zero since γ is continuous. □

Proof of Theorem 3.6
Since the upper envelope F (i.e. $F(y) = \sup_{f\in\mathcal{F}} |f(y)|$) satisfies the conditions (H.2), (H'.2), and (H.4), this remains true for every function $f \in \mathcal{F}$. Theorem 3.5 then implies that, for any $f \in \mathcal{F}$,

$$-\gamma^{f}(a) = \lim_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{x\in S} |r^{f}_{n}(x) - r^{f}(x)| > a\right).$$

Therefore,

$$-\gamma_{\mathcal{F}} = -\inf_{f\in\mathcal{F}} \gamma^{f}(a) \leq \liminf_{n\to\infty} \frac{1}{nh_{n,2}} \log P\left(\sup_{f\in\mathcal{F}}\sup_{x\in S} |r^{f}_{n}(x) - r^{f}(x)| > a\right).$$

The lower bound is then verified.

On the other hand, observe that

$$\sup_{f\in\mathcal{F}}\sup_{x\in S}|r_n^f(x)-r^f(x)| \le \max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\sup_{f\in V_{f_i}(\epsilon)}\sup_{x\in S}|r_n^f(x)-r^f(x)|$$

where

$$V_{f_i}(\epsilon)=\{f\in\mathcal{F}:\|f-f_i\|_\infty<\epsilon\}.$$

Then, we have

$$\begin{aligned}\sup_{f\in\mathcal{F}}\sup_{x\in S}|r_n^f(x)-r^f(x)| \le{}& \max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\sup_{f\in V_{f_i}(\epsilon)}\sup_{x\in S}|r_n^f(x)-r_n^{f_i}(x)|\\ &+\max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\sup_{x\in S}|r_n^{f_i}(x)-r^{f_i}(x)|\\ &+\max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\sup_{f\in V_{f_i}(\epsilon)}\sup_{x\in S}|r^{f_i}(x)-r^f(x)|.\end{aligned}$$

Note that

$$|r_n^f(x)-r_n^{f_i}(x)|\le\|f-f_i\|_\infty \qquad \text{and} \qquad |r^{f_i}(x)-r^f(x)|\le\|f-f_i\|_\infty.$$

It follows then, that

$$\sup_{f\in\mathcal{F}_1}\sup_{x\in S}|r_n^f(x)-r^f(x)| \le \max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\sup_{x\in S}|r_n^{f_i}(x)-r^{f_i}(x)|+2\epsilon.$$

Let x_j^n be as in the proof of Theorem 3.5. Then, we have

$$\sup_{x\in S}|r_n^{f_i}(x)-r^{f_i}(x)|\le \max_{1\le j\le m_n}|r_n^{f_i}(x_j^n)-r^{f_i}(x_j^n)|+\max_{1\le j\le m_n}\sup_{a_{j-1}^n\le x<a_j^n}|r^{f_i}(x)-r^{f_i}(x_j^n)|.$$

Obviously, if r^F is a continuous function, then, for any $f\in\mathcal{F}$, r^f is also a continuous function. Using the continuity of the functions r^{f_i}, which is actually a uniform continuity, we obtain, for n large enough,

$$\max_{1\le j\le m_n}\sup_{a_{j-1}^n\le x<a_j^n}|r^{f_i}(x)-r^{f_i}(x_j^n)|\le\epsilon.$$

Therefore,

$$\sup_{f\in\mathcal{F}}\sup_{x\in S}|r_n^f(x)-r^f(x)| \le \max_{1\le i\le \mathcal{H}_{\mathcal{F}}(\epsilon)}\max_{1\le j\le m_n}|r_n^{f_i}(x_j^n)-r^{f_i}(x_j^n)|+3\epsilon.$$

Consequently, we have

$$P\left(\sup_{f\in\mathcal{F}}\sup_{x\in S}|r_n^f(x)-r^f(x)|>a\right)\le\sum_{i=1}^{\mathcal{H}_{\mathcal{F}}(\epsilon)}\sum_{j=1}^{m_n}P(|r_n^{f_i}(x_j^n)-r^{f_i}(x_j^n)|>a-3\epsilon).$$

The rest of the proof is similar to that of Theorem 3.5. □

Proof of Theorem 3.7
Denote by $g_n^{\mathcal{P}_n}$ the histogram density estimator at stage n, based on the partition $\mathcal{P}_n$. We have for any $x \in \mathbb{R}$

$$P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right)\geq P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right) \tag{3.20}$$

In addition, Let $k, l \in \mathbb{Z}$ such that $A_{n,j}^0 \subset A_{n,k} \subset A_{n,l}^1$. For $x \in A_{n,j}^0$, we have

$$\begin{aligned} g_n^{\mathcal{P}_n^0}(x) = \frac{m(A_{n,k})}{m(A_{n,j}^0)}\frac{1}{nm(A_{n,k})}\sum_{i=1}^n \mathbb{1}_{A_{n,j}^0}(X_i) &\leq \frac{m(A_{n,k})}{m(A_{n,j}^0)}\frac{1}{nm(A_{n,k})}\sum_{i=1}^n \mathbb{1}_{A_{n,k}}(X_i)\\ &\leq \frac{m(A_{n,l}^1)}{m(A_{n,j}^0)}\frac{1}{nm(A_{n,k})}\sum_{i=1}^n \mathbb{1}_{A_{n,k}}(X_i). \end{aligned}$$

In the same way, we obtain

$$\begin{aligned} g_n^{\mathcal{P}_n}(x) = \frac{m(A_{n,l}^1)}{m(A_{n,k})}\frac{1}{nm(A_{n,l}^1)}\sum_{i=1}^n \mathbb{1}_{A_{n,k}}(X_i) &\leq \frac{m(A_{n,l}^1)}{m(A_{n,k})}\frac{1}{nm(A_{n,l}^1)}\sum_{i=1}^n \mathbb{1}_{A_{n,l}^1}(X_i)\\ &\leq \frac{m(A_{n,l}^1)}{m(A_{n,j}^0)}\frac{1}{nm(A_{n,l}^1)}\sum_{i=1}^n \mathbb{1}_{A_{n,l}^1}(X_i). \end{aligned}$$

Set now

$$\alpha_n^0 = \min_{j\in\mathbb{Z}} m(A_{n,j}^0), \quad \alpha_n^1 = \min_{j\in\mathbb{Z}} m(A_{n,j}^1),$$
$$\beta_n^0 = \max_{j\in\mathbb{Z}} m(A_{n,j}^0) \quad \text{and} \quad \beta_n^1 = \max_{j\in\mathbb{Z}} m(A_{n,j}^1).$$

It follows then, for any $\mathcal{P}_n \in \mathcal{F}_n$, that

$$\frac{\alpha_n^0}{\beta_n^1} g_n^{\mathcal{P}_n^0}(x) \leq g_n^{\mathcal{P}_n}(x) \leq \frac{\beta_n^1}{\alpha_n^0} g_n^{\mathcal{P}_n^1}(x).$$

Thus,

$$\inf_{\mathcal{P}_n\in\mathcal{F}_n}|g_n^{\mathcal{P}_n}(x)-g_X(x)| \geq \max\left\{\frac{\alpha_n^0}{\beta_n^1}g_n^{\mathcal{P}_n^0}(x)-g_X(x), g_X(x)-\frac{\beta_n^1}{\alpha_n^0}g_n^{\mathcal{P}_n^1}(x)\right\}.$$

Consequently,

$$\begin{aligned} &\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right)\\ &\qquad\geq \max\left\{\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\frac{\alpha_n^0}{\beta_n^1}g_n^{\mathcal{P}_n^0}(x)-g_X(x)>a\right),\right. \qquad (3.21)\\ &\qquad\qquad \left.\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(g_X(x)-\frac{\beta_n^1}{\alpha_n^0}g_n^{\mathcal{P}_n^1}(x)>a\right)\right\}. \end{aligned}$$

Now we will use the Gärtner-Ellis Theorem in order to obtain a lower bound for

$$\liminf_{n\to\infty} \frac{1}{n\beta_n^1} \log P\left(\frac{\alpha_n^0}{\beta_n^1} g_n^{\mathcal{P}_n^0}(x) - g_X(x) > a\right).$$

Let $A_{n,j}^0$ be the element of partitions $\mathcal{P}_n^0$ which contains x and set

$$\begin{aligned}\Psi_{n,x}(t) = E\left[\exp\left\{tn\beta_n^1 \frac{\alpha_n^0}{\beta_n^1} g_n^{\mathcal{P}_n^0}(x)\right\}\right] &= E\left[\exp\left\{t\frac{\alpha_n^0}{m(A_{n,j}^0)}\sum_{i=1}^n \mathbb{1}_{A_{n,j}^0}(X_i)\right\}\right] \\ &= \left[E\left[\exp\left\{t\frac{\alpha_n^0}{m(A_{n,j}^0)}\mathbb{1}_{A_{n,j}^0}(X)\right\}\right]\right]^n \\ &= \left[\Phi_{n,x}(t)\right]^n.\end{aligned}$$

We have

$$\Phi_{n,x}(t) = 1 + \left(\exp\left\{\frac{\alpha_n^0}{m(A_{n,j}^0)}t\right\} - 1\right)\int_{A_{n,j}^0} g_X(z)dz.$$

The conditions (F.1) and (F.2) imply that

$$\lim_{n\to\infty}\left(\exp\left\{\frac{\alpha_n^0}{m(A_{n,j}^0)}t\right\} - 1\right)\int_{A_{n,j}^0} g_X(z)dz = 0.$$

Then, by the fact that $\log(1+u) \simeq u$ in the neighborhood of Zero, we have

$$\lim_{n\to\infty}\frac{1}{n\beta_n^1}\log\Psi_{n,x}(t) = \lim_{n\to\infty}\left(\exp\left\{\frac{\alpha_n^0}{m(A_{n,j}^0)}t\right\} - 1\right)\frac{1}{\beta_n^1}\int_{A_{n,j}^0} g_X(z)dz.$$

Note that the mean value Theorem gives

$$\lim_{n\to\infty}\frac{1}{m(A_{n,j}^0)}\int_{A_{n,j}^0} g_X(z)dz = g_X(x). \tag{3.22}$$

Therefore, the conditions (F.2) and (F.3) give

$$\lim_{n\to\infty}\frac{1}{n\beta_n^1}\log\Psi_{n,x}(t) = \alpha(e^t-1)g_X(x).$$

Since the function $t \to \alpha(e^t-1)g_X(x)$ is finite and differentiable everywhere, the Gärtner-Ellis Theorem gives the following lower bound

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\frac{\alpha_n^0}{\beta_n^1}g_n^{\mathcal{P}_n^0}(x) - g_X(x) > a\right) &\geq -\inf_{u>g_X(x)+a}\sup_{t\in\mathbb{R}}\left\{ut - \alpha(e^t-1)g_X(x)\right\} \\ &= -\beta^+_{\alpha,g_X(x)}(a).\end{aligned} \tag{3.23}$$

We show in a similar way that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(g_X(x)-\frac{\beta_n^1}{\alpha_n^0}g_n^{\mathcal{P}_n^1}(x)>a\right) &\geq -\inf_{u<g_X(x)-a}\sup_{t\in\mathbb{R}}\left\{ut-(e^{\frac{t}{\alpha}}-1)g_X(x)\right\}\\ &= -\beta^-_{\alpha,g_X(x)}(a)\end{aligned} \tag{3.24}$$

Combining (3.21), (3.23) and (3.24), we obtain

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right) &\geq -\min\left\{\beta^+_{\alpha,g_X(x)}(a),\beta^-_{\alpha,g_X(x)}(a)\right\}\\ &= -\beta_{\alpha,g_X(x)}(a).\end{aligned} \tag{3.25}$$

Moreover, the combination of statements (3.20) and (3.25) imply then that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right) &\geq -\inf_{x\in\mathbb{R}}\beta_{\alpha,g_X(x)}(a)\\ &= -\beta_\alpha^g(a).\end{aligned}$$

This achieves the proof of Theorem 3.7. □

Proof of Corollary 3.3
Note that, for any $g_X\in\Theta$, we have

$$\begin{aligned}&\inf_{g_n\in\mathcal{G}_n}\sup_{g_X\in\Theta}\frac{1}{n\beta_n^1}\log P\left(\sup_{x\in\mathbb{R}}|g_n(x)-g_X(x)|>a\right)\\ &\qquad\geq\inf_{g_n\in\mathcal{G}_n}\frac{1}{n\beta_n^1}\log P\left(\sup_{x\in\mathbb{R}}|g_n(x)-g_X(x)|>a\right)\\ &\qquad\geq\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|g_n^{\mathcal{P}_n}(x)-g_X(x)|>a\right).\end{aligned} \tag{3.26}$$

Then by Theorem 3.7 and the statement (3.26) we have

$$\begin{aligned}\liminf_{n\to\infty}\inf_{g_n\in\mathcal{G}_n}\sup_{g_X\in\Theta}\frac{1}{n\beta_n^1}\log P\left(\sup_{x\in\mathbb{R}}|g_n(x)-g_X(x)|>a\right) &\geq -\inf_{g_X\in\Theta}\beta_\alpha^g(a)\\ &= -\beta_\alpha^\Theta(a).\end{aligned}$$

The result is then proven. □

Proof of Theorem 3.8
As in the proof of Theorem 3.7, we have, for $x\in A^0_{n,j}\subset A_{n,k}\subset A^1_{n,l}$,

$$\frac{\sum_{i=1}^n f(Y_i)\mathbb{1}_{A^0_{n,j}}(X_i)}{\sum_{i=1}^n\mathbb{1}_{A^1_{n,l}}(X_i)}\leq r_n^{\mathcal{P}_n}(x)\leq\frac{\sum_{i=1}^n f(Y_i)\mathbb{1}_{A^1_{n,l}}(X_i)}{\sum_{i=1}^n\mathbb{1}_{A^0_{n,j}}(X_i)}.$$

Thus,

$$\inf_{\mathcal{P}_n\in\mathcal{F}_n}|r_n^{\mathcal{P}_n}(x)-r(x)|\geq\max\left\{\frac{\sum_{i=1}^n f(Y_i)\mathbb{1}_{A^0_{n,j}}(X_i)}{\sum_{i=1}^n\mathbb{1}_{A^1_{n,l}}(X_i)}-r(x),\right.$$
$$\left. r(x)-\frac{\sum_{i=1}^n f(Y_i)\mathbb{1}_{A^1_{n,l}}(X_i)}{\sum_{i=1}^n\mathbb{1}_{A^0_{n,j}}(X_i)}\right\}.$$

Consequently,

$$\begin{aligned}
&P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|r_n^{\mathcal{P}_n}(x)-r(x)|>a\right)\\
&\geq\max\left\{P\left(\frac{1}{n\beta_n^1}\sum_{i=1}^n\left(f(Y_i)\mathbb{1}_{A^0_{n,j}}(X_i)-(r(x)+a)\mathbb{1}_{A^1_{n,l}}(X_i)\right)>0\right),\right. \qquad (3.27)\\
&\qquad\left. P\left(\frac{1}{n\beta_n^1}\sum_{i=1}^n\left(f(Y_i)\mathbb{1}_{A^1_{n,l}}(X_i)-(r(x)-a)\mathbb{1}_{A^0_{n,j}}(X_i)\right)<0\right)\right\}.
\end{aligned}$$

Set now

$$\Psi_{n,x,a}(t)=E\left[\exp\left\{t\sum_{i=1}^n\left(f(Y_i)\mathbb{1}_{A^0_{n,j}}(X_i)-(r(x)+a)\mathbb{1}_{A^1_{n,l}}(X_i)\right)\right\}\right].$$

Since the (X_i,Y_i) are i.i.d. then,

$$\Psi_{n,x,a}(t)=\left[E\left[\exp\left\{t\left(f(Y)\mathbb{1}_{A^0_{n,j}}(X)-(r(x)+a)\mathbb{1}_{A^1_{n,l}}(X)\right)\right\}\right]\right]^n=\left[\Phi_{n,x,a}(t)\right]^n.$$

Thus, we have

$$\begin{aligned}
\Phi_{n,x,a}(t)&=1+\int_{-\infty}^{\infty}\int_{A^1_{n,l}}\left(\exp\left\{t\left(f(y)\mathbb{1}_{A^0_{n,j}}(z)-(r(x)+a)\right)\right\}-1\right)g(z,y)dzdy\\
&=1+\int_{-\infty}^{\infty}\int_{A^0_{n,j}}\left(\exp\left\{t(f(y)-r(x)-a)\right\}-1\right)g(z,y)dzdy\\
&\quad+(e^{-t(r(x)+a)}-1)\int_{A^1_{n,l}\setminus A^0_{n,j}}g_X(z)dz.
\end{aligned}$$

As in proof of Theorem 3.7, it follows, by conditions (F.1)-(F.3), (H.2) and (H'.2), that

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n\beta_n^1}\log\Psi_{n,x,a}&=\alpha\int_{-\infty}^{\infty}\left(\exp\left\{t(f(y)-r(x)-a)\right\}-1\right)g(x,y)dy\\
&\quad+(e^{-t(r(x)+a)}-1)(1-\alpha)g_X(x)\\
&:=I^+_{\alpha,x,a}(t).
\end{aligned}$$

Conditions (H.2), (H'.2) and (H.3) imply that $I^+_{\alpha,x,a}$ is finite and differentiable everywhere. Then the Gärtner-Ellis theorem gives

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\frac{1}{n\beta_n^1}\sum_{i=1}^n\Big(f(Y_i)\mathbb{1}_{A^0_{n,j}}(X_i)-(r(x)+a)\mathbb{1}_{A^1_{n,l}}(X_i)\Big)>0\right)\\ \geq -\inf_{u>0}\sup_{t\in\mathbb{R}}\left\{ut-I^+_{\alpha,x,a}(t)\right\}\\ = -\gamma^+_{\alpha,x}(a).\end{aligned}\tag{3.28}$$

In a similar way we show that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\frac{1}{n\beta_n^1}\sum_{i=1}^n\Big(f(Y_i)\mathbb{1}_{A^1_{n,l}}(X_i)-(r(x)-a)\mathbb{1}_{A^0_{n,j}}(X_i)\Big)<0\right)\\ \geq -\inf_{u<0}\sup_{t\in\mathbb{R}}\left\{ut-I^-_{\alpha,x,a}(t)\right\}\\ = -\gamma^-_{\alpha,x}(a).\end{aligned}\tag{3.29}$$

Combining (3.27), (3.28) and (3.29), we obtain, for any $x\in\mathbb{R}$,

$$\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|r_n^{\mathcal{P}_n}-r(x)|>a\right)\geq-\gamma_{\alpha,x}(a).\tag{3.30}$$

Note that, for any $x\in\mathbb{R}$

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|r_n^{\mathcal{P}_n}-r(x)|>a\right)\\ \geq\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}|r_n^{\mathcal{P}_n}-r(x)|>a\right).\end{aligned}\tag{3.31}$$

It follows then that

$$\begin{aligned}\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|r_n^{\mathcal{P}_n}-r(x)|>a\right)\geq-\inf x\in\mathbb{R}\gamma_{\alpha,x}(a)\\ :=\gamma^r_\alpha(a).\end{aligned}$$

Thus, we finished the proof. □

Proof of Corollary 3.4
Note that, for any $r\in\mathcal{R}$, we have

$$\begin{aligned}\liminf_{n\to\infty}\inf_{r_n\in\mathcal{R}_n}\sup_{r\in\mathcal{R}}\frac{1}{n\beta_n^1}\log P\left(\sup_{x\in\mathbb{R}}|r_n(x)-r(x)|>a\right)\\ \geq\liminf_{n\to\infty}\inf_{r_n\in\mathcal{R}_n}\frac{1}{n\beta_n^1}\log P\left(\sup_{x\in\mathbb{R}}|r_n(x)-r(x)|>a\right)\\ \geq\liminf_{n\to\infty}\frac{1}{n\beta_n^1}\log P\left(\inf_{\mathcal{P}_n\in\mathcal{F}_n}\sup_{x\in\mathbb{R}}|r_n^{\mathcal{P}_n}(x)-r(x)|>a\right).\end{aligned}$$

It follows, by Theorem 3.8 that,

$$\liminf_{n\to\infty} \inf_{r_n\in\mathcal{R}_n} \sup_{r\in\mathcal{R}} \frac{1}{n\beta_n^1} \log P\left(\sup_{x\in\mathbb{R}} |r_n(x) - r(x)| > a\right) \geq -\inf_{r\in\mathcal{R}} \gamma_\alpha^r(a).$$

The proof is then achieved. $\square$

Chapitre 4

Some functional large deviations principles in nonparametric function estimation

In this paper we investigate functional large deviation behaviours of some nonparametric function estimates. Towards this end, we define a process and study its large deviation behaviour in the space $L_1 \times L_1 \times L_1$ with respect to the weak convergence topology. As by-products, we derive large deviation principles in the L_1-space equipped with the weak convergence topology simultaneously for several density and regression estimators built up using the delta-sequence estimation method.

4.1 Introduction

Let $(X_n, Y_n)_{n\geq 1}$ be a sequence of i.i.d. $\mathbb{R}^2$-valued random vector defined on a probability space $(\Omega, \mathcal{A}, P)$. Denote by g the joint density function of the random vector (X_1, Y_1) and by g_X and g_Y the marginal density functions of X_1 and Y_1 respectively. A sequence of real functions $\delta_m(x, y)$ defined on $\mathbb{R}^2$ is called a delta-sequence whenever the following condition holds

$$(\mathbf{H_0}) \text{ for any real } \mathcal{C}^\infty\text{−function } \varphi, \quad \lim_{m\to\infty} \int \delta_m(x, y)\varphi(y)dy = \varphi(x).$$

The delta-sequence technique is a way to study simultaneously properties of estimates pertaining to several usual methods in nonparametric function estimation. Introduced by Watson & Leadbetter (1965)(79; 80), the delta-sequence method relative to density estimation was studied among others by Földes & Révész (1974)(33) and Walter & Blum (1979)(76). On the basis of a sample $(X_1, ..., X_n)$, the delta-sequence estimator of the density function g_X, at stage n, is defined, for any $x \in \mathbb{R}$, by

$$g_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_{m_n}(x, X_i).$$

Similarly, taking a real function f, we define the delta-sequence estimator of the f-indexed regression function $m_f(x) := E[f(Y)|X = x]$, at stage n, as any estimator, based on the sample $(X_1, Y_1), ..., (X_n, Y_n)$, of the form

$$m_n^f(x) = \frac{r_n^f(x)}{g_n(x)},$$

where

$$r_n^f(x) = \frac{1}{n}\sum_{i=1}^{n} f(Y_i)\delta_{m_n}(x, X_i).$$

Choices of the Delta-sequence $\delta_m(x, y)$ allow to recover many usual estimators. Examples include :

-*Kernel estimates.* Let K be a real function defined on $\mathbb{R}$ and let $(m_n)_{n\geq 1}$ be a sequence of smoothing parameters. Set

$$\delta_{m_n}(x, y) = m_n K\left(m_n(x - y)\right).$$

It is well known that if K is a Parzen-Rozenblatt kernel, then δ_{m_n} is a delta-sequence and g_n and r_n^f are respectively the kernel density and regression estimators.

-*Histogram estimates.* Let $\mathcal{P}_n = \{A_{n,j} : j \in J_n\}$ be a partition of the real line such that $|J_n| = m_n$, $\max_{j\in J_n} \lambda(A_{n,j}) \to 0$ and $n \min_{j\in J_n} \lambda(A_{n,j}) \to \infty$ as $n \to \infty$. We obtain the classical histogram and regressogram estimators by taking

$$\delta_{m_n}(x, y) = \sum_{j\in J_n} \frac{1}{\lambda(A_{n,j})} \mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y),$$

with λ the Lebesgue measure.

-*Orthogonal Series estimates.* Let $(e_k)_{k\geq 1}$ be a complete orthonormal system on the domain (a, b) (eventually the whole real line) consisting of eigenfunctions of a compact operator on the space $L^2(a, b)$. Let

$$\delta_m(x, y) = \sum_{k=1}^{m} e_k(x)e_k(y), \quad (x, y) \in (a, b)^2.$$

We can see in Walter (1965)(77) that the δ_m's are in fact delta-sequences. We refer to Walter & Blum (1979)(76) for an extensive list of delta-sequence examples in density estimation.

The main gaol of this work is to establish some functional large deviation principles (LDP in short) in nonparametric function estimation setting. In this respect, consider the vector process

$$W_n(x) = (r_n^{\varphi}(x), r_n^{\phi}(x), g_n(x)),$$

where φ and ϕ are two positive real functions. Obviously, the process $W_n(x)$ is closely related to the estimates introduced above and properties of these estimates will be deduced as by-products.

Notice that the LDP relative to the process W_n holds, under some regularity conditions, in the product space $\mathcal{X} \equiv L_1 \times L_1 \times L_1$ equipped with the weak topology. Recall that the topological dual of the space L_1 is the space L_∞ and that the L_1-weak topology is the weakest topology that makes continuous the linear forms in L_1, i.e., the forms defined by

$$\begin{aligned} Q \ : \ & L_1 \longrightarrow \mathbb{R} \\ & p \longrightarrow < p, k >_{dx} := \int p(x)k(x)dx, \end{aligned}$$

for any L_∞-function k. The process W_n is said to satisfy a LDP in the topological space $\mathcal{X}$ with the speed ϵ_n and the rate function I whenever the function I is lower semicontinuous and

– for any open set $O \subset \mathcal{X}$,

$$\liminf_{n\to\infty} \epsilon_n \log P(W_n \in O) \geq - \inf_{x\in O} I(x),$$

– for any closed set $F \subset \mathcal{X}$,

$$\limsup_{n\to\infty} \epsilon_n \log P(W_n \in F) \leq - \inf_{x\in F} I(x).$$

Below in the next section, we identify the normalization sequence (ϵ_n) as well as the rate function I corresponding to each studied case.

The index function f allow to study simultaneously properties of several estimates. The first example is given by the most classical regression function estimator where f stands as the identity function. Whenever $f(y) = \mathbb{1}_A(y)$ is the indicator function of a set A, it follows that $m_n^f(x)$ is the estimator of the conditional probability measure of the event $\{Y_1 \in A\}$ given $X_1 = x$. Similarly, one can define estimates for the conditional variance and conditional moments. The large deviation principles for all these estimates may be derived from Theorem 4.1, below, which established the LDP for the process W_n, by making use of the contraction principle (see, for instance, Dembo & Zeitouni (1998)(23), page 126). Note that the proof of Theorem 4.1 uses the classical Gärtner-Ellis Theorem in topological vector spaces of infinite dimension.

There exists an extensive large deviations literature involving many areas of probability and statistics. We refer to the books of Dembo & Zeitouni (1998)(23), Deuschel & Stroock (1989)(24) and Dupuis & Ellis (1997) (29) and the references therein for an account of results and applications. In nonparametric function estimation setting, several results have been obtained these last years. We refer to Louani (1998, 2003)(52; 49) where Chernoff type pointwise and L_∞ results have been stated for the kernel and orthogonal series density estimators. Similarly, Louani (1999)(53) established results for the conditional empirical process and the Nadaraya-Watson estimate of the regression function. The L_1 large deviation point of view has been considered in several papers. We quote the results obtained by Beirlant *et al.* (2001)(3) for the L_1-error of the histogram based density estimator. Louani (2000, 2005)(54; 55) studied the deviation of the kernel density estimator and considered uniformity over classes of kernels, classes of density functions and ranges of smoothing parameters. The L_1 functional aspect of the kernel density estimator with respect to the weak topology has been studied by Lei *et al.* (2003)(47). Note also that Louani & Ould Maouloud (2007)(50) studied some functional aspects related to the regression function estimator with model selection applications.

4.2 Results

4.2.1 Hypotheses and notations

In order to give our results, we introduce some notations. From now on, for any $h \in L_\infty$, set

$$l(h) = \log \int_{-\infty}^{\infty} e^{h(x)} g_X(x) dx.$$

Consider, for any $k_1, k_2, k_3 \in L_\infty$, the function

$$\Lambda_{\phi,\varphi}(k_1, k_2, k_3) = \log \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp\left\{k_1(x)\varphi(y) + k_2(x)\phi(y) + k_3(x)\right\} g(x, y) dx dy.$$

The Legendre transforms of the functions l and Λ are defined, for any $c \in L_1$, by

$$l^*(c) = \sup_{h \in L_\infty} \left\{ \int_{-\infty}^{\infty} h(x)c(x)dx - l(h) \right\},$$

and, for any $a, b, c \in L_1$, by

$$\Lambda^*_{\phi,\varphi}(a,b,c) = \sup_{k_1,k_2,k_3 \in L_\infty} \left\{ \int_{-\infty}^{\infty} (k_1(x)a(x) + k_2(x)b(x) + k_3(x)c(x))dx \right. \\ \left. - \Lambda_{\phi,\varphi}(k_1,k_2,k_3) \right\}$$

respectively. Set now the assumptions upon the delta-sequence δ_m, the distribution of the random vector (X_1, Y_1) and the functions ϕ and φ needed to prove our results.

($\mathbf{H_1}$) For any function $k \in L_\infty$,

$$\lim_{m \to \infty} \int_{-\infty}^{\infty} \delta_m(x, X_1)k(x)dx = k(X_1) \qquad \text{in probability.}$$

($\mathbf{H_2}$) For any $l \in \mathbb{N}$ and any m large enough, there exists a number $M_l > 0$, independent of m, such that

$$\mathbb{1}_{\{|z| \leq M_l\}} \int_{|x| > 2M_l} \delta_m(x,z)dx \leq \frac{1}{2l^2}.$$

($\mathbf{H_3}$) For any $t > 0$, we have

$$\int_{-\infty}^{\infty} e^{t\phi(y)} g_Y(y)dy < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} e^{t\varphi(y)} g_Y(y)dy < \infty.$$

($\mathbf{H_4}$) For any $t > 0$, we have

$$\int_{-\infty}^{\infty} \phi(y)e^{t\phi(y)} g_Y(y)dy < \infty \quad \text{and} \quad \int_{-\infty}^{\infty} \varphi(y)e^{t\varphi(y)} g_Y(y)dy < \infty.$$

($\mathbf{H_5}$) There exists a real number $\alpha > 1$, such that

$$\int_{\infty}^{\infty} e^{(\phi(y)+\varphi(y)+1)^\alpha} g_Y(y)dy < \infty.$$

4.2.2 Discussion of conditions

We discuss the conditions $\mathbf{H_1} - \mathbf{H_2}$ and consider examples of delta-sequences for which clearly they hold. In the first place, we treat the kernel and histogram method cases. We give sufficient conditions upon the kernels and partitions that allow to have the conditions $\mathbf{H_1} - \mathbf{H_2}$ satisfied. In this respect, Let K be a real function and let $(h_n)_{n\geq 1}$ be a sequence of positive real number. Introduce the following conditions

$(\mathbf{K_1})$ $\quad h_n \to 0 \quad$ and $\quad nh_n \to \infty \quad$ as $\quad n \to \infty$.

$(\mathbf{K_2})$ $\quad K \geq 0 \quad$ and $\quad \int K(x)dx = 1$.

Consider the kernel delta-sequence given by

$$K_n(x,y) = \frac{1}{h_n} K\left(\frac{x-y}{h_n}\right).$$

Proposition 4.1. *Suppose that the kernel K satisfies the conditions $(\mathbf{K_1})-(\mathbf{K_2})$. Then the conditions $\mathbf{H_1} - \mathbf{H_2}$ hold true for the delta-sequence K_n.*

Let now $\mathcal{P}_n = \{A_{n,j} : j \in J_n\}$ be a partition of the real line and consider the following condition

$(\mathbf{C})$ $\quad \max_{j\in J_n} \lambda(A_{n,j}) \to 0\;$ and $\; n \min_{j \in J_n} \lambda(A_{n,j}) \to \infty\;$ as $\quad n \to \infty$.

Consider the histogram delta-sequence given by

$$P_n(x,y) = \sum_{j\in J_n} \frac{\mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y)}{\lambda(A_{n,j})}.$$

Proposition 4.2. *Suppose that the condition $\mathbf{C}$ is satisfied. Then the conditions $\mathbf{H_1} - \mathbf{H_2}$ hold true for the delta sequence P_n.*

Remarks. Conditions $(\mathbf{H_3})$ and $(\mathbf{H_4})$ are of classical nature in large deviations, they guaranty the existence and differentiability of the Laplace transform function Λ associated to the process W_n, which is needed to use the abstract Gärtner-Ellis Theorem.

The differentiability is taken here in the Gâteaux sense. The functions l^* and Λ^* are the Legendre Transforms of the functions l and Λ respectively. Note that the function $l^*(c)$ identically equals the Kullback-Leibler information $R(c\|g_X)$ defined by

$$R(c\|g_X) := \begin{cases} \int_{-\infty}^{\infty} c(x)\log\left(\frac{c(x)}{g_X(x)}\right)dx & \text{if } c \text{ is a density and } c \ll g_X, \\ \infty & \text{elsewhere.} \end{cases} \tag{4.1}$$

4.2.3 Results

The main result of this work is given in the following theorem. It establishes a LDP for the vector process W_n.

Theorem 4.1. *Suppose that the conditions* $(\mathbf{H_0}) - (\mathbf{H_5})$ *are satisfied. Then the process* W_n *satisfies a large deviation principle in the space* $\mathcal{X}$ *equipped with the weak topology with the speed* $1/n$ *and the good rate function* $\Lambda^*_{\phi,\varphi}$.

The first corollary gives a large deviations principle for the density estimator g_n in the space L_1. This result is an extension of the previous one given by Lei *et al.* (2003) in the case of kernel density estimation. This corollary is derived from Theorem 2.1 by the contraction principle and a suitable choice of a continuous transform.

Corollary 4.1. *Under the conditions (H.1) and (H.2), the process* g_n *satisfies a large deviations principle in the space* L_1 *equipped with the weak topology with the speed* $1/n$ *and the good rate function* l^* *given by*

$$l^*(c) = R(c\|g_X).$$

The next corollary deals with a LDP for the regression function estimate considered on a compact set T.

Corollary 4.2. *Suppose that conditions* $\mathbf{H_1}$, $\mathbf{H_2}$ *are satisfied and that the function* f *is bounded. Then, the process* m_n^f *satisfies a large deviations principle in the space* $L^1(T)$ *with the speed* $1/n$ *and the rate function* Γ^*, *defined by*

$$\Gamma^*(d) = \inf\left\{\gamma^*(d\times c, c) : c \textit{ is a density and } c \ll g_X\right\}.$$

where

$$\gamma^*(a,c) = \inf_{b\in L^1} \Lambda^*_{f^+,f^-}(a+b,b,c).$$

4.3 Proofs

Proof of Theorem 4.1. We split up the proof into three steps. In the first one, we calculate the logarithm Laplace transform function $\Lambda_{\phi,\varphi}$ associated to the vector process $W_n \equiv (r_n^{\phi}, r_n^{\varphi}, g_n)$, and show that this function is finite and Gâteaux differentiable everywhere. The second step deals with the exponential tightness of the process W_n. We conclude in the third step by using the abstract Gärtner-Ellis Theorem.

Step 1. The logarithm Laplace transform function is defined, for any $k_1, k_2, k_3 \in L_\infty$, by

$$\Lambda_{\phi,\varphi}(k_1, k_2, k_3) := \lim_{n\to\infty} \frac{1}{n} \log E\left[\exp\left\{n\left\langle (k_1,k_2,k_3),(r_n^{\phi}, r_n^{\varphi}, g_n)\right\rangle\right\}\right].$$

where $< .,. >$ denotes the inner product. Since the random vectors $(X_i, Y_i)_{1\le i\le n}$ are independent and identically distributed, we obtain

$$\begin{aligned} &E\left[\exp\left\{n\left\langle (k_1,k_2,k_3),(r_n^{\phi}, r_n^{\varphi}, g_n)\right\rangle\right\}\right] \\ &\quad = \left[E\left[\exp\left\{\int_{-\infty}^{\infty} \left(\phi(Y)k_1(z) + \varphi(Y)k_2(z) + k_3(z)\right)\delta_{m_n}(z,X)dz\right\}\right]\right]^n. \end{aligned}$$

By the conditions $(\mathbf{H_1})$ and $(\mathbf{H_3})$ and the dominated convergence Theorem, it follows that

$$\begin{aligned} &\lim_{n\to\infty} E\left[\exp\left\{\int_{-\infty}^{\infty} \left(\phi(Y)k_1(z) + \varphi(Y)k_2(z) + 1\right)\delta_{m_n}(z,X)dz\right\}\right] \\ &\qquad = E\left[\exp\left\{\phi(Y)k_1(X) + \varphi(Y)k_2(X) + k_3(X)\right\}\right]. \end{aligned}$$

Indeed, by the condition $(\mathbf{H_1})$, we have

$$\begin{aligned} &\lim_{n\to\infty} \int_{-\infty}^{\infty} \left(\phi(Y)k_1(z) + \varphi(Y)k_2(z) + k_3(z)\right)\delta_{m_n}(z,X)dz \\ &\qquad = \phi(Y)k_1(X) + \varphi(Y)k_2(X) + k_3(X) \quad \text{ in probability }, \end{aligned}$$

and, by the condition $(\mathbf{H_3})$, it follows that

$$\begin{aligned} &E\left[\exp\left\{\int_{-\infty}^{\infty} \left(\phi(Y)k_1(z) + \varphi(Y)k_2(z) + 1\right)\delta_{m_n}(z,X)dz\right\}\right] \\ &\qquad = E\left[\exp\left\{\phi(Y)\|k_1\|_\infty + \varphi(Y)\|k_2\|_\infty + \|k_3\|_\infty\right\}\right] < \infty. \end{aligned}$$

Therefore, for any $k_1, k_2, k_3 \in L_\infty$, we have

$$\begin{aligned} \Lambda^*_{\phi,\varphi}(k_1,k_2,k_3) :&= \lim_{n\to\infty} \frac{1}{n} \log E\left[\exp\left\{n\left\langle (k_1,k_2,k_3),(r_n^{\phi}, r_n^{\varphi}, g_n)\right\rangle\right\}\right] \\ &= \log E\left[\exp\left\{\phi(Y)k_1(X) + \varphi(Y)k_2(X) + k_3(X)\right\}\right]. \end{aligned}$$

Note that the conditions $(\mathbf{H_3}) - (\mathbf{H_4})$ imply that the function $\Lambda_{\phi,\varphi}$ is finite and Gâteaux differentiable everywhere.

Step 2. In order to establish the exponential tightness of the processes $(r_n^{\phi}, r_n^{\varphi}, g_n)$ in $\mathcal{X}$, we have to show that, for any $L > 0$, there exists a compact set Γ_L, such that

$$\limsup_{n\to\infty} \frac{1}{n} \log P\left((r_n^{\phi}, r_n^{\varphi}, g_n) \in \Gamma_L^c\right) \leq -L. \tag{4.2}$$

To this end, consider the real number $\alpha > 1$ of condition $(\mathbf{H_5})$. Thus, we have

$$A = \int_{-\infty}^{\infty} \exp\left\{(\phi(y) + \varphi(y) + 1)^{\alpha}\right\} g_Y(y)dy < \infty.$$

Set

$$F(x) = \int_{-\infty}^{\infty} \exp\left\{(\phi(y) + \varphi(y) + 1)^{\alpha}\right\} g(x, y)dy.$$

Since F is a positive integrable function and assuming the condition $(\mathbf{H_2})$, we can find, for any positive integer $l > \log A$, a positive real number $M_l > 0$ such that

$$\int_{|x|>M_l} F(x)dx < e^{-2l^2}(e^l - 1), \tag{4.3}$$

and

$$\mathbb{1}_{\{|z|\leq M_l\}} \int_{|x|>2M_l} \delta_m(x, z)dx \leq \frac{1}{2l^2}. \tag{4.4}$$

Set

$$K_l = \left\{h \in L_1 : \int_{|x|>2M_l} h(x)dx \leq \frac{1}{3l}\right\},$$

and

$$\Gamma_L = \bigcap_{l\geq L} K_l \times \bigcap_{l\geq L} K_l \times \bigcap_{l\geq L} K_l.$$

To establish the compactness of the set Γ_L, it is enough to show that $\bigcap_{l\geq L} K_l$ is a compact set in L_1. In this respect, let $\epsilon > 0$, and observe that there exists a positive number M_l such that, for any $h \in \bigcap_{l\geq L} K_l$,

$$\int_{|x|>2M_l} h(x)dx < \epsilon.$$

It results, via Prohorov Theorem, that $\bigcap_{l\geq L} K_l$ is a compact set in L_1 for the weak topology.

We turn now to the proof of the statement 4.2. Observe that

$$\begin{aligned}P\left((r_n^{\phi}, r_n^{\varphi}, g_n) \in \Gamma_L^c\right) &\leq \sum_{l \geq L} P\left(\int_{|x|>2M_l} (r_n^{\phi}(x) + r_n^{\varphi}(x) + g_n(x))dx > \frac{1}{l}\right) \\ &\leq \sum_{l \geq L} P\left(2l^2 \sum_{i=1}^{n} (\phi(Y_i) + \varphi(Y_i) + 1)\right. \\ &\left. \times \int_{|x|>2M_l} \delta_{m_n}(x, X_i)dx > 2nl\right)\end{aligned}$$

Using Tchebycheff inequality and the fact that the random vectors $(X_i, Y_i)_{i \geq 1}$ are i.i.d., we obtain

$$\begin{aligned}&P\left(2l^2 \sum_{i=1}^{n} (\phi(Y_i) + \varphi(Y_i) + 1) \int_{|x|>2M_l} \delta_{m_n}(x, X_i)dx > 2nl\right) \\ &\leq e^{-2nl}\left[E\left[\exp\left\{2l^2(\phi(Y) + \varphi(Y) + 1)\int_{|x|>2M_l} \delta_{m_n}(x, X)dx\right\}\right]\right]^n \\ &= e^{-2nl}\left[\int\int \exp\left\{2l^2(\phi(y) + \varphi(y) + 1)\int_{|x|>2M_l} \delta_{m_n}(x, z)dx\right\} g(z, y)dzdy\right]^n.\end{aligned}$$

Thus,

$$\begin{aligned}&P\left((r_n^{\phi}, r_n^{\varphi}, g_n) \in \Gamma_L^c\right) \leq \\ &\sum_{l \geq L} e^{-2nl}\left[\int_{-M_l}^{M_l}\int_{-\infty}^{\infty} \exp\left\{2l^2(\phi(y) + \varphi(y) + 1)\int_{|x|>2M_l} \delta_{m_n}(x, z)dx\right\} g(z, y)dzdy\right. \\ &\left.+ \int_{|z|>M_l}\int_{-\infty}^{\infty} \exp\left\{2l^2(\phi(y) + \varphi(y) + 1)\int_{|x|>2M_l} \delta_{m_n}(x, z)dx\right\} g(z, y)dzdy\right]^n.\end{aligned}$$

Using the condition ($\mathbf{H_2}$) and the fact that $\int_{|x|>2M_l} \delta_{m_n}(x, z)dx \leq 1$, we obtain

$$\begin{aligned}P\left((r_n^{\phi}, r_n^{\varphi}, g_n) \in \Gamma_L^c\right) &\leq \sum_{l \geq L} e^{-2nl}\left[\int_{-M_l}^{M_l}\int_{-\infty}^{\infty} \exp\left\{(\phi(y) + \varphi(y) + 1)\right\} g(z, y)dzdy\right. \\ &\left.+ \int_{|z|>M_l}\int_{-\infty}^{\infty} \exp\left\{2l^2(\phi(y) + \varphi(y) + 1)\right\} g(z, y)dzdy\right]^n \\ &\leq \sum_{l \geq L} e^{-2nl}\left[\int_{-M_l}^{M_l} F(x)dx\right. \\ &\left.+ \int_{|z|>M_l}\int_{-\infty}^{\infty} \exp\left\{2l^2(\phi(y) + \varphi(y) + 1)\right\} g(z, y)dzdy\right]^n\end{aligned}$$

Let $\beta > 1$ such that $1/\alpha + 1/\beta = 1$, Combining the statements (3.3),(3.4) and the

Young inequality ($uv \leq u^\alpha/\alpha \;+\; v^\beta/\beta \leq u^\alpha + v^\beta$ for $u, v > 0$), it follows that

$$
\begin{aligned}
P\left((r_n^\phi, r_n^\varphi, g_n) \in \Gamma_L^c\right) &\leq \sum_{l \geq L} e^{-2nl} \left[\int_{-M_l}^{M_l} F(z)dz + e^{2^\beta l^{2\beta}} \int_{|z|>M_l} F(z)dz\right]^n \\
&= \sum_{l \geq L} e^{-2nl} \left[A + \left(e^{2^\beta l^{2\beta}} - 1\right) \int_{|z|>M_l} F(z)dz\right]^n \\
&\leq \sum_{l \geq L} e^{-2nl} \left[A + \left(e^{2^\beta l^{2\beta}} - 1\right) e^{-2^\beta l^{2\beta}} (e^l - A)\right]^n \\
&\leq \sum_{l \geq L} e^{-nl} \\
&\leq 2e^{-nL}.
\end{aligned}
$$

This achieves the proof of the statement (4.2).

Step 3. Since the Logarithm Laplace transform associate to the process W_n is finite and Gâteaux differentiable everywhere, and the process is exponentially tight, then according to the abstract Gärtner-Ellis, the Process W_n satisfies a large deviations principle with the speed $1/n$ and the rate function $\Lambda^*_{\phi,\varphi}$.

Proof of Corollary 4.1. The proof of this corollary is a direct application of the contraction principle. In this respect, the functions φ and ϕ are chosen to be null. As matter of fact, we consider therefore the process $W_n = (0, 0, g_n)$. The conditions $(\mathbf{H_1})-(\mathbf{H_2})$ and the obvious fact that the null function satisfies the conditions $(\mathbf{H_3}) - (\mathbf{H_4})$, imply that the conditions of Theorem 2.1 are satisfied. Consequently, $(0, 0, g_n)$ satisfies a large deviations principle in $\mathcal{X}$ with the speed $1/n$ and the good rate function

$$
\begin{aligned}
\Lambda^*_{0,0}(a, b, c) = \sup_{k_1,k_2,k_3 \in L_\infty} \Bigg\{ \int a(x)k_1(x)dx &+ \int b(x)k_2(x)dx \\
&+ \int c(x)k_3(x)dx - \log \int e^{k_3(x)} g_X(x)dx \Bigg\}.
\end{aligned}
$$

Obviously, we have

$$
\Lambda^*_{0,0}(0, 0, c) = \sup_{k_3 \in L_\infty} \left\{ \int c(x)k_3(x)dx - \log \int e^{k_3(x)} g_X(x)dx \right\},
$$

and

$$
\Lambda^*_{0,0}(a, b, c) = \infty \quad \text{if} \quad (a, b) \neq (0, 0).
$$

In other words, the finiteness domain $\mathcal{D}_{\Lambda^*_{0,0}}$ of the function $\Lambda^*_{0,0}$ as a subset of $\{0\} \times \{0\} \times L_1$ is a closed set of the space $\mathcal{X}$. Therefore, by Lemma 4.5.1 in (Dembo &

Zeitouni (1998)(23), page 118), the process $(0,0,g_n)$ satisfies a large deviations principle in $\{0\}\times\{0\}\times L_1$ with the speed $1/n$ and the rate function $\Lambda^*_{0,0}$. Consider now the map

$$\begin{aligned}\mathcal{H} \ &: \ \{0\}\times\{0\}\times L_1 \longrightarrow L_1\\ &\qquad (0,0,c) \quad \longrightarrow c\end{aligned}$$

which, obviously, is continuous. Since $\mathcal{H}(0,0,g_n)=g_n$, the process g_n satisfies a large deviation principle in L_1 with the speed $1/n$ and the rate function l^*, where

$$l^*(c)=\Lambda^*_{0,0}(0,0,c).$$

Finally, the proof of the statement (4.1) is given in Dupuis & Ellis (1997)(29). □

Proof of Corollary 4.2. Since $|f(y)|\le M$, $f^+(y)\le M$ and $f^-(y)\le M$ and then they satisfy the conditions $(\mathbf{H_3})-(\mathbf{H_6})$. In first place we will prove that (r_n^f,g_n) satisfies a LPD in $L^1(T)\times L^1(T)$. In this order, consider the following application

$$\begin{aligned}\mathcal{H}_1 \ &: \ L^1(T)\times L^1(T)\times L^1(T)\longrightarrow L^1(T)\times L^1(T)\\ &\quad (a,b,c)\longrightarrow (a-b,c).\end{aligned}$$

Obviously this application is continuous and transforms the process $(r_n^{f^+},r_n^{f^-},g_n)$ into the process (r_n^f,g_n), i.e., $\mathcal{H}_1(r_n^{f^+},r_n^{f^-},g_n)=(r_n^f,g_n)$. Making use of the contraction principle, it follows that the process (r_n^f,g_n) satisfies a LDP with the rate function

$$\begin{aligned}\Lambda^*_f(d,c) &= \inf\{\Lambda_{f^+,f^-}(a,b,c)\ :\ a-b=d\}\\ &= \inf_{b\in L^1(T)}\Lambda_{f^+,f^-}(b+d,b,c)\\ &= \gamma^*(d,c).\end{aligned}$$

Consider now the subset $\mathcal{B}\subset L^1(T)\times L^1(T)$ defined as follows

$$(a,c)\in\mathcal{B}\Leftrightarrow\ c\ge 0\ \text{ and }\ |a(x)|\le Mc(x)\quad\text{almost everywhere}$$

The set $\mathcal{B}$ is a closed set. Indeed, let $(a_n,c_n)\in\mathcal{B}$ be a sequence that converge to (a,b). Let us verify that $(a,c)\in\mathcal{B}$. Since $c_n\ge 0$ and $c_n\to c$, then $c\ge 0$ a.e. It remains to verify that $\frac{|a|}{c}\le M$ a.e. Suppose the opposite, i.e., $\frac{|a|}{c}\not\le M$ a.e. Without loss of generality, we can the suppose that there exist $\epsilon>0$ and a Borel set A (with $\lambda(A)>0$) such that

$$a(x)>(M+\epsilon)c(x),$$

whenever $x\in A$. It follows that

$$\int_A a(x)dx\ >\ (M+\epsilon)\int_A c(x)dx. \tag{4.5}$$

We consider two cases
Suppose that $\int_A c(x)dx = 0$. Since $(a_n, c_n) \in \mathcal{B}$, it follows that

$$|a_n(x)| \leq Mc_n(x) \quad \text{a.e.}$$

Thus,

$$-M \int c_n(x)\mathbb{1}_A(x)dx \leq \int a_n(x)\mathbb{1}_A(x)dx \leq M \int c_n(x)\mathbb{1}_A(x)dx.$$

Therefore, since $\int c_n(x)\mathbb{1}_A(x)dx \to \int c(x)\mathbb{1}_A(x)dx = 0$ and $\int a_n(x)\mathbb{1}_A(x)dx \to \int a(x)\mathbb{1}_A(x)dx$, it results that

$$\int a(x)\mathbb{1}_A(x)dx = 0.$$

This contradicts the statement (4.5).
Suppose now that $\int c(x)\mathbb{1}_A(x)dx > 0$. From the previous case, it follows that $\int a(x)\mathbb{1}_A(x)dx > 0$. Since $a_n \to a$, then for any $\alpha > 0$, there exists n_0, such that

$$\text{for any } \; n \geq n_0, \quad \int a_n(x)\mathbb{1}_A(x)dx \geq (1-\alpha)\int a(x)\mathbb{1}_A(x)dx. \tag{4.6}$$

Combining the statements (4.6) and (4.5), it results that

$$\text{for any } \; n \geq n_0, \quad \int a_n(x)\mathbb{1}_A(x)dx \geq (1-\alpha)(M+\epsilon)\int c(x)\mathbb{1}_A(x)dx.$$

Choosing $\alpha \leq \epsilon/2(M+\epsilon)$, we obtain

$$\text{for any } \; n \geq n_0, \quad \int a_n(x)\mathbb{1}_A(x)dx \geq (M+\frac{\epsilon}{2})\int c(x)\mathbb{1}_A(x)dx. \tag{4.7}$$

Similary, since $c_n \to c$, for any $\sigma > 0$, there exists n_1, such that

$$\text{for any } \; n \geq n_1, \quad \int c_n(x)\mathbb{1}_A(x)dx \leq (1+\sigma)\int c(x)\mathbb{1}_A(x)dx. \tag{4.8}$$

The statement (4.8), combined with the statment (4.7), implies that, for $n \geq \max\{n_0, n_1\}$,

$$\int a_n(x)\mathbb{1}_A(x)dx \geq \frac{M+\frac{\epsilon}{2}}{1+\sigma}\int c_n(x)\mathbb{1}_A(x)dx.$$

A choice of $\sigma \leq \epsilon/(4M+\epsilon)$ gives

$$\int a_n(x)\mathbb{1}_A(x)dx \geq (M+\frac{\epsilon}{4})\int c_n(x)\mathbb{1}_A(x)dx. \tag{4.9}$$

The statement (4.9) contradicts the fact that $(a_n, c_n) \in \mathcal{B}$. We achieved the the proof of the closeness of the set $\mathcal{B}$.

Since the function f is bounded, it is obvious that $P\left((r_n^f, g_n) \in \mathcal{B}\right) = 1$. Moreover, the set $\mathcal{B}$ is closed in $L^1(T) \times L^1(T)$. Therefore, by Lemma 4.5.1 page 118 in (23), it results that (r_n^f, g_n) satisfies a LDP in $\mathcal{B}$ with the same speed and rate function.

Consider now the following application

$$\begin{aligned} \mathcal{H}_2 \quad : \quad & \mathcal{B} \longrightarrow L^1(T) \\ & (a,c) \longrightarrow H_2(a,c), \end{aligned}$$

where

$$H_2(a,c)(x) = \begin{cases} \frac{a(x)}{c(x)} & \text{if } b(x) > 0 \\ 0 & \text{elsewhere.} \end{cases}$$

For $c \in L^1(T)$, set $S(c) = \{x \in T : c(x) > 0\}$. In order to verify that H_2 is continuous, it is enough to verify that, for any $L^\infty(T)$-function k, the application $H_k : \mathcal{B} \longrightarrow \mathbb{R}$, defined by

$$H_k(a,c) = \int_{S(c)} \frac{a(x)}{c(x)} k(x)dx$$

is continuous. For $(a_1, c_1), (a_2, c_2) \in \mathcal{B}$, we have

$$\begin{aligned} |H_k(a_1,c_1) - H_k(a_2,c_2)| &= \left| \int_{S(c_1)} \frac{a_1(x)}{c_1(x)} k(x)dx - \int_{S(c_2)} \frac{a_2(x)}{c_2(x)} k(x)dx \right| \\ &\leq 2M\|k\|_\infty \lambda(S(c_1)\Delta S(c_2)) \\ &+ \left| \int_{S(c_1)\cap S(c_2)} \left(\frac{a_1(x)}{c_1(x)} - \frac{a_2(x)}{c_2(x)} \right) k(x)dx \right|, \end{aligned}$$

where $A\Delta B$ denotes the symmetric difference between the set A and the set B, i.e., $A\Delta B = A \cap B^c \cup A^c \cup B$. Set now $S_\delta(c_1) = \{x \in T : c_1(x) > \delta\}$. It follows that

$$\begin{aligned} |H_k(a_1,c_1) - H_k(a_2,c_2)| \leq 2M\|k\|_\infty \lambda(S(c_1)\Delta S(c_2)) + 2M\|k\|_\infty \lambda(S(c_1)/S_\delta(c_1)) \\ + \frac{1}{\delta} \left| \int_{S_\delta(c_1)} (a_1(x) - a_2(x))\, k(x)dx \right| \\ + \frac{M}{\delta} \left| \int_{S_\delta(c_1)} (c_1(x) - c_2(x))\, k(x)dx \right|, \end{aligned}$$

for any $\delta > 0$, we have Thus, for $(a,c) \in \mathcal{B}$ fixed and $\epsilon > 0$ we can find a neighborhood V of (a,b) such that

$$(a', c') \in V \quad \Rightarrow \quad |H_k(a',c') - H_k(a,c)| < \epsilon.$$

The remainder of the proof follows from the contraction principle since H_2 is continuous.

□

Proof of Proposition 4.1. In order to prove the condition $\mathbf{H_1}$, It suffices to show that

$$\lim_{n\to\infty}\int\int K_n(x,y)\,|k(x)-k(y)|\,g_X(y)dxdy \;=\; 0. \tag{4.10}$$

Observe that

$$\begin{aligned}
&\int\int K_n(x,y)\,|k(x)-k(y)|g_X(y)dxdy\\
&\qquad = \int\int K_n(x,y)\,|k(x)g_X(y)-k(y)g_X(y)|\,dxdz\\
&\qquad \le \int\int K_n(x,y)\,|k(x)g_X(y)-k(x)g_X(x)|\,dxdz\\
&\qquad + \int\int K_n(x,y)\,|k(x)g_X(x)-k(y)g_X(y)|\,dxdz\\
&\qquad \le \|k\|_\infty\int\int K_n(x,y)\,|g_X(y)-g_X(x)|\,dxdz\\
&\qquad + \int\int K_n(x,y)\,|k(x)g_X(y)-k(y)g_X(y)|\,dxdz.
\end{aligned} \tag{4.11}$$

It follows from the statement (4.11) that it is enough to show that for any integrable function h,

$$\lim_{n\to\infty}\int\int K_n(x,y)\,|h(x)-h(y)|\,dxdy = 0. \tag{4.12}$$

Subsequently, for any $\epsilon>0$, there exists a continuous function h^* with support $[-R,R]$, such that

$$\int |h(x)-h^*(x)|dx<\epsilon.$$

Since $\int K_n(x,y)dx=\int K_n(x,y)dy=1$, then we have

$$\begin{aligned}
\int\int K_n(x,y)\,|h(x)-h(y)|\,dxdy \;&\le\; \int\int K_n(x,y)\,|h(x)-h^*(x)|\,dxdy\\
&+ \int\int K_n(x,y)\,|h^*(x)-h^*(y)|\,dxdy\\
&+ \int\int K_n(x,y)\,|h^*(y)-h(y)|\,dxdy\\
&= \int |h(x)-h^*(x)dx\\
&+ \int\int K_n(x,y)\,|h^*(x)-h^*(y)|\,dxdy\\
&+ \int |h^*(y)-h(y)|dy\\
&\le 2\epsilon+\int\int K_n(x,y)\,|h^*(x)-h^*(y)|\,dxdy.
\end{aligned}$$

Therefore, it suffices to show that

$$\lim_{n\to\infty} \int\int P_n(x,y)\,|h^*(x) - h^*(y)|\,dxdy = 0. \tag{4.13}$$

The remainder of the proof is similar to the proof of Lemma 1 in Devroye (1983)(25).

In order to verify the condition $\mathbf{H_2}$, observe that, since the kernel K is tight, then for any $l > 0$, there exists $M_l > 0$ such that

$$\int_{|u|>M_l} K(u)du \;<\; \frac{1}{2l^2}.$$

Since $h_n \to 0$ as $n \to \infty$ then, for n large enough

$$\begin{aligned}\int_{|x|>2M_l} \mathbb{1}_{\{|z|\le M_l\}} K\left(\frac{x-z}{h_n}\right) dx \;&\le \int_{|u|>M_l} \mathbb{1}_{\{|z|\le M_l\}} K(u)du \\ &< \frac{1}{2l^2}.\end{aligned}$$

□

Proof of Proposition 4.2. Recall that in the histogram case, the delta-sequence is defined by

$$P_n(x,y) = \sum_{j\in J} \frac{\mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y)}{\lambda(A_{n,j})}.$$

In order to prove the condition $\mathbf{H_1}$, observe first that

$$\int P_n(x,y)dx = \int P_n(x,y)dy = 1.$$

Thus, for any L_∞-function k, we have

$$\int P_n(x,X_1)k(x)dx \;-\; k(X_1) \;=\; \int P_n(x,X_1)(k(x) - k(X_1))dx.$$

Similarly as in the proof of Proposition 4.1, it is enough to show, for any continuous function h^* with support $[-R,R]$, that

$$\lim_{n\to\infty} \int\int P_n(x,y)\,|h^*(x) - h^*(y)|\,dxdy = 0. \tag{4.14}$$

Observe that

$$\int\int P_n(x,y)\,|h*(x) - h^*(y)|\,dxdy = \sum_{j\in J} \int_{A_{n,j}} \int_{A_{n,j}} \frac{|h^*(x) - h^*(y)|}{\lambda(A_{n,j})} dxdy. \tag{4.15}$$

The continuity of the function h^* over its bounded support, which is actually a uniform continuity, implies that there exists a real number $\sigma(\epsilon) > 0$ such that $|h^*(x) - h^*(y)| < \epsilon$, whenever $|x - y| < \sigma(\epsilon)$. Set

$$J_n = \{j \in J \ : \ A_{n,j} \cap [-R, R] \neq \emptyset\}.$$

It follows then that

$$\int\int P_n(x, y) \, |h * (x) - h^*(y)| \, dxdy = \sum_{j \in J_n} \int_{A_{n,j}} \int_{A_{n,j}} \frac{|h^*(x) - h^*(y)|}{\lambda(A_{n,j})} dxdy.$$

Let $n_0 \in \mathbb{N}$ be such that, for any $n \geq n_0$, $\max_{j \in J} \lambda(A_{n,j}) < \sigma(\epsilon)$. Since

$$\bigcup_{j \in J} A_{n,j} \subseteq [-R - \sigma(\epsilon), R + \sigma(\epsilon)],$$

then we have

$$\sum_{j \in J_n} \int_{A_{n,j}} \left(\int_{A_{n,j}} \frac{|h^*(x) - h^*(y)|}{\lambda(A_{n,j})} dx \right) dy \leq 2\epsilon(R + \sigma(\epsilon)). \tag{4.16}$$

Combining the statements (4.15) and (4.16), we obtain the result (4.14). This achieves the proof of the condition $\mathbf{H_1}$.

To prove the condition $\mathbf{H_2}$, we observe, since

$$\lim_{n \to \infty} \max_{j \in J} \lambda(A_{n,j}) \ = \ 0.$$

that, for any $l > 0$, there exists $M_l > 0$ such that, for n large enough,

$$\max_{j \in J} \lambda(A_{n,j}) < M_l.$$

Therefore, we have

$$P_n(x, z) \mathbb{1}_{\{|x| > 2M_l\}} \mathbb{1}_{\{|z| \leq M_l\}} \ = \ 0.$$

This statement implies that the condition $\mathbf{H_2}$ holds true. □

Chapitre 5

$L^1(\mathbb{R})$-functional moderate deviations principle for the histogram density estimate

This paper establishes a moderate deviations principle for the histogram in the $L^1(\mathbb{R})$-space equipped with the weak topology. Subsequently, we derive a Chernoff type result for the L^1-moderate deviation for the histogram.

5.1 Introduction

Let $X_1, X_2, ...X_n$ be a sequence of independent and identically distributed random variables defined on a probability space $(\Omega, \mathcal{A}, P)$. Let f be the density function of the random variable X_1. Let $\mathcal{P}_n = \{A_{n,j} : j \in J\}$ be a partition of the real line. The histogram estimator (called also histogram) of the density function based on the sample $(X_1, X_2, ..., X_n)$, at stage n, is defined, for any $x \in \mathbb{R}$, by

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_n(x, X_i),$$

where

$$\delta_n(x, y) = \sum_{j \in J} \frac{\mathbb{1}_{A_{n,j}}(x)\mathbb{1}_{A_{n,j}}(y)}{\lambda(A_{n,j})}.$$

Here $\mathbb{1}_{A_{n,j}}$ and λ denote respectively the indicator function of the set $A_{n,j}$ and the Lebesgue measure on $\mathbb{R}$.

It is seen in chapter 4 that the process f_n satisfies a large deviations principle with the Kullback information as a rate function. The main goal of this paper is to establish a moderate deviations principle for the centered process $(f_n - E(f_n))$ in the space L^1 equipped with the weak topology. More explicitly, we identify two sequences $\epsilon_n \to 0$, $b_n \to \infty$ as $n \to \infty$ as well as a lower semi-continuous function I on L^1, such that the process $\frac{n}{b_n}(f_n - E(f_n))$ satisfies a large deviation principle with the speed ϵ_n and the rate function I, i.e.,

– For any closed set $F \subset L^1$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n)) \in F\right) \;\leq\; -\inf_{g\in F}\; I(g),$$

– For any open set $O \subset L^1$,

$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n)) \in O\right) \;\geq\; -\inf_{g\in O}\; I(g),$$

Recall that the topological dual of the space L^1 is the space L^∞ and that the weak topology on L^1 is the weakest topology that makes continuous the linear forms in L^1, i.e., the forms defined by

$$\begin{aligned} Q \;:\; & L^1 \longrightarrow \mathbb{R} \\ & g \longrightarrow < g, k >_{dx} := \int g(x)k(x)dx, \end{aligned}$$

for any $k \in L^\infty$.

There exists an extensive moderate deviations literature in probability and statistics, we cite de Acosta (1992) (18) who establishes a moderate deviations principle for the sum of i.i.d. random vector taking values in a Banach space. Considering the sum of an i.i.d. sequence of random vectors taking values in a Banach space, Ledoux (1992) (44) gives, for large classes of normalizing sequences $(b_n)_n$, a necessary and sufficient condition for the moderate deviations upper bound. de Acosta (1997) (20) obtains lower bounds for moderate deviations of empirical measures of a Markov chain with general state space under the assumption of ergodicity of degree 2. de Acosta & Chen, Xia (1998) (21) obtain upper bounds for moderate deviations of sums of Markov chains and for empirical measures of Markov chains. Hu, Yijun & Lee, Tzong-Yow (2001) (42) establish moderates deviations principles of sums of i.i.d. random vectors taking values in a separable Banach space, generalizing thus the related results of Borovkov & Mogulskii (1980) (9) where moderate deviations results for empirical measures in a Hausdorff space are stated. Gao, Fuqing (2006) (36) gives a functional moderate deviation principle for random processes with stationary and independent increments. We cite also Chen, Xia & Guillin (2004) (14) where a moderate deviation principle for Harris recurrent Markov chains is established and a functional law of the iterated logarithm is obtained. As results involving nonparametric estimation we cite Mokkadem *et al.* (2005) (58), who study the large deviations behaviour of the kernel estimator of a density function, when the kernel takes negative values. They establish large and moderate deviations principles for the kernel estimators of partial derivatives of a density function. In Mokkadem *et al.* (2005) (57), the authors establish that, when the kernel estimator of the mode of a probability density is known to fulfill a central limit theorem, it also satisfies a moderate deviations principle. They apply this result to the analysis of confidence intervals for the mode. Gao, Fuqing (2003) (35) establishes a Chernoff type result for the uniform moderate deviations and a pointwise moderate deviation principle for a kernel density estimator.

5.2 Results

In order to set the needed conditions, introduce some notations and definitions. From now on, we denote by μ the probability measure associate to the density f and by μ_n the empirical measure associated the sample $X_1, X_2, ..., X_n$, i.e., for any borel set $B \subset \mathbb{R}$,

$$\mu_n(B) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_B(X_i).$$

Consider the map H from $L^2(\mu)$ to L^1 such that $H(h) := hf$, where $L^2(\mu)$ denotes the space of functions square integrable with respect to μ. Set $\mathcal{H}_\mu := H(L^2(\mu))$ the range of $L^2(\mu)$ by the transformation H. We equip the space $\mathcal{H}_\mu$ with the scalar product $\langle\ ,\ \rangle_{\mathcal{H}_\mu}$ and the underlaying norm $\|\ \|_{\mathcal{H}_\mu}$ defined as follows, for $g_1 = H(h_1), g_2 = H(h_2) \in \mathcal{H}_\mu$,

$$\langle g_1, g_2\rangle_{\mathcal{H}_\mu} := \langle h_1, h_2\rangle_\mu = \int h_1 h_2 d\mu.$$

In this way the space $\mathcal{H}_\mu$ is a Hilbert space and its closed unit ball $K := \{g \in \mathcal{H}_\mu : \|g\|_{\mathcal{H}_\mu} \leq 1\}$ is weakly compact in L^1.

In the following we give respectively the normalized logarithm Laplace transform limit associate with the centered process $nb_n^{-1}(f_n - Ef - n)$ and its Legendre transform. For any L^∞-function k, set

$$\Lambda(k) \ = \ \frac{1}{2}\int_{-\infty}^{\infty} k^2(x)f(x)\ dx,$$

and denote by Λ^* the Legendre transform of Λ i.e., for any L^1-function g,

$$\Lambda^*(g) = \sup_{k\in L^\infty}\left\{\int_{-\infty}^{\infty} g(x)k(x)dx - \frac{1}{2}\int_{-\infty}^{\infty} k^2(x)f(x)dx\right\}.$$

We identify in Lemma 1 the function Λ^* and its effective domaine.

Set now the assumptions upon the the partition $\mathcal{P}_n$, the sequence $(b_n)_n$ necessary to establish our results.

$(\mathbf{H_1})$ $nb_n^{-1} \to \infty$ and $\sqrt{n}b_n^{-1} \to 0$ as $n \to \infty$.

$(\mathbf{H_2})$ $\max_{j\in J} \lambda(A_{n,j}) \to 0$ and $n \min_{j\in J} \lambda(A_{n,j}) \to \infty$

$(\mathbf{H_3})$ There exists a sequence $(r_n)_n$ such that $\lim_{n\to\infty} r_n = \infty$ and

– $\dfrac{n\ r_n}{b_n^2\ \min_j \lambda(A_{n,j})} \ \longrightarrow\ 0$ as $n \to \infty$.

– $\limsup\limits_{n\to\infty} n^2 b_n^{-2} P(|X_1| > r_n) \ < \ \infty$.

Remark The condition $(\mathbf{H_1})$ is of classical nature in the moderate deviations literature. The condition $(\mathbf{H_3})$ ensure somehow the exponential tightness of the process $nb_n^{-1}(f_n - Ef_n)$ and holds true for $r_n = n/b_n$ whenever :
$n^2 b_n^{-3}(\min_j \lambda(A_{n,j}))^{-1} \to 0$ as $n \to \infty$,
$P(|X| > M) = O(M^{-2})$ as $M \to \infty$.

The following lemma and theorem give respectively an identification of the rate function and a moderate deviations principle for the process $(f_n - Ef_n)$.

Lemma 5.1. *For any $g \in L^1$, we have*

$$\Lambda^*(g) = \begin{cases} \frac{1}{2}\|g\|^2_{\mathcal{H}_\mu} & \text{if } g \in \mathcal{H}_\mu, \\ \infty & \text{elswhere} \end{cases}$$

The following theorem gives a moderate deviations principle for the process $(f_n - E(f))$.

Theorem 5.1. *Under Conditions $(\mathbf{H_1})$-$(\mathbf{H_3})$, we have*

– *for any open set $O \subset L^1$,*

$$\liminf_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n)) \in O\right) \ \geq\ -\inf_{g\in O} \Lambda^*(g) \ :=\ \Lambda^*(O),$$

– *for any closed set $F \subset L^1$,*

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n)) \in F\right) \ \leq\ -\inf_{g\in F} \Lambda^*(g) \ :=\ -\Lambda^*(F).$$

Corollary 5.1. *Under hypotheses of Theorem 5.1, we have for any $r > 0$*

$$\lim_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}\|f_n - E(f_n)\|_{L^1} > r\right) \; = -\frac{r^2}{2}.$$

5.3 Proof

Proof of Theorem 5.1. Suppose that $g \in \mathcal{H}_\mu$, then there exists $h \in L^1(\mu)$ such that $g = hf$. It follows that

$$\begin{aligned}\Lambda^*(g) - \frac{1}{2}\|g\|^2_{\mathcal{H}_\mu} &= \sup_{k\in L^\infty}\left\{\int kh\; d\mu - \frac{1}{2}\int k^2\; d\mu - \frac{1}{2}\int h^2\; d\mu\right\}\\ &= -\frac{1}{2}\inf_{k\in L^\infty}\int (h-k)^2\; d\mu.\end{aligned}$$

Setting $h_n \min(h, n)$, it follows that, for any n, $h_n \in L^\infty$ and consequently,

$$0 \;\le\; \inf_{k\in L^\infty}\int (h-k)^2\; d\mu \;\le\; \int (h-h_n)^2\; d\mu.$$

Considering the limit $n \to \infty$ in the last statement, we obtain

$$\inf_{k\in L^\infty}\int (h-k)^2\; d\mu \;=\; 0,$$

since $h \in L^2(\mu)$. It follows that, for any $g \in \mathcal{H}_\mu$

$$\Lambda^*(g) = \frac{1}{2}\|g\|^2_{\mathcal{H}_\mu}.$$

In order to complete the proof, it is enough to show that $\Lambda^*(g) = \infty$ for $g \notin \mathcal{H}_\mu$. Observe that $\Lambda^*(g) = \Lambda^*(|g|)$. Indeed, for any $k \in L^\infty$, we have

$$\begin{aligned}\int g(x)k(x)dx \;-\; \frac{1}{2}\int k^2(x)f(x)dx \;&=\; \int |g(x)|l(x)dx \;-\; \frac{1}{2}\int l^2(x)f(x)dx\\ &\le\; \Lambda^*(|g|),\end{aligned}$$

where $l = k\mathbb{1}_{\{g\ge 0\}} - k\mathbb{1}_{\{g<0\}}$. It follows that $\Lambda^*(g) \le \Lambda^*(|g|)$. Similarly, we have $\Lambda^*(|g|) \le \Lambda^*(g)$. Suppose now that $g \notin \mathcal{H}_\mu$, then $|g| \notin \mathcal{H}_\mu$. Consequently, $|g|$ is not absolutely continuous with respect to f and therefore, there exists a borel set A such that $\int_A f(x)dx = 0$ and $\int_A |g(x)|dx > 0$. For $k = t\mathbb{1}_A$, we have

$$\int |g(x)|k(x)dx - \frac{1}{2}\int k^2(x)f(x)dx = t\int_A |g(x)|dx \le \Lambda^*(g).$$

Considering the limit $t \to \infty$, we obtain $\Lambda^*(|g|) = \infty$ and consequently, $\Lambda^*(g) = \infty$. □

Proof of Theorem 5.1. The proof is based on the abstract gärtner-Ellis Theorem. For any L^∞-function k, set

$$\Lambda(k) = \lim_{n\to\infty} \frac{n}{b_n^2} \log E\left[\exp\left\{b_n\langle f_n - E(f_n), k\rangle_{dx}\right\}\right],$$

where $\langle h, k\rangle := \langle h, k\rangle_{dx} = \int h(x)k(x)dx$. We have

$$\begin{aligned} E\left[\exp\left\{b_n\langle f_n - E(f_n), k\rangle\right\}\right] &= \exp\left\{-b_n\langle E(f_n), k\rangle\right\} E\left[\exp\left\{b_n\langle f_n, k\rangle\right\}\right] \\ &= \exp\left\{-b_n\langle E\left(\delta_n(.,X_1)\right), k\rangle\right\} E\left[\exp\left\{b_n\langle f_n, k\rangle\right\}\right]. \end{aligned} \tag{5.1}$$

Since the random variables $X_1, X_2, ..., X_n$ are i.i.d., it follows that

$$\begin{aligned} E\left[\exp\left\{b_n\langle f_n, k\rangle\right\}\right] &= E\left[\exp\left\{\frac{b_n}{n}\sum_{i=1}^n \langle \delta_n(.,X_i), k\rangle\right\}\right] \\ &= \left[E\left[\exp\left\{\frac{b_n}{n}\langle \delta_n(.,X_1), k\rangle\right\}\right]\right]^n \\ &= \left[\Psi_n(k)\right]^n, \end{aligned} \tag{5.2}$$

where

$$\Psi_n(k) = E\left[\exp\left\{\frac{b_n}{n}\langle \delta_n(.,X_1), k\rangle\right\}\right].$$

From the combination of the statements (5.1) and (5.2), we have

$$\frac{n}{b_n^2}\log E\left[\exp\left\{b_n\langle f_n - E(f_n), k\rangle\right\}\right] \;=\; -\frac{n}{b_n}\langle E\left(\delta_n(.,X_1)\right), k\rangle \;+\; \frac{n^2}{b_n^2}\log \Psi_n(k). \tag{5.3}$$

Taylor series expansion of the exponential function gives

$$\begin{aligned} \Psi_n(k) \;=\; 1 \;+\; \frac{b_n}{n}E\left[\langle \delta_n(.,X_1), k\rangle\right] \;&+\; \frac{b_n^2}{2n^2}E\left[\langle \delta_n(.,X_1), k\rangle^2\right] \\ &+\; o\left(\frac{b_n^2}{n^2}E\left[\langle \delta_n(.,X_1), k\rangle^2\right]\right). \end{aligned} \tag{5.4}$$

The condition (H.1), the statement (5.4) together with the fact that $\langle \delta_n(.,X_1), k\rangle \leq \|k\|_\infty$ give

$$\begin{aligned} \log \Psi_n(k) \;\simeq\; \frac{b_n}{n}E\left[\langle \delta_n(.,X_1), k\rangle\right] \;&+\; \frac{b_n^2}{2n^2}E\left[\langle \delta_n(.,X_1), k\rangle^2\right] \\ &+\; o\left(\frac{b_n^2}{n^2}\right), \end{aligned} \tag{5.5}$$

since $\log(1+u) \simeq u$ whenever u is in the neighborhood of zero. Now since $E\left[\langle \delta_n(.,X_1), k\rangle\right] = \langle E[\delta_n(.,X_1)], k\rangle$, we obtain by the combination of the statements (5.3) and (5.5)

$$\frac{n}{b_n^2}\log E\left[\exp\left\{b_n\langle f_n - E(f_n), k\rangle\right\}\right] \;\simeq\; \frac{1}{2}E\left[\langle \delta_n(.,X_1), k\rangle^2\right] + \; o\left(1\right). \tag{5.6}$$

Observe that

$$\lim_{n\to\infty} \langle \delta_n(.,X_1), k\rangle_{dx} = k(X_1) \text{ in probability .} \tag{5.7}$$

Indeed,

$$\begin{aligned} E[|\langle \delta_n(.,X_1), k\rangle_{dx} - k(X_1)|] &\leq \int\int \delta_n(x,y)|k(x)f(y) - k(y)f(y)|dxdy \\ &\leq \|k\|_\infty \int\int \delta_n(x,y)|f(x) - f(y)|dxdy \\ &\quad + \int\int \delta_n(x,y)|f'(x) - f'(y)|dxdy. \end{aligned}$$

Therefore, in order to establish (5.7), it is enough to show that, for any L^1-function h

$$\lim_{n\to\infty} \int\int \delta_n(x,y)|h(x) - h(y)|dxdy = 0.$$

Let $h \in L^1$ and $\epsilon > 0$, then there exists a continuous g with bounded support C such that

$$\int |h(x) - g(x)|dx < \epsilon.$$

It follows then that

$$\begin{aligned} \int\int \delta_n(x,y)|h(x) - h(y)|dxdy &\leq \int\int \delta_n(x,y)|h(x) - g(x)|dxdy \\ &\quad + \int\int \delta_n(x,y)|g(x) - g(y)|dxdy \\ &\quad + \int\int \delta_n(x,y)|g(y) - h(y)|dxdy \\ &< 2\epsilon + \int_C\int_C \delta_n(x,y)|g(x) - g(y)|dxdy, \end{aligned}$$

since $\int \delta_n(x,y)dx = \int \delta_n(x,y)dy = 1$. Moreover, we have

$$\begin{aligned} \int_C\int_C \delta_n(x,y)|g(x) - g(y)|dxdy &= \sum_j \int_{A_{n,j}\cap C}\int_{A_{n,j}\cap C} \lambda(A_{n,j})^{-1}|g(x) - g(x)|dxdy \\ &\leq \lambda(C)w_g(h_n), \end{aligned}$$

where

$$w_g(h_n) = \sup_{x,y\in C,\ |x-y|<h_n} |g(x) - g(y)|,$$

and $h_n = \max_j \lambda(A_{n,j})$. The continuity of the function g which is actually a uniform continuity over the compact set C implies that $w_g(h_n) \to \infty$ since $h_n \to 0$ as $n \to \infty$. This concludes the proof of (5.7) since the choice of $\epsilon > 0$ is arbitrary.

From the statement (5.7), it follows that

$$\lim_{n\to\infty} E\left[\langle \delta_n(.,X_1),k\rangle_{dx}^2\right] \;=\; E[k^2(X_1)] \;=\; \int_{-\infty}^{\infty} k^2(x)f(x)dx.$$

Therefore, we have

$$\Lambda(k) \;=\; \frac{1}{2}\int_{-\infty}^{\infty} k^2(x)f(x)dx.$$

The normalized logarithm Laplace transform Λ, associated with the process $\frac{n}{b_n}(f_n - E(f_n))$, is clearly finite and Gâteaux differentiable everywhere. The lower bound then holds.

To complete the proof it remains to establish the upper bound. Let r_n be as in the condition ($\mathbf{H_3}$) and set $\Pi_{r_n} = \{A_{n,j} : A_{n,j} \cap [-r_n, r_n] \neq \emptyset\}$. We show first that, for any $\alpha > 0$,

$$\limsup_{n\to\infty} \frac{n}{b_n^2}\log P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B)-\mu(B)| > \alpha\right) \;\leq\; -\frac{\alpha^2}{2}. \tag{5.8}$$

By a double use of Scheffé Theorem, we obtain

$$\begin{aligned} P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B)-\mu(B)| > \alpha\right) \;&\leq\; P\left(\sup_{B\in\Pi_{r_n}} |\mu_n(B)-\mu(B)| > \frac{b_n\alpha}{2n}\right) \\ &\leq\; \sum_{B\in\mathcal{E}_{r_n}} P\left(|\mu_n(B)-\mu(B)| > \frac{b_n\alpha}{2n}\right) \\ &\leq\; N(\mathcal{E}_{r_n}) \sup_{B\in\mathcal{B}(\mathbb{R})} P\left(|\mu_n(B)-\mu(B)| > \frac{b_n\alpha}{2n}\right), \end{aligned} \tag{5.9}$$

where $\mathcal{E}_{r_n}$ is the set of all possible unions of elements of the partition Π_{r_n}, $N(\mathcal{E}_{r_n})$ denotes its cardinality and $\mathcal{B}(\mathbb{R})$ denotes the borel σ-field on $\mathbb{R}$.

Now we will give an exponential estimation of the deviations $P\left(|\mu_n(B)-\mu(B)| > a\right)$ for any borel set $B \subset \mathbb{R}$. Observe that

$$|\mu_n(B)-\mu(B)| = \max\{\mu_n(B)-\mu(B)\;,\;\mu(B)-\mu_n(B)\}$$

Thus,

$$\begin{aligned} p_B^n(a) &= P\left(\max\{\mu_n(B)-\mu(B)\ ,\ \mu(B)-\mu_n(B)\} > a\right) \\ &\leq 2\max\{p_{B,+}^n(a),\ p_{B,-}^n(a)\}. \end{aligned} \tag{5.10}$$

where

$$p_B^n(a) = P(|\mu_n(B)-\mu(B)| > a),$$

$$p_{B,+}^n(a) = P(\mu_n(B)-\mu(B) > a)$$

and

$$p_{B,-}^n(a) = P(\mu(B)-\mu_n(B) > a).$$

Using Tchebycheff inequality, we obtain

$$\begin{aligned} p_{B,+}^n(a) &= P\left(t\sum_{i=1}^n \mathbb{1}_B(X_i) > n(a+\mu(B))t\right) \\ &\leq \exp\{-n(a+\mu(B))t\}E\left[\exp\left\{t\sum_{i=1}^n \mathbb{1}_B(X_i)\right\}\right]. \end{aligned}$$

Since the random variables $X_1, ..., X_n$ are i.i.d., it follows that

$$p_{B,+}^n(a) \leq \exp\{-n(a+\mu(B))t\}\left(E\left[\exp\left\{t\mathbb{1}_B(X)\right\}\right]\right)^n. \tag{5.11}$$

Since

$$\begin{aligned} E\left[\exp\left\{t\mathbb{1}_B(X)\right\}\right] &= \int_{-\infty}^{\infty} e^{t\mathbb{1}_B(x)}f(x)dx \\ &= (e^t-1)\mu(B)+1, \end{aligned}$$

it follows then from the statement (5.11) that, for any $t>0$,

$$p_{B,+}^n(a) \leq \exp\{-n\phi_{t,\mu(B)}^+(a)\},$$

where

$$\phi_{t,b}^+(a) = (a+b)t - \log(1+b(e^t-1)).$$

Therefore,

$$p_{B,+}^n(a) \leq \exp\{-n\phi_{\mu(B)}^+(a)\}, \tag{5.12}$$

where

$$\phi_b^+(a) = \sup_{t>0}\varphi_{t,b}^+(a).$$

Similarly, we obtain an analogous version of the statement (5.12) for $p_{B,-}^n(a)$, thus we obtain

$$p_{B,+}^n(a) \leq \exp\{-n\phi_{\mu(B)}^-(a)\}, \tag{5.13}$$

where

$$\phi_b^-(a) = \sup_{t>0} \varphi_{t,b}^-(a),$$

and

$$\varphi_{t,b}^-(a) = (a-b)t - \log(1 + b(e^{-t} - 1)).$$

Setting now

$$\phi_b(a) \;=\; \min\{\phi_b^+(a)\ ,\ \phi_b^-(a)\},$$

we obtain by the combination of the statements (5.10), (5.12) and (5.13),

$$p_B^n(a) \;\leq\; \exp\{-n\phi_{\mu(B)}(a)\}.$$

Thus,

$$\sup_{B\in\mathcal{B}(\mathbb{R})} p_B^n(a) \;\leq\; 2\exp\{-n\phi(a)\}, \tag{5.14}$$

where

$$\phi(a) \;=\; \inf_{0\leq b\leq 1} \phi_b(a).$$

Replacing a by $\frac{b_n\alpha}{2n}$ we obtain by the combination of the statements (5.9) and (5.14)

$$\begin{aligned} P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B) - \mu(B)| > \alpha\right) \;&\leq\; 2N(\mathcal{E}_{r_n})\exp\left\{-n\phi(\frac{b_n\alpha}{2n})\right\} \\ &= 2N(\mathcal{E}_{r_n})\exp\left\{-n\,g(\frac{b_n\alpha}{n})\right\}, \end{aligned}$$

where $g(\alpha) = \phi(\alpha/2)$. The function g is the rate function introduced by Louani (2000) (54) page 3. This function is continuously differentiable and in neighborhood of 0, and has the following shape

$$g(\alpha) \;=\; \frac{\alpha^2}{2}(1 \;+\; o(1)).$$

Consequently, we have

$$\begin{aligned} \limsup_{n\to\infty} \frac{n}{b_n^2}\log P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B) - \mu(B)| > \alpha\right) &\leq\; -\lim_{n\to\infty}\frac{n^2}{b_n^2}g(\frac{b_n\alpha}{n}) \\ &\quad + \lim_{n\to\infty}\frac{2n}{b_n^2}N(\mathcal{E}_{r_n}). \\ &= -\frac{\alpha^2}{2}, \end{aligned} \tag{5.15}$$

since $N(\mathcal{E}_{r_n})$ is at most equal to $2^{N(\Pi_{r_n})}$ and $N(\Pi_{r_n}) \leq \frac{2r_n}{\min_j \lambda(A_{n,j})}$. This achieves the proof of the statement (5.8).

Let F be a closed set in L^1. If $\Lambda^*(F) = 0$, the upper bound is obvious. Suppose that $0 < r < \Lambda^*(F)$. Setting $K = \left\{ g \in H_\mu : (\int \frac{g^2(x)}{f(x)} dx)^{1/2} \leq 1 \right\}$, K is a compact set in L^1 and $\sqrt{2r}K \cap F = \emptyset$. Therefore, there exists δ such that $\sqrt{2r}K + \delta A \cap F = \emptyset$, where A is the closed unit ball in L^1 with respect to the norm $\| \ \|_{L^1}$. It follows that

$$\begin{aligned} P(\frac{n}{b_n}(f_n - E(f_n)) \in F) &\leq P(\frac{n}{b_n}(f_n - E(f_n)) \in (\sqrt{2r}K + \delta A)^c) \\ &\leq P(q(\frac{n}{b_n}(f_n - E(f_n))) > \sqrt{2r}), \end{aligned} \tag{5.16}$$

where q is the Minkowski functional of the convex symmetric set $K + \frac{\delta}{\sqrt{2r}}A$, i.e.,

$$q(g) := \inf \left\{ \beta : g \in \beta(K + \frac{\delta}{\sqrt{2r}}A) \right\}.$$

The presence of the term $\frac{\delta}{\sqrt{2r}}A$ and the fact that $K \subset A$ ensure that the norm q is equivalent to the norm $\| \ \|_{L^1}$. Moreover, we have $q(g) \leq \frac{\sqrt{2r}}{\delta}\|g\|_{L^1}$.

Observe that we can write

$$\begin{aligned} f_n - E(f_n) &= \sum_{B \in \Pi_{r_n}} (\mu_n B - \mu(B)) \frac{\mathbb{1}_B}{\lambda(B)} + \sum_{B \in \mathcal{P}_n / \Pi_{r_n}} (\mu_n B - \mu(B)) \frac{\mathbb{1}_B}{\lambda(B)} \\ &:= T_n + R_n. \end{aligned}$$

Thus

$$q(f_n - E(f_n)) \leq q(T_n) + q(R_n) \leq q(T_n) + \frac{\sqrt{2r}}{\delta}\|R_n\|_{L^1}.$$

therefore, for any $0 < \epsilon < 1$, we have

$$P\left(\frac{n}{b_n}(f_n - E(f_n)) \in F\right) \leq P\left(\frac{n}{b_n} q(T_n) > (1-\epsilon)\sqrt{2r}\right) + P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right).$$

Since T_n takes values in a finite dimension subspace of L^1, then there exist $k_1, ...k_l \in L^\infty$ such that $\max_j \|k_j\|_\infty \leq 1$ and $(1-\epsilon)q(T_n) \leq \max_j \langle k_j, T_n \rangle \leq \|T_n\|_{L^1}$. It follows then that

$$\begin{aligned}
P\left(\frac{n}{b_n}(f_n - E(f_n)) \in F\right) \;&\leq\; P\left(\frac{n}{b_n}\|T_n\|_{L^1} > (1-\epsilon)^2\sqrt{2r}\right) \\
&\quad +\; P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right) \\
&\leq\; P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B)-\mu(B)| > (1-\epsilon)^2\sqrt{2r}\right) \\
&\quad +\; P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right).
\end{aligned} \tag{5.17}$$

Using the statement (5.8), we obtain

$$\limsup_{n\to\infty} \frac{n}{b_n^2}\log P\left(\frac{n}{b_n}\sum_{B\in\Pi_{r_n}} |\mu_n(B)-\mu(B)| > (1-\epsilon)^2\sqrt{2r}\right) \;\leq\; (1-\epsilon)^2 r. \tag{5.18}$$

On the other hand, we have

$$\|R_n\|_{L^1} \;\leq\; \mu_n(|x|>r_n) \;+\; \mu(|x|>r_n).$$

Using Tchebycheff inequality,it follows, for any $t>0$, that

$$\begin{aligned}
P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right) \;&\leq\; P\left(\sum_{i=1}^{n}(\mathbb{1}_{|X_i|>r_n} + \mu(|x|>r_n)) > \epsilon\delta b_n\right) \\
&\leq \exp\{-b_n\epsilon\delta\}\exp\{nt\mu(|x|>r_n)\}\left[E\exp\{t\mathbb{1}_{|X_i|>r_n}\}\right]^n
\end{aligned} \tag{5.19}$$

Moreover, we have

$$E\exp\{t\mathbb{1}_{|X_i|>r_n}\} = \exp\{\log(1+(e^t-1)\mu(|x|>r_n))\}.$$

Taking $t=\log 2$, we obtain

$$E\exp\{t\mathbb{1}_{|X_i|>r_n}\} \leq \exp\{\mu(|x|>r_n)\}.$$

Thus, we obtain

$$P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right) \leq \exp\{-b_n\epsilon\delta\}\exp\{(1+\log 2)n\mu(|x|>r_n)\}.$$

Therefore, combining the the last inequality together with $(\mathbf{H_1})$ and $(\mathbf{H_3})$, we obtain

$$\begin{aligned}
\limsup_{n\to\infty}\frac{n}{b_n^2}\log P\left(\frac{n}{b_n}\|R_n\|_{L^1} > \epsilon\delta\right) &\leq \limsup_{n\to\infty}\left[-nb_n^{-1}\epsilon\delta \;+\; (1+\log 2)n^2 b_n^{-2}P(|X|>r_n)\right] \\
&= -\infty.
\end{aligned} \tag{5.20}$$

The combination of the statements (5.17), (5.18) and (5.20) yields

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n) \in F\right) \leq -(1-\epsilon)^2 r.$$

Since the choices of $0 < \epsilon < 1$ and $0 < r < \Lambda^*(F)$ are arbitrary we obtain

$$\limsup_{n\to\infty} \frac{n}{b_n^2} \log P\left(\frac{n}{b_n}(f_n - E(f_n) \in F\right) \leq -\Lambda^*(F),$$

by considering first the limit $\epsilon \to 0$ and subsequently the limit $r \to \Lambda^*(F)$. □

Proof of Corollary 5.1.

Since the closed ball $\{g \in L^1 : \|g\|_{L^1} \leq r\}$ is convex, then it is also weakly closed (Proposition A.4). Theorem 5.1 yields

$$-\Lambda^*(\|g\|_{L^1} > r) \leq \liminf_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}\|f_n - E(f_n)\|_{L^1} > r\right).$$

For any $a > r$, $\|af\|_{L^1} = a > r$, it follows that $\Lambda^*(\|g\|_{L^1} > r) \leq \Lambda(af) = \frac{a^2}{2}$. Therefore, for any $a > r$, we have

$$-\frac{a^2}{2} \leq \liminf_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}\|f_n - E(f_n)\|_{L^1} > r\right).$$

Thus, we obtain

$$-\frac{r^2}{2} \leq \liminf_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}\|f_n - E(f_n)\|_{L^1} > r\right).$$

For the upper bound, we have

$$\|f_n - E(f_n)\|_{L^1} \leq \sum_{B\in\Pi_{r_n}} |\mu_n(B) - \mu(B)| + \mu_n(|x| > r_n) + \mu(|x| > r_n).$$

Similarly as in the proof of Theorem 5.1 (statement (5.8)), we state that

$$\limsup_{n\to\infty} nb_n^{-2} \log P\left(nb_n^{-1}\|f_n - E(f_n)\|_{L^1} > r\right) \leq -\frac{r^2}{2}.$$

□

Annexe A

Topologie faible

Dans les chapitre 4 et 5, nous avons considéré l'espace L^1 muni de la topologie faible, ainsi dans cette annexe nous rappelons quelques définitions et outils de la topologie faible.

A.1 Cadre des espaces de Banach

Soit $\mathcal{X}$ un espace de Banach et soit $\mathcal{X}^*$ son dual algébrique, i.e., l'espace de toute les applications linéaires $\varphi : \mathcal{X} \longrightarrow \mathbb{R}$ (appelées formes linéaires). On considère le produit scalaire $\langle\ ,\ \rangle : \mathcal{X} \times \mathcal{X}' \longrightarrow \mathbb{R}$ défini, pour $(x, \varphi) \in \mathcal{X} \times \mathcal{X}'$, par $\langle x, \varphi\rangle = \varphi(x)$.

Définition A.1. *Le dual topologique de $\mathcal{X}$, noté $\mathcal{X}'$ est l'espace de toute les formes linéaires continues sur $\mathcal{X}$. La topologie faible sur $\mathcal{X}$, notée $\sigma(\mathcal{X}, \mathcal{X}')$, est la topologie la moins fine rendant continues toutes les applications $\varphi : \mathcal{X} \longrightarrow \mathbb{R}$, $\varphi \in \mathcal{X}'$.*

Proposition A.1. *La topologie $\sigma(\mathcal{X}, \mathcal{X}')$ est séparée.*

Proposition A.2. *Soit $x \in \mathcal{X}$. On obtient une base de voisinages de x pour la topologie faible $\sigma(\mathcal{X}, \mathcal{X}')$ en considérant les ensembles de la forme*

$$V_{I,\epsilon}(x) := \{y \in \mathcal{X} : |\langle y - x, \varphi_i\rangle| < \epsilon \ \forall i \in I\},$$

où I est fini, $\varphi_i \in \mathcal{X}'$ et $\epsilon > 0$.

Proposition A.3. *Soit $(\mathcal{Y}, \mathcal{T})$ un espace topologique et soit $f : (\mathcal{Y}, \mathcal{T}) \longrightarrow (\mathcal{X}, \sigma(\mathcal{X}, \mathcal{X}'))$ une application. Pour que f soit continue il faut et il suffit que pour tout $\varphi \in \mathcal{X}'$ l'application $\varphi \circ f : (\mathcal{Y}, \mathcal{T}) \longrightarrow \mathbb{R}$, $y \longrightarrow \langle f(y), \varphi\rangle$ est continue.*

Toute partie fermée pour la topologie faible $\sigma(\mathcal{X}, \mathcal{X}')$ est fermé pour la topologie forte (topologie de la norme). La réciproque est fausse en dimension infinie. Toutefois les deux notions coincident pour les ensembles convexes.

Proposition A.4. *Soit $C \subset \mathcal{X}$ un ensemble convexe. Alors C est fermé pour la topologie faible si et seulement s'il est fermé pour la topologie forte.*

Proposition A.5. *Soit $f : \mathcal{X} \longrightarrow]-\infty, \infty]$ une fonction convexe semi-continue inférieurement par rapport à la topologie forte. Alors f est semi-continue inférieurement pour la topologie faible. En particulier, si $x_n \to x$ pour $\sigma(\mathcal{X}, \mathcal{X}')$, alors*

$$f(x) \leq \liminf_{n\to\infty} f(x_n).$$

Pour $x \in \mathcal{X}$ on définit $J_x : \mathcal{X}' \longrightarrow \mathbb{R}$, $\varphi \longrightarrow J_x(\varphi) := \langle \varphi, x \rangle$ et on définit l'injection canonique $J : \mathcal{X} \longrightarrow \mathcal{X}''$, $x \longrightarrow J_x$.

Définition A.2. *On dit que $\mathcal{X}$ est réflexif si $J(\mathcal{X}) = \mathcal{X}''$*

Proposition A.6. *Toute partie convexe fermée et bornée dans un espace réflexif $\mathcal{X}$ est compacte pour la topologie faible $\sigma(\mathcal{X}, \mathcal{X}')$.*

A.2 Les espaces L^p

A.2.1 Cas $1 < p < \infty$

Théorème A.1. *(de représentation de Riesz.) L'espace L^p est réflexif et pour tout $\varphi \in (L^p)'$, il existe un unique $u \in L^q$ (avec $1/p + 1/q = 1$) tel que*

$$\langle \varphi, f \rangle = \int uf \quad \forall f \in L^p.$$

De plus on a,

$$\|\varphi\|_{(L^p)'} = \|u\|_{L^q}.$$

Ce théorème permet d'identifier le dual topologique de L^p à l'espace L^q.

A.2.2 Cas $p = 1$

Théorème A.2. *(de représentation de Riesz.) Pour tout $\varphi \in (L^1)'$, il existe un unique $u \in L^\infty$ tel que*

$$\langle \varphi, f \rangle = \int uf \quad \forall f \in L^1.$$

De plus on a,

$$\|\varphi\|_{(L^1)'} = \|u\|_{L^\infty}.$$

Ce théorème permet d'identifier $(L^1)'$ à l'espace L^∞.

A.2.3 Cas $p = \infty$

L'espace $(L^\infty)'$ contient strictement l'espace L^1. En effet, il existent des formes linaires sur L^∞ qui ne s'écrivent pas sous la forme $f \longrightarrow \int fu$ avec $u \in L^1$. Par conséquent, L^1 n'est pas réflexif et les boules fermées bornées de L^1 ne sont pas compactes pour la topologie faible $\sigma(L^1, L^\infty)$.

Perspectives de recherches

En conclusion de cette thèse nous donnons quelques perspectives de recherches pour des travaux futurs.

En extension aux résultats obtenus dans le premier chapitre, il est intéressant de faire le même travail pour la déviation en norme infinie $\left\{\|r_n^f - r^f\|_\infty : f \in \mathcal{F}\right\}$ et aussi d'améliorer l'application pour définir un critère de selection de modèles opérationnel.

Une autre question qui suscite notre intérêt est celle relative à l'estimateur de la fonction de densité par la méthode des séries orthogonales. Soit $\mathcal{X}$ un intervalle de $\mathbb{R}$ et soit $\{\phi_m\}_{n\geq 1}$ un système orthonormal complet de l'espace $L^2(X)$. Soit $X_1, X_2, ..., X_n$ une suite de variables aléatoires indépendantes et identiquement distribuées de densité commune $f \in L^2(X)$. L'estimateur de la fonction de densité par la méthode des séries orthogonales est défini par

$$f_n(x) = \frac{1}{n}\sum_{i=1}^{n} \delta_n(x, X_i),$$

où

$$\delta_n(x, X_i) = \sum_{j=1}^{m_n} \phi(x)\phi(X_i).$$

Ici m_n désigne une suite d'entiers tendant vers l'infini avec n. Louani (2003) (49) a établit un résultat de type chernoff pour la déviations en norme L^∞ de cet estimateur par rapport à la densité sous-jacente. Le but recherché sera d'étudier l'aspect fonctionnel lié aux grandes déviations dans l'espace L^2 pour cet estimateur.

Étendre le résultat de déviations modérées obtenu dans le dernier chapitre, en premier, à l'estimateur à noyau de la fonction de densité et ensuite à l'estimateur par la méthode des delta-suites est une question qui suscite notre intérêt. Plus précisément, étant donné que la limite des fonctions génératrice de moments associée à l'estimateur de la densité ne depends pas de la méthode d'estimation utilisée, on s'attend à ce que le résultat obtenu dans le cadre de l'estimation par la méthode des histogrammes reste valable pour d'autres méthodes d'estimation.

Bibliographie

[1] Abou-Jaoudé, S. (1976). Sur une condition nécessaire et suffisante de L_1-convergence presque complète de l'estimateur de la partition fixe pour une densité. *C. R. Acad. Sci. Paris Sér. A-B*, 283(16):Aii, A1107–A1110.

[2] Bahadur, R. R. (1971). *Some limit theorems in statistics.* Society for Industrial and Applied Mathematics, Philadelphia, Pa. Conference Board of the Mathematical Sciences Regional Conference Series in Applied Mathematics, No. 4.

[3] Beirlant, J., Devroye, L., Györfi, L. & Vajda, I. (2001). Large deviations of divergence measures on partitions. *J. Statist. Plann. Inference*, 93(1-2):1–16.

[5] Berrahou, N. (2003). Principe de grandes déviations pour l'estimateur de la densité par la méthode des delta-suites. *C. R. Math. Acad. Sci. Paris*, 337(5):347–352.

[6] Berrahou, N. (2006). Principe de grandes déviations uniforme pour l'estimateur de la densité par la méthode des delta-suites. *C. R. Math. Acad. Sci. Paris*, 343(9):595–600.

[4] Berrahou, N. & Louani, D. (2006). Efficiency of some tests when testing symmetry hypothesis. *J. Nonparametr. Stat.*, 18(7-8):465–482 (2007).

[7] Bleuez, J. & Bosq, D. (1976). Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales. *C. R. Acad. Sci. Paris Sér. A-B*, 282(17):Aiii, A1023–A1026.

[8] Bleuez, J. & Bosq, D. (1979). Conditions nécessaires et suffisantes de convergence de l'estimateur de la densité par la méthode des fonctions orthogonales. *Rev. Roumaine Math. Pures Appl.*, 24(6):869–886.

[9] Borovkov, A. A. & Mogulskii, A. A. (1980). Probabilities of large deviations in topological spaces. II. *Sibirsk. Mat. Zh.*, 21(5):12–26, 189.

[11] Bosq, D. (1978). Estimation de la densité par projection sur un sous-espace de dimension finie. *Portugal. Math.*, 37(1-2):93–111 (1981).

[10] Bosq, D. & Lecoutre, J. (1987). *Théorie de l'estimation fonctionnelle.* Economica.

[12] Bryc, W. & Dembo, A. (1995). On large deviations of empirical measures for stationary gaussian processes. stoch. *Proc. Appl.*, 58:23–34.

[13] Cencov, N. (1962). Evaluation of an unknown distribution density from observations. *Soviet. Math*, 3:1559–1562.

[14] Chen, X. & Guillin, A. (2004). The functional moderate deviations for Harris recurrent Markov chains and applications. *Ann. Inst. H. Poincaré Probab. Statist.*, 40(1):89–124.

[15] Chernoff, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statistics*, 23:493–507.

[16] Collomb, G. (1979). Conditions nécessaires et suffisantes de convergence uniforme d'un estimateur de la régression, estimation des dérivées de la régression. *C. R. Acad. Sci. Paris Sér. A-B*, 288(2):A161–A163.

[17] Cramér, H. (1938). Sur un nouveau théorème-limite de la théorie des probabilités. *Dans Actualités Scientifiques et Indistruelles, number 736*, pages 5–23. Hermann.

[18] de Acosta, A. (1992). Moderate deviations and associated Laplace approximations for sums of independent random vectors. *Trans. Amer. Math. Soc.*, 329(1):357–375.

[19] de Acosta, A. (1994). Large deviations for vector-valued Lévy processes. *Stochastic Process. Appl.*, 51(1):75–115.

[20] de Acosta, A. (1997). Exponential tightness and projective systems in large deviation theory. *Dans Festschrift for Lucien Le Cam*, pages 143–156. Springer, New York.

[21] de Acosta, A. & Chen, X. (1998). Moderate deviations for empirical measures of Markov chains : upper bounds. *J. Theoret. Probab.*, 11(4):1075–1110.

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

[23] Dembo, A. & Zeitouni, O. (1998). *Large deviations techniques and applications*, volume 38 de *Applications of Mathematics (New York).* Springer-Verlag, New York, second édition.

[24] Deuschel, J.-D. & Stroock, D. W. (1989). *Large deviations*, volume 137 de *Pure and Applied Mathematics*. Academic Press Inc., Boston, MA.

[25] Devroye, L. (1983). The equivalence of weak, strong and complete convergence in L_1 for kernel density estimates. *Ann. Statist.*, 11(3):896–904.

[26] Devroye, L. (1987). *A course in density estimation*, volume 14 de *Progress in Probability and Statistics*. Birkhäuser Boston Inc., Boston, MA.

[27] Devroye, L. & Györfi, L. (1985). *Nonparametric density estimation*. Wiley Series in Probability and Mathematical Statistics : Tracts on Probability and Statistics. John Wiley & Sons Inc., New York. The L_1 view.

[28] Devroye, L. P. (1978). The uniform convergence of the Nadaraya-Watson regression function estimate. *Canad. J. Statist.*, 6(2):179–191.

[29] Dupuis, P. & Ellis, R. S. (1997). *A weak convergence approach to the theory of large deviations*. Wiley Series in Probability and Statistics : Probability and Statistics. John Wiley & Sons Inc., New York. , A Wiley-Interscience Publication.

[30] Einmahl, U. & Mason, D. M. (2000). An empirical process approach to the uniform consistency of kernel-type function estimators. *J. Theoret. Probab.*, 13(1):1–37.

[31] Esaulova, V. (2004). Large deviations of L_1-error of empirical measures on partitions. *Statist. Probab. Lett.*, 69(4):439–445.

[32] Fix, E. & Hodges, J. (1951). Discriminatory analysis, nonparametric discrimination. *USAF School of Aviation Medicine, Randolph Field, Tex., Project*, pages 21–49.

[33] Földes, A. & Révész, P. (1974). A general method for density estimation. *Stud. Sci. Math. Hungar.*, 9:81–92 (1975).

[34] Gamboa, F., Rouault, A. & Zani, M. (1999). A functional large deviations principle for quadratic forms of Gaussian stationary processes. *Statist. Probab. Lett.*, 43(3):299–308.

[35] Gao, F. (2003). Moderate deviations and large deviations for kernel density estimators. *J. Theoret. Probab.*, 16(2):401–418.

[36] Gao, F. (2006). Moderate deviations and functional limits for random processes with stationary and independent increments. *Sci. China Ser. A*, 49(12):1753–1767.

[37] Geffroy, J. (1974). Sur l′estimation d′une densité dans un espace métrique. *C. R. Acad. Sci. Paris Sér. A*, 278:1449–1452.

[38] Geffroy, J. (1980). Étude de la convergence du régressogramme. *Publ. Inst. Statist. Univ. Paris*, 25(1-2):41–56.

[39] Greblicki, W. (1974). Asymptotically optimal algorithms for recognition and identification under probabilistic conditions. *Prace Nauk. Inst. Cybernet. Techn. Politech. Wrocław.*, (18 Ser. Monograf. No. 3):90.

[40] Györfi, L. (2004). Large deviations of Hellinger distance on partitions. *Dans Mathematics and computer science. III*, Trends Math., pages 531–537. Birkhäuser, Basel.

[41] Hu, S. H. (1993). A large deviation result for the least squares estimators in nonlinear regression. *Stochastic Process. Appl.*, 47(2):345–352.

[42] Hu, Y. & Lee, T.-Y. (2003). Moderate deviation principles for trajectories of sums of independent Banach space valued random variables. *Trans. Amer. Math. Soc.*, 355(8):3047–3064 (electronic).

[43] Lecoutre, J. (1982). Contribution à l'estimation non paramétrique de la régression. *Thèse de Doctorat de l'Université de Pierre et Marie Curie (Paris VI.*

[44] Ledoux, M. (1992). Sur les deviations moderees des sommes de variables aleatoires vectorielles independantes de meme loi. *Ann. Inst. H. Poincare Probab. Statist.*, 28(2):267–280.

[45] Lei, L. (2006). Large deviations of the kernel density estimator in $L^1(\mathbb{R}^d)$ for reversible Markov processes. *Bernoulli*, 12(1):65–83.

[46] Lei, L. & Wu, L. (2005). Large deviations of kernel density estimator in $L^1(\mathbb{R}^d)$ for uniformly ergodic Markov processes. *Stochastic Process. Appl.*, 115(2):275–298.

[47] Lei, L., Wu, L. & Xie, B. (2003). Large deviations and deviation inequality for kernel density estimator in $L_1(\mathbf{R}^d)$-distance. *Dans Development of modern statistics and related topics*, volume 1 de *Ser. Biostat.*, pages 89–97. World Sci. Publ., River Edge, NJ.

[48] Loftsgaarden, D. O. & Quesenberry, C. P. (1965). A nonparametric estimate of a multivariate density function. *Ann. Math. Statist.*, 36:1049–1051.

[51] Louani, D. (1997). Principes de grandes déviations pour l'estimateur à noyau de la densité de probabilité. *C. R. Acad. Sci. Paris Sér. I Math.*, 324(5):569–572.

[52] Louani, D. (1998). Large deviations limit theorems for the kernel density estimator. *Scand. J. Statist.*, 25(1):243–253.

[53] Louani, D. (1999). Some large deviations limit theorems in conditional nonparametric statistics. *Statistics*, 33(2):171–196.

[54] Louani, D. (2000). Large deviations for the L_1-distance in kernel density estimation. *J. Statist. Plann. Inference*, 90(2):177–182.

[49] Louani, D. (2003). Large deviations results for orthogonal series density estimators and some applications. *Math. Methods Statist.*, 12(2):177–196.

[55] Louani, D. (2005). Uniform L_1-distance large deviations in nonparametric density estimation. *Test*, 14(1):75–98.

[50] Louani, D. & Ould Maouloud, S. (2007). Un principe de grandes déviations fonctionel en estimation non-paramèrique. *C. R. Math. Acad. Sci. Paris*, 344(10):645–650.

[56] Mogul′skiĭ, A. A. (1976). Large deviations for the trajectories of multidimensional random walks. *Teor. Verojatnost. i Primenen.*, 21(2):309–323.

[58] Mokkadem, A., Pelletier, M. & Worms, J. (2005a). Large and moderate deviations principles for kernel estimation of a multivariate density and its partial derivatives. *Aust. N. Z. J. Stat.*, 47(4):489–502.

[57] Mokkadem, A., Pelletier, M. & Worms, J. (2005b). A large deviations upper bound for the kernel mode estimator. *Teor. Veroyatn. Primen.*, 50(1):189–200.

[59] Nadaraja, È. A. (1964). On a regression estimate. *Teor. Verojatnost. i Primenen.*, 9:157–159.

[60] Nadaraja, È. A. (1965). On non-parametric estimates of density functions and regression. *Teor. Verojatnost. i Primenen.*, 10:199–203.

[61] Nikitin, Y. (1995). *Asymptotic efficiency of nonparametric tests.* Cambridge University Press, Cambridge.

[62] Osmoukhina, A. V. (2001). Large deviations probabilities for a test of symmetry based on kernel density estimator. *Statist. Probab. Lett.*, 54(4):363–371.

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

[64] Prakasa Rao, B. L. S. (1983). *Nonparametric functional estimation.* Probability and Mathematical Statistics. Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York.

[67] Révész, P. (1968). *The laws of large numbers.* Probability and Mathematical Statistics, Vol. 4. Academic Press, New York.

[65] Révész, P. (1971). Testing of density functions. *Period Math. Hungar.*, 1(1):35–44.

[66] Révész, P. (1972). On empirical density function. *Period. Math. Hungar.*, 2:85–110. Collection of articles dedicated to the memory of Alfréd Rényi, I.

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

[69] Royall, R. (1966). A class of nonparametric estimates of a smooth regression function. *Ph. D. dissertation. Stanford University.*

[70] Sanov, I. N. (1961). On the probability of large deviations of random variables. *Dans Select. Transl. Math. Statist. and Probability, Vol. 1*, pages 213–244. Inst. Math. Statist. and Amer. Math. Soc., Providence, R.I.

[71] Scott, D. W. (1992). *Multivariate density estimation.* Wiley Series in Probability and Mathematical Statistics : Applied Probability and Statistics. John Wiley & Sons Inc., New York. Theory, practice, and visualization, A Wiley-Interscience Publication.

[72] Stone, C. J. (1977). Consistent nonparametric regression. *Ann. Statist.*, 5(4):595–645. With discussion and a reply by the author.

[73] Tiago de Oliveira, J. (1963a). Estatistica de densidades. *Resultados Assintoticos. Rev. Fac. Ci. Univ. Lisboa, ser. A*, 9:65–171.

[74] Tiago de Oliveira, J. (1963b). Estatistica de densidades. *Resultados Assintoticos. Rev. Fac. Ci. Univ. Lisboa, ser. A*, 9:111–206.

[75] Tukey, J. (1961). Curves as parameters and touch estimation. *Proc ; of the 4th Berkeley Symp. on Math. Stat. Prob.*, pages 681–694.

[76] Walter, G. & Blum, J. (1979). Probability density estimation using delta sequences. *Ann. Statist.*, 7(2):328–340.

[77] Walter, G. G. (1965). Expansions of distributions. *Trans. Amer. Math. Soc.*, 116:492–510.

[78] Wang, B. Z. (1989). The exponential rate of the large deviation probability for the nearest neighbor estimate of a regression function. *J. Systems Sci. Math. Sci.*, 9(2):175–192.

[81] Watson, G. S. (1964). Smooth regression analysis. *Sankhyā Ser. A*, 26:359–372.

[79] Watson, G. S. & Leadbetter, M. R. (1964a). Hazard analysis. I. *Biometrika*, 51:175–184.

[80] Watson, G. S. & Leadbetter, M. R. (1964b). Hazard analysis. II. *Sankhyā Ser. A*, 26:101–116.

[82] Worms, J. (2001). Moderate deviations of some dependent variables. I. Martingales. *Math. Methods Statist.*, 10(1):38–72.

Résumé

Cette thèse traite quelques aspects fonctionnels et non fonctionnels des grandes déviations et des déviations modérées en estimation fonctionnelle. Nous avons introduit dans la première partie un processus qui nous a permis de traiter de façon unifiée l'estimation de la fonction de densité et de la fonction de régression en utilisant plusieurs méthodes d'estimation. Plus explicitement, des principes de grandes déviations fonctionnels et non fonctionnels et des résultats de type Chernoff ponctuels et uniformes ont été obtenus. Dans un premier lieu nous avons établi un principe fonctionnel de grandes déviations pour l'estimateur par la méthode du noyau de la fonction de régression indexé par une famille de fonction vérifiant les conditions du théorème d'Arzèla-Ascoli. Ces résultats ont été utilisés pour définir un critère de sélection de modèles. Par la suite, dans la deuxième partie, nous nous sommes intéressé à l'estimation de la fonction de densité et de la fonction de régression par la méthode des histogrammes et nous avons obtenu des principes de grandes déviations ponctuels, des résultats de type Chernoff ponctuels et uniformes pour ces estimateur ainsi que des résultats de type minimax. Enfin dans les deux dernières parties, nous avons établi des principes fonctionnels de grandes déviations dans l'espace L^1 pour les estimateurs par la méthode des delta-suites des fonctions de densité et de régression ainsi qu'un principe de déviations modérées dans L^1 pour l'estimateur de la fonction de densité par la méthode des histogrammes.

Abstract

This thesis deals with functional and nonfunctional aspects of large and moderate deviation in nonparametric function estimation. We introduce a process which allows to derive in an unified way pointwise and functional large deviation principles and pointwise and uniform Chernoff type results for the density and regression function estimates. In first step, we establish a functional large deviations principle for the kernel estimate of the regression function indexed by a family of functions satisfying Arzèla-Ascoli conditions. Subsequently, the results have been applied to define a large deviations selection model criterion. In the second part, we investigate large deviation behaviours related to the density and regression function estimates using the histogram method. Here, large deviation principles, ponitwise and uniform chernoff type results as well as minimax bounds for both estimates have been stated. In the two last chapters, functional large deviation principles in the L^1-space endowed with the weak topology for the delta sequences density and regression function estimates as well as a functional moderate deviations principle for the histogram density estimate have been stated.